SBORNÍK

prací účastníků vědecké konference doktorského studia
Fakulta informatiky a statistiky
Vysoké školy ekonomické v Praze

Vědecký seminář se uskutečnil dne 11. února 2016
pod záštitou děkana FIS doc. RNDr. Luboše Marka, CSc.

Sestavení sborníku
prof. Ing. Petr Doucek, CSc.
proděkan pro vědu a výzkum

Nakladatelství Oeconomica – Praha 2016

ISBN 978-80-245-2136-7

OBSAH

Předmluva

V únoru roku 2016 se konal tradiční seminář „Den doktorandů“ Fakulty informatiky a statistiky VŠE v Praze. Seminář slouží jako platforma pro presentaci výsledků vědecké a odborné práce studentů všech doktorských oborů fakulty. Pro mnohé z nich je to první vystoupení před odbornou veřejností, na němž získávají cenné zkušenosti a tříbí si tak jak formulace svých názorů a hypotéz, tak i argumenty na jejich podporu. Nedílnou součástí jejich vystoupení je pak presentace závěrů výzkumné práce a diskuse jejích výsledků. Vlastní příprava vystoupení pak umožní doktorandům zvykat si na fakt, že výsledky své dlouhodobé práce musí stěsnat do relativně krátkého úseku presentace, v níž by měli být schopni jasně, srozumitelně, přehledně a poutavým způsobem přednést to, co vybádali.

Do dvacátého prvního ročníku „Dne doktorandů“ se přihlásilo celkem deset účastníků. Z toho na oboru „Aplikovaná informatika“ to bylo sedm příspěvků, na oboru „Ekonometrie a operační výzkum“ dva příspěvky a na oboru „Statistika“ jeden příspěvek. V letošním roce se semináře zúčastnil i zahraniční účastník z Centra informačních technologií Fakulty řízení a informatiky ze Žilinské univerzity v Žilině. Semináře se zúčastnili studenti obou forem studia, jak presenční, tak i kombinované.

Vzhledem k malému počtu přihlášených prací v programu „Ekonometrie a operační výzkum“ byla vytvořena jedna sekce pro obory „Ekonometrie a operační výzkum“ a „Statistika“.

Za práci v komisích chci poděkovat všem jejím členům, kteří pracovali pod vedením předsedů. Na oboru „Aplikovaná informatika“, který je součástí i stejnojmenného studijního oboru, byl předsedou jako již v několika uplynulých ročnících doc. Ing. Vojtěch Svátek, Dr. (KIZI) a členové komise byli v tomto roce prof. Ing. Zdeněk Molnár, CSc. (KIT) a doc. Ing. Vlasta Střížová, CSc. (KSA). V případě spojené komise pro příspěvky z oboru „Ekonometrie a operační výzkum“ a příspěvky z oboru „Statistika“, které zastřešuje společný studijní program „Kvantitativní metody v ekonomice“ byl předsedou prof. Ing. Roman Hušek, CSc. (KEKO), členy jeho komise pak byli prof. Ing. Hana Řezanková, CSc. (KSTP) a doc. RNDr. Ing. Michal Černý, Ph.D. (KEKO).

Komise pečlivě sledovaly předvedené výkony a stanovily následující pořadí nejlepších:

Studijní program – Aplikovaná informatika

1. místo: Ing. Pavel Jašek: Využití modelu celoživotní hodnoty zákazníka v rámci řízení výkonnosti maloobchodních společností

2. místo: Ing. Ludmila Malinová: Sociální a manažerská kybernetika v rámci analyzování organizace

3. místo: Mgr. Jindřich Mynarz: Generating examples of paths summarizing RDF datasets

Studijní program – Kvantitativní metody v ekonomice

1. místo: Ing. Roman Hajtmanek: Rovnomerné maticové rozvrhy v intuitívnej fuzzy logike

Oceněným studentům blahopřeji a věřím, že získané zkušenosti uplatní ve své další práci, ať už vědecké nebo v praxi. Uznání také patří všem vědeckým a pedagogickým pracovníkům, školitelům doktorandů, kteří se „Dne doktorandů“ zúčastnili a kteří svým vedením a radami byli nápomocni při zpracování a presentaci příspěvků.

Zvláštní poděkování pak patří studijní referentce doktorského studia paní Jitce Krajíčkové, díky níž byl seminář skvěle organizačně zajištěn, dále paní Petře Šarochové za administrativní podporu akce a Mgr. Lee Nedomové za práci při editaci a sestavení tohoto sborníku.

Prof. Ing. Petr Doucek, CSc,
Proděkan pro vědu a výzkum

Využití modelu celoživotní hodnoty zákazníka v rámci řízení výkonnosti maloobchodních společností

Pavel Jašek

pavel.jasek@vse.cz

Doktorand oboru Aplikovaná informatika

Školitel: doc. Ing. Ota Novotný, Ph.D., (ota.novotny@vse.cz)

Abstrakt: Příspěvek shrnuje problematiku aplikace modelu celoživotní hodnoty zákazníka v rámci řízení výkonnosti maloobchodních společností. Jsou diskutována specifika těchto modelů, odlišnosti maloobchodního odvětví od tradičních vertikál, kde se predikce budoucí hodnoty zákazníka používá a velký důraz je kladen na praktickou aplikaci v kontextu operativního i strategického řízení výkonnosti, obzvláště v rámci řízení marketingu. Informatické dopady na potřebné datové zdroje a integrační služby jsou analyzovány zejména z pohledu možného rozšíření modelů o nefinanční složky, které mohou také demonstrovat hodnotu zákazníka pro společnost.

Oproti tradičnímu využití CLV v telekomunikacích a finančním odvětví jsou maloobchodní společnosti specifické zejména nekontrakčním prostředím. Podle toho je nutné opatrně volit i statistické modely pro odhad CLV a umět je správně posuzovat. Právě vznikající metodika pro využití CLV k řízení výkonnosti firmy si dává za cíl uchopit tuto problematiku a vytvořit velmi praktické operativní i strategické postupy, metody, procesy a techniky, které budou určeny zejména marketingovým manažerům v maloobchodním odvětví.

Klíčová slova: Celoživotní hodnota zákazníka, Řízení vztahu se zákazníky, Řízení podnikové výkonnosti, Řízení marketingu, Hodnota zákaznické báze

Úvod

V konceptu strategické orientace na zákazníka, jak ji popisuje Fader (2012), je orientace na řízení budoucí hodnoty zákazníků důležitým prostředkem pro správné nastavení marketingového řízení. Tento příspěvek se zabývá oblastí analýzy a řízení výkonnosti obchodních společností, konkrétně ve vztahu k životní hodnotě zákazníka a metodám a postupům jejího stanovení na základě analýzy zákaznických dat. Vzhledem k intenzivnímu využívání dat z různých součástí informačního systému firmy i dalších součástí ICT, jde o specifickou aplikaci podnikové informatiky sloužící ke zlepšení výkonnosti a účinnosti marketingových a obchodních procesů společností.

Práce je zaměřena na maloobchodní společnosti, zejména ty působící v digitálním prostředí (např. Internetové obchody). Podstatnou otázkou pro tyto společnosti je, zda se vůbec mají dlouhodobou hodnotou svých zákazníků zabývat, zda pro ně jsou věrní zákazníci důležití pro celkové výsledky podnikání a zda změnou přístupu k řízení marketingových aktivit dosáhnou vyšší účinnosti. Potřeba jasně metodicky uchopit problematiku dlouhodobé hodnoty zákazníků v rámci řízení výkonnosti společnosti vychází z Fader (2012).

Předmětná oblast se soustřeďuje na pojem životní hodnoty zákazníka ve vztahu k modelování hodnoty různých segmentů zákazníků i individuálnímu profilování a skórování. K této problematice se váže také modelování loajality zákazníků a odhad ztráty zákazníka. Na tyto aktivity budou využita všechna transakční data o zákazníkovi integrovaná s daty ze systémů pro řízení vztahu se zákazníky (Customer Relationship Management, dále jako CRM), z dalších nástrojů pro Customer Intelligence a také z nástrojů pro webovou analytiku. Jedním z důvodů využití více zdrojů dat je multikanálové prostředí, se kterým společnosti doposud pracovaly roztříštěně.

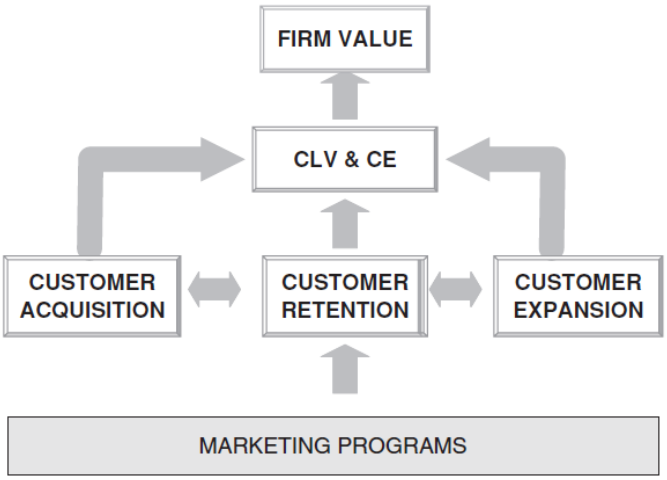
Tyto postupy a nástroje jsou stěžejní pro řízení společnosti ze dvou pohledů:

- Ve vztahu k celkové výkonnosti firmy – jaký vliv má současná či budoucí hodnota zákazníka na finanční výsledky společnosti?
- Ve vztahu k individuálním metrikám marketingu – jak lze pomocí odhadu budoucí hodnoty zákazníka lépe řídit marketingové aktivity (např. pro akviziční kampaně nových zákazníků či cross-sellingové taktiky)?

Obchodní společnosti doposud nebyly typickými uživateli pokročilých analytických metod pro řízení hodnoty zákazníka a je možné o nich nalézt pouze málo zmínek v literatuře. Známy jsou aplikace v rámci telekomunikačních a finančních společností. Proto je jedním z cílů tohoto příspěvku a navazující disertační práce přispět k aplikaci nových metod do tržních vertikál, jež mají málo zkušeností se systematickým přístupem k řízení výkonnosti.

Customer Liftime Value

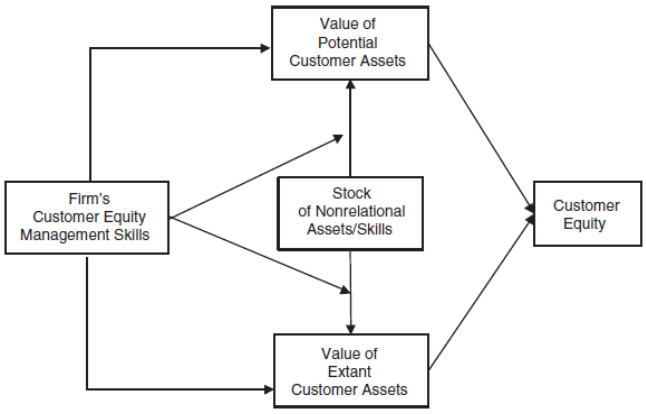
Celoživotní hodnota zákazníka (dlouhodobá hodnota zákazníka, Customer Lifetime Value, dále jako CLV) je definována jako čistá současná hodnota budoucích zisků od konkrétního zákazníka (Fader, 2012). Obrázek 1 vymezuje vztah mezi marketingovými aktivitami pro získávání, aktivaci a retenci zákazníků, dopadem na jejich budoucí hodnotu CLV a CE a následně také celkovou hodnotou společnosti.



Obrázek 1: Konceptuální rámec využití modelu CLV Zdroj: Gupta et al. (2006)

Customer Equity

Hodnota zákaznické báze (Zákaznický kapitál, Customer Equity, dále jako CE) je dle zjednodušené definice součet dílčích dlouhodobých hodnot zákazníků přes celou zákaznickou bázi (Fader, 2012). Bližší definice pojmů CLV a CE uvádí např. Gupta et al. (2006). Obrázek 2 zobrazuje konceptuální model pro řízení celkové hodnoty zákaznické báze s nejdůležitějšími faktory rozdělenými na současnou hodnotu zákazníků a budoucí potenciální hodnotu, kterou může společnost od zákazníků očekávat.



Obrázek 2: Konceptuální model řízení Hodnoty zákaznické báze Zdroj: Hogan, Lemon & Rust (2002)

Takovéto strategické a systematické využívání dat o zákaznících by dle Davenport (2007) mělo vést také k dosažení konkurenční výhody, a tím pádem je důležitým tématem pro řízení marketingových a obchodních činností v rámci obchodních řetězců.

Současné využití CLV ve firmách

Řízení výkonnosti firmy ve vztahu k měření a analýze dlouhodobé hodnoty zákazníka je tradičně používáno v rámci telekomunikačních společností a finančních institucí (např. pro produkt povinného ručení automobilů). Jak uvádí Machander (2012), z důvodu velké saturace trhu pro tyto společnosti nabývá na důležitosti navyšování tržeb plynoucí ze stávající zákaznické báze.

Telekomunikační společnosti jsou poměrně specifické vzhledem ke svému byznys modelu. Významným a praktickým ukazatelem pro měření situace na telekomunikačním trhu je ukazatel *ARPU* (average revenue per unit, průměrný výnos na uživatele), kterým si operátoři segmentují např. podle jednotlivých služeb (data, SMS, hlasové služby). Modely budoucí profitability zákazníka pomáhají telekomunikačním společnostem určit náklady spojené s provozem zákazníka, což slouží také jako podklad pro výpočty finanční návratnosti marketingových kampaní. Appeltauer (2011) zdůrazňuje, že analytický přístup je v oboru telekomunikací nutnost a nutný podklad pro rozhodování. V rámci řízení výkonnosti marketingových kampaní v online prostředí doporučuje nejen měření současného plnění cílů, ale sledování predikované hodnoty cíle s pomocí ukazatele *run rate*.

Šídlo (2002) uvádí, že pro finanční instituce je typický posun od produktové profitability k profitabilitě zákazníků. Budoucí hodnota zákazníka je v mnoha situacích snadno definovatelná či odhadnutelná (např. pro hypotéky). Velkou roli hrají také retenční programy, zejména pro bankovní domy, které nemají příliš konkurenceschopnou nabídku a zákazníci jsou ochotni přecházet ke konkurenci (např. spořicí účty nových internetových bank).

Pro maloobchodní společnosti často platí následující podmínky, které definují výrazně odlišné požadavky na modely CLV v této tržní vertikále:

- B2C (business-to-customer) prostředí přímo propojující koncové zákazníky se společností,
- nekontrakční vztah (non-contractual settings), ve kterém zákazník není povinen volit pro svůj nákup právě jednu společnost a lze očekávat nižší kvalitu prediktivních modelů (např. v kontrastu s telekomunikačními společnostmi, kde může být zákazník smluvně vázán na roky dopředu),
- neklubový vztah (non-membership status), kdy zákazníci nebývají členy věrnostních klubů, které by je výrazně motivovaly měnit své nákupní chování,
- pokračující neurčitost (always-a-share, opak lost-for-good), v rámci níž zákazník, který přestal nakupovat, se ještě může vrátit k nakupování u společnosti i za několik let,
- kontinuální prostředí (continuos buying), kdy má zákazník možnost nakupovat kdykoliv, cokoliv, opakovaně, klidně několikrát denně,
- proměnlivá hodnota útrat (variable-spending environment), kdy široké portfolio produktů s různými cenami dává základ velkému rozptylu hodnot, mezi kterými nemusí být přímá souvislost a
- anonymnost prostředí, kdy zákazník nemusí být přesně identifikován v rámci svých více nákupů a společnost má jen omezené možnosti tuto identifikaci nahrazovat.

CLV a CE pro strategické řízení výkonnosti firmy

Berger (2006) se zabývá propojením hodnoty zákazníka s výkonností celého podniku. Dopady tohoto propojení na ohodnocení diskutoval Bauer (2003). Berger (2002) uvádí vazbu CLV a CE na koncept řízení hodnoty zákazníků (Customer Value Management). Lewis (2015) dokonce uvádí využití při oceňování podniku během akvizice společností (Mergers and Acquisitions).

Dle Lewise (2006) spočívají praktické aplikace této problematiky v dílčích analýzách a metodách řízení obchodní výkonnosti firem. Účinnější investice do marketingových kanálů jsou jednou z přímých aplikací řízení obchodní výkonnosti.

CLV v kontextu operativního řízení společnosti

Operativní využití v rámci CRM zahrnuje možné marketingové aktivity, které vedou k akceschopnosti a byznys pravidlům pro práci se změnami odhadů CLV. Využití CLV v marketingovém řízení popisuje Karlíček a kolektiv (2014).

Dříve popisované modely hodnoty zákazníka nereflektují digitální marketingové kanály a měkká data. Zejména nepopisují, jak k současné či budoucí hodnotě zákazníka mohou být připojeny informace získané o spokojenosti zákazníků (např. přes koncept Net Promoter Score), data o chování zákazníka na webových stránkách, data ze sociálních sítí. Tzv. „e-hodnota“ zákazníka se v současné praxi bere pouze z online kanálů a funguje izolovaně od interakcí zákazníka např. v pobočkové síti obchodního řetězce.

Nutnost zapojení více atributů o zákaznících zdůrazňuje Chen (2003), který také uvádí nutnost integrace více proměnných dohromady v rámci jednoho modelu hodnoty zákazníka. Přestože sjednocení hodnoty není jednoduché, je důležité pro další vývoj určit, zda se aspekty z online kanálů mají promítat do celofiremní hodnoty zákazníka.

Využití modelů CLV a CE v Řízení byznys informatiky

V rámci práce na zobecněných řešeních v řízení provozu a rozvoje IT, resp. podnikové informatiky, je koncept celoživotní hodnoty zákazníka zapracován jako objekt F052 do systému MBI (Management byznys informatiky)1. Tabulka 1 ukazuje část MBI faktoru, který definuje efekty, výhody a uplatnění faktoru, a také určuje problémy, omezení a otázky uplatnění faktoru.

Tabulka 1: Positivní efekty a možné problémy CLV a CE Zdroj: Autor

Positiva	Negativa
• Zvýšení ekonomických efektů díky přesnějšímu vyhodnocování zákazníků a jejich potřeb. • Efektivnější příprava, přesnější zaměření marketingových akcí a jejich vyšší přínosy pro podnik. • CE: Získání komplexní informace o hodnotě zákaznické báze a jejích potenciálních přínosů pro podnik.	• Nedaří se vždy dosáhnout správné segmentace zákazníků, kritéria je v konkrétním prostředí obtížné nastavit. • Nejsou dostupné informace o profitabilitě zákazníků. • Problematická dostupnost informací o všech zákaznících na potřebné úrovni detailu. • Ekonomická náročnost pořízení dat o celé zákaznické bázi.

Modely pro predikci CLV

Pro predikci CLV existuje mnoho modelů. Gupta et al. (2006) modely rozděluje do šesti kategorií:

- RFM modely (Recency, Frequency, Monetary value). Skórovací modely, které predikují aktivitu zákazníka během jednoho následujícího období, bez určení finanční hodnoty. Komponenty RFM jsou silnými prediktory pro odhad budoucího zákaznického chování.
- Pravděpodobnostní modely využívají statistické stochastické procesy k určení, zda bude zákazník v příštím období aktivní.
- Ekonometrické modely slouží k modelování zákaznické akvizice, aktivace a retence, a je možné je tedy spojit s predikcí hodnoty zákazníka.
- Dynamické modely perzistence analyzují chování jednotlivých složek CLV jako součásti dynamického systému.
- Informatické modely (ve smyslu dolování znalostí z databází, strojového učení a neparametrických statistických metod) a jejich kombinace s předchozími způsoby.
- Difuzní a růstové modely, které odhadují přírůst nových zákazníků s využitím agregovaných dat.

V této práci jsou využity modely z tří těchto kategorií: Status Quo, Pareto/NBD, Markovovy řetězce a VAR modely.

Status quo je příkladem intuitivního naivního přístupu k odhadu celoživotní hodnoty zákazníka. Model očekává, že výnosnost klienta bude pořád stejná po určitý jednoduše určený počet let, i když jde o nekontrakční prostředí. V literatuře se model využívá jako základní očekávání (baseline) pro složitější modely, aby bylo možné při dosažení jen nepatrně lepších výsledků u složitějšího modelu

[1] Viz http://mbi.vse.cz/public/cs/obj/FACTOR-114

vyvažovat nejen kvalitu predikce, ale také procesní a datovou náročnost komplexnějšího modelování. Wübben a Wangenheim (2008) takovéto využití popisují a očekávají dobré fungování pouze v krátkém období predikce.

Pareto/NBD (Pareto/Negative Binomial Distribution) model odhaduje, jak dlouho bude daný zákazník aktivní a kolikrát během této doby nakoupí. Protože samostatně neobsahuje stanovení finanční hodnoty útraty, bývá tento model doplněn o navazující Gamma-Gamma model. Existují také alternativní modely jako BG/NBD, které se odlišují zejména předpoklady a definicí aktivního zákazníka na konci sledovaného období. (Fader, Hardie & Lee, 2005)

Kombinace Markovových řetězů a rozhodovacích stromů (CART) potřebuje RFM komponenty a nějaká demografická data, aby bylo možné kategorizovat zákazníky. Vytvoří se rozhodovací strom, který rozdělí zákazníky do skupin podle výnosnosti, neboť ziskovost v rámci Monetary value je vysvětlovaná proměnná. Pak se podle Markov Switching modelu určují pravděpodobnosti přechodů zákazníků mezi jednotlivými výdělečnými stavy. Výsledkem je odhad CLV na zákazníka. (Pauwee, Putten a Wezel, 2007)

Vector autoregressive (VAR) modely jsou autoregresivní modely (Villanueva et al., 2008), které zahrnují časová zpoždění, kategorie akvizičních zdrojů pro získání zákazníků a interakce mezi nimi (Vraná a Jašek, 2015).

Způsoby měření přesnosti a kvality modelů

Srovnávání různých modelů pro odhad CLV je náročné z důvodu odlišného zaměření jednotlivých metod a jejich výstupních metrik. Hledá se nejmenší společný prvek, ve kterém je možné modely posuzovat. Využívají se dvě úrovně srovnávání:
1) srovnání CLV za všechny zákazníky dohromady,
2) úspěšnost výběru nejziskovějších zákazníků.

Aby byly modely co nejvíce srovnatelné, je vhodné:

- predikovat vývoj CLV všech stávajících zákazníků,
- vybrat určité krátké období pro odhadovanou dobu (např. 1 rok) místo spoléhání na celoživotní predikci,
- při odhadu modelů ponechat validační vzorek a
- ignorovat odhad nově příchozích zákazníků.

Berka (2005) uvádí, že v případě numerické predikce se jako kritérium správnosti obecně používají např. střední kvadratická chyba (mean squared error, MSE), odmocnina ze střední kvadratické chyby (root mean squared error, RMSE), střední absolutní chyba (mean absolute error, MAE), relativní kvadratická chyba (relative squared error, RSE), nebo korelační koeficient ρ.

Donkers, Verhoef a Jong (2007) pro srovnání predikce CLV různými modely využívají střední absolutní chybu (MAE, mean absolute error) v % z průměrné hodnoty CLV, střední kvadratickou chybu (RMSE, root mean square error) v % z průměrné hodnoty CLV.

Schopnost modelů najít nejvýnosnější zákazníky je dalším důležitým kritériem. Donkers, Verhoef a Jong (2007) popisuje metriku Hit rate, kdy rozdělí zákazníky na čtvrtiny podle výše CLV a srovnávají s reálným rozdělením jako v klasifikační úloze. Pro jemnější a praktičtější využití je vhodné rozdělit zákazníky na 10 % nejziskovějších + 90 % zbývajících, a analyzovat Hit rate na těchto dvou skupinách.

Ke srovnání Hit rate se využijí data z klasifikační tabulky (známá také jako classification matrix, confusion table apod.). Na základě dat z tabulky se určí sensitivita, specificita, přesnost (Acc, accuracy, správnost) a F1 skóre, dle (Berka, 2005). Výslednou vzorovou šablonu pro vyplnění výsledků porovnání různých modelů uvádí tabulka 2.

Tabulka 2: Prázdná šablona pro vyhodnocení kvality modelů CLV Zdroj: Autor

Model CLV	Chybovost		Kvalita Hit rate			
	MAPE	RMSE	Sensitivita	Specificita	Přesnost	F1 Skóre
Status Quo						
Pareto/NBD						
Markovovy procesy + CART						
VAR modely						

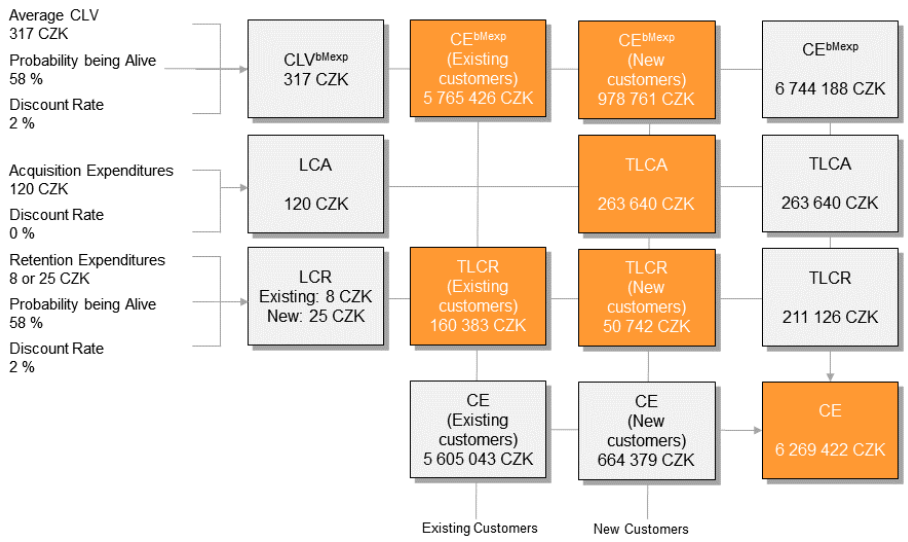
Hodnota zákaznické báze (Customer Equity)

Pro strategické rozmýšlení o kvalitě zákaznické báze musí být připraven vhodný způsob reportování Hodnoty zákaznické báze (CE). Wiesel, Skiera & Villanueva (2008) navrhnuli inovativní pohled Customer Equity Statement, který v jejich práci vychází z veřejných informací a uvádí jej na příkladu společnosti Netflix. Vhodnost tohoto přístupu podporuje i práce autorů Gupta, Lehmann & Stuart (2004), jejichž odhad CE z dostupných výročních zpráv veřejně obchodovatelných společností byl velmi blízko skutečné tržní hodnotě většiny zkoumaných společností. Nutno podotknout, obdobně jako upozorňuje Wiesel, Skiera & Villanueva (2008), že tato aproximace je méně vhodná pro firmy v nekontrakčním prostředí z důvodu těžkého odhadu stávajících a ztracených zákazníků v určitém čase.

Kumar & George (2007) rozlišují mezi dvěma přístupy k měření CE podle nároků na datové zdroje, vypočítané metriky, výchozí předpoklady a úroveň agregace, přičemž oba způsoby vedou k jiným metodám maximalizace hodnoty zákazníků:

- v neagregovaném pohledu je celoživotní hodnota zákazníků maximalizována přes aplikaci strategií zaměřených na jednotlivé zákazníky, pod což autoři řadí optimální alokaci zdrojů, analýzu sekvenčních nákupů a řízení vah mezi investicemi do akvizice a retence zákazníků,
- v agregovaném pohledu je cílem zlepšovat faktory působící na maximalizaci CE.

Customer Equity Statement, jak jej představili Wiesel, Skiera & Villanueva (2008) také pomáhá s vyjádřením takových faktorů, včetně odhadů celoživotní hodnoty, rozlišení výdajů na akvizici a retenci klientů. Schematickým znázorněním veškeré metriky zpřístupňuje k manažerskému rozhodování.



Obrázek 3: Customer Equity Statement pro analyzovaná data internetového obchodu se zdravotními a zábavnými produkty. Jde o hodnoty ke čtvrtému kvartálu 2013. CE = customer ekvity (Hodnota zákaznické báze), CLV^{bMExp} = customer lifetime value before marketing expenditures (Celoživotní hodnota zákazníka před uplatněním marketingových výdajů), LCA = lifetime acquisition expenditures (celoživotní akviziční náklady), LCR = lifetime retention expenditures (celoživotní retenční náklady), TLCA = total lifetime acquisition expenditures (suma celoživotních akvizičních nákladů), and TLCR = total lifetime retention expenditures (suma celoživotních retenčních nákladů).

Zdroj: Jašek (2015), volně dle Wiesel, Skiera & Villanueva (2008)

Obrázek 3 zobrazuje stav zákaznické báze k poslednímu kvartálu roku 2013 pro zkoumaný internetový obchod (Jašek, 2015). Customer Equity Statement zde představuje odhad hodnoty zákaznické báze pro následující 3 roky s odhadem průměrných akvizičních nákladů 120 CZK, retenčních nákladů 8 CZK na rok pro stávající zákazníky a se stanovenou diskontní sazbou 2 %. Model ukazuje průměrnou hodnotě CLV 445 CZK pro nové zákazníky, což je o 40 % vyšší než celková průměrná hodnota CLV. Důležitým ukazatelem pro strategické marketingové řízení je fakt, že noví zákazníci mají tvořit jen 11 % celkové hodnoty CE v příštích 3 letech.

Návrh metodiky pro využití CLV k řízení výkonnosti firmy

Cílem navazující disertační práce je vytvořit metodiku pro řešení problematiky dlouhodobé hodnoty zákazníka v rámci řízení výkonnosti maloobchodních společností. Tato práce shrnuje očekávaný obsah metodiky:

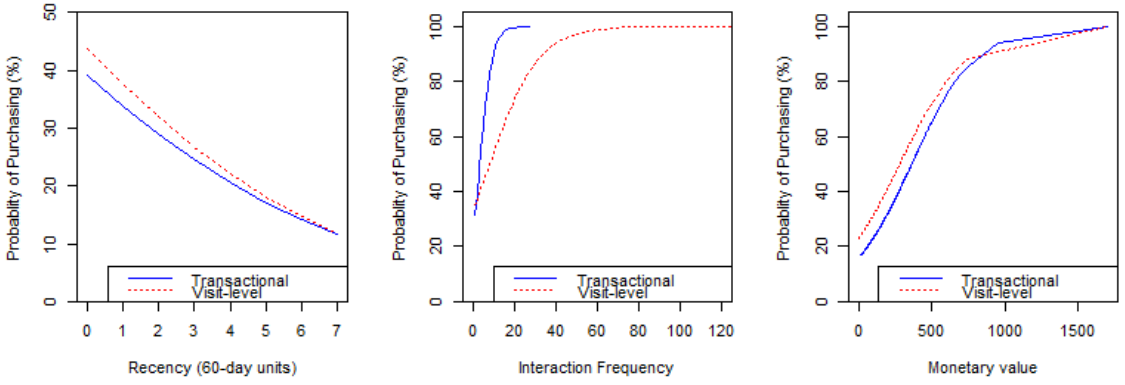
- Souhrn doporučených praktik a postupů pro aplikaci modelů CLV a CE do řízení výkonnosti celé společnosti, s důrazem na marketingové řízení; na základě dat uložených a zpracovávaných v informačních systémech společnosti.
- Metodika bude zahrnovat několik metod, jak přistupovat k predikcím individuální hodnoty CLV a jak využívat CE. Podle typu společnosti a dalších definovaných pravidel budou stanoveny zásady vhodného užití.
- Techniky, které budou součástí metod, budou upřesňovat aplikaci statistických metod pro konkrétní situaci, definovat vazby mezi dílčími fázemi metody (např. pro využití CLV v rámci optimalizace marketingového rozpočtu, pro využití CE v rámci odhadu hodnoty zákaznické báze v případě agresivnějšího růstu počtu zákazníků apod.).

Z této motivace je zřejmé, že důraz musí být kladen na správné podchycení veškerých datových zdrojů, integraci těchto dat v informačních systémech firmy a přípravu potřebných výstupů, aby byly predikované hodnoty k dispozici ve správný čas a ve správných systémech, aby bylo možné hodnoty aplikovat jak ve strategickém řízení, tak i v operativním řízení marketingu a dalších odvětvích. V této práci je záběr pouze na operativní řízení.

Datové a informační zdroje v informačních systémech firmy

Základem pro výpočet CLV a CE jsou veškerá transakční data zákazníků, která typicky bývají uložena v systémech CRM nebo ERP (Enterprise Resource Planning). Vzhledem k důrazu na ziskovost zákazníků jsou integrována také data z manažerského účetnictví nebo přímo z účetnictví společnosti kvůli správné alokaci nákladů a výnosů.

Zapojení nefinančních dat do modelů celoživotní hodnoty zákazníka předpokládá v digitálním marketingu integraci dat z webové analytiky (Novotný a Jašek, 2013). Integrace signálů o návštěvnosti webu či mobilní aplikace pak může být vhodným vstupem např. do RFM modelů, jak ukazuje obrázek 4. Rozšíření transakčních dat jde zejména poznat v dopadu frekvence interakce zákazníka se společností a z toho vyplývající pravděpodobností opakovaného nákupu.



Obrázek 4: Vizuální srovnání pravděpodobnosti opakovaného nákupu využití transakčních dat a údajů doplněných o návštěvnost na webu, s využitím tří logistických modelů Buy ~ Recency, Frequency a Monetary Value. Zdroj: Jašek (2015).

Využití v marketingovém řízení

Pro praktické dennodenní řízení marketingových aktivit je podstatné stanovit procesy, do kterých bude využití CLV a CE zapojeno. Výsledná metodika bude obsahovat právě souhrn doporučených praktik a postupů pro aplikaci modelů. Zároveň k tomu musí být data připravená a vhodně integrovaná s cílovými systémy (např. e-mailing, správa reklamních systémů). V této části jsou uvedeny praktické příklady:

- Rosset (2002) reaguje na využití hodnoty zákazníka pro zlepšení udržení vztahů se zákazníky.
- Manuální vyhodnocení akvizičních kanálů podle zpětného pohledu, odkud byli získáni kvalitní zákazníci a úprava nastavení rozpočtů do marketingových kanálů podle výsledků.
- Metody (např. ve smyslu klasifikačních regresních modelů nebo rozhodovacích stromů) na hledání faktorů ovlivňujících vyšší hodnotu.
- Snížení slevových incentiv na segmentu zákazníků, kde to není tolik potřeba (Hallberg, 1995). Gupta et al. (2006) uvádí síťové efekty zákazníků, kteří dobrovolně a bez náhrady šíří dobré jméno firmy. Techniky word-of-mouth popisuje jako dvakrát hodnotnější než získávání návštěvníků přes klasické marketingové aktivity.
- Selekce nejziskovějších zákazníků pro výběr omezených marketingových aktivit (např. up-sellingových soutěží). Do této kategorie patří také analýza důvodů, proč se předpoveď klasifikace nepodařila.
- Shlukování individuálních výsledků je zajímavou metodou, při které se neposuzují individuální zákazníci, ale homogenní skupiny zákanzíků podobných také přes odhad budoucí hodnoty. Jak ve své práci uvádí Jašek (2015), zapojení shlukové analýzy pomáhá stanovit kvalitu akvizičních marketingových kanálů. Výsledky takové analýzy jsou vidět v tabulce 4.

Tabulka 3: Průměrné hodnoty CLV (AVG_{CLV}), směrodatná odchylka hodnoty zákazníků (SD_{CLV}) a celková hodnota shluků (CE_{group}) vytvořených hierarchickým shlukováním. Hodnoty jsou uvedeny v EUR.
Zdroj: Jašek (2015)

Shluk	Počet zákazníků (podíl z celku)	Počet zákazníků podle akvizičního zdroje					CE_{group}	AVG_{CLV}	SD_{CLV}
		(none)	cpc	email	organic	referral			
1	2 208 (15 %)	0	0	0	2 208	0	62 146	28,15	18,47
2	4 618 (31 %)	0	0	0	4 618	0	27 452	5,94	3,52
3	4 969 (33 %)	0	0	0	0	4 969	54 016	10,87	12,26
4	2 941 (20 %)	2 941	0	0	0	0	39 393	13,39	16,68
5	259 (2 %)	0	242	17	0	0	5 296	20,45	21,20

Závěr

Příspěvek zdůraznil význam celoživotní hodnoty zákazníka v rámci operativního i strategického řízení výkonnosti maloobchodních společností. Jak bylo uvedeno v kapitole Současné využití CLV ve firmách, oproti tradičnímu využití CLV v telekomunikacích a finančním odvětví je maloobchod dost specifický zejména nekontrakčním prostředím. Podle toho je nutné opatrně volit i statistické modely pro odhad CLV a umět je správně posuzovat (což bylo diskutováno v kapitole Způsoby měření přesnosti a kvality modelů).

Právě vznikající metodika pro využití CLV k řízení výkonnosti firmy si dává za cíl uchopit tuto problematiku a vytvořit velmi praktické operativní i strategické postupy, metody, procesy a techniky, které budou určeny zejména marketingovým manažerům v maloobchodním odvětví.

Jak uvádí podkapitola Datové a informační zdroje v informačních systémech firmy, je potřeba pracovat nejen s tradičními transakčními záznamy k zákazníkům, ale využívat také moderní datové zdroje, např. údaje z webové analytiky, protože se ukazuje, jak přidávají nový pohled do odhadů budoucí hodnoty zákazníků.

Další výzkum by měl být směrován na praktická srovnávání metod a modelů vhodných pro maloobchodní odvětví, na kvalitativní marketingový průzkum mezi vedoucími pracovníky marketingu a na získání případových studií pro demonstraci přínosů modelů CLV a CE v maloobchodním odvětví.

Literatura

APPELTAUER, Roman, 2011. *Web Analytics v korporaci*. In: Web Analytics Wednesday. [Online] 29. červen 2011. Dostupné z http://www.webanalyticswednesday.cz/wp-content/uploads/2011/07/web-analytics-v-korporaci.pdf

BAUER, Hans H., HAMMERSCHMIDT, Maik a BRAEHLER, Matthias, 2003. The Customer Lifetime Value Concept and its Contribution to Corporate Valuation. *Yearbook of Marketing and Consumer Research*, Vol. 1. Berlín: Duncker und Humblot. ISSN: 1612-9814.

BERGER, Paul D. et al., 2002. Marketing Actions and the Value of Customer Assets A Framework for Customer Asset Management. *Journal of Service Research*. roč. 5, č. 1, s. 39–54

BERGER, Paul. D. et al., 2006. *From Customer Lifetime Value to Shareholder Value: Theory, Empirical Evidence, and Issues for Future Research. Journal of Service Research* [online]. roč. 9, č. 2, s. 156–167 ISSN 1094-6705. Dostupné z: doi:10.1177/1094670506293569

DAVENPORT, Tomas a HARRIS, Jeanne, 2007. *Competing on Analytics: The New Science of Winning*. 1. vydání. Boston: Harvard Business School Press, 2007. str. 240. ISBN-13: 978-1422103326.

DONKERS, Bas, VERHOEF, Peter C., de JONG, Martijn G., 2007. Modeling CLV: A test of competing models in the insurance industry. In: *Quantitative Marketing and Economics*, Vol. 5, vydání 2, s. 163-190.

FADER, Peter, 2012. *Customer Centricity: Focus on the Right Customers for Strategic Advantage* (Wharton Executive Essentials). Philadelphia: Wharton Digital Press. ISBN 1613630166.

FADER, P., HARDIE, B., a LEE, K., 2005. *A Note on Implementing the Pareto/NBD Model in MATLAB*. Dostupné z http://www.brucehardie.com/notes/008/pareto_nbd_MATLAB.pdf

GUPTA et al., 2006. Modeling Customer Lifetime Value. *Journal of Service Research* [online]. roč. 9, č. 2, s. 139–155 ISSN 1094-6705. Dostupné z: doi:10.1177/1094670506293810

HALLBERG, G. a OGILVY, D., 1995. *All Consumers are Not Created Equal: The Differential Marketing Strategy for Brand Loyalty and Profits*. 1. vydání. Wiley, 1995. str. 336. ISBN-13: 978-0471120049.

CHEN, Zhan a DUBINSKY, Alan, 2003. A conceptual model of perceived customer value in e-commerce: A preliminary investigation. In: *Psychology & Marketing*. Vol. 20, vydání 4, s. 323 - 347. Wiley Online Library [online]. [7. duben 2013]. Dostupné z: http://onlinelibrary.wiley.com/doi/10.1002/mar.10076/abstract

JAŠEK, Pavel, 2015. Advances in Performance Management Using Customer Equity. In: *IDIMT-2015 (Information Technology - Human Values, Innovation and Economy), Poděbrady*, 09.09.2015 – 11.09.2015. Linz: Trauner Verlag.

KARLÍČEK a kolektiv, 2014. Role marketingu ve firmách. 1. vyd. Zlín: Radim Bačuvčík – VeRBuM, 2014. 106 s. ISBN 978-80-87500-56-9.

KUMAR, V. a GEORGE, Morris, 2007. Measuring and Maximizing Customer Equity: a Critical Analysis. In: *Journal of the Academy of Marketing Science*. Vyd. 35, č. 2, s. 157 – 171. Dostupné z: doi:10.1007/s11747-007-0028-2.

LEWIS, Michael, 2006. Customer acquisition promotions and customer asset value. *Journal of Marketing Research*. Vol. 63, s. 195–203.

LEWIS, Michael, 2015. Incorporating dynamics in customer lifetime value models. In: *Handbook of Customer equity,* V. Kumar and Denish Shah eds. Edward-Elgar Publishing, MA. ISBN: 978-1781004975.

HOGAN, J. E. a RUST, R. T., 2002. Customer Equity Management: Charting New Directions for the Future of Marketing. *Journal of Service Research* [online]. roč. 5, č. 1, s. 4–12 ISSN 1094-6705. Dostupné z: doi:10.1177/1094670502005001002

MACHANDER, Aleš, 2012. *Aplikace CVM v telekomunikační společnosti*. Diplomová práce. Praha: Vysoká škola ekonomická, 2012. 120 str.

NOVOTNÝ, Ota a JAŠEK, Pavel, 2013. Enhancing Customer Lifetime Value with Perceptual Measures Contained in Enterprise Information Systems. In: *7th International Conference on Research and Practical Issues of Enterprise Information Systems*. Linz: Trauner Verlag.

PAUWEE, Peter, van der PUTTEN, Peter a van WEZEL, Michiel, 2007. DTMC: An Actionable e-Customer Lifetime Value Model Based on Markov Chains and Decision Trees. In: *Proceedings of the ninth international conference on Electronic commerce*, s. 253 – 262. ISBN: 978-1-59593-700-1.

ROSSET, Saharon et al., 2002. Customer Lifetime Value Modeling and its use for Customer Retention Planning. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, s. 332 – 340. ISBN:1-58113-567-X.

ŠÍDLO, Petr, 2002. *Hodnota zákazníka a její řízení ve finančnictví*. IT System, 5/2002. ISSN: 1802-615X. Dostupné z: http://www.systemonline.cz/clanky/hodnota-zakaznika-a-jeji-rizeni-ve-financnictvi.htm

VRANÁ, Lenka a JAŠEK, Pavel, 2015. Persistence Models for Customer Equity. In: *The 9th International Days of Statistics and Economics, Prague, September 10-12, 2015*.

WIESEL, Thorsten, SKIERA, Bernd a VILLANUEVA, Julián, 2008. Customer Equity: An Integral Part of Financial Reporting. In: *Journal of Marketing*: Vol. 72, 2, s. 1-14.

WÜBBEN, Markus a von WANGENHEIM, Florian, 2008. Instant Customer Base Analysis: Managerial Heuristics Often „Get it Right“. In: *Journal of Marketing*, Vol. 72, s. 82 – 93. ISSN: 0022-2429.

JEL Classification: M21, C52, M31

Summary

Application of Customer Lifetime Value Model in Performance Management of E-commerce

The paper summarizes application of Customer Lifetime Value Models (CLV) in performance management of retail companies. Specifics of such models are discussed, as well as differences from other industries that already benefit from CLV models. Focus is made on practical application in operative and strategic performance management and marketing management in particular. Various impacts on information technologies, such as data requirements and data integration processes are analyzed from the perspective of possible extension of these models towards nonfinancial factors for customer value.

Retail companies differ from traditional use of CLV for telecommunication and financial industries mainly due to non-contractual business settings. For that reason, specific statistical models and methods for predicting and evaluation of the prediction of CLV are required. A methodology that is being created within the doctoral dissertation thesis of the author aims at clarifying this customer-centric concept and provide retail industry with practical processes, methods and techniques on both operative and strategic level.

Keywords: Customer Lifetime Value, Customer Relationship Management, Corporate Performance Management, Marketing Management, Customer Equity

Aplikace principů COBIT 5 v prostředí cloud computingu

Soňa Karkošková

xkars05@vse.cz

Doktorand oboru Aplikovaná informatika

Školitel: doc. Jiří Feuerlicht, Ph.D., (jiri.feuerlicht@vse.cz)

Abstrakt: Governance podnikových informačních technologií se rozvinula do globálně uznávaného rámce COBIT 5 publikovaného asociací Information Systems Audit and Control Association (ISACA). COBIT 5 definuje dobré praktiky pro efektivní governance a management podnikového IT. Cloud computing je stále se rozvíjející oblast poskytování IT ve formě služeb přes Internet. Cloud přináší řadu benefitů, ale také nových výzev, rizik a úkolů, které jsou patrné zejména v oblasti IT Governance. Tato práce se soustředí na aplikaci principů COBIT 5 v prostředí cloud computingu.

Klíčová slova: COBIT 5, Cloud computing, IT governance, IT management.

Úvod

Současné obchodní prostředí se vyznačuje vysokou komplexností a tlakem na organizace, aby byly schopné rychle a pružně reagovat na měnící se podmínky a požadavky na trzích. Pro udržení si konkurenceschopného postavení je nutné mít implementovány takové metody, procesy a politiky, které umožní flexibilně realizovat či měnit obchodní procesy s ohledem na zabezpečení efektivního dosažení podnikových cílů. Podpora obchodních procesů by byla prakticky neudržitelná bez přímé podpory informačních technologií, které se během svého vývoje staly integrální součástí podniku. Aby IT poskytovalo požadovanou hodnotu nejen organizaci samotné, ale i zákazníkům a dalším zainteresovaným stranám, musí být vyvažován finanční i nefinanční přínos IT s riziky, které s ním souvisí. S rozvojem řízení podnikového IT vznikala řada rámců a doporučení v oblasti IT Governance jako je Control Objectives for Information and Related Technology (COBIT), Information Technology Infrastructure Library (ITIL), ISO 38500 pro podnikové IT governance a řada dalších (Jäntti, a další, 2015). Důležitost IT Governance jako podmnožiny podnikového governance se soustředí zejména na oblast řízení výkonnosti a rizik informačních technologií a snaží se zajistit, aby IT produkovalo očekávanou obchodní hodnotu při optimalizaci nákladů a rizik (Orozco, a další, 2015). Vzhledem k rostoucí závislosti úspešnosti podnikání na informačních technologiích poskytovaných interními či externími dodavateli a nárůstu objemu informací generovaných podnikem nutných pro podporu podnikání, je zavedení některého z rámců pro řízení IT nezbytným krokem. Jednou z nových a stále se rozvíjejících oblastí informačních technologií je poskytování IT zdrojů ve formě služeb přes síť (Wattal, a další, 2014). Cloud computing s sebou nese určité výhody, ale pojí se s ním řada oblastí, které vyžadují pečlivé zvážení jako například bezpečnost, dostupnost dat, fyzické umístnění dat a jejich přenositelnost, vlastnictví dat, legislativní požadavky a realizovatelnost auditu či životaschopnost poskytovatele (Vitti, a další, 2014), (Rao, a další, 2015), (Bailey, a další, 2014), (Ravi, a další, 2015). Rizika spojená s používáním služeb v prostředí cloudu se mohou odlišovat v závislosti na modelu služeb a modelu nasazení. Governance se v prostředí cloud computingu musí přizpůsobit specifickým charakteristikám, aby služby poskytovaly požadovaný přínos v souladu s podnikovými cíli. Stejně tak jako tradiční IT governance, tak i governance pro prostředí cloudu vyžaduje definování cílů, politik, rolí a odpovědností v závislosti na zralosti a komplexitě organizace, neboť využití cloudových služeb ovlivňuje obchodní procesy kritické pro dosažení podnikových cílů (Bailey, a další, 2014). Součástí cloud governance by mělo být nejen řízení rizik, ale i zajištění optimálního využití a řízení různých forem IT poskytovaných v různých formách a to ať externě, tak interně (Bailey, a další, 2014). Jedním z široce akceptovaných rámců IT governance je COBIT 5 (Youssfi, a další, 2014). COBIT 5 sjednocuje dobré praktiky strategického a operativního řízení IT do podoby komplexního rámce governance a managementu podnikových informačních technologií s cílem dosáhnout strategických obchodních cílů při optimálním využití IT zdrojů (ISACA, 2012). Ačkoliv COBIT 5 v zásadě není určen pro cloud computing, praktiky, principy a procesy v něm obsažené lze přizpůsobit také pro řízení služeb cloud computingu.

V této práci diskutuji COBIT 5 a jeho praktickou aplikovatelnost v prostředí cloud computingu a to jak oblast IT governance, tak i oblast IT managementu.

Související práce

Aby přínosy z informačních technologií převážily náklady s nimi spojené, musí být v organizaci implementovaná jasně definovaná governance a takové řídící praktiky, které umožnují sladit IT s obchodními cíli (Jäntti, a další, 2015). Existují autoři, kteří se zabývají přizpůsobením tradičních IT Governance rámců jako jsou COBIT, ITIL nebo ISO pro prostředí cloud computingu (Bailey, a další, 2014) (Stanley, 2014). Bailey (Bailey, a další, 2014) se domnívá, že aplikace IT governance principů pro řízení cloudových služeb, je nezbytná pro jejich optimální adaptaci v podniku a podporu podnikových procesů, řízení rizik, dosažení podnikových cílů a dosažení očekávané hodnoty. Bailey představuje IT Cloud Governance Dial, který identifikuje pět oblastí IT Governance důležité při adopci cloudu, jejichž výstupy jsou aplikovatelné jako základ pro IT governance rámec modifikovaný pro prostředí cloudu.

Iniciativou cloud computingu je zabývá řada velkých organizací, které vytvořily a vydaly standardy a doporučení pro oblast cloud computingu a to zejména National Institute of Standards and Technology (NIST, 2013), konsorcium the Open Goup (the Open Group, 2013), Cloud Standards Customer Council (Cloud standards customer council, 2015) a další. Řada vědeckých publikací věnující se rámci COBIT je však postavena na analýze a implementaci jeho předchozí verze 4.1 nebo se vůbec nevěnují aplikaci COBIT 5 v prostředí cloud computingu. Kromě příručky Controls and Assurance in the Cloud: Using COBIT 5 je nedostatek vědeckých článků a publikací, které by se věnovaly synergii rámce COBIT 5 a cloud computingu. Z tohoto důvodu je tento článek založen hlavně na publikacích asociace ISACA a to COBIT 5: A business framework for Governance and Management of enterprise IT (ISACA, 2012), COBIT 5 Enabling Processes (ISACA, 2012), webové stránky COBIT 5 (ISACA, 2014), (ISACA, 2015) a Controls & Assurance in the Cloud: Using COBIT 5 (ISACA, 2014).

Aplikace COBIT 5 v prostředí cloud computingu

Mezinárodní asociace Information Systems Audit and Control Association (ISACA) vydala příručku Controls and Assurance in the Cloud: Using COBIT 5 obsahující praktické kroky k realizaci auditu, kontroly a řízení rizik a zajištění informační bezpečnosti v cloudu postavené na využití COBIT 5 (ISACA, 2014). Příručka poskytuje následující oblasti pro efektivní správu a řízení rizikových otázek v prostředí cloud computingu (ISACA, 2014):

- governance a management,
- otázky bezpečnosti a zajištění,
- posuzování způsobilosti procesů,
- scénáře rizik cloudu a kontrolní seznam podnikového cloud risk managementu,
- praktický přístup k měření ROI v prostředí cloud computingu.

Z hlediska sedmi COBIT 5 enablerů je zapotřebí se v prostředí cloud computingu věnovat řešení následujících otázek a úkolů (ISACA, 2014):

- Enabler principů, politik a rámců
    - Oblast bezpečnostní politiky a její transparentnosti. Zde je nutno společně s poskytovatelem cloudových služeb dojít k dohodě obsahující záruky, že poskytovatel aplikuje bezpečnostní mechanismy na takové úrovni, které jsou vyžadovány spotřebitelem. Tyto bezpečnostní mechanismy a jejich úrovně musí být definovány v SLA.
    - Oblast shody s legislativními požadavky. Patří sem nutnost vyřešení požadavků na dostupnost metod zabezpečujících legislativní ochranu citlivých dat, jako jsou osobní údaje, údaje z platebních médií, finanční reporty nebo geografické uložení dat. Současně je potřeba mít k dispozici data pro provedení auditu.
- Enabler procesů
    - Oblast řízení bezpečnosti. Je zapotřebí prověřit, zda poskytovatel má implementovány procesy pro detekci bezpečnostních narušení jako je například detekce vniknutí neau-

torizovaného uživatele do systému. Poskytovatel musí provádět bezpečnostní logování aktivit cloudu, které musí být dostupné spotřebiteli pro potřeby auditu.

- Enabler organizačních struktur
    - Oblast opatrnosti a pečlivosti s jakou vlastník serveru přistupuje k ochraně dat spotřebitele. Jedná se zejména o model veřejného cloudu a uložení citlivých dat v cloudovém prostředí. Zde je nutno získat od poskytovatele informace o všech s ním spolupracujících externích firmách, které by se potenciálně při své činnosti u poskytovatele mohly dostat do kontaktu s těmito citlivými daty spotřebitele. Spotřebitel musí mít možnost se rozhodnout, zda bude těmto externím firmám důvěřovat. Spotřebitel musí mít v tomto případě mít k dispozici informace, kde jsou jeho citlivá data fyzicky uložena.
- Enabler kultury, etiky a chování
    - Oblast životaschopnosti poskytovatele. Je nutno zvážit rizika spojená s možným ukončením činnosti poskytovatele. Je nutno provést analýzu rizik a rozhodnout se zda a která data mohou být u daného poskytovatele uložena a jaké náklady budou spojené s případným nouzovým transferem dat a zda bude takový transfer realizovatelný.
    - Oblast prověřování ostatních spotřebitelů. Je vhodné mít údaje o tom, kteří další spotřebitelé cloudových služeb sdílejí stejné cloudové servery nebo datová úložiště. Spotřebitel by měl prověřit reputaci těchto spotřebitelů.
- Enabler informací
    - Oblast vlastnictví dat. V některých smlouvách o využití cloudových služeb může být dohodnuto, že vlastníkem dat uložených v cloudovém prostředí je poskytovatel služeb. Poskytovatel může v těchto případech ve smlouvě vyžadovat definovat poplatek za převedení dat zpět ke spotřebiteli v případě ukončení využívání služby.
    - Oblast ochrany záznamů pro forenzní audity. Spotřebitel musí zvážit dostupnost dat a záznamů, tak aby byly případně k dispozici pro forenzní audit. Data spotřebitele jsou obecně uložená na konkrétním serveru společně s daty jiných spotřebitelů a migrována mezi různými servery často velmi geograficky vzdálenými. Může tedy nastat situace, že pro konkrétní časový interval data pro audit nebudou k dispozici. Může se případně i stát, že místní úřad zabaví server za účelem soudního zajištění dat některého ze spotřebitelů i s daty všech ostatních spotřebitelů, kteří sdílí stejný server.
    - Oblast nakládání s daty. Jedním z důsledků dynamické škálovatelnosti, kdy aplikace a datové soubory jsou přeneseny z jednoho serveru na druhý, může být to, že původní aplikace a jejich data zůstanou na originálním datovém serveru a nesmí být ihned vymazána. Například bezpečnostní logy, které jsou spojeny s danou aplikací a daným serverem mohou být vymazány až po jejich uložení do zálohovacího systému pro účely budoucího bezpečnostního auditu. Nicméně aplikace a data spotřebitele musí být po provedené záloze a v časovém intervalu daném v SLA spolehlivě vymazána. To dává spotřebiteli jistotu, že jeho citlivá data se nedostanou k nepovolaným osobám. Stejně tak musí být ve smlouvě jasně definováno, že po ukončení smluvního vztahu s poskytovatelem budou data spotřebitele okamžitě zničena.
- Enabler služeb, infrastruktury a aplikací
    - Oblast fyzického uložení dat. Z technologického hlediska, data mohou být fyzicky uložena na kterémkoliv serveru poskytovatele. Toto uložení se může dynamicky měnit. Mohou však existovat legislativní nebo data governance požadavky vyžadující aby určitá data spotřebitele byla uložena v dané geografické lokalitě. Tato lokalita může být definována jako země nebo určitá oblast území, jako datové centrum nebo dokonce jako server.
    - Oblast smíšení dat. Pokud více spotřebitelů sdílí stejnou aplikaci na stejném serveru, pak jednotlivé datové soubory této aplikace mohou obecně obsahovat data více než jednoho spotřebitele. V případě, že spotřebitel vyžaduje v SLA, že jeho data nesmí být smíšená s daty jiných spotřebitelů, musí poskytovatel pomocí vhodné technologie zabezpečit, že jednotlivé datové soubory obsahují data pouze daného spotřebitele. V některých případech může být tato separace provedena tak, že data spotřebitelů jsou kódována a každý se spotřebitelů používá jiný kód.

  - Oblast řízení identit a přístupu. Při jakékoliv manipulaci s daty spotřebitele pomocí některé z aplikací musí být vždy zřejmé, který fyzický uživatel s těmito daty manipuloval. A to ať ve smyslu vytvoření změny nebo výmazu dat či pouhého prohlížení dat. To znamená, že každý uživatel musí být jednoznačně identifikován a na základě této identifikace mu pro dané aplikace musí být přiřazena role, která umožní tomuto uživateli pouze takové datové manipulace vyplývající z jeho pracovního zařazení ve společnosti spotřebitele nebo některé jeho partnerské společnosti.
  - Oblast disaster recovery. Scénáře zálohování a havarijní obnovy dat spotřebitele jsou v prostředí cloud poskytovány jako služby. Poskytovatel služby však nemusí být tím, kdo tyto služby skutečně realizuje, neboť tyto služby mohou být outsourcovány na třetí strany. V některých případech mohou tyto třetí strany fyzické provádění dat znovu outsourcovat. V prostředí veřejného cloudu je zapotřebí, aby ve smlouvě byly otázky testování, obnovitelnosti dat nebo požadavky na časové lhůty obnovení dat detailně definovány.
- Enabler lidí, zkušeností a kompetencí
  - Oblast závislosti na proprietárním API poskytovatele. Pokud spotřebitel využívá takové služby, ke kterým je přístup možný pouze proprietárního API poskytovatele, je zapotřebí zvážit riziko v případě nutnosti změny poskytovatele. Náklady na převod dat z aplikace s proprietárním API jsou obvykle mnohem vyšší než převod dat z prostředí standardních neproprietárních API.

Governance a management v prostředí cloud computingu

Rozhodnutí podniku, že bude využívat cloudové služby, kterými nahradí některé ze služeb IT, má vždy vliv na s těmito službami spojené obchodní procesy. Governance a management procesy musí být přizpůsobeny novým podmínkám tak, aby po přechodu na cloudové služby byla zajištěna kontinuita výkonnosti a požadavků na shodu v závislosti na platných podnikových směrnicích a cílech vycházejících z potřeb zainteresovaných stran. Hodnota, kterou má přinést využití cloudových služeb musí být jasně definovaná, akceptovaná a měřitelná. To napomáhá určit priority ohledně využití cloudových služeb a stanovit metriky pro měření odchylek skutečně dosažené hodnoty od plánované hodnoty. Vzájemná souhra mezi cíli adopce cloudu a podnikovými cíli je rozhodující pro efektivní řízení rizik cloudu a pro očekávanou redukci nákladů. Prostředí cloudu přináší do podnikových procesů nová rizika, u nichž musí podnik vždy provádět analýzy, zda příslušné riziko je vždy v akceptovatelných hranicích.

Důvody pro adopci cloudových technologií jsou obecně (ISACA, 2014):

- získání konkurenční výhody,
- proniknutí na nové trhy,
- zvýšení úrovně existujících produktů a služeb,
- udržení si stávajících zákazníků.
- zvýšení produktivity,
- redukce nákladů,
- vývoj produktů a služeb, jejichž realizace je možná pouze v cloudovém prostředí,
- překonání geografických bariér.

Dalším důležitým důvodem je obava ze ztrát, které by mohly postihnout podnik, pokud by konkurenční podnik akceptoval cloudové prostředí mnohem dříve a rychleji. Přestože důvody pro zavedení cloudových služeb jsou zřejmé a široce uznávané, má adopce cloudu obvykle za následek vznik konfliktních situací. Dochází k narušení stávající podnikové kultury, kdy zaměstnanci společnosti ještě nejsou dostatečně vyškoleni v oblasti cloudových řešení. Dále může docházet ke konfliktům mezi cloudově orientovanými procesy a klasickými stávajícími procesy a ani organizační struktura ještě není schopná dostáhnout maximálních efektivnosti z využití cloudových služeb. Současně se zaváděním cloudových služeb nebo s přechodem na cloudové služby je potřeba systematicky vyhodnocovat připravenost podniku na adopci cloudového prostředí. Cílem je jednak odstranění konfliktních oblastí a přesnější ohodnocení nákladů a rizik spojených s přechodem na cloudové služby. Proces vyhodnocování připravenosti by měl obsahovat (ISACA, 2014):

- nové politiky a procedury, které popisují adopci, řízení a z hlediska podniku správné využití cloudového prostředí, a zejména přijetí procesů týkajících se rozšiřování stávajícího portfolia služeb a změnového řízení,
- reingeneering existujících procesů užívající tradiční IT služby, tak aby byly schopny akceptovat nové aktivity spojené s užíváním cloudových služeb,
- transformace stávající organizační struktury tak, aby byla podporovat nové prostředí cloudu,
- vyškolení zaměstnanců na nové prostředí k zajištění potřebné podnikové kultury a chování, tak aby pochopili důvody přechodu na nová podniková řešení a pro úspěšnou adopci těchto řešení,
- odborné školení zaměstnanců, aby byli schopni porozumět a plně podporovat nové cloudové prostředí na všech potřebných úrovních.

Přechod na cloudové služby se samozřejmě musí odrazit v potřebě doplnění stávajících procesů governance a managementu podnikového IT, tak aby byly schopny efektivně působit i v podmínkách cloudu. Týká se to zejména potřeby reakce na zvýšené riziko v prostředí cloudových služeb zahrnující otázky bezpečnosti, shody s platnou legislativou, ochrany soukromí, projektovou činností a spolupráce s partnery. Dále je zapotřebí zajistit kontinuitu kritických obchodních procesů, jejichž data se v prostředí cloud computingu mohou rozšířit za hranice stávajícího datového centra. Je zapotřebí modifikovat procesy související se změnovým řízením, neboť flexibilita a škálovatelnost cloudového prostředí vyžaduje nové přístupy k této problematice. V prostředí cloudu a s tím související legislativní požadavky si vyžadují vytvoření nových procesů pro jejich splnění.

Rozšíření IT governance o principy cloud governance umožnuje získat hodnotu z adopce služeb cloudu prostřednictvím realizace přínosů cloudových služeb definovaných a odsouhlasených před započetím migrace do cloudu, řízení rizik během využívání cloudových služeb a zabezpečení optimálního využití a řízení zdrojů a to jak interních, tak i outsourcovaných od externích poskytovatelů. Cloud computing umožnuje efektivnější realizaci a podporu operativních činností a přináší nové obchodní příležitosti. Rizika spojená s využíváním cloudových služeb bez governance procesů mohou být značně vysoká a je potřeba je cíleně řídit, jinak by ve své podstatě mohly vést k poklesu flexibility obchodních procesů a tím ke snížení obchodní agility. COBIT 5 rozlišuje mezi následujícími třemi IT kategoriemi rizik (ISACA, 2014):

- Riziko realizace IT benefitů. Tato kategorie je interpretována jako riziko ztráty příležitosti využití technologií ke zvýšení efektivnosti a účinnosti obchodních procesů nebo jako enabler nové obchodní iniciativy.
- Riziko IT programu nebo projektu. Tato kategorie je spojena s přínosem IT k novému nebo vylepšenému obchodnímu řešení realizovaného obvykle ve formě projektu nebo programu jako součást investičního portfolia.
- Riziko fungování IT a dodávek služeb. Tato kategorie je spojena se všemi aspekty obchodních činností podniku a to jako obvyklá výkonnost IT systému a služeb, která se může ve svém důsledku projevit silným poškozením hodnoty podniku

COBIT 5 poskytuje soubor metrik pro monitorování a řízení aktivit souvisejících s nasazením a využíváním cloudových služeb, který je postavený na procesních praktikách, detailním popisu vztahů mezi vstupy a výstupy jednotlivých procesů, RACI diagramu a na kvantifikaci capability modelu hodnocení specifických procesů (ISACA, 2014).

V prostředí COBIT 5 vychází cloud governance opět z definované kaskády cílů, na jejímž vrcholu stojí potřeby zainteresovaných stran mapující se do třinácti podnikových cílů, které jsou následně mapovány do třinácti IT cílů a to včetně cloudových služeb. Ty se následně mapují do cílů enablerů. Potřeby zainteresovaných stran jsou ovlivněny například změnami obchodního prostředí, legislativními požadavky či novými technologiemi. Podnikové a IT cíle jsou rozčleněny do čtyř oblastí finanční dimenze, dimenze zákazníků, interní dimenze a dimenze učení se a růst, které reflektují dimenze dle balanced scorecard.

NIST pro prostředí cloud computingu definuje pět zainteresovaných stran, které se účastní na transakcích, realizaci procesů či plní související úkoly a identifikuje jejich vzájemné působení (NIST, 2013). Těmito zainteresovanými stranami jsou cloud spotřebitel, cloud poskytovatel, cloud auditor, cloud

broker a cloud zprostředkovatel. V prostředí COBIT 5 je cloud computing považován jako rozšíření IT aktiv a tudíž musí být identifikovány všechny relevantní zainteresované strany, jichž se přijetí cloud služeb dotýká. Pro detailní analýzu a identifikaci potřeb zainteresovaných stran v prostředí cloud computingu jsou zásadní následující COBIT 5 procesy a k nim vztahující se procesní praktiky (ISACA, 2014):

- EDM02 Zajištění realizace přínosů (Ensure Benefits Delivery)
  - EDM02.02 Evaluate the governance system – podpora vrcholového managementu, přizpůsobení řídící struktury, zavedení procesů a praktik, přidělení pravomocí a odpovědností, zajištění podnikové kultury podporující přijetí nových informačních technologií včetně vyškolení zaměstnanců, zabezpečení potřebných a relevantních informací pro podporu rozhodování
- APO02 Řízení strategie (Manage Strategy)
  - APO02.02 Assess the current environment capabilities and performance – porozumění podnikové architektury v souvislosti s IT, posouzení výkonnosti interních IT capabilit a externích IT služeb a zhodnocení poskytovatelů služeb včetně finančního dopadu, potenciálních nákladů a přínosů využívání externích služeb
  - APO02.03 Define the target IT capabilities – definování požadovaných IT capabilit a IT služeb na základě zhodnocení současného IT prostředí a moderních informačních technologií v souladu s podnikovou architekturou
  - APO02.04 Conduct a gap analysis – identifikace rozdílů mezi aktuálním a požadovaným stavem IT capabilit a IT služeb, zvážení kritických faktorů úspěchu provádění strategie a upřesnění definice požadovaných IT capabilit a IT služeb včetně formulace potenciálně dosažitelné honoty a benefitů
  - APO02.05 Define the strategic plan and road map – vytvoření strategického plánu pro dosažení požadovaných IT capabilit a IT služeb obsahující definici IT cílů, formulace investičních programů do IT včetně způsobu získávání zdrojů, financování, identifikace souvisejících rizik, nákladů, organizačních změn a regulatorních požadavků, vytvoření projektového plánu na nižší úrovni detailu a převedení cílů do měřitelných ukazatelů
- APO05 Řízení portfolia (Manage Portfolio)
  - APO05.01 Establish the target investment mix – identifikace nákladů, finančních a nefinančních přínosů investic do IT, očekávané návratnosti investic v průběhu celého ekonomického životního cyklu podnikového IT
  - APO05.03 Evaluate and select programmes to fund – hodnocení a prioritizace investičních programů, vyčlenění finančních prostředků a iniciace zvolených programů
- APO06 Řízení rozpočtů a nákladů (Manage Budget and Costs)
  - APO06.01 Manage finance and accounting – zavedení jednotných finančních a účetních standardů
  - APO06.02 Prioritise resource allocation – implementace rozhodovacího procesu prioritizace IT zdrojů a pravidel rozhodování o způsobu alokace finančních prostředků
  - APO06.03 Create and maintain budgets – vytvoření rozpočtů obsahujících očekávané náklady na IT v závislosti na portfoliu investičních programů do IT
  - APO06.04 Model and allocate costs – kategorizace nákladů na IT, identifikace platebních modelů služeb a vytvoření transparentního modelu nákladů na IT umožňujícího identifikovat skutečné využití služeb a s nimi spojených poplatků a předpovídat náklady na IT
  - APO06.05 Manage costs – zavedení procesu řízení nákladů na IT, jejich monitorování, identifikace odchylek a jejich dopadů na obchodní procesy
- APO12 Řízení rizik (Manage Risk)
  - APO12.03 Maintain a risk profile – správa registru rizik obsahujícího popis jednotlivých rizik, očekávané frekvence jejich výskotu, potenciální dopad a reakce na riziko, identifikace IT zdrojů a IT služeb nezbytných pro zajištění požadované realizace obchodních procesů a jejich omezení

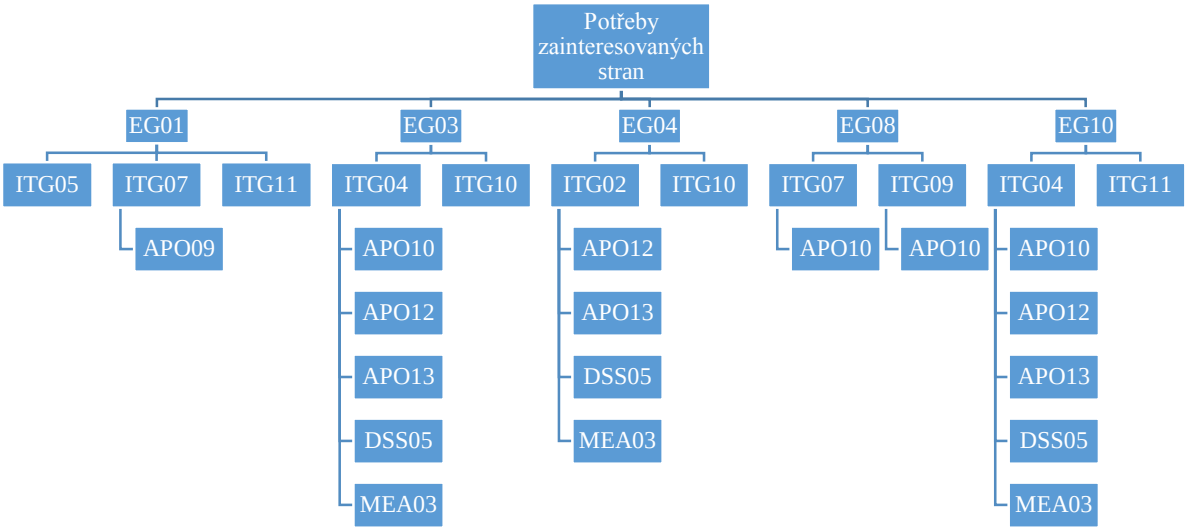
- BAI01 Řízení programů a projetů (Manage Programmes and Projects)
    - BAI01.02 Initiate a programme – zahájení programu cílem potvrdit očekávané přínosy přijetí cloud služeb
    - BAI01.03 Manage stakeholder engagement – identifikace, analýza, zapojení a řízení zainteresovaných stran, analýza zájmů a požadavků zainteresovaných stran, zabezpečení aktivní výměny přesných a konzistentních informací mezi zainteresovanými stranami ve správný čas
    - BAI01.04 Develop and maintain the programme plan – definice a zdokumentování všech projektů z hlediska jejich rozsahu, vstupů, výstupů, dosažení cílů a hodnoty IT investic pro zainteresované strany, zajištění souladu se strategickými cíli podniku
- BAI03 Řízení identifikace a budování řešení (Manage Solutions Identification and Build)
    - BAI03.01 Design high-level solutions – navržení IT řešení, které budou schopny na vysoké úrovni zabezpečit podporu obchodních procesů a dosažení podnikových cílů v souladu s podnikovou architekturou, IT strategií, byznys strategií a regulatorními požadavky
- MEA03 Monitorování, hodnocení a posouzení shody s externími požadavky (Monitor, Evaluate and Assess Compliance With External Requirements)
    - MEA03.01 Identify external compliance requirements – identifikace a monitorování změn legislativních, regulatorních a dalších externích požadavků ve shodě s IT
    - MEA03.02 Optimise response to external requirements – pravidelně přezkoumávat a upravovat politiky, principy, standardy, procedury a metodiky ve shodě s legislativními, regulatorními a dalšími externími požadavky, proškolení relevantních zaměstnanců
    - MEA03.03 Confirm external compliance – potvrzení shody politik, principů, standardů, procedur a metodik s legislativními, regulatorními a dalšími externími požadavky
    - MEA03.04 Obtain assurance of external compliance – realizovat pravidelné přezkoumání o souladu politik, principů, standardů, procedur a metodik a posouzení míry shody, prohlášní poskytovatelů IT služeb třetích stran o úrovni jejich souladu s platnými zákony a předpisy

Mezi podnikové cíle a s nimi související IT cíle obzvláště důležité pro prostředí cloud computingu patří (ISACA, 2014):

- EG01 Hodnota obchodních investic pro zainteresované strany (Stakeholder value of business investments)
    - ITG05 Realizace benefitů z investic do IT a z portfolia služeb (Realised benefits from IT-enabled investments and services portfolio)
    - ITG07 Zajištění IT služeb v souladu s obchodními požadavky (Delivery of IT services in line with business requirements)
    - ITG11 Optimalizace IT aktiv, IT zdrojů a IT capabilit (Optimisation of IT assets, resources and capabilities)
- EG03 Řízení obchodních rizik (Managed business risk)
    - ITG04 Řízení rizik vyplývající z využití IT (Managed IT-related business risk)
    - ITG10 Bezpečnost inforamcí a bezpečnost provozu infrastruktury a aplikací (Security of information, processing infrastructure and applications)
- EG04 Shoda s legislativními předpisy a oborovými standardy (Compliance with external laws and regulations)
    - ITG02 Shoda IT s externími právními předpisy a regulacemi a podopra IT pro zajištění shody byznysu s externími právními předpisy (IT compliance and support for business compliance with external laws and regulations)
    - ITG10 Bezpečnost inforamcí a bezpečnost provozu infrastruktury a aplikací (Security of information, processing infrastructure and applications)
- EG08 Obchodní agilita (Agile responses to a changing business environment)
    - ITG07 Zajištění IT služeb v souladu s obchodními požadavky (Delivery of IT services in line with business requirements)

  - ITG09 IT agilita (IT agility)
- EG10 Optimalizace nákladů IT služeb (Optimisation of service delivery costs)
  - ITG04 Řízení rizik vyplývající z využití IT (Managed IT-related business risk)
  - ITG11 Optimalizace IT aktiv, IT zdrojů a IT capabilit (Optimisation of IT assets, resources and capabilities)

Pro prostředí cloud computingu se k výše uvedeným IT cílům vztahují procesy znázorněné na Obrázku 1 (ISACA, 2014):



Obrázek 1: COBIT 5 procesy vztahující se k IT cílům v prostředí cloud computingu. Zdroj: (ISACA, 2014)

Proces APO09 Řízení dohod o službách (Manage Service Agreements) zabezpečuje efektivní využití IT služeb definovaných v katalogu služeb, jejich identifikaci, specifikaci, návrh, publikování a monitorování úrovně služeb a ukazatele výkonnosti. Na základě informací obsažených v katalogu služeb napomáhá definovat a připravit dohody o poskytování služeb, provádět pravidelné přezkoumávání dohod o službách a jejich revizi v případě potřeby. Proces APO10 Řízení dodavatelů (Manage Suppliers) zabezpečuje, aby služby od všech dodavatelů splňovaly požadavky podniku, výběr dodavatelů, řízení vztahů s dodavateli včetně řízení rizik, řízení smluv, monitorování výkonu dodavatele a dodržování poskytování dohodnuté úrovně služeb. Cílem procesu APO12 Řízení rizik (Manage Risk) je průběžně identifikovat, posoudit a snížit rizika související s IT v rámci úrovně tolerance stanovených výkonným vedením společnosti. Cílem procesu APO13 Řízení bezpečnosti (Manage Security) je definovat, provozovat a monitorovat systém pro řízení bezpečnosti informací. Proces DSS05 Řízení bezpečnosti služeb (Manage Security Services) obsahuje praktiky vztahující se k zajištění síťové a komunikační bezpečnosti, řízení a správě uživatelských identit a přístupových oprávnění a to tak, aby bezpečnostní riziko bylo pro podnik přijatelné a v souladu s bezpečnostní politikou. Proces MEA03 Monitorování, hodnocení a posouzení shody s externími požadavky (Monitor, Evaluate and Assess Compliance With External Requirements) zabezpečuje, aby legislativní, regulatorní a další externí požadavky byly ve shodě s IT a umožňuje realizovat pravidelné přezkoumání interních politik, principů, standardů, procedur a metodik a posouzení míry jejich shody s legislativními požadavky.

Závěr

Informační technologie jsou pro podnik důležitým aktivem zejména ze strategického hlediska. Přijetí cloud computingových služeb spolu se zavedením procesů cloud governance může zvýšit hodnotu, kterou tato aktiva podniku přináší. Správné přijetí cloudových služeb vyžaduje stanovení řídících procesů a kontrolních metrik, které zabezpečí optimální vyvážení přínosů a rizik spojených s využitím cloud služeb. Přijetí těchto služeb výrazně ovlivňuje fungování podniku a to nejen z hlediska přijetí změn v organizační struktuře ale také v nastavení nových politik a podnikové kultury, tak aby podporovala adopci nových technologických řešení. Adopce cloud služeb je spojená také s nutností redefinovat stávající obchodní procesy, aby byly plně kompatibilní s cloudovým prostředím a zabezpečo-

valy dosažení potřeb zainteresovaných stran. Potřeby zainteresovaných stran jsou dle kaskády cílů definované v COBIT 5 mapovány do celopodnikových cílů. COBIT 5 definuje procesy, které umožňují vyhodnocovat potřeby zainteresovaných stran, změny v obchodním prostředí, výkonnost IT procesů a schopnost dosažení očekávaných výstupů cloudovými informačními technologiemi. Výsledkem je definice příležitostí spojených s adopcí cloudových služeb a zhodnocení a optimalizace hodnoty služeb IT v cloudovém prostředí oproti identifikovaným rizikům generovaným nasazením cloudového prostředí. COBIT 5 propojuje potřeby zainteresovaných stran s obchodními a IT cíli a cíli enablerů a poskytuje procesní capability model, který umožňuje hodnocení výkonnosti procesů a dosažených cílů prostřednictvím souboru měřitelných ukazatelů. COBIT 5 není vyvinut primárně pro specifické prostředí cloud computingu, ale řada principů či dobrých praktik COBIT 5 je aplikovatelná i pro cloud. COBIT 5 umožňuje realizovat governance a management IT i pro prostředí cloud computingu, kdy výchozím krokem je zhodnocení požadavků a potřeb zainteresovaných stran. COBIT 5 napomáhá organizaci implementovat opakovatelně využitelné procesy a nastavit příslušnou úroveň kontroly v závislosti na míře rizik IT a tak optimalizovat obchodní hodnotu celého podnikového IT.

Analýza aplikovatelnosti principů COBIT 5 v prostředí cloud computingu je součástí výzkumu metodických rámců IT governance a management. Tento výzkum uskutečňovaný na Katedře informačních technologií VŠE je spojen s realizací disertační práce Správa a řízení služeb cloud computingu. Další výzkum se bude týkat možností přizpůsobení a rozšíření vybraných procesů COBIT 5 pro v disertační práci navrhovaný cloud computing management, který bude obsahovat model řízení životního cyklu služeb cloud computingu z pohledu spotřebitele služeb cloud computingu.

Literatura

Bailey, Elana a Becker, Jack D. 2014. *A Comparison of IT Governance and Control Frameworks in Cloud Computing.* Savannah : Twentieth Americas Conference on Information Systems, 2014.

Cloud standards customer council. 2015. CSCC Deliverables:. *CSCC Resource Hub.* [Online] 2015. http://www.cloud-council.org/resource-hub.htm.

ISACA. 2014. About COBIT 5. *COBIT 5 an ISACA framework.* [Online] ISACA, 2014. https://cobitonline.isaca.org/about.

ISACA. 2012. *Cobit 5 Enabling Processes.* ISACA, 2012. ISBN 978-1-60420-241-0.

ISACA. 2015. COBIT 5 Publications Directory. *COBIT 5.* [Online] Information Systems Audit and Control Association, 2015. http://www.isaca.org/COBIT/Pages/Product-Family.aspx.

ISACA. 2012. *COBIT 5: A business framework for Governance and Management of enterprise IT.* United States of America : ISACA, 2012. ISBN 978-1-60420-237-3.

ISACA. 2014. *Controls & Assurance in the Cloud: Using COBIT 5.* ISACA, 2014. ISBN 978-1-60420-464-3.

Jäntti, Marko a Hotti , Virpi. 2015. *Defining the relationships between IT service management and IT service governance* . [prod.] Springer US. Information Technology and Management, 2015. ISSN: 1385-951X.

NIST. 2013. *NIST Cloud Computing Standards Roadmap.* National Institute of Standards and Technology, 2013.

Orozco, Jorge, a další. 2015. *A Framework of IS/Business Alignment Management Practices to Improve the Design of IT Governance Architectures.* [Vol. 10, No. 4; 2015] [prod.] Canadian Center of Science and Education. International Journal of Business and Management, 2015. ISSN 1833-3850.

Rajmohan, B a Balashankar, M. 2014. *The Cloud and SOA - Creating the Architecture for Today and Future.* [Volume 1, Issue 6, Dec-Jan, 2014] International Journal of Research in Engineering & Advanced Technology, 2014. ISSN: 2320 - 8791.

Rao, N.Venkateswara a Srinivasu, Dr. N. 2015. *A Review on Cloud Data Integrity Auditing Methods and Protocols.* [Volume 2 Issue 7 pp. 25 - 30] International Journal of Engineering Technology Science and Research, 2015. ISSN 2394 – 3386.

Ravi, T.N. a Sankar, Sharmila. 2015. *Measuring the Security Compliance Using Cloud Control Matrix.* [prod.] IDOSI Publications. Middle-East Journal of Scientific Research 23 (8): 1797-1803, 2015, 2015. ISSN 1990-9233.

Stanley, Ronald. 2014. *Cocreation and Implementing ITIL Service Management in the Cloud: A Case Study.* [pp 61-71] Springer Netherlands, The 8th International Conference on Knowledge Management in Organizations, 2014. ISBN: 978-94-007-7286-1.

The Open Group. 2013. Cloud Computing White Papers. *the Open Group Cloud Compuitng.* [Online] 2013. http://www.opengroup.org/cloud/whitepapers/.

Vitti, Pedro Artur Figueiredo, a další. 2014. *Current Issues in Cloud Computing Security and Management.* [pp. 36 - 42] Lisbon, Portugal : SECURWARE 2014 : The Eighth International Conference on Emerging Security Information, Systems and Technologies, 2014. ISBN: 978-1-61208-376-6.

Wattal, S, a další. 2014. *Cloud computing - An emerging trend in information technology.* [pp. 168 - 173] [prod.] IEEE. 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014. INSPEC Accession Number: 14210923 .

Youssfi, Karim, Boutahar, Jaouad a Elghazi, Elghazi. 2014. *A Tool Design of Cobit Roadmap Implementation.* [Vol. 5, No. 7, pp 86 - 94] (IJACSA) International Journal of Advanced Computer Science and Applications, 2014.

JEL Classification: M15

Summary

Application of COBIT 5 principles in the cloud computing environment

Governance of enterprise information technology has evolved into a globally recognized framework COBIT 5 published by Information Systems Audit and Control Association (ISACA). COBIT 5 defines good practices for effective governance and management of enterprise IT. Cloud computing is an emerging area of providing IT resources as a service over a network. Cloud brings many benefits, but also new challenges, risks and issues that are particularly apparent in IT Governance. This paper focuses on the application of COBIT 5 principles in the cloud computing environment.

Keywords: COBIT 5, Cloud computing, IT governance, IT management

Sociální a manažerská kybernetika v rámci analyzování organizace

Ludmila Malinová

ludmila.malinova@vse.cz

Doktorand oboru Aplikovaná informatika

Školitel: doc. Ing. Zora Říhová, CSc., (zora.rihova@vse.cz)

Abstrakt: Článek představuje výběr a shrnutí určitých vybraných metodologií a přístupů, které slouží k analýze organizace z pohledu kybernetických principů a respektování sociálního faktoru v organizaci. V úvodu článku se čtenář dozvídá o nutnosti uvažovat nad každou organizací z hlediska základních kybernetických předpokladů (komplexita, otevřenost, dynamičnost, přijímání vstupů a výstupů, je možné pozorovat zpětnovazební smyčky, možnost využívání mentálních modelů, rekurzivita atd.) a také lehce naznačuje dělení současné moderní kybernetiky. Článek se zaměřuje také na přehled historického vývoje v oblasti systémových metodologií. V další části článku je rozebrána problematika kybernetických přístupů k analýze organizace, kdy nejvýznamnější posun v chápání zaznamenala tzv. manažerská kybernetika, která byla definována S. Beerem (1972) a která je založena na modelu životaschopného systému (VSM), jde o uplatnění kybernetických principů (rekurze, adaptivní homeostáze, nebezpečí redundance potenciálních příkazů, týmová synergie) do manažerské praxe. Následně byl tento model mnohokrát upravován a dal vzniknout dalším přístupům jako například podniková či organizační kybernetika, které jsou v článku také přiblíženy.

V závěru článku je nastíněna připravovaná metodologie MALINA (Management And Labor Internal Network Analysis). Tato metodologie se zabývá návrhem oblastí v rámci analýzy organizace (informace, znalosti, komunikace, využití ICT, spokojenost zaměstnanců, atd.), které mají určitou vypovídající schopnost o současném stavu organizace. Součástí metodologie je návrh nástroje pro vyhodnocení, který obsahuje také modulární dotazník.

Klíčová slova: kybernetika, manažerská kybernetika, organizační kybernetika, systémové metodologie, metodologie MALINA

Úvod

Systémové přístupy a moderní kybernetika mají stále velký potenciál, co se týče jejich možného použití. Nejde pouze o teorie, ale o principy, modely a metodiky, které mohou být a měly by být přenášeny do organizační praxe. V případě, že se snažíme provádět v organizaci změny, měli bychom znát samotnou podstatu systému organizace a odhalit možné problémy, které hrozí při zavádění nových změn (např. nasazení nového informačního systému). Organizaci však není možné jednoduše popsat modelem, který bude vždy pravdivý a funkční a stejně tak jednoznačně daná metodika nebude fungovat při změně podmínek v organizaci. Změna může nastat, jak v oblasti zaměstnanců, kteří v organizaci pracují, tak změna může přijít z okolí organizace. Současné trendy a snahy o nejrůznější optimalizace a poradenství v rámci strategického řízení organizací by měly respektovat právě komplexní povahu každé organizace a její jedinečnost.

V zahraniční praxi se setkáváme s poradci, kteří nabízí přímo poradenské služby pod hlavičkou např. kybernetiky druhého řádu, organizační či manažerské kybernetiky, atd. Tito poradci se od běžných poradců, se kterými se setkáváme v našich organizacích, liší především přístupem a způsobem myšlení. Jejich postup je opačný, ti, kteří se zabývají kybernetikou či systémovými vědami se snaží pochopit systém a navrhnout určitá řešení na míru dané organizace. Naopak klasičtí poradci mají svá ověřená řešení, do kterých se snaží za každou cenu napasovat danou organizaci a často dochází k opomínání důležitých specifik takové organizace.

V tomto článku je uveden přehled metodik a jejich možného použití z různých hledisek jejich účelu. Článek se zabývá dvěma hledisky, a to systémovými vědami a kybernetickými přístupy. V závěru článku bude nastíněn postup vytvářené metodiky MALINA (*Management And Labor Internal Network Analysis*), jejíž součástí je nástroj pro vyhodnocení úrovně sociální a manažerské kybernetiky v organizaci.

Přehled vývoje metodologií v oblasti systémových metodologií

Systémová dynamika (SD) – založená J. W. Forresterem v roce 1965. Metodologie byla navržená pro matematické modelování chování organizací. SD se potýká s komplexními problémy za použití stavů a toků v určitých částech systému a jejich vztahy. (Mildeová, 2008)

Metodologie měkkých systémů (SSM) – vytvořená P. Checklandem v roce 1969. Metodologie se vztahuje k měkkým problémům v procesu modelování organizace. Pomáhá definovat problémy a připravuje modelové situace a možné změny. (Checkland, 1999)

Kritické systémové heuristiky (CSH) – vytvořena W. Ulrichem v roce 1983. Tato metodologie je založená na hranicích systému, které vedou k identifikaci chybných úvah v organizaci, které jsou jiné než očekávané. (Ulrich, 2005)

Návrh vedený klientem (CLD) – vytvořeno F. Stowellem roku 1993. Metodologie se zaměřuje na systemické a kontextuální problémy analýzou klientů a zaměstnanců v organizacích. Obsahuje také sadu nástrojů a technik pro definici problémů (Stowell, West, 1994)

ETHICS – vytvořena E. Mumfordovou roku 1995. Tato metodologie byla výchozí pro několik dalších přístupů. Metodologie připouští, že problémy spojené s implementací informačních systémů nevychází pouze z technologických omezení, ale velmi často spíše ze strany sociálního faktoru v organizacích. (Mumford, 1996)

Kontextové systémové dotazování – např. Socio-technický toolbox vytvořený P. Bednarem, M. Sadok a Shidrovou v roce 2014. Jde o komplexní analytický nástroj připravený pro systémového analytika. Tato analýza je založena na know-how zaměstnanců a také na jejich tužbách nejen na informacích a znalostech. Tato sada nástrojů je rozdělena do 8 různých systémových oblastí. Je zde dostupných 27 analytických nástrojů a vice než 30 šablon pro analýzu organizace. Oblasti jsou konzistentní a připravené pro komplexní systémovou analýzu, tak že je možné si sadu nástrojů připravit dle aktuálních potřeb dané organizace. (Bednar and Welch, 2014; Bednar, Sadok, Shiderova, 2014; Bednar, Sadok, 2015)

Kybernetické přístupy k popisu organizace

Klasická kybernetika se vyvíjela od 50. letech minulého století v rámci skupiny okolo Norberta Wienera, který je uznáván jako zakladatel kybernetiky jako vědy, která se zabývala řízením v umělých, ale i živých systémech. Toho kybernetici dosahovali pomocí modelů, které byly inspirovány z nejrůznějších vědeckých oblastí. Postupně však vznikala potřeba popisovat také systémy značně neurčité a neustále se měnící, tedy komplexní. Kybernetici brzy dospěli k názoru, že takové systémy není možné popsat modelem a není možné je jednoznačně řídit, aniž bychom se snažili porozumět danému systému a jeho prvkům per se. (Rosický, 2009)

Tyto snahy vykrystalizovaly ve vznik tzv. kybernetiky druhého řádu, která již neakcentuje systémy oddělené, ale zabývá se určitým bodem v daném systému, a tedy konkrétním prvkem/uživatelem, kterého nazývá pozorovatel. Každý pozorovatel má jiný úhel pohledu na stejný systém, jinou formu znalostí a zkušeností, často má k dispozici i jiné informace o daném systému. Proto pokud se pokusíme popsat organizaci z pohledu top manažera, získáme úplně jiný pohled, než pokud se budeme snažit popsat organizaci z pohledu běžného zaměstnance. (Rosický, 2001; Rosický, 2009)

Tato forma neurčitosti vede k problémům, které vznikají především v komunikaci, efektivitě práce či v rámci implementace různých změn v organizaci, prací s informacemi nebo dokonce spokojenosti daného jedince. (Rosický, 2001)

Velmi významným autorem v této oblasti je Stafford Beer, který většinu svého profesního života věnoval právě kybernetice druhého řádu a v podstatě se stal poradcem pro organizace a zabýval se změnami právě z pohledu organizací jako komplexních systémů. Vytvořil také model, který byl v oblasti managementu průlomový, poukázal na aspekty komplexního řízení organizací a nezbytnosti brát v úvahu nejen technické, ale také sociální faktory organizace. Model nazval „the viable system model“ (dále jen VSM) v překladu model životaschopného systému. Tento model byl prvně představen v knize *Brain of the firm* z roku 1972. (Beer, 1981)

Dělení kybernetiky

- *Kybernetika I. řádu* - Původní vědecká disciplína pod názvem kybernetika byla vytvořena v roce 1948 Norbertem Wienerem a skupinou, která ho obklopovala a má svůj základ v knize Kybernetika aneb Řízení a sdělování u organismů a strojů. Jde především o propojení znalostí a informací z dostupných informací z nejrůznějších vědních oborů a snaha popsat a pochopit komplexní systémy. To je tedy, to co dnes známe pod pojmem kybernetika prvního řádu, a také to, co si většina lidí představuje pod pojmem kybernetika a její matematicko-logické modely. (Valée, 2003; Potocan, Mulej, Kajzer, 2005)
- *Kybernetika II. řádu* - Hans von Foerster představil v 70. letech kybernetiku druhého řádu, která brala v úvahu rychle se měnící prostředí a globalizaci, která ho donutila převrátit pohled kybernetiky, kdy pozorujeme zvenku určitý systém na kybernetiku, kde je ústřední postavou pozorovatel systému a ptáme se jakou má znalost a zkušenost v rámci daného systému a jak vnímá působení ostatních. Jednoduše by se dalo říci, že začal brát do úvahy lidský faktor a tedy sociální složku systémů, což přidává další dimenzi komplexním systémům a vede k holistickému chápání systému i s jeho vnitřními (neovlivnitelnými) vlivy. (Foerster, 1974)
- *Kybernetika III. řádu* - Okolo roku 2000 vznikla definice kybernetiky třetího řádu, která se vrací zpět k původním dílům z roku 1951 a jejím autorem je Vallée. Pojetí kybernetiky třetího řádu má za cíl explicitní syntézu pozorování, rozhodování, ovlivňování a klade důraz právě na fázi rozhodování. (Potocan, Mulej, Kajzer, 2005)

Šest základních atributů kybernetiky

Novodobá kybernetika je brána jako synergický celek a neměla by být redukována pouze na zpětnovazební smyčky a nejrůznější typy modelů. Kybernetické pojetí ovlivňování, řízení a vedení prvků systému, událostí a procesů obsahuje šest následujících atributů, které by měly být vždy brány v úvahu (Potocan, Mulej, Kajzer, 2005):

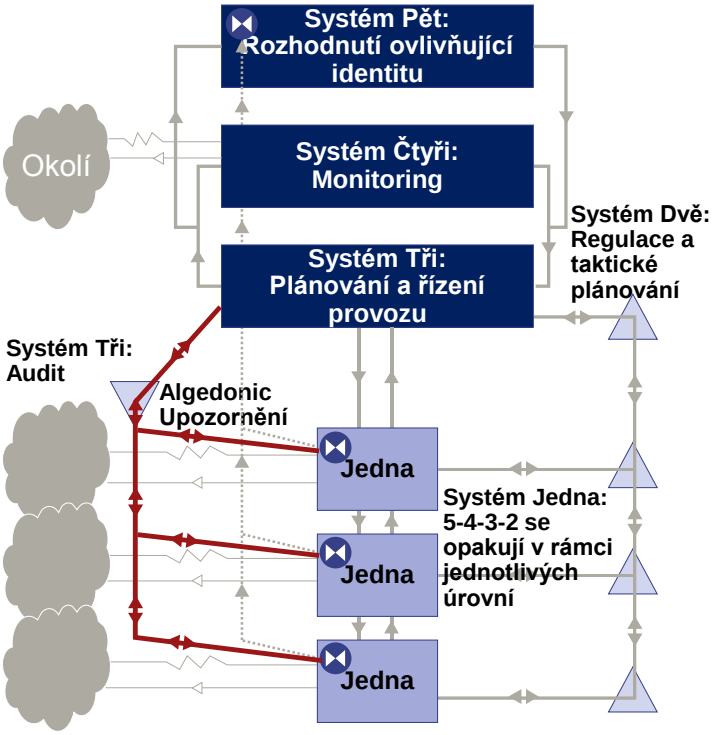
- *Komplexita* – mnohovztažnost, a to jak interní, tak externí vazby,
- *Otevřenost* – existují vztahy s okolím a ostatními systémy,
- *Dynamičnost* – existuje možnost změny, zahrnuje jak pozorovatele, tak ty, kdo rozhodují i samotné aktéry, o kterých je rozhodováno,
- *Přijímá vstupy a vytváří výstupy* – vliv na systém mají spíše informace než pouze materiální a energetické toky,
- *Toky jsou ovlivňovány zpětnovazebními smyčkami* – stabilizuje a zjednodušuje negativní zpětná vazba a posiluje pozitivní zpětná vazba,
- *Mentální modelování* – explicitně nebo implicitně dle vybraného úhlu pohledu.

Manažerská kybernetika - Model životaschopného systému (Viable system model)

Model životaschopného systému (VSM) byl vytvořen Staffordem Beerem za účelem analýzy organizace z pohledu kybernetiky druhého řádu. Představen byl tento model v knize *Brain of the firm – The Managerial Cybernetics of Organization* z roku 1972. Šlo o průlom v manažerském myšlení, kdy se S. Beer začal zabývat kybernetickým přístupem, kdy v organizaci není možné vše plánovat a organizovat dle předem daných a jednoznačných schémat a modelů, protože spousta věcí v organizaci je jednoduše organizována sama za účelem určité míry samoorganizace, která je úzce spjata se sociálním faktorem každé jedné organizace. S. Beer poukazuje na to, že dobrý manažer by měl být spíše facilitátorem než diktátorem a měl by se snažit věci pouze usměrňovat v rámci určitých pravidel, tak aby byly zajištěny základní funkce systému (organizace). (Beer, 1981; Beer, 1985)

Tento model položil základ manažerské kybernetice a je hluboce založen právě na principu rekurze, kdy je možné rekurzivně popsat nižší i vyšší systém v rámci dané organizace dle určitých pravidel. Model organizace je zde představen na principu fungování lidského mozku a jsou zde definovány jednotlivé systémy, které by měla každá organizace (i její jednotlivé části) obsahovat. (Beer, 1981, Rudall, 2002)

Popis systémů VSM (Beer, 1981; Ríos, 2010):



Obrázek 5 – VSM dle S. Beera Zdroj: Beer, 1981

Systém 1 – produkce zboží nebo služeb, tvoří primární aktivity organizace, dle Obrázku 1, jde o tři oddělení.

Systém 2 – zajišťuje harmonické fungování jednotlivých Systémů 1

Systém 3 – optimalizuje funkce všech Systémů 1

Systém 4 – monitorování celkového prostředí organizace, jeho cílem je připravenost na změnu, která může vzejít z okolního prostředí.

Systém 5 – zajišťuje normativní rozhodování a odpovědností za definování etiky, vize a identity organizace. Předává kritické informace pro životaschopnost systému.

Komunikační kanály jsou odpovědné za propojení všech systémů a funkcí, ale také za propojení organizace s okolním prostředím. Speciální kanály jsou algedonické (algedonic – definováno v rámci knihy Brain of organization – z řeckého slova – v překladu „pain“ and „pleasure“), které sbírají a předávají do Systému 5 informace kritické pro životaschopnost systému. (Beer, 1981)

K pochopení modelu VSM používá S. Beer vysvětlení čtyř kybernetických principů, na kterých je tento model založen (Beer, 1981; Rudall, 2002):

- *Princip rekurzivity* – každý živoucí systém obsahuje rekurzivní subsystémy, které je možné popsat podobným principem, a zároveň má takový systém, také systém nadřazený do kterého patří a který tvoří jeho okolí.
- *Princip adaptivní homeostáze* – živý systém je systém, který využívá principu homeostáze, stabilní systém je mrtvý systém. Živé systémy jsou v neustále se měnícím dynamickém prostředí a okolo rovnovážného stavu pouze oscilují.
- *Princip redundance potenciálních příkazů* – autorita pro předávání příkazů by neměl vycházet z řetězce příkazů, ale především dle relevantních informací.
- *Princip týmové synergie* – využití týmové práce a jejích synergií a upouštění od hierarchií a plochých organizačních struktur.

Model byl aplikován ve více než 150 organizacích po celém světě a účastníci označili zkušenost se zaváděním modelu za stimulující a vzrušující zážitek. Tento model byl využíván a modifikován v rámci

nejrůznějších systémových či kybernetických teorií a stal se základem i pro dva následující přístupy. (Rudall, 2002)

Případová studie v rámci IT organizace, která byla zpracována Davidem Wastellem, který se zabýval poradenstvím v rámci nejrůznějších organizací a rozhodl se využít VSM pro oddělení prodeje, kde se organizace snažila proniknout na zahraniční trhy po celé Evropě. A po analýze pomocí VSM šlo 75 % prodejů do zahraničí, a to ve velmi krátké době. V rámci projektu byl naprogramován informační systém pro distribuci v Evropě, a to přímo pracovníky prodeje, což vedlo ke značné autonomii tohoto oddělení a především možností rozšířit prodeje po Evropě. Pomocí VSM byl vytvořen konceptuální model týmů a oddělení, jako kritické se ukázaly vazby a vztahy s ostatními odděleními. Také se pomocí analýzy povedlo zjistit, že rozhodování ohledně produkce je často neformální, dle komunikace vedoucího s konkrétním klientem, a to např. bez znalosti využití současných výrobních kapacit. Jednoznačným závěrem byla formalizace vztahů mezi Prodejem a Výrobou a druhým velkým úspěchem bylo vyvinutí informačního systému pro prodej v Evropě, přímo v rámci daného oddělení. (Wastell, 1997)

Organizační kybernetika

Organizační kybernetika (OC) je úzce spojena s Beerovou teorií životaschopných systémů. Organizační kybernetika pomáhá navrhnout a diagnostikovat svou organizaci. Definuje se organizační identita, účel a hranice, provádí celým procesem vytváření struktury a provádí detailní strukturu jednotlivých komponent organizace z pohledu životaschopnosti. Organizační kybernetika spadá do kybernetiky třetího řádu, kdy se navrací k původní myšlence VSM a přidává novodobé pohledy dle potřeb a požadavků současného managementu organizací. (Ríos, 2010)

Vzhledem ke komplexitě jaké organizace dnes musí čelit, je třeba přidělit určitou kapacitu v rámci organizace k řešení různých problémů vyvstávajících právě z povahy komplexních systémů. V praxi jsou velmi často rozšířené řízení podniků pomocí nejrůznějších modelů, např. business modely, ovšem management založený na takovýchto modelech má vždy určitá omezení vyplývající z hranic a limitů daných modelů. (Ríos, 2010)

Tento rámec nejdříve definuje identitu a účel celé organizace, dále dochází k analýze struktury, po které vzniknou menší části systému organizace a tzv. suborganizace, pro každou z nich se určuje jejich prostředí, dále jsou zjišťovány podmínky fungování definovaných systémů a dle Beera jsou tyto podmínky adekvátně popsány a jsou řízeny. Pro každou z definovaných systémů (rekurze) je určována soudržnost jejích interních funkcí. Rámec také obsahuje tzv. organizační patologie. Jde o předem definované problémy, které mají svůj původ v základních principech kybernetiky. Tyto patologie jsou dále analyzovány v každé části analýzy organizace. (Ríos, 2010)

Organizační kybernetika se zabývá následujícími aspekty (Ríos, 2010):

- *Životaschopnost* – schopnost údržby své samostatné existence – přežít i přes změny v okolí. Musí zde existovat schopnost seberegulace, adaptace a učení se a také evoluce.
- *Rozmanitost* – indikuje úroveň komplexity systému, odkazuje na počet možných stavů a aktuálního či potenciálního chování v rámci dané situace.
- *Ashbyho zákon nezbytné rozmanitosti* – pouze rozmanitost může zničit/absorbovat rozmanitost.
- *Conan-Asbyho teorém* – v kontextu vyhodnocení rozhodnutí manažerů je kvalita těchto rozhodnutí limitována kvalitou modelů, které používají.
- *Model životaschopného systému (VSM)* – v rámci tohoto modelu byly stanoveny nezbytné a dostačující podmínky pro životaschopnost organizace. Jde o soustavu jednotlivých systémů a definovaných vztahů a funkcí mezi systémy a jejich okolím.
- *Rekurzivní charakter VSM* – všechny životaschopné systémy obsahují další systémy, tím je zajištěno jejich přežití a většinou jde o rekurzi určitých podmínek systému.
- *Organizační patologie* - V rámci VSM je definováno pět systémů, kde každý zastává svou funkci (více o těchto funkcích v kapitole Manažerská kybernetika - Model životaschopného systému (Viable system model)). Pokud některý z těchto definovaných systémů nefunguje nebo funguje špatně, jsou generovány určité problémy, které byly rozděleny do tří skupin: strukturální, funkcionální a informační patologie.

Autor tohoto konceptu, Dr. José Pérez Ríos, vytvořil také software, který je možné pro využití daného rámce použít k usnadnění práce. Software se jmenuje VSMod a je k dispozici zdarma po registraci, pokud je používán pro výuku a výzkum. Umožňuje namodelovat organizaci rekurzivně v rámci principů organizační kybernetiky a následně se zaměřit na problémy v jednotlivých identifikovaných systémech či subsystémech. (Ríos, 2015)

Podniková kybernetika

Význam a role podnikové kybernetiky (BC) byly definovány v článku „Business cybernetics – a provocative suggestion“ (Potocan a kol, 2005). Co se týče podnikové kybernetiky, spadá v rámci dělení do kybernetiky třetího řádu. V podstatě jde o spojení systémových teorií a kybernetiky v uceleném pohledu na podnik. Přestože systémové teorie a kybernetika vznikaly ve stejném čase, vzplál mezi nimi boj ve snaze odlišit se a poukázat na vlastní přínos a úhel pohledu. Později se však začaly prolínat a ve snaze o holistický pohled na danou problematiku vznikaly synergie. Snahou autorů, Potocan, Mulej a Kajzer, bylo využít vše, čeho tyto vědní disciplíny dosáhly a pokusit se je aplikovat na podnikovou praxi. (Potocan, Mulej, Kajzer, 2005)

Podniková kybernetika je nabízena také jako poradenská činnost, pro organizace, které chtějí zavádět v organizacích změny. Ve světě je běžné, že poradenské společnosti nabízejí holistický a ucelený pohled na organizaci a na základě takové analýzy jsou schopni manažerům doporučit správné postupy či podpořit jejich rozhodování právě uceleným pohledem na současnou situaci v organizaci. Jde o pohled na nejrůznější organizační problémy ve formě určitého workflow podpořeného různými nástroji (šablony, tabulky, myšlenkové mapy, atd.). Příkladem může být management kampaní, datová analýza, Business Intelligence, analytické CRM, atd. (Jammalamadaka, 2002)

Hlavní význam podnikové kybernetiky je v položení základních otázek, co je pro danou organizaci nezbytné pro holistické pochopení konceptu určité části organizace nebo určitého zkoumaného podnikového systému. Nezbytnou součástí je také vytvoření dialektického systému v dané organizaci, kde se střetávají názory a pohledy jednotlivých zaměstnanců v organizaci. Většinou jde o tým vybraných specialistů, kteří by si měli být v rámci týmu rovni. (Potocan, Mulej, Kajzer, 2005)

Role podnikové kybernetiky vychází z pojetí staré i moderní kybernetiky (viz dělení kapitola Dělení kybernetiky). Snahou podnikové kybernetiky je brát organizace nebo lidi jako podnikové systémy s následujícími charakteristikami (Potocan, Mulej, Kajzer, 2005):

- Podnikový systém je komunikační síť, ve které si jednotlivci či skupiny vyměňují informace
- Podnikový systém je systém nebo síť aktivit, ve kterých probíhá transformace vstupů na výstupy
- Podnikový systém je společenský systém, který má společenské úlohy a k nim také zodpovědnosti.
- Lidský jedinec vykazuje z pohledu podnikového systému více méně stejné atributy jako organizace

Co se týče podnikového systému a vztahu podnikové kybernetiky, hlavním přínosem je pohled na organizace holisticky. Podniková kybernetika se zaměřuje především na identifikaci účelu, obsahu, metodologie a podmínek užití, potřeb a možností daného podnikového systému (jakož i samotných uživatelů). (Potocan, Mulej, Kajzer, 2005)

V případě podnikové školy je objektem denodenní podnikový život. Dialektický systém zde tvoří nabídka všech nabízených předmětů, systémem je zde úhel pohledu, který je prezentován v rámci jednotlivých kurzů a ten je ještě modelován výukovými materiály, které jsou studentům k dispozici. Komunikace, která zde probíhá a jak jsou předávány informace, probíhá vždy pomocí modelů, ať už pomocí obrázků, knih, vzorců nebo jazykem, který užívá zažité fráze z daného oboru. Obě strany se snaží obsáhnout a popsat určitou část reality a pochopit daný objekt. Realita však obsahuje daleko větší úsek, než je jedinec schopen popsat nebo pochopit. Redukce probíhá na obou stranách. Otázkou je, do jaké míry redukujeme realitu a do jaké míry jsme schopni pochopit, to co je nám prezentováno/redukováno a jak se s tím v realitě vyrovnáme a co s danými informacemi uděláme, to už je problém, který řeší každý jedinec individuálně dle svých dalších schopností a znalostí, a to je problém se kterým pomáhá podniková kybernetika. (Potocan, Mulej, Kajzer, 2005)

Druhý příklad použití podnikové kybernetiky může být například podnikový proces. Objekt, který je brán v úvahu, je zde předmět podnikání organizace. Dialektický systém je tvořen podnikovými funkcemi a synergiemi, které v organizaci existují, tak aby zabezpečili zhotovení zakázek a dodrželi smluvní plnění, kterými je organizace vázaná. Systémem je zde úhel pohledu vybrané podnikové funkce, např. výroba, prodej, účetnictví, atd. Modelem je zde organizační struktura či struktury, které jsou dány normami nebo tak jak jsou modelovány podnikové procesy. (Potocan, Mulej, Kajzer, 2009)

Nastínění přístupu metodologie MALINA

V dnešní době je pro manažery v organizacích důležitá otázka práce se zaměstnanci, jakožto sociální složkou. Dostáváme se do konfliktu se znalostmi lidí, předáváním informací, komunikací, využíváním ICT prostředků v organizaci a celkově s měkkými faktory, které všechny ovlivňují práci lidí a tím i úspěch projektu, potažmo celé organizace.

Velice často se setkáváme s nepochopením základních vztahů, které v organizaci panují ze strany managementu, a to i přesto, že se někteří manažeři snaží porozumět. Snahou sociální a manažerské kybernetiky v organizaci by měla být možnost, jak poukázat na odlišnosti v jednání a smýšlení zaměstnanců, které vedou k určitým důsledkům. Pokud by existoval nástroj, který by dokázal ohodnotit současnou situaci z několika pohledů a navrhnout manažerovi konkrétní řešení, velmi často by se podařilo odchytit samotné příčiny neúspěchu v organizacích. Právě k tomu by měla posloužit navrhovaná metodologie MALINA - *Management And Labor Internal Network Analysis* (pracovní název), která se zabývá analýzou interních vztahů mezi managementem a pracovníky. Dle své povahy bude možné metodologii nasadit v rámci všech typů společností, ale v rámci disertační práce bude zpracování metodologie orientováno na malé a střední podniky.

Cíl metodologie

Obsah metodologie:

- Definované oblasti zájmu, které metodologie pokrývá
- Sběr dat pomocí modulárního dotazníku (součástí metodologie bude i nastavení logiky závislostí dle odpovědí respondentů v rámci možných výsledků)
- Nástroj pro vyhodnocení – sjednocení dat z dotazníků a jejich analýza a vyhodnocení
- Závěry a doporučení vyplývající z vyplněných dotazníků
- Ověření metodiky:
    - Pilotní projekt – ověření správnosti logiky závislostí
    - Případové studie – organizace z několika různých oblastí
    - Vyhodnocení výsledků společně s managementem daných společností

Výsledná metodologie by neměla být pouze teoretickým rámcem, ale měla by také umožňovat použití v závislosti na navrhované realizaci pomocí dotazníků (dotazníky budou anonymní) a nástroje pro vyhodnocení programovaného v prostředí MS Excel za pomoci VBA (Visual Basic for Application). V rámci dotazníku bude možné otázky a oblasti parametrizovat a také samotné výsledky bude mít možnost manažer filtrovat z různých pohledů. Samotná logika metodologie tedy bude ukryta v rámci nástroje pro vyhodnocení. Výsledkem by mělo být také ohodnocení daných oblastí, aby měl manažer představu, jak si organizace v daných oblastech aktuálně počíná. Vzhledem k tomu, že nejde o jednoznačný model – vyplývající z kybernetiky druhého a třetího řádu. Je možné dotazník používat opakovaně, s předpokladem, že pokaždé dostaneme trochu jiné odpovědi, dle aktuálního stavu zaměstnanců, kteří jej budou vyplňovat, a stavu organizačního systému. Tím je také možné si dopad určitých rozhodnutí u zaměstnanců ověřovat.

Způsob organizace sběru a získávání experimentálních dat

V rámci výběru organizací, které budou ochotné se podílet na výzkumu, bude osloveno větší množství firem, které svým rozsahem spadají do kategorie SME (malé a střední podniky) na kterých by se daný nástroj a jeho fungování prověřilo. Ve vybraných organizacích budou rozeslány dotazníky a následně budou odeslány zpět řešiteli, který ve výše uvedeném nástroji provede vyhodnocení, a to především

z důvodu úplné anonymizace odpovědí. Po vyhodnocení bude zaslán report manažerům, se kterými bude následně domluveno interview, na kterém budou jednotlivé výsledky a doporučení validovány.

Využití a přínos výsledků

Sociální kybernetika se zabývá pohledem zaměstnanců a manažerská kybernetika se zabývá pohledem manažerů, a tedy řízením organizace. Hlavním cílem je tedy tvorba metodiky pro ohodnocení organizace z několika úhlů pohledu na základě odpovědí zaměstnanců z dotazníků a jejich interpretace k doporučením, které poskytne nástroj, který provede samotné vyhodnocení dotazníků a aplikuje vyvinutou metodiku. Další funkcí nástroje je kromě výsledného pohledu na organizaci také možnost volby různých pohledů, ať nad jednotlivými odděleními či pozicemi, tak z pohledu různých věkových kategorií, atd.

Plánované využití je pro široký management organizací, ať už na úrovni top managementu či na nižších úrovních řízení.

Hlavní přínos by měl být v lepším pochopení fungování vztahů v organizaci, prací s informacemi a znalostmi zaměstnanců, podpora komunikace a prostředí a především vhodná podpora ICT prostředky.

Závěr

Z přehledu metodologií zabývajících se systémovým myšlením se v čase ukazuje jednoznačná potřeba, která vede od analyzování tvrdých systémů přes měkké systémy až po kombinaci obojího. V současné době je patrná snaha o přístupy, které na organizaci pohlížejí komplexně, ale zároveň umožňují určitou míru volnosti, a to dle potřeb dané organizace. Co se týče kybernetických přístupů, je zde tato potřeba patrná už od samého počátku, kdy byl v roce 1972 vytvořen model VSM (model životaschopného systému), kde S. Beer jednoznačně poukazuje na principy rekurzivity, adaptivity, nutnosti týmových synergií a hrozbu redundance příkazů v organizaci. Tento model byl následně přejímán a upravován, dá se říci až do dnešní doby a většina kybernetických přístupů z něj stále vychází.

Navrhovaná metodologie MALINA by měla posunout vědu v oblasti sociální a manažerské kybernetiky a jejich vztahu k současným organizacím a poukázat na možnosti, které dnešním manažerům nabízí. Tato metodologie by měla být unikátní v tom, že bude nabídnuta přímo formou dotazníku s nástrojem k vyhodnocení, což významně ušetří jak finanční a časové náklady na analýzu, ale zároveň je dotazník replikovatelný a je možnost sledovat vývoj úrovně organizace v čase. Metodologie by tak měla pomoci managementu při řízení organizací a podpořit manažerské rozhodování.

Literatura

BEDNAR, P.; SADOK, M., 2015. Socio-Technical Toolbox for Business Analysis in practice. DOI 10.1007/978-3-319-07040-7_21

BEDNAR, P.; SADOK, M., SHIDEROVA, V., 2014 Socio-Technical Toolbox for Business Analysis in Practice. Smart Organizations and Smart Artifacts. ISBN 978-3-319-07040-7.

BEDNAR, P.; WELCH, Ch. E., 2014. Contextual inquiry and socio-technical practice. Kybernetes. DOI:10.1108/K-07-2014-0156.

BEER, S., 1981. Brain of the firm: the managerial cybernetics of organization. 2d ed. New York: J. Wiley, xiii, 417 p. ISBN 0471276871.

BEER, Stafford. Diagnosing the system for organizations. New York: Wiley, 1985, xiii, 152 p. ISBN 0471906751.

FRANCOIS, C. a kol, 2004. International Encyclopedia of Systems and Cybernetics, K.D. Saur, Munchen.

FOERSTER, H., 1974. Cybernetics of cybernetics, Urbana, IL. Dostupné online: http://faculty.stevenson.edu/jlombardi/pdf's/cybernetics/cybernetics_cybernetics_hvf.pdf [naposledy navštíveno dne 26.1.2016].

CHECKLAND, P., 1999. Soft Systems Methodology: a 30-year retrospective

JAMMALAMADAKA, S., 2002. Business Cybernetics. Dostupné online: www.business-cybernetics.com [naposledy navštíveno dne 26.1.2016].

MILDEOVÁ, S.; VOJTKO, V., 2008. Systémová dynamika. 2. vyd. Praha (System Dynamics, 2nd edition): Oeconomica. ISBN 978-80-245-1448-2.

MUMFORD, E., 1996. Systems Design: Ethical Tools for Ethical Change, Baskingstoke, England: Macmillan Press.

POTOCAN, V.; MULEJ, M.; KAJZER, S., 2005. Business cybernetics: a provocative suggestion, Kybernetes, Vol 34 Iss 9/10

POTOCAN, V.; MULEJ, M., 2009. Business cybernetics – provocation number two, Kybernetes, Vol 38 Iss 1/2

RÍOS, J. P., 2010. Models of organizational cybernetics for diagnosis and design, Kybernetes, Vol 39 Iss 9/10.

RÍOS, J. P., 2015. Organizational Cybernetics. Dostupné online: http://www.organizationalcybernetics.org/ [naposledy navštíveno dne 26.1.2016].

ROSICKÝ, A., 2001. Information and social systems evolution. In The Praxis of Cybernetics and the Cybernetics of Praxis. Vancouver: ASC, p. 29.

ROSICKÝ, A., 2009. Informace a systémy: základy teorie pro úspěšnou praxi. Praha: Oeconomica. ISBN 978-80-245-1629-5.

RUDALL, B. H., 2002. The viable system model (VSM) in management, Kybernetes, Vol. 31 Iss: 1

RUDALL, B. H., 2002. Future trends in cybernetics, Kybernetes, Vol. 31 Iss: 2

STOWELL, F., WEST, D., 1994. Client-Led Design: A Systemic Approach to Information Systems. Definition, London: McGraw-Hill.

ULRICH, W., 2005. A brief introduction to critical systems heuristics (CSH). Web site of the ECOSENSUS project, Open University, Milton Keynes, UK, 14 October 2005, Dostupné online: http://www.ecosensus.info/about/index.html [naposledy navštíveno dne 26.1.2016].

VALLÉE, R., 2003. History of Cybernetics (in EOLSS Encyclopedia of Life Support Systems. Dostupné online: www.eolss.net [naposledy navštíveno dne 26.1.2016].

WASTELL, D., 1997. Just a thought: VSM Case study. Dostupné online: http://www.triarchypress.net/a-vsm-case-study.html [naposledy navštíveno dne 26.1.2016].

JEL Classification: M15, D83, J53, O32

Summary

Social and managerial cybernetics in the context of analyzing organization

Abstract: The article presents a selection and summary of certain applied methodologies and approaches used to analyze the organization in terms of cybernetic principles and respecting the social factor in the organization. In the introduction to the article, the reader learns about the necessity to think about every organization in terms of basic cybernetics principels (complexity, openness, dynamism, receiving inputs and outputs, possibility to observe the feedback loops, the possibility of using mental models, recursion etc.) and also slightly introduce historical types of cybernetics. The article also presents an overview of the historical development of the system methodologies. The next part of the article elaborates the issue of cybernetic approaches to the analysis of the organization. The most important shift in the understanding of management is Management Cybernetics, which was defined by S. Beer (1972) and is based on the Viable System Model (VSM). VSM applies various cybernetic principles (recursion, adaptive homeostasis, the potential danger of commands redundancy, team synergies) to managerial practice. Subsequently, this model was reviewed several times and was a basis for other approaches such as Business or Organizational Cybernetics, that are also introduced in the article.

In conclusion, the article outlines the suggested methodology called MALINA (Management And Labor Internal Network Analysis). This methodology describes areas in the analysis of the organization (information, knowledge, communication, ICT utilization, staff satisfaction, etc.) that are meaningful for evaluation of the current state of the organization. Part of the methodology is an evaluation tool that includes modular employee's questionnaire.

Keywords: cybernetics, managerial cybernetics, organizational cybernetics, system methodologies, MALINA methodology

Generating examples of paths summarizing RDF datasets

Jindřich Mynarz
jindrich.mynarz@vse.cz
Ph.D. student of Applied computer science
Thesis advisor: doc. Ing. Vojtěch Svátek, Dr., (svatek@vse.cz)

Abstract: As datasets become too large to be comprehended directly, a need for data summarization arises. Data summary can present typical patterns commonly found in a dataset, from which a high-level understanding of the data can be obtained. Nonetheless, such abstract understanding can be improved by providing concrete examples of the summary patterns. If possible, the chosen examples should be diverse and representative of the patterns they instantiate. In this paper, we present 3 methods for generating examples of patterns discovered in RDF datasets. The patterns we consider are the most frequent path graphs that consist of classes or data types connected by RDF properties. We propose an RDF/S vocabulary for describing these path graphs and their instances. The presented methods for generating path examples include random, distinct, and representative selection and are based on randomization, diversification, and clustering. Partial implementation of these methods is described along with the trade-offs considered during development.

Keywords: summarization, diversity, Resource Description Framework

Introduction

Many datasets nowadays are too large to comprehend by examining all their data. Fortunately, most large datasets exhibit regular structure. Partial comprehension of such datasets can be derived from the common patterns manifested in their structure. In a way, a dataset's structure provides a high-level summary of the dataset's content. Nevertheless, telling dataset structure is somewhat more difficult with data formalized using the Resource Description Framework (RDF) (Cyganiak, 2014). RDF is a graph data model without a schema fixed up-front that decomposes data into binary relations represented as RDF triples. While traditional relational databases enforce ex-ante schema, schema in RDF datasets often emerges only ex-post as a result of combination of several RDF vocabularies and their specific way of use. Schema of RDF datasets is thus descriptive rather than prescriptive.

Since schemata of RDF datasets are not formalized before use, they typically need to be discovered from data via statistical inference. An example of a tool that discovers and summarizes structure of RDF datasets is LODSight (Dudáš, Svátek, and Mynarz, 2015). LODSight offers a node-link visualization of the most common path graphs found in a summary of an RDF dataset. Path graphs are trees with *n* nodes, such that 2 nodes have vertex degree 1 and *n - 2* nodes have vertex degree 2, so that the graph can be drawn with all its nodes and edges lying in a single straight line (Gross and Yellen, 2006, p. 18). The nodes in these RDF paths in LODSight are types (either classes or data types), while the edges are predicates connecting instances of the types in the summarized dataset. This description of paths corresponds to the definition 6 from Troullinou et al. (2015b). For example, a path can connect the class `gr:BusinessEntity`[2] via the predicate `schema:address` to the class `schema:PostalAddress` that is connected to the data type `xsd:string` via the predicate `schema:streetAddress`. This path connects companies to their postal addresses that have literal street addresses.

Empirical schema of RDF datasets embodied in the LODSight's paths may be rather abstract. In order to recover some of the understanding obtained by examining the data itself, it is possible to present representative examples instantiating the patterns found in a dataset. Providing several diverse examples of a chosen pattern can complement the high-level understanding of a dataset with inductive understanding following from the examples. LODSight has a basic functionality to generate examples of RDF paths. The examples are retrieved in a non-random fashion from an RDF store using the default sort order for SPARQL (Harris and Seaborne, 2013) results, so that for each query the same

[2] All namespace prefixes (`gr:`, `schema:`, etc.) used in this paper can be resolved to their corresponding namespace IRIs by using http://prefix.cc.

examples are retrieved. In this paper we present an improvement of this functionality. It is implemented in a library that generates random and illustrative examples of RDF paths. In the following sections we present the methods used for example selection and their partial implementation.

Before proceeding with our work, we consider the research related to it. While a lot of related effort is dedicated to summarization of RDF datasets, generating examples from summaries is rarely covered. In cases the research on RDF summarization is concerned with selection of the most representative features of the analysed datasets, it does so predominantly on the level of patterns and not pattern instances (e.g., Presutti et al., 2011; Troullinou et al., 2015a; Troullinou et al., 2015b). Similarly to our work, Presutti et al. (2011) also propose a vocabulary for describing paths in RDF datasets.[3] Nevertheless, to our knowledge the work of Dudáš, Svátek, and Mynarz (2015), on which this research builds, is the first that is concerned with representative instances of patterns from RDF datasets.

Example generation methods

The library for generating examples of RDF paths that is described in this paper offers 3 methods for example selection of increasing sophistication, including random selection, distinct selection, and representative selection. An integral part of these methods is the representation of RDF paths, for which we propose the RDF Path vocabulary. Additionally, a fundamental part of the latter 2 methods is similarity computation. We describe both these parts before we explain how the example generation methods work.

RDF Path vocabulary

We designed the RDF Path Vocabulary as a mean to formalize the description of RDF paths. The development of a custom vocabulary was motivated by the goal of improving the representation of RDF paths. In particular, there was a need to distinguish path nodes instantiating classes from literal nodes instantiating data types. Additionally, the representation of the examples of RDF paths needed to be extensible; e.g., to allow attaching labels (`rdfs:label`) to nodes.

The result is a simple vocabulary[4] that is formalized using RDF Schema (Brickley and Guha, 2014). Paths are represented as instances of the `Path` class. Each path has at least 1 edge that instantiates the `Edge` class. Path is connected to its edges via the `edges` property. In order to maintain the order of edges, the object of this property is a collection (`rdf:Seq`) that wraps the edges constituting the path. RDF collections are structured as linked lists, which incurs some difficulty in their processing. However, in our case we use their JSON-LD (Sporny et al., 2014) serialization as regular arrays. Each edge has a starting and ending node (linked via the `start` and `end` properties, respectively) and a predicate that connects the nodes (linked by the `edgeProperty` property). Attentive readers may notice a degree of isomorphism between the path edges and the RDF reification vocabulary (Manola and Miller, 2004). In fact, the `Edge` class is derived as a subclass of `rdf:Statement`, while its properties are designed as subproperties of the properties used for reifying RDF statements. Both the start and the end node of each edge must instantiate either a class or a data type. Each instantiated class and data type must be explicitly typed as either `rdfs:Class` or `rdfs:Datatype`. Paths must be continuous, so that the edge's start node must be equal to the end node of the preceding edge. Both paths and their examples are represented as instances of the `Path` class, but they differ in the representation of nodes. While path nodes are identified by blank nodes used as existentially quantified variables, nodes in path examples are concrete resources. To ensure the future stability of the vocabulary's IRI, we registered it as a persistent IRI in the w3id.org namespace.

[3] http://www.ontologydesignpatterns.org/ont/lod-analysis-path.owl

[4] See https://raw.githubusercontent.com/jindrichmynarz/rdf-path-examples/master/resources/rdf_path.ttl for Turtle serialization of the vocabulary.

Similarity computation

A key component of the non-random example selection methods is the computation of similarity between examples of paths. Since the examples instantiate the same path, their edge properties are identical and hence do not need to be compared. Similarity of path examples is computed as a mean average of the similarities of the paths' corresponding pairs of constituent nodes. The way of node similarity computation is determined by the inferred type of node. Similarity is always normalized to the $[0,1]$ interval, such that 0 is used for completely dissimilar resources, while 1 is used for equivalent resources. All implemented similarity metrics are symmetric, so that $sim(A,B) = sim(B,A)$.

In general, there are 2 kinds of node types: referents and literals. Referents are resources identified via IRI or a blank node and literals are textual resources, such as strings or numbers. Computation of similarity between literals is chosen based on their data types that are either explicitly provided in the source data or inferred. Data type can be inferred by matching a literal to the regular expression constraining the lexical space of the data type. If unequal data types are to be compared, we try to get their lowest common ancestor in the type hierarchy. If a common ancestor is found, then it determines what similarity metric is used. We amended the hierarchy of XML Schema[5] data types to be rooted in `xsd:string`. This way, the similarity computation defaults to string comparison if the data types of the compared operands do not match. For example, `xsd:negativeInteger` and `xsd:long` are compared as `xsd:integer`, which is lowest common ancestor they share.

Similarity metrics are chosen to be appropriate for the data types of the compared literals. For instance, numeric literals may be compared using a different metric than string literals. Similarity of ordinal values, such as numbers, is normalized by *"reducing each value to the same scale in proportion to the size of the considered space"* (Euzenat and Shvaiko, 2013, p. 86) (i.e. the difference between maximum and minimum value). Some data types can be further decomposed and their similarity can be computed from their constituent parts. For example, URLs can be split into domain names, protocols, port numbers etc., and their similarity can be derived from similarities of their parts.

Similarity of non-identical referents (resources identified by IRIs or blank nodes) is computed as similarity of their descriptions. If identifiers of the referents are identical, then their similarity is 1. Otherwise, the referents are compared by value, i.e. their representation in RDF, which recursively includes representations of the linked resources if available. We use an adjusted form of Jaccard index to aggregate the similarity of a pair of resources (11111111). We sum the similarities of objects of properties found in both compared resource descriptions and divide it by the number of RDF triples found in the resources' descriptions, in which the resource is in the subject position.

$$J(A,B) = \frac{\sum_{i \in A \cap B} sim(A_i, B_i)}{|A \cup B|} \quad (1)$$

Co11mputation of similarity of referents is recursive. If referents are found as part of resources' descriptions, their similarity is in turn computed as the similarity of their descriptions. This way of similarity computation can be considered as bottom-up, since the similarity of referents proceeds from the similarity of literals found in the referents' descriptions. Should either of the properties shared in the compared resources have multiple objects, their similarity is computed for each combination of the objects and aggregated to maximum. Overall, computation of pairwise similarities between path examples has approximately quadratic complexity (more precisely $O(\frac{n(n-1)}{2})$) based on the number of combinations.

Type inference used in the similarity computation supports referents and several data types defined by the XML Schema, including `xsd:decimal`, `xsd:date`, `xsd:dateTime`, `xsd:duration`, `xsd:boolean`, and `xsd:anyURI`, defaulting to `xsd:string`. Default similarity measure for strings is the Jaro-Winkler metric, which computes edit distance of the compared strings while taking into account their length. If the compared literals have the English language tag, then the Porter

[5] http://www.w3.org/TR/xmlschema-2

stemmer is applied to normalize them before the comparison. Similarity of `xsd:anyURI` is computed as a weighted arithmetic mean of similarities of the constituent URI parts measured by the Jaro-Winkler metric, so that higher weight is given to more prominent parts of URIs, such as the domain name. Similarity of ordinal values, such as `xsd:decimal` or `xsd:date`, is calculated as their absolute difference normalized by the maximum spread of values of the compared property:

$$sim(A, B) = \frac{|A - B|}{\max() - \min()} \quad (2)$$

Having discussed the representation of RDF paths using the RDF Path vocabulary and the similarity computation we turn to discuss the example selection methods built on top of these foundations.

Random selection

Our baseline approach is random selection of examples of RDF paths. Using this method we retrieve *n* random examples of the provided RDF path. This selection is based on random sort order of the retrieved path examples. Since the selection is random, the results may include near duplicates or outliers that do not represent well what the analysed dataset contains. Hence, we propose more sophisticated methods for example selection.

Distinct selection

The distinct selection method attempts to maximize the diversity of the selected examples. Using this method we want to select *n* examples such that their mutual similarity is minimal. In other words, *"given a set P of n items, we aim at selecting k items out of them, such that, the average pairwise distance between the selected items is maximized"* (Drosou and Pitoura, 2009, p. 1). However, selecting the most diverse subset of items is known as a NP-hard problem. Thus, solutions to this task need to be approximated by using heuristics. We chose the greedy construction heuristic (Drosou and Pitoura, 2009), since it achieves good diversity in short execution time. Using this heuristic we start by selecting 1 random example, then pair it with its most dissimilar example, and continue adding examples that have the lowest total similarity with the already included ones. To compute the similarity between path examples we use the similarity measure described above. Using this method we produce examples that provide a broader coverage of an RDF path in the processed dataset than the random selection provides.

Representative selection

In addition to the chosen examples being diverse we want them to be representative. The selection of examples with maximum diversity may not be the most representative of the summarized dataset. It may include misleading outliers, since they maximize the overall dissimilarity of the selected examples. Instead, we want to pick examples that are both diverse and representative of the most common kinds of data in the analysed dataset. This reformulation of the example selection task leads to clustering, so that *"the goal now is to identify and recommend a set of representative items, one for each cluster, so that the average distance of each item to its representative is minimized"* (Castells et al., 2015, p. 897). One suitable method for the selection of representative examples is k-medoids clustering. This type of clustering *"searches for* k *'representative' objects called medoids, which minimize the average dissimilarity of all objects of the data set to the nearest medoid"* (Kaufman and Rousseeuw, 1987). It offers efficient means for computing clusters from pairwise similarities and produces medoids that can be considered the most representative members of the generated clusters. In fact, medoid is the *"most centrally located object in a cluster"* (Park and Jun, 2009). We select the generated medoids as the most representative examples of a given RDF path.

Implementation of example generation

Having described the methods employed for example selection, we now detail how we implemented these methods. Currently, the implementation is only partial and lacks, most notably, the representative selection method. Generation of examples of RDF paths was implemented as

a ClojureScript library.[6] ClojureScript compiles to JavaScript, which makes it possible to use the library from a regular JavaScript application, such as the current LODSight visualizer. The library is released as open source under the terms of the Eclipse Public License 1.0.

We adopted JSON-LD as the data exchange format. The library expects the RDF paths to be provided in JSON-LD syntax. The generated examples are too returned in JSON-LD. Using JSON-LD allows us to manipulate paths and path examples as native JavaScript data structures. Moreover, JSON-LD enables us to harness standard operations defined by the JSON-LD API (Longley et al. 2014), such as the compaction that makes it possible to coerce heterogeneous RDF graph into a regular data structure with predictable nesting and attribute names.

Path examples are retrieved from SPARQL endpoints using the SPARQL 1.1 Protocol.[7] The queried SPARQL endpoint must allow for cross-domain requests by providing the appropriate cross-origin resource sharing HTTP headers.[8] The library requests the data in the N-Triples[9] syntax, which is easy to parse due to its simplicity. The obtained payload is converted to JSON-LD and compacted into a more predictable structure. The described transformations are executed by the jsonld.js library.[10] We experimented with retrieving data directly in the JSON-LD syntax, which would allow us to use JSON-P[11] requests to simplify cross-domain data retrieval, but we encountered inconsistencies in serialization of IRIs and blank nodes in our test RDF store (OpenLink Virtuoso[12]), such that it was not possible to tell them apart from literals, so we resorted back to using N-Triples.

All of the presented methods use random selection in SPARQL. The random method directly uses the examples generated by a SPARQL query with random order. The other methods fetch a sample of path examples out of which *n* required examples are selected. The size of the sample can be configured via a sampling factor, which is used to multiply the desired number of examples *n* to obtain the size of the retrieved sample. For example, if *n* is 5 and the sampling factor is set to 20, then 100 path examples are retrieved, which are narrowed down to 5 examples using the above-described methods.

Surprisingly, using random selection in SPARQL is not trivial. The specification of SPARQL defines the `RAND()` function that generates random numbers from the interval [0,1). However, the specification does not prescribe how the function should be evaluated. In many cases, SPARQL query engines evaluate `RAND()` only once per query execution, so that multiple uses of this function in a single query return the same value. This makes it unsuitable for random selection of SPARQL result bindings directly during query execution. A common work-around for this issue consists of generating a random offset that is used in a query to retrieve a random subset of the ordered result set. A downside of this solution is that the bindings matching the query must be counted beforehand, so that the offset can be generated lower than the number of matches minus the desired result set size. This adds an overhead of an additional query that must be executed. Nevertheless, we found that OpenLink Virtuoso can be forced to evaluate `RAND()` per each result binding if the `DISTINCT` solution modifier is used. We adopted this solution for retrieving random results, while being aware that it is specific to a particular software, and hence prevents the developed prototype to be used with RDF stores that exhibit different behaviour of `RAND()`. Nevertheless, we tested that the random selection also works with Apache Jena Fuseki,[13] which produces random results even without the `DISTINCT` modifier.

If either distinct or representative selection method is used, an additional SPARQL query is issued to retrieve a k-hop graph neighbourhood of the path example's nodes. K-hop graph neighbourhood of a node is a subgraph that includes adjacent nodes that are at most *k* hops away. For example, 2-hop

[6] https://github.com/jindrichmynarz/rdf-path-examples
[7] http://www.w3.org/TR/sparql11-protocol
[8] http://www.w3.org/TR/cors
[9] http://www.w3.org/TR/n-triples
[10] https://github.com/digitalbazaar/jsonld.js
[11] http://json-p.org
[12] http://virtuoso.openlinksw.com
[13] https://jena.apache.org/documentation/fuseki2

neighbourhood contains RDF triples in which a node is in the subject position (1st hop) plus the triples, in which subjects are the objects of the triples included in the first hop (2nd hop). The obtained dataset is used for computing the similarity of referents. Similarly to path data, we request the data in N-Triples, convert it to JSON-LD, and force its compact representation by using JSON-LD compaction.

Since the library compiles to JavaScript, example selection can be executed on the client-side in browsers of users interacting with a dataset summary using LODSight visualizer. Because of this target execution environment, we had to consider several performance trade-offs. Due to a limited processing power that can be expected in web browsers, we curtail the resources' use of similarity computation by limiting the sampling factor (20 is used as the default) and the number of hops in the graph neighbourhood considered in the computation (1-hop neighbourhood is retrieved). We also optimize the similarity function via memoization to speed up repeated access to pairwise similarities. If possible, ClojureScript transducers are used for chained sequence processing to avoid the realization of intermediate sequences. The complete process of generating path examples is executed asynchronously in order not to block the user experience when interacting with a dataset summary.

Conclusion and future work

We presented 3 methods for generating examples of RDF paths that build progressively on each other. The baseline method uses a simple random selection. The more sophisticated methods filter a random sample down to the most salient examples. Diverse selection attempts to maximize dissimilarity of the chosen examples to cover a broader variety of instances of a given RDF path. Representative selection goes further by picking examples that are both diverse and representative of the provided path.

We described partial implementation of the proposed methods along with a discussion of trade-offs that guided the implementation decisions. We learnt that semantic web technologies, such as JSON-LD, are still brittle, and their standards are poorly implemented.[14] Moreover, when processing RDF outside of SPARQL, there is an apparent "impedance mismatch" similar to the one between relational databases and object-oriented programming, that is caused by the RDF data model that does not match the data structures available in common programming languages.

The future work of this research lies primarily in completing the implementation of the representative selection method and in evaluation and comparison of all of the devised example generation methods. To improve the similarity computation we intend to weight RDF properties by their selectivity to distinguish properties carrying most information from noise. It is likely that improvements can be also achieved by experimenting with different similarity aggregation functions besides the mean average that is currently used. We also consider extending the proposed methods from path graphs to general graphs patterns, which would require to use topological sort for graph data to produce deterministic order. Once we complete the outlined features of the example generation library, we plan to integrate it into the LODSight visualizer.

References

BRICKLEY, Dan; GUHA, R.V., 2014. *RDF Schema 1.1* [online]. W3C Recommendation 25 February 2014. W3C [cit. 2016-01-03]. Available from WWW: http://www.w3.org/TR/rdf-schema

CASTELLS, Pablo; HURLEY, Neil J.; VARGAS, Saul, 2015. Novelty and diversity in recommender systems. In RICCI, Francesco, ROKACH, Lior, SHAPIRA, Bracha (eds.). *Recommender systems handbook*. 2nd ed. Berlin; Heidelberg, Springer, pp. 881–918. DOI 10.1007/978-1-4899-7637-6_26.

CYGANIAK, Richard; WOOD, David; LANTHALER, Markus, 2014. *RDF 1.1 concepts and abstract syntax* [online]. W3C Recommendation 25 February 2014. W3C [cit. 2016-01-02]. Available from WWW: http://www.w3.org/TR/rdf11-concepts

DROSOU, Marina; PITOURA, Evaggelia, 2009. *Comparing diversity heuristics* [online]. Technical Report TR-2009-05. Ioannina: University of Ioannina, 2009 [cit. 2015-12-29]. Available from WWW: http://cs.uoi.gr/tech_reports/publications/TR-2009-05.pdf

[14] For example, we had to patch the N-Quads parser of jsonld.js to allow comments in data.

DUDÁŠ, Marek; SVÁTEK, Vojtěch; MYNARZ, Jindřich, 2015. Dataset summary visualization with LODSight. In GANDON, Fabien [et al.] (eds.). *The Semantic Web: ESWC 2015 Satellite Events, Portorož, Slovenia, May 31 – June 4, 2015, Revised Selected Papers*. Berlin; Heidelberg: Springer, pp. 36–40. Lecture notes in computer science, vol. 9341. ISBN 978-3-319-25639-9.

EUZENAT, Jérôme; SHVAIKO, Pavel, 2013. *Ontology matching*. Berlin; Heidelberg: Springer. ISBN 978-3-642-38721-0.

GROSS, Jonathan L.; YELLEN, Jay, 2006. *Graph theory and its applications*. 2nd ed. Boca Raton (FL): CRC Press. ISBN 1-58488-505-X.

HARRIS, Steve; SEABORNE, Andy (eds.), 2013. *SPARQL 1.1 Query Language* [online]. W3C Recommendation 21 March 2013. W3C [cit. 2015-12-29]. Available from WWW: http://www.w3.org/TR/sparql11-query

KAUFMAN, Leonard; ROUSSEEUW, Peter J., 1987. Clustering by means of medoids. In DODGE, Yadolah (ed.). *Statistical data analysis based on the L1-norm and related methods*. Amsterdam: North Holland; Elsevier, pp. 405–416.

LONGLEY, David [et al.], 2014. *JSON-LD 1.0 processing algorithms and API* [online]. W3C Recommendation 16 January 2014. W3C [cit. 2015-12-29]. Available from WWW: http://www.w3.org/TR/json-ld-api

MANOLA, Frank; MILLER, Frank, 2004. *RDF primer* [online]. W3C Recommendation 10 February 2004. W3C [cit. 2016-01-05]. Available from WWW: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

PARK, Hae-Sang; JUN, Chi-Hyuck, 2009. A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*. Vol. 36, pp. 3336–3341. DOI 10.1016/j.eswa.2008.01.039.

PRESUTTI, Valentina [et al.], 2011. Extracting core knowledge from linked data. In *Proceedings of the Second International Workshop on Consuming Linked Data, Bonn, Germany, October 23, 2011* [online]. Aachen: RWTH Aachen University [cit. 2016-01-26]. CEUR workshop proceedings, vol. 782. Available from WWW: http://ceur-ws.org/Vol-782/PresuttiEtAl_COLD2011.pdf

SPORNY, Manu [et al.], 2014. *JSON-LD 1.0: A JSON-based serialization for linked data* [online]. W3C Recommendation 16 January 2014. W3C [cit. 2015-12-29]. Available from WWW: http://www.w3.org/TR/json-ld

TROULLINOU, Georgia [et al.], 2015a. RDF Digest: efficient summarization of RDF/S KBs. In GANDON, Fabien [et al.] (eds.). *The semantic web: latest advances and new domains. Proceedings of the 12th European Semantic Web Conference, ESWC 2015, Portoroz, Slovenia, May 31 – June 4, 2015*. Berlin; Heidelberg: Springer, pp. 119–134. Lecture notes in computer science, vol. 9088. DOI 10.1007/978-3-319-18818-8_8.

TROULLINOU, Georgia [et al.], 2015b. Ontology understanding without tears: the summarization approach. *Semantic Web: Interoperability, Usability, Applicability.* Online version of a paper under review. Available from WWW: http://www.semantic-web-journal.net/content/ontology-understanding-without-tears-summarization-approach. ISSN 1570-0844.

JEL Classification: C88

Acknowledgements: This research was supported by the VŠE IGA F4/90/2015. The author would like to thank to Paolo Tomeo and Václav Zeman for suggesting improvements to the presented work.

Shrnutí

Generování příkladů cest sumarizujících RDF datasety

Se zvyšováním rozsahu datových sad se snižuje možnost jim přímo porozumět a vzrůstá proto potřeba sumarizace dat. Shrnutí dat může nabývat podoby typických vzorů, které se v datech nacházejí. Na základě těchto vzorů je možné datům porozumět na obecné úrovni. Nicméně, toto shrnutí dat může být obohaceno konkrétními příklady vybraných vzorů. Je-li to možné, příklady vzorů poskytnuté uživateli by měly dobře reprezentovat vzory, které ilustrují. V tomto článku představujeme 3 metody pro generování příkladů vzorů nalezených v datových sadách ve formátu RDF. Tyto metody zahrnují náhodný výběr, výběr rozdílných příkladů a výběr reprezentativních příkladů. Prezentované metody výběru jsou založené na náhodnosti, diverzifikaci a clusterování.

Kromě návrhu metod text popisuje kompromisy zvažované při jejich implementaci. Soustředíme se na vzory v podobě cest RDF grafy, které se skládají z tříd nebo datových typů, které jsou propojeny RDF predikáty. V článku je prezentován RDF/S slovník pro popis těchto grafových cest a jejich instancí.

Klíčová slova: sumarizace, diverzita, Resource Description Framework

Využití analýz obsahu Voice of Customer v marketingu

Lucie Šperková

lucie.sperkova@vse.cz

Doktorand oboru Aplikovaná informatika

Školitel: doc. Ing. Ota Novotný, Ph.D, (novotny@vse.cz)

Abstrakt: Článek prezentuje široký pohled na problematiku Voice of Customer a Word of Mouth a jejich analýz pomocí opinion miningu. Cílem textu je rešerše literatury zaměřující se na současný přístup k hodnocení Word of Mouth na marketingovém poli působnosti doplněna o rešerši literatury řešící problematiku opinion miningu a sentiment analýz. Studie ukazuje jakousi mezeru mezi marketingovým pojetím Voice of Customer a informatickým pojetím jeho analýz. Článek popisuje současný stav problematiky a navrhuje další postup k překlenutí této mezery. Text je teoretickým podkladem pro další výzkum integrace dat Voice of Customer do procesů Business Intelligence. Součástí článku je srovnání úspěšnosti metod a algoritmů pro analýzy sentimentu.

Klíčová slova: Voice of Customer, Word of Mouth, opinion mining, sentiment analýzy, marketing, Business Intelligence

Úvod

Současný marketing klade důraz na zákazníka a v prostředí internetu se zaměřuje na vytváření a následné vyhodnocování interakcí. Současné přístupy uchovávají data o zákaznících a marketingových aktivitách v relačních databázích z CRM nebo webových analytických nástrojů jako Google Analytics. Spotřebitelé ovšem na internetu čím dál více sdílí své názory, které mohou výrazně obohatit stávající zákaznické analýzy. Otázkou ovšem zůstává, jak takové množství nestrukturovaných dat integrovat do současných procesů. Manuální metody marketingového výzkumu jsou těžko škálovatelné, pomalé a žádají si o nahrazení automatizovanými data-driven modely. Data generovaná z internetové komunikace skrývají obrovský potenciál pro získávání znalostí, zákaznickou analytiku a efektivnější přístup k marketingu. Analýzy obsahu Voice of Customer dat mohou pomoci najít prioritní klienty, identifikovat zákazníkovy potřeby, problémy týkající se produktů a služeb, zákazníkův sentiment, zákazníkovu hodnotu, zefektivňovat komunikaci mezi zákazníkem a společností, zlepšit kvalitu produktů a služeb, maximalizovat profitabilitu podniku. Výsledky těchto úsilí pak vytváří příležitost upravit marketingové praktiky tak, aby byl zaujat správný přístup těm správným zákazníkům konzistentně přes všechny prodejní kanály. Článek se zabývá současnými přístupy k takovýmto datům z marketingového hlediska a dále se zaměřuje na analýzy obsahu a sentimentu těchto textů. V druhé části potom navrhuje přístup integrace těchto dat do Business Intelligence a jejich využití v marketingu.

Vymezení pojmu Voice of Customer a Word of Mouth

Sdílení zákazníkova chování konání skrz aplikace a sociální sítě různého charakteru a komunikace se společnostmi je na marketingovém poli známo jako Word-of-Mouth (WoM) nebo také Voice of Customer (VoC) či Customer Voice, přesto tyto pojmy nejsou zaměnitelné. Z rešerše literatury lze vyvodit, že pojem Voice of Customer je pojmu Word-of-Mouth nadřazený, přestože pojem WoM se objevuje v literatuře mnohem dříve a často je těmito pojmy chápáno to samé.

Word-of-Mouth marketing je velmi oblíbenou marketingovou metodou, kterou se zabývalo mnoho publikací už v době, kdy neměl s online marketingem nic společného (např. Katz & Lazarsfeld, 1955; Richins, 1983). Anderson (1998) definoval WoM jako přenášení nějaké emoce na základě spokojenosti zákazníka s daným produktem či službou. Helm a Schlei (1998) uvádějí, že WoM není pojem pro komunikaci jako takovou, ale o její monitorování: „Monitorování ústní komunikace, pozitivní i negativní, mezi zákazníky, dodavateli, nezávislými experty o značce, produktech, službách a dalších věcech týkajících se vybrané firmy."

Literatura je celkem bohatá na definice tohoto pojmu. Asi nejobecnější definici lze najít v článku Huang et al. (2012): „WoM je proces, kdy jednotlivec předává svůj osobní postoj jiné osobě."

Hu, Pavlou a Zhang (2006) definují, co může být subjektem tohoto postoje: WoM je veškerá neformální komunikace směrovaná dalším konzumentům o vlastnictví, používání nebo charakteristikách daného produktu či služby nebo jejich prodejců, tedy společnosti a značky.

Hennig-Thurau et al. (2004, p. 39) definují, kdo je tím jednotlivcem a jaký postoj předává. Zároveň už do definice přidává i platformu, skrz kterou je komunikace předávána: „Jakékoliv pozitivní nebo negativní prohlášení učiněné potenciálním, skutečným nebo bývalým zákazníkem o výrobku nebo společnosti, které je dostupné velkému množství lidí a institucí přes internet."

Bronner a de Hoog (2010) Henning-Thurauovu definici pozměnili tak, aby zachycovala nynější a budoucí vývoj komunikace více komplexně: Jakékoliv tvrzení – negativní, pozitivní nebo neutrální, řečeno současnou, potenciální nebo bývalou zúčastněnou stranou o výrobku, službě, společnosti či osobě, dostupné většímu množství lidí, institucí nebo organizací skrz digitálně propojené platformy.

Poslední dvě definice už jsou charakteristiky další evoluční fáze WoM a tou je electronic Word-of-Mouth (eWoM) (Choi & Scott, 2012), v literatuře nazývané také jako digitální (Hu, Pavlou & Zhang, 2006) nebo online (Wu & Zheng, 2012). Tato internetová verze WoM se stává majoritním informačním zdrojem pro zákazníky. Právě eWoM vypovídá často nejlépe o tom, co si lidé o dané firmě a jejích produktech skutečně myslí, a jaké jsou jejich reálné důvody, proč u dané společnosti nakupují nebo právě nakupovat nechtějí. Takové WoM Helm a Schlei (1998) definují jako spontánní WOM. Alternativou spontánního WoM pak míní WoM organické - doporučování dané značky či produktu spokojenými zákazníky. Jakýkoliv konzument na světě se může připojit na internet a přečíst si názory jiných. S internetem pak také dramaticky roste i rozsah této komunikace generující obrovské množství nestrukturovaných dat. Amplifikovaný WoM je výsledkem marketingových aktivit, které cíleně podporují šeptandu mezi lidmi (Helm & Schlei, 1998).

V roce 2004 vznikla oficiální obchodní asociace věnovaná WoM marketingu (WOMMA). Asociace vytvořila terminologický rámec pro popis a měření WoM pro komerční subjekty zabývající se media, brand nebo WoM marketingem. WOMMA (2005) definuje WoM jako: „Akt, kdy spotřebitel vytváří a / nebo distribuuje marketingově relevantní informaci jiným spotřebitelům."

Takzvaný hlas zákazníka, pojem, který se prvně objevuje až v (Griffin & Hauser, 1993) v rámci zlepšování kvality (Quality Function Deployment), je pojmem pro zastřešení zákaznických potřeb. Zákaznická potřeba je zde chápána jako zákazníkův vlastní popis toho, jaké vlastnosti by měl daný product či služba mít, aby bylo dosaženo jeho spokojenosti. Cílem miningu VoC je pak dle Griffin a Hauser (1993) pochopení zákazníkových potřeb a jejich transformace do klíčových funkčních požadavků. Tyto požadavky se objevují právě v různých formách z různých zdrojů:

- ve strukturované formě jako
    - hodnocení pomocí bodových škál
    - výsledky strukturovaných dotazníků
- v nestrukturované formě
    - přímá zpětná vazba společnosti (emaily, hovory, formuláře)
    - výsledky řízených výzkumů (narativní příběhy)
    - spontánní WoM a organické WoM ve formě eWoM, tedy hodnocení z internetových zdrojů jako jsou fóra, blogy, sociální sítě apod.

Zjednodušeně tedy lze říci, že pokud analyzujeme veškerý obsah, který uživatel napsal či řekl, analyzujeme tak jeho „hlas", tedy VoC. Jeho vliv na další účastníky tohoto „hovoru" lze pak řadit mezi WoM.

Dnes má VoC velmi významné důsledky pro širokou škálu manažerských činností, jako je:

- budování značky a pověsti,
- zvyšování konverzí, tedy prodejních transakcí
- získávání a udržení zákazníků,
- vývoj produktu,
- zajištění kvality.

Současné přístupy k analýzám WoM

Vznik internetu a poté různorodých sociálních médií jako blogy, mikroblogy, sociální sítě, fóra, online recenze atd. a především růst popularity e-commerce, byl pro výzkum WoM významným milníkem a dal tak vzniknout jeho elektronické verzi. Dellacoras et al. (2007) poznamenali, že praxe recenzování produktů online výrazně zvyšuje potenciál pro empirické porozumění WoM marketingu. Zatímco artikulovaná hodnocení zmizí krátce potom, co byla vyřčena, a je tedy velmi obtížné je zachytit a analyzovat, online tvrzení přetrvávají dlouho poté, co byla napsána. Breazeale (2009) uvádí, že "digitální platformy mění naše chápání a podstatu významu eWoM: „WoM již nemizí okamžitě a není nutně spontánní. Online recenze jsou snadno přístupné dlouhotrvající záznamy názorů, dostupných každému zdarma, přestože je třeba brát ohled na ověření jejich zdrojů.“

Využití eWoM je popsáno spousty články především z oblasti marketingu. Velká část literatury se zaměřuje na dopad WoM a jeho vliv:

- na nákupní rozhodování a prodeje
- na kvalitu produktu a služeb
- na zákazníkovo chování a postoje
- na zákaznickou loajalitu
- na implikace do marketingu
- na měření WoM

ale také

- na user experience a service experience (př. Pai et al. 2012)
- na kontroly kvality (př. Ashton et al., 2014)
- v sociálních sítích (př. Wu and Zheng 2012)

Mnoho dosud kvantitativně i kvalitativně provedených studií využívá vysoké množství eWoM dat a informací o přispívajících uživatelích na webových platformách a zkoumá důležitost a role vlivu online hodnocení produktů na prodeje a zákazníkovo budoucí rozhodování. Senecal a Nantel (2004) ukázali, že subjekty, které hledaly doporučení, měly tendenci tato doporučení následovat. Většina kvantitativních výzkumů ovšem vůbec nebere v potaz nestrukturovaná data a jejich obsah a kontext, jsou zaměřena pouze na zákaznická hodnocení pomocí škál (např. pět hvězdiček z pěti). Chevalier a Mayzlin (2006) zkoumají vliv zákaznických recenzí na relativní prodeje knih na stránkách Amazonu a Barns and Nobles. Zjistili, že zvýšení průměrného hodnocení vede ke zvýšení relativních prodejů a zároveň, že jednohvězdičkové hodnocení má vyšší dopad než pětihvězdičkové.

Nicméně výsledky jsou různé. Některé výzkumy podporují tvrzení, že online hodnocení má významný vliv na prodeje a zákazníkovo rozhodování (Chevalier & Mayzlin 2006; Seneca & Nantel 2004), zatímco jiní tento pohled napadají (Chen et al. 2004, Godes & Mayzlin 2004). Rozdílné závěry jsou také v případě počtu recenzí a jejich průměrného hodnocení. Někteří tvrdí, že počet recenzí má na prodeje vliv (Liu, 2006) nebo alespoň dodatečný vliv vedle samotných recenzí (Wu & Zheng, 2012), jiní tvrdí, že významné je hodnocení těchto recenzí (Dellacoras et al., 2007).

Hu, Pavlou a Zhang (2006) jdou ještě dál a zkoumají vliv těchto hodnocení na kvalitu produktu. Na datech společnosti Amazon provedli průzkum, zda online hodnocení produktů pomocí ordinární škály skutečně odráží kvalitu těchto produktů. Z empirického testování rozdělení hodnocení zjistili, že 53 % produktů má bimodální nenormální rozdělení ve tvaru U. Pro tyto produkty průměrné skóre nutně neodráží skutečnou kvalitu produktu a může tak poskytnout mylné informace.

Zákaznická hodnocení ovšem mohou být kromě číselného hodnocení také v podobě komentáře nebo obojího najednou. Robson et al. 2013 se navíc ptají, co znamená pětihvězdičkové hodnocení, které není nikde jasně definováno (například jak zákazník určuje hranici mezi hvězdičkami). Zároveň může spotřebitel dát plné hvězdičkové hodnocení, ale k němu napsat komentář hodnotící negativní aspekty produktu či naopak. Obecně jsou tedy i recenzní hodnocení nestrukturované a nesystematické. Proto je vhodné analyzovat i textové komentáře těchto hodnocení, kdy zákazník odůvodňuje svou známku pro daný produkt. Bez plného porozumění WoM nelze hodnotit jeho skutečný dopad. Takových výzkumů ovšem není mnoho. Tsang a Prendergast (2009) ukázali, že zákaznické komentáře měly silnější dopad na ovlivnění nákupního rozhodování a vnímanou důvěryhodnost nežli hodnocení pomocí ordinální

škály. Podobně, Chevalier a Mayzlin (2006) poukázali, že i když spotřebitelé přezkoumali hodnocení pomocí hvězdiček, také četli a aplikovali informace v uvedených komentářích. Zároveň zjistili, že spotřebitelé mají tendence komentovat spíše pozitivně než negativně, což bylo potvrzeno i výzkumem (Robson et al. 2013), kteří k této analýze využili textmingového nástroje Leximancer. Nicméně se ale v obou výzkumech narazilo na zjištění, že negativní hodnocení a komentáře mají větší sílu přesvědčit zákazníka produkt nekupovat, než mají pozitivní komentáře či hodnocení na postrčení k nákupnímu rozhodnutí a generování zisků.

Interpretace VoC

Interpretaci VoC lze dělit typicky na automatizovanou a ne-automatizovanou (manuální) analýzu obsahu. Dle (Robson et al., 2013) je manuální analýza případ lidské klasifikace textu pomocí předem daného klasifikačního systému. Příkladem může být marketingový specialista procházející jednotlivé online nástroje (diskuzní fóra, profily na Facebooku, apod.) a vytvářející si poznámky o průběhu marketingových aktivit, které později využije pro vyhodnocení těchto aktivit vzhledem k nastaveným cílům. Automatickou analýzu provádí informační technologie. Automatický sběr dat může být realizován přímým uložením provedených akcí do databáze nebo umělým aktérem naprogramovaným pro sběr dat na vybraných webových stránkách, tzv. konektor neboli crawler. Je potřeba určit datové zdroje a jakým způsobem bude sběr proveden.

Toto rozdělení je ovšem velmi zjednodušené. V praxi i po automatizovaném zpracování a vyhodnocení dat následuje vždy individuální posouzení výsledků daným specialistou. Manuální sběr, zpracování a vyhodnocení dat je zde chápáno bez pomoci příslušných informatických nástrojů zmiňovaných v rámci pokročilých metod automatizovaného přístupu. Jedná se tedy pouze o individuální sběr a posouzení s využitím základních nástrojů jako např. kancelářský balík Microsoft Office.

Každá z těchto metod má své výhody a omezení a rozhodnutí, kterou použít, záleží na specifikaci a cíli daného projektu. Konkrétně automatizované analýzy mohou vést k vyšší spolehlivosti zjištění, stejně jako větší možnosti zpracovávat velké objemy textových dat. Ovšem jsou omezeny ve své schopnosti odhalit komunikativní záměr použití slov nebo symbolické významy slov. Touto problematikou se zabývá celý obor opinion miningu.

Obsah VoC

Produkt či služba, na který je vznášen požadavek, je vždy zařaditelný do nějaké hierarchie nebo rodiny produktů či služeb. Dle Peng et al. (2012) zákazníkův požadavek na novou vlastnost typicky obsahuje minimálně následující metadata:

- Textový popis žádoucí funkcionality nebo aktuálního problému (událost)
- Související produkty nebo služby (produkt)
- Informace o zákazníkovi: jméno, jaké produkty již vlastní, do jakého segmentu trhu patří, velikost firmy,… (klient)

Obsah VoC sdělení tedy většinou zahrnuje úvod, vysvětlení, popis zkušenosti s produktem nebo službou, srovnání s jinými produkty či službami, ohodnocení a doporučení. Tento obsah může být buď pozitivní nebo negativní a tedy pro podnik buď příznivý, nebo poškozující. VoC tvrzení tedy obsahuje jak subjekt, tak slova jeho hodnocení (názorová slova nebo také slova nesoucí sentiment). Subjekt indikuje společnost nebo její produkt či službu, zatímco slova ohodnocení, nejčastěji adjektiva či adverbia, jsou používána autorem k popisu subjektu, který diskutuje. Evaluace produktů a služeb je často vyjádřena sentimentem, který nesou hodnotící slova.

Obsah VoC může obsahovat jeden, ale i více subjektů. Subjekt, na který je manažer společnosti či konzument zaměřen může být celý popisný odstavec v textu. Ačkoliv jsou počítače schopny najít, jak subjekt, tak hodnotící slova v obsahu, nejsou schopny rozlišit, které hodnocení patří kterému subjektu, a to může způsobit problém s budoucí analýzou textu (Pang & Lee, 2008).

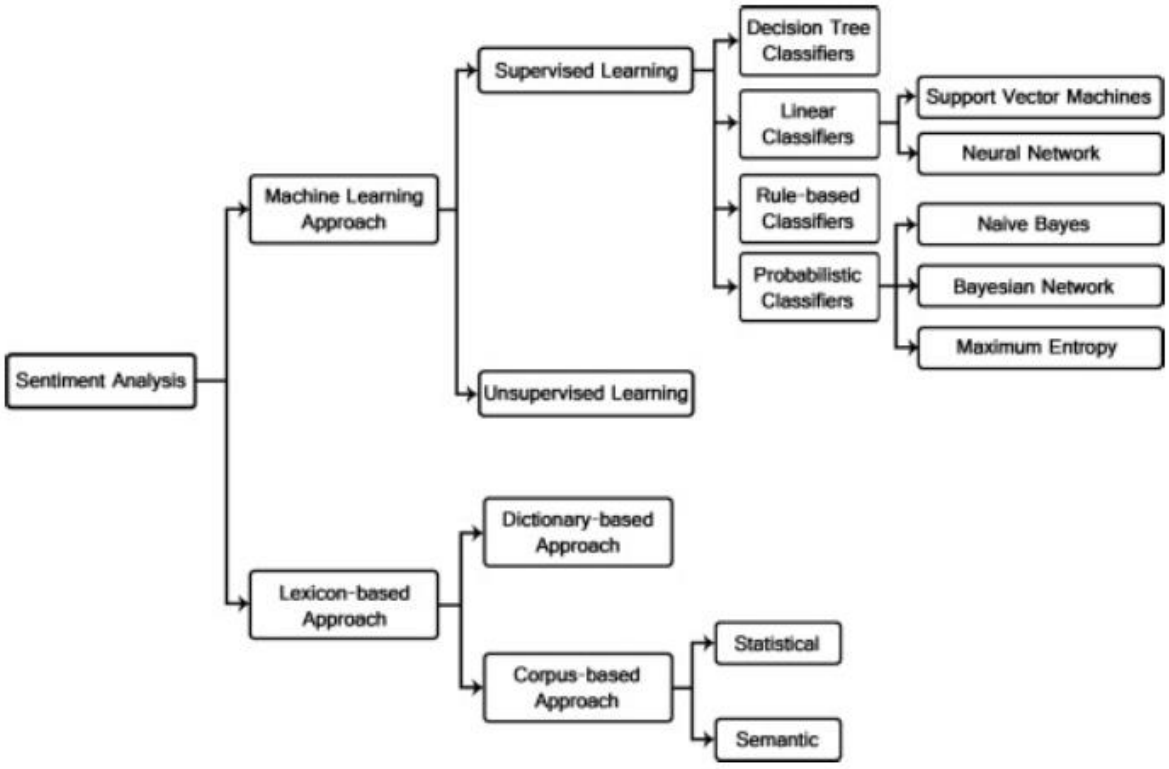
Mnoho dosud provedených studií se také zabývá detekcí tzv. aspektu neboli tématu (př. Titov & McDonald, 2008). Aspektem lze chápat určité téma, pod které lze zařadit dané pojmy v konkrétní doméně, které se často vyskytují spolu. Např. v doméně restaurace mohou být identifikována témata jako

jídlo a pití, atmosféra, personál, cena, atd., pod které lze poté řadit jednotlivá slova, která dané téma reprezentují – v rámci jídla a pití to bude např. omáčka, rýže, kořeněný, láhev apod. Sentiment pak může být přiřazován celému tématu.

Význam z toho obsahu je pak dolován pomocí metod opinion miningu. Opinion mining a analýzy sentimentu lze chápat jako synonyma. V literatuře je tento proces známý také jako analýza subjektivity nebo také extrakce hodnocení (v angličtině appraisal extraction) (Pang & Lee, 2008). Jedná se o proces sledování a extrahování znalosti veřejnosti o nějakém produktu či problému. Je to aplikace získávání znalostí z textu pomocí technik zpracování přirozeného jazyka, matematické lingvistiky a dolování dat k určení objektivní a subjektivní informace, kterou poskytují ostatní (Khan et al., 2009).

Metody pro analýzy sentimentu

Existují dva základní přístupy k analýzám sentimentu: založené na strojovém učení (Machine Learning) a založené na slovní zásobě (lexicon based). Metody řazené pod tyto přístupy ukazuje Obrázek 1.



Obrázek 6: Metody klasifikace sentimentu Zdroj: Medhat et al., 2014

Přístupy na základě Machine Learningu používají klasifikační techniky ke klasifikaci textu. Metody založené na slovní zásobě využívají sentiment definovaný ve slovnících s názorovými slovy (opinion words) pro rozeznání polarity. Přiřazují sentiment skóre k názorovým slovům, která pak najdou v datech, a určí, jak pozitivní, negativní nebo objektivní slovo obsažené ve slovníku je. Některé výzkumy tyto metody pro získání lepší výkonnosti kombinují. Pod sémantické přístupy lze zařadit

- Latentní sémantické asociace,
- Latentní Dirichletovy alokace,

Pod statistické pak

- Latentně proměnné modely, jako skryté Markovské modely (HMM),
- Podmíněná náhodná pole (CRF),
- Bodová vzájemná informace (PMI).

Díky velkému množství technik spousta výzkumníků experimentuje s různými algoritmy a srovnává je mezi sebou. Dohromady bylo v rešerši indikováno kolem padesáti výzkumů. V tomto článku ukazuji pouze některé výsledky těchto srovnání v Tabulce 1.

Tabulka 4: Využité techniky pro sentiment analýzy a jejich přesnost v dřívějších výzkumech Zdroj: autor

Autoři	Zdroj dat	Technika
Pang et al. (2002)	filmová review	NB (81,5%), ME (81%), SVM (82,9%)
Turney (2002)	filmová review	PMI (65,8%)
Dave et al. (2003)	produktová review	SVM (85,8 - 87,2%), NB (81,9 - 87%)
Hu & Liu (2004)	produktová review	slovník (84%)
Abbasi et al. (2008)	web fórum	SVM, entropie vážená genetickým algoritmem + selekce příznaků (95,55%)
Khan et al. (2011)	filmová review	slovník (86,6%)
Zhang et al. (2011)	Twitter	SVM + slovník (85,4%)
Mudinas et al. (2012)	produktová a filmová review	ML + slovník (82,3%)
O'Kefee & Koprinska (2006)	filmová review	NB, SVM + metoda vážení příznaků (87,15%)
Socher et al. (2013)	filmová review	Recursive Neural Tensor Network + rozhodovací strom (80,7%)

Techniky supervizovaného strojového učení ukázaly ve výzkumech relativně lepší výsledky než nesupervizované na slovní zásobě založené metody. Nicméně i tyto metodu jsou důležité, protože supervizované učení vyžaduje velké množství trénovacích dat, což je velmi náročný úkol oproti získávání nových neoznačených dat s rostoucím množstvím internetové komunikace. Pokud jsou trénovací data nedostatečná, může klasifikace pak selhat. Většina domén, kromě často analyzovaných filmových recenzí, také postrádají označená tréninková data; v tomto případě jsou nesupervizované metody pro vývoj aplikací velmi užitečné. Většina výzkumů potvrzuje, že Support Vectore Machine klasifikátor (SVM) má relativně lepší výkon než jiné algoritmy. Názorová slova obsažena ve slovnících jsou pro metody založené na slovní zásobě velmi důležitá.

V případě, že slovník obsahuje málo slov nebo je nedůkladný, může být snížen výkon analýz. Dalším významným problémem tohoto přístupu je závislost polarity sentimentu na konkrétní doméně nebo kontextu u mnoha slov, které aktuální slovníky neumí zachytit. Tento problém se snaží aktuálně řešit mnoho článků. Hlavní výhodou hybridního přístupu pomocí kombinace slovníkových metod a metod strojového učení je možnost získat to nejlepší z obou přístupů s vysokou přesností.

Návrh využití analýz VoC v marketingu

Z provedené rešerše vyplynulo, že existuje jistá mezera mezi analýzami VoC pro marketingové účely a sentiment analýzami prováděnými především na poli informatiky, přestože se obě oblasti zabývají stejnými daty. Problém je, že marketéři se zaměřují spíše na kvantitativní metriky množství a významu recenzí a jejich vliv na prodeje (opět vyjádřeny v tvrdých ukazatelích) než na metriky, které by vycházely ze samotného textu. U článků, které se týkají přímo analýz textu, se výzkumníci zase více zabývají úspěšnosti algoritmů, které k analýzám využili, než nějakým konkrétním využití toho, co díky analýze zjistili. Co se týče financí, tak sociální média a fóra byla již využívána jako indikátor veřejného vnímání akciového trhu. Například Antweiler a Frank (2004) zjistili, že celkový sentiment může pomoci předvídat volatilitu trhu, a že rozptyl ratingových kategorií je spojen se zvýšeným objemem obchodování. U marketingu ovšem praktické využití stále není úplně odkryto. Přesto existují některé snahy systémově zahrnout sentiment analýzy do nákupního rozhodování. Například Popescu a Etzioni (2005) navrhli nesupervizovaný systém jak extrahovat důležité vlastnosti produktu pro potenciální zá-

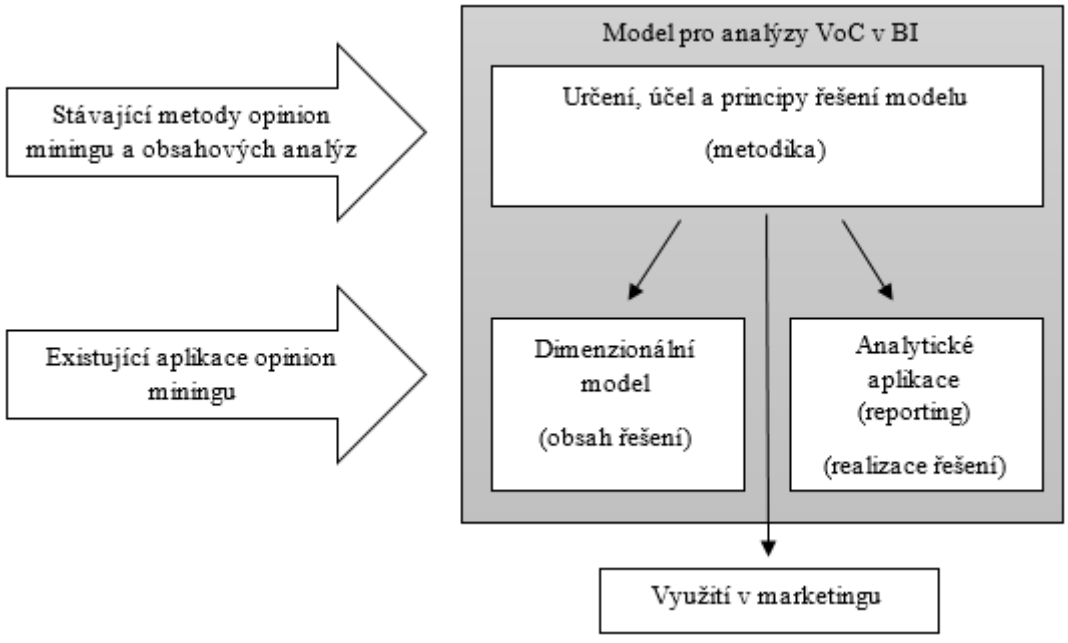
kazníky. Liu et al. (2005) navrhli rámec, který srovnává zákaznické názory na konkurenční produkty. Tyto snahy mohou sloužit jako prvotní výpočetní infrastruktura pro sociálními médii řízené Business Intelligence.

Procesy zahrnutí VoC do BI

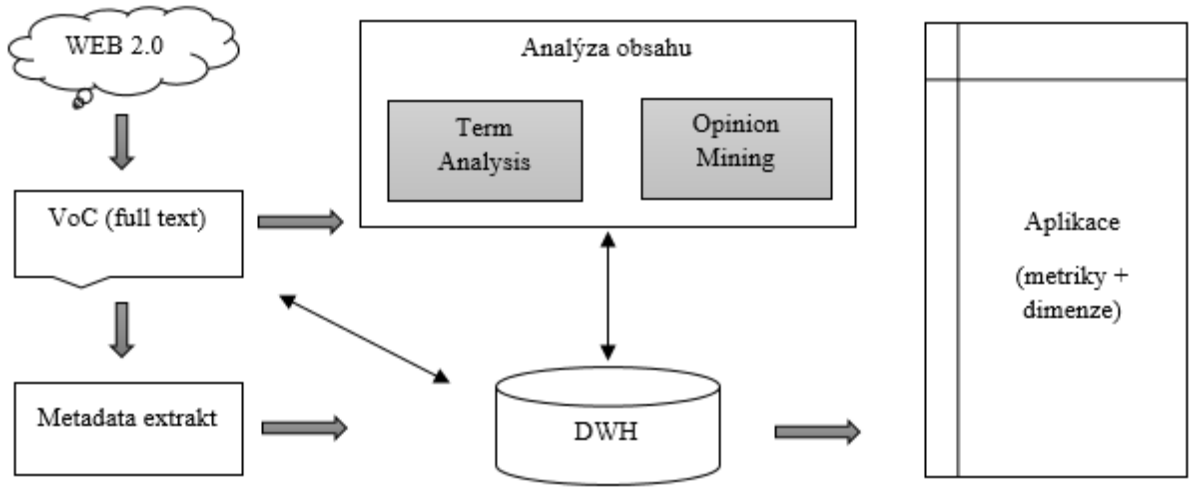
V rámci rešerše byli také identifikováni autoři, kteří se snaží analýzy VoC integrovat do jakéhosi rámce či metodiky, která by měla být nějakým způsobem využitelná v Business Intelligence. Chau a Xu (2012) navrhli rámec pro analýzu obsahu z blogů pro Business Intelligence. Hybridní rámec integrující doménovou znalost s daty řízenými přístupy pro analyzování semistrukturovaných zákaznických požadavků zase vytvořili (Peng et al., 2012). Metodologii pro zpřesnění textminingového a opinionminingového nástroje o doménově specifické znalosti pro získávání přesnějšího sentimentu z dat skrz různé nástroje Web 2.0 vytvořili (Varadarajan & Soundarapandian, 2013).

Jako nejsofistikovanější přístup k integraci strukturovaných a nestrukturovaných dat považuji článek Baars & Kamper (2008). Ti přístupy zasazují do třívrstvého rámce Business Intelligence: 1) komponenta fungující jako můstek přidaná mezi logickou a prezentační vrstvu, která spojuje funkcionalitu vyhledávání a navigace v datech, 2) middleware hub pro výměnu výsledků a šablon mezi analytickými systémy a komponentami distribuce znalosti, 3) nástroje pro analýzu obsahu a extrakci metadat. Nedostatkem je, že se zaměřují pouze na interní data společnosti, mezi zdroje nezahrnují obsah VoC z internetových zdrojů.

Pro potřeby analýz obsahu VoC skrze procesy BI, kde je možné získané informace z VoC kombinovat z již existujícími strukturovanými daty v datovém skladu, je potřeba definovat model, který bude sloužit jako referenční model pro budoucí realizaci řešení. Takový model zobrazuje Obrázek 2. Nutnou podmínkou řešení je jeho jednoduchost a přenositelnost. Předpokladem je využití stávajících metod a technik k analýzám obsahu a sentimentu příspěvků v rámci analytických procesů nad datovým skladem. Jednou z možností je jejich aplikační napojení, kdy se tyto výpočty budou dít mimo základní ETL procesy a ukládat zpětně do skladu, kde budou výsledná data kombinována s daty strukturovanými získanými např. z CRM (Obrázek 3).



Obrázek 7: Referenční model pro VoC analýzy v BI Zdroj: autor



Obrázek 8: Návrh integrace VoC do BI Zdroj: autor

Možné využití analýz VoC v marketingu

Protože cílem budoucího výzkumu je integrace dat VoC do procesu Business Intelligence, je nutné definovat, k jakému účelu budou tato data sbírána a analyzována, tedy k jakým konkrétním akcím budou použita. Tento návrh se zaměřuje zejména na zákaznickou analytiku:

- Detekce produktů, o kterých zákazník mluví pozitivně a následné zacílení reklamy. V tomto případě lze využít metod cross-sellingu i up-sellingu v případě, že známe identitu zákazníka a ten už u nás nakoupil. Jedná se o retenční aktivitu.
- Predikce, zda zákazník zůstane dále zákazníkem nebo odejde. Řízení churn managemetu.
- Segmentace zákazníku pro cílení reklamy.
- Identifikace sentimentu přímé komunikace pro prioritní vyřizování zákaznických požadavků.
- Zacílení na přátele zákazníků, kteří mluví pozitivně o produktu (v rámci sociálních sítí).
- Korelace vývoje počtu nových zákazníků s celkovou náladou VoC.
- Korelace vývoje tržeb zákazníků s celkovou náladou VoC.
- Analýza ovlivnění spočítané CLV zákazníka dle toho, co píše.
- Přerozdělení rozpočtu do kampaní dle vypočítané atribuce přirozeného WoM.
- Detekce takzvaných opinion holderů (ti, co šíří názory a ostatní je poslouchají) a následné zacílení reklamy.

Závěr

V rámci článku byla zjištěna jistá mezera mezi přístupy k datům VoC ze strany informatiky a marketingu. Marketérům chybí zaměření se na samotné analýzy obsahu textu automatickou cestou, informatikům, kteří tyto analýzy provádět umí, zase chybí konkrétní aplikace do marketingu. Existuje jen málo studií snažící se systémově zahrnout obsahové analýzy do marketingových aktivit. Jako překlenutí této mezery navrhuje autorka integraci VoC do procesů Business Intelligence, kde v rámci jeho analytické části je potřeba definovat potřebné ukazatele a dimenze, které budou sloužit pro marketingové účely. Článek zároveň navrhuje některé možnosti konkrétních analýz. Pro potřeby budoucího výzkumu byl navržen referenční model, na kterém bude postavena konkrétní podrobná metodika a realizace. Vstupem budou již existující techniky, algoritmy a metody obsahových analýz, zejména sentimentu, jejichž komparativní analýza je také součástí článku.

Literatura

ABBASI, Ahmed; CHEN, Hsinchun; SALEM, Arab. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums.ACM Transactions on Information Systems (TOIS), 2008, 26.3: 12.

ANDERSON, Eugene W. Customer satisfaction and word of mouth. Journal of service research, 1998, 1.1: 5-17.

ANTWEILER, Werner; FRANK, Murray Z. Is all that talk just noise? The information content of internet stock message boards. The Journal of Finance, 2004, 59.3: 1259-1294.

ASHTON, Triss; EVANGELOPOULOS, Nicholas; PRYBUTOK, Victor R. Quantitative quality control from qualitative data: control charts with latent semantic analysis. Quality & Quantity, 2015, 49.3: 1081-1099.

BAARS, Henning; KEMPER, Hans-George. Management support with structured and unstructured data—an integrated business intelligence framework. Information Systems Management, 2008, 25.2: 132-148.

BONE, Paula Fitzgerald. Word-of-mouth effects on short-term and long-term product judgments. Journal of business research, 1995, 32.3: 213-223.

BREAZEALE, Michael. Word of mouse: An assessment of electronic word-of-mouth research. International Journal of Market Research, 2009, 51: 297-318.

BRONNER, Fred; DE HOOG, Robert. Consumer-generated versus marketer-generated websites in consumer decision making. International Journal of Market Research, 2010, 52.2: 231-248.

DAVE, Kushal; LAWRENCE, Steve; PENNOCK, David M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In:Proceedings of the 12th international conference on World Wide Web. ACM, 2003. p. 519-528.

DELLAROCAS, Chrysanthos; ZHANG, Xiaoquan Michael; AWAD, Neveen F. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive marketing, 2007, 21.4: 23-45.

GODES, David; MAYZLIN, Dina. Using online conversations to study word-of-mouth communication. Marketing science, 2004, 23.4: 545-560.

GRIFFIN, Abbie; HAUSER, John R. The voice of the customer. Marketing science, 1993, 12.1: 1-27.

HELM, S.; SCHLEI, J. Referral potential–potential referrals. An investigation into customers' communication in service markets. In: Track 1 Market Relationships, Proceedings 27th EMAC Conference, Marketing Research and Practice. 1998. p. 41-56.

HENNIG-THURAU, Thorsten, et al. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?. Journal of interactive marketing, 2004, 18.1: 38-52.

HU, Minqing; LIU, Bing. Mining and summarizing customer reviews. In:Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004. p. 168-177.

HU, Nan; PAVLOU, Paul A.; ZHANG, Jennifer. Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication. In: Proceedings of the 7th ACM conference on Electronic commerce. ACM, 2006. p. 324-330.

HUANG, Eric H., et al. Improving word representations via global context and multiple word prototypes. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 2012. p. 873-882.

CHAU, Michael; XU, Jennifer. Business intelligence in blogs: Understanding consumer interactions and communities. MIS quarterly, 2012, 36.4: 1189-1216.

CHEN, Pei-Yu; WU, Shin-yi; YOON, Jungsun. The impact of online recommendations and consumer feedback on sales. ICIS 2004 Proceedings, 2004, 58.

CHEVALIER, Judith A.; MAYZLIN, Dina. The effect of word of mouth on sales: Online book reviews. Journal of marketing research, 2006, 43.3: 345-354.

CHOI, Jae H.; SCOTT, Judy E. Electronic word of mouth and knowledge sharing on social network sites: a social capital perspective. Journal of theoretical and applied electronic commerce research, 2013, 8.1: 69-82.

KHAN, Aurangzeb; BAHARUDIN, Baharum; KHAN, Khairullah. Sentiment classification from online customer reviews using lexical contextual sentence structure. In: Software Engineering and Computer Systems. Springer Berlin Heidelberg, 2011. p. 317-331.

KHAN, Khairullah, et al. Mining opinion from text documents: A survey. In: Digital Ecosystems and Technologies, 2009. DEST'09. 3rd IEEE International Conference on. IEEE, 2009. p. 217-222.

LAZARSFELD, Paul F.; KATZ, Elihu. Personal influence: the part played by people in the flow of mass communications. Glencoe, Illinois, 1955.

LIU, Bing; HU, Minqing; CHENG, Junsheng. Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th international conference on World Wide Web. ACM, 2005. p. 342-351.

MEDHAT, Walaa; HASSAN, Ahmed; KORASHY, Hoda. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 2014, 5.4: 1093-1113.

MUDINAS, Andrius; ZHANG, Dell; LEVENE, Mark. Combining lexicon and learning based approaches for concept-level sentiment analysis. In:Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining. ACM, 2012. p. 5.

O'KEEFE, Tim; KOPRINSKA, Irena. Feature selection and weighting methods in sentiment analysis. In: Proceedings of the Australasian document computing symposium. 2009. p. 67-74.

PAI, Mao-Yuan, et al. Electronic word of mouth analysis for service experience. Expert Systems with Applications, 2013, 40.6: 1993-2006.

PANG, Bo; LEE, Lillian. Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2008, 2.1-2: 1-135.

PANG, Bo; LEE, Lillian; VAITHYANATHAN, Shivakumar. Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002. p. 79-86.

PENG, Wei, et al. Mining the "voice of the customer" for business prioritization.ACM Transactions on Intelligent Systems and Technology (TIST), 2012, 3.2: 38.

POPESCU, Ana-Maria; NGUYEN, Bao; ETZIONI, Oren. OPINE: Extracting product features and opinions from reviews. In: Proceedings of HLT/EMNLP on interactive demonstrations. Association for Computational Linguistics, 2005. p. 32-33.

RICHINS, Marsha L. Negative word-of-mouth by dissatisfied consumers: A pilot study. The journal of marketing, 1983, 68-78.

ROBSON, Karen, et al. Making sense of online consumer reviews a methodology. International Journal of Market Research, 2013.

SENECAL, Sylvain; NANTEL, Jacques. The influence of online product recommendations on consumers' online choices. Journal of retailing, 2004, 80.2: 159-169.

SOCHER, Richard, et al. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP). 2013. p. 1642.

TITOV, Ivan; MCDONALD, Ryan. Modeling online reviews with multi-grain topic models. In: Proceedings of the 17th international conference on World Wide Web. ACM, 2008. p. 111-120.

TSANG, Alex SL; PRENDERGAST, Gerard. Is a "star" worth a thousand words? The interplay between product-review texts and rating valences.European Journal of Marketing, 2009, 43.11/12: 1269-1280.

TURNEY, Peter D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002. p. 417-424.

VARADARAJAN, S. & SOUNDARAPANDIAN. Maximizing Insight from Unstructured Data. Business Intelligence Journal, 2013, 18.3: 17–26.

WORD OF MOUTH MARKETING ASSOCIATION a OTHERS, 2005. WOMMA terminology framework. Framework [online]. Dostupné z: http://scholar.google.com/scholar?hl=en&btnG=

Search&q=intitle:WOMMA+Terminology+Framework#0\nhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:WOMMA+terminology+framework#0\nhttp://i-wisdom.typepad.com/iwisdom/files/womma_term_framework.pdf

WU, Weifang; ZHENG, Rong. The impact of word-of-mouth on book sales: review, blog or tweet?. In: Proceedings of the 14th Annual International Conference on Electronic Commerce. ACM, 2012. p. 74-75.

ZHANG, Ley, et al. Combining lexiconbased and learning-based methods for twitter sentiment analysis. HP Laboratories, Technical Report HPL-2011, 2011, 89.

JEL Classification: M3

Summary

Utilization of Voice of Customer Content Analyses in Marketing

Paper presents a broad view of the issue of Voice of the Customer and Word of Mouth and their analysis using opinion mining. It examines current approaches to evaluation of Word of Mouth in marketing field and also solves the issue of sentiment analyses. The study shows a gap between the marketing concept of Voice of Customer and informatics concept of its analysis. Marketers lack focus on the actual content analysis of the VoC in automatic way; IT professionals who can perform these analyses missing a specific application in marketing. There are few studies attempting to systematically include content analysis in marketing activities. As a bridging this gap, the author proposes the integration of VoC to Business Intelligence processes, where in its analytical part is needed to define the necessary indicators and dimensions that will be used for marketing purposes. The article suggests some possible specific applications of VoC content analyses for marketing activities. For future research reference model was designed for further design of specific detailed methodology and implementation. Existing techniques, algorithms and methods of content analyses, especially sentiment analyses will be the input. The comparative analysis of opinion mining methods and techniques shows that there are many possibilities how to perform this analyses. Author proposes to perform analyses outside the ETL processes and integrate the results to data warehouse with the possible access to full-text.

Keywords: Voice of Customer, Word of Mouth, opinion mining, sentiment analyses, marketing, Business Intelligence

Zaměstnanci v globálních modelech poskytování IT služeb formou outsourcingu

Jindra Tumová

Jindra.tumova@gmail.com

Doktorand oboru Aplikovaná informatika

Školitel: doc. Ing. Zora Říhová, CSc.

Abstrakt: Rostoucí konkurenční prostředí spolu s rychlými změnami technologií nutí společnosti optimalizovat provozy podpůrných služeb a soustředit se na vlastní obchod/činnosti. I přes to, že společnosti se stávají závislými na informačních technologiích více a více, část informačních technologií se stává komoditou a naopak část inspiruje k novým obchodům a formám působení společností. Nadnárodní společnosti poskytující služby informačních technologií formou ousourcingu neustále vyvíjí globální modely poskytování těchto služeb tak, aby byly schopny nabídnout jak industrializované IT služby z globálních produkčních center, tak i specifické služby vyžadující vysoké zákaznické porozumění. Poskytování IT služeb z globálních center formou outsourcingu má dopady na zaměstnance, jak v globálním tak lokálním produkčním centru a to jak na požadavky jednotlivých IT profesí, tak na dovednosti a znalosti IT profesionála a požadované profese.

Klíčová slova: outsourcing, globální model, zaměstnanci, metodika, ITIL, komodita, industrializace,

Úvod

Ekonomická situace, společenské změny společně s rychlými technologickými změnami staví společnosti před otázku, jaký business model bude co nejúspěšnější a jak restrukturalizovat řízení a provoz ke zvýšení konkurenceschopnosti. (Tříska 2001) 82% respondentů – světových business managerů odpovědělo, že pozice IT v předpokládané restrukturalizaci je prioritní. Významné světové společnosti transformovaly či restrukturalizovaly IT funkce: vytvořením sdílených center, kompetenčních center, centralizací, přesunem výkonu IT do levnějších – nízkonákladových lokalit či outsourcingem IT funkcí.

Potenciální benefity outsourcingu, jako je snížení nákladů, jejich snadné řízení, predikovatelnost a transparentnost, zvýšení kvality poskytovaných služeb včetně specifických znalostí, zlepšení business modelu a soustředění na vlastní business, zvýšení flexibility či realizaci specifických projektů formou outsourcingu natolik již dnes převyšují možná rizika, že na růst služeb poskytovaných formou outsourcingu připadá 60 % veškerého růstu globálního trhu IT (Noska 2013).

Outsourcing podpůrných částí organizace, jako je IT, s sebou nese příležitost využít jak nejmodernějších technologií, tak i ostatních trendů - například využití globálních center pro standardní část služeb, přesunutí části služeb za hranice státu, avšak v rámci kontinentu (nearshoring), využití zdrojů a přesun služeb mimo kontinent (offshoring) či kombinovat předcházející včetně subdodavatelů a vlastních zdrojů k nejrychlejšímu dosažení cíle.

Dle (Sokolovsky 2010) „Rostoucí dostupnost outsourcingových a offshoringových možností zvyšuje atraktivnost těchto modelů.

- Mnoho IT služeb je nabízeno externími poskytovateli IT služeb a tudíž je lze nakupovat externě.
- IT služeb jsou stále častěji dodávány a nabízeny ve standardizované, strukturované, předdefinované formě. Projekty a aplikace již nestojí v centru podpory IT. Technické detaily zůstávají skryty. Vývojáři se mohou zaměřit na obchodní problémy a specifikaci hardware a software mohou přenechat dodavatelům.
- IT je ve stále větší míře dokonce schopna spoluutvářet obchodní procesy svých klientů. Změny v pracovních postupech klientů jdou ruku v ruce se změnami IT. Předpokladem tohoto trendu je však vysoký stupeň zralosti poskytovatele IT služeb.

Vedle těchto výhod existuje celá řada tržních trendů, které urychlují rozdělení na IT supply a IT demand.“

Vzhledem k trendu využití nových modelů poskytování IT služeb je hlavními cíli článku:

- popsat globální modely poskytování IT služeb, jejich vývoj od poskytování služeb z různých lokalit, off shoring, k plnému globálnímu modelu,
- definovat základní pravidla služeb poskytovaných lokálně a globálně,
- charakterizovat vývoj požadavků na zaměstnance během posledních 20ti let, požadavků v globálních modelech a požadavků při IT službě jako komoditě
- popsat dopady na zaměstnance při přesunu služeb do globálního produkčního centra.

V práci je použita jak světová literatura, citace článků odborné IT komunity, výzkumy vybraných světových agentur, tak i praktické zkušenosti autorky z vedení lokálního produkčního centra společnosti Siemens a přesunu/poskytování služeb do globálních center.

Globální modely poskytování IT služeb

Neustálé zvyšování konkurence, globalizace trhů i globální působení společností vedou jak ke změnám poskytování služeb, tak především k jednotnému konceptu řízení, jednotným produktům a službám, versus zvláštnosti přístupu k zákazníkům. (Voříšek 2003) uvádí, že pro globálně působící firmy je typický přechod od mechanistického k holistickému přístupu k organizaci a řízení podniku.

Při globalizaci společností CIO uvažovali o globální expanzi IT částí a stěžejním bodem pro řešení byla governance. Rostly obavy, že bez důkladně postaveného řízení, rozhodování, dohledu a dokonce i jednoduchém přehledu o IT se rychle IT organizace stane zmatenou.

Dle (Pastore 2008) zásadní pozornost v globální správě je model řízení. Má IT organizace fungovat centrálně, lokálně nebo kombinovaně? Není dokonalý model, model, který vyhovuje v jedné fázi životního cyklu jedné organizaci, může být nepraktický pro jinou organizaci. Nicméně, většina z dotázaných CIO doporučuje použití centralizovaného modelu. To znamená, že firemní CIO a užší vedení jsou odpovědné za rozhodnutí, jako je IT strategie, priority projektu, výběr systémů a metodiky vývoje aplikací.

Příklady modelů řízení IT:

- centrální model řízení IT – lokální či dceřiné společnosti plně naplňují IT strategii, metodologii, a standardy mateřské společnosti - např. ve společnostech Siemens, CitiBank, T-Mobile
- lokální model IT – lokální či dceřiné společnosti nemají společnou strategii, je plně na lokálním CIO jaké technologie, metodologie budou pro IT použity pro podporu a naplnění lokálních obchodních cílů např. společnosti Continental, Reifeisenbank.
- kombinovaný model řízení – lokální či dceřiné společnosti využívají některé služby mateřské společnosti, ale lokální strategii, výber systémů, metodologií je plně na lokálním CIO, příkladem může být např. Česká spořitelna, člen finanční skupiny Erste Group.

Poskytování služeb z různých lokalit - vývoj

Všichni hlavní poskytovatelé IT služeb vyvíjejí komplexní globální delivery modely, které využívají globálních delivery center poskytujících služby napříč zeměmi. Na druhé straně je stále potřeba poskytovat IT služby také na lokální úrovni.

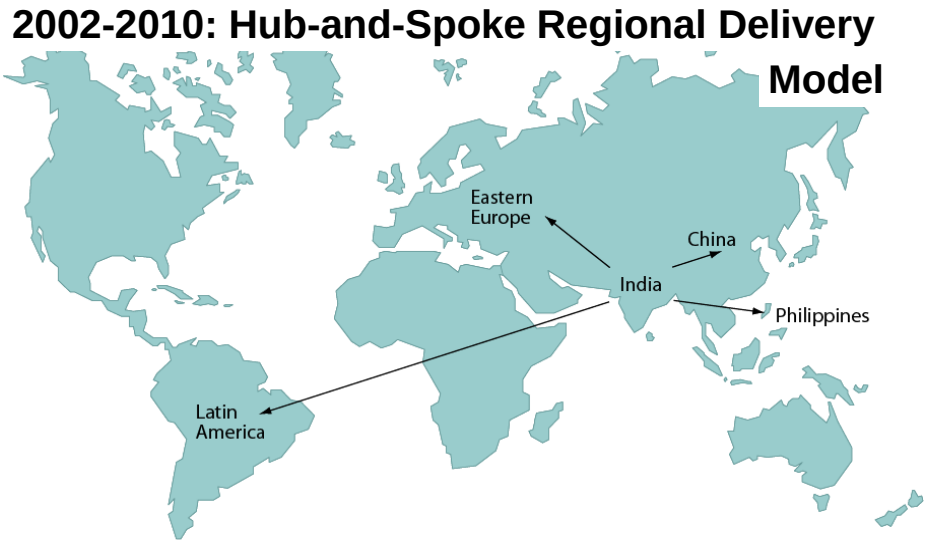
Využití globálních zdrojů se stalo nevyhnutelným, pokud poskytovatel služeb chce vytvářet služby na základě nejnižších nákladů. To znamená, že poskytovatelé služeb potřebují vybudovat delivery schopnosti v zemích s nízkými náklady.

Po řadu let bylo dodávání IT služeb prováděno jednotkami umístěnými v zemi nebo dokonce v místě zákazníka. Ještě na konci minulého století, kdy více a více společností částečně využívalo offshore lokalit k obsluze svých zákazníků na základě nižších nákladů, bylo úkolem dodávající jednotky v zemi zákazníka starat se o management těchto offshore služeb. (Salonen 2010) toto nazývá ‘Point-to-Point’ offshore model.

Obrázek 9:Point to Point offshore model (Salonen 2010)

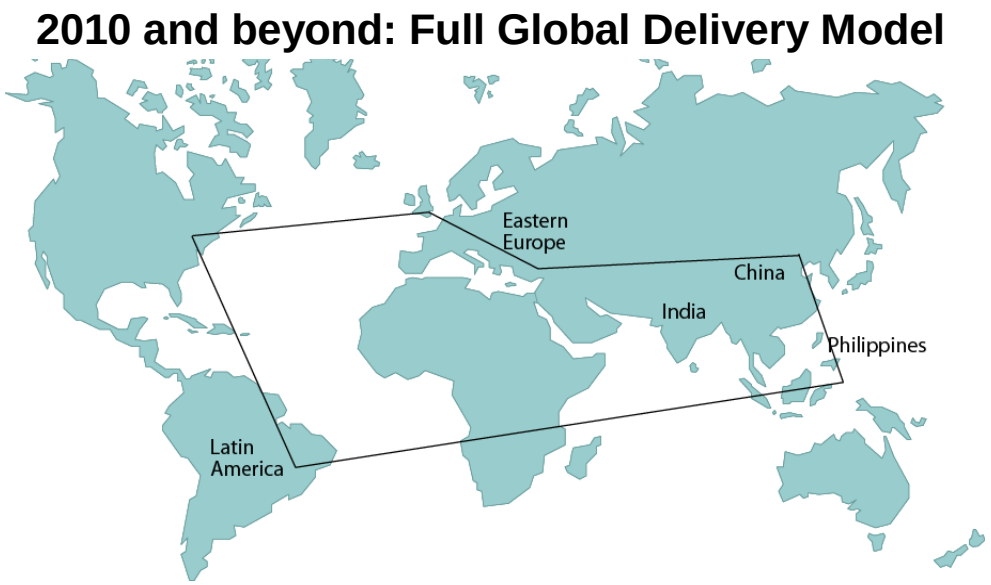
Tento model, založený především na nízkých cenách se ukázal jako nedostatečný - projekty i služby se prodražují následkem nedostatečného řízení. Na přelomu století, kdy se hlavní tržní trend industrializace IT služeb stal více a více dominantní se ukázalo, že je potřeba tento trend aplikovat na offshore lokality tak, aby globální model řízení využíval jak nízkonákladové lokality, tak v těchto lokalitách budoval střediska poskytující industrializované služby. Zatímco global delivery management se v minulosti zaměřoval na levnou pracovní sílu, hodnota je stále více viděna v efektivních procesech a v přidané hodnotě duševního vlastnictví které může být přesouváno napříč zeměmi.

Následkem toho poskytovatelé IT služeb začali budovat delivery centra v nízkonákladových lokalitách, které byly schopny dodávat standardní služby do mnoha různých zemí. Aby mohli definovat standardní služby, procesy a nástroje mnoho poskytovatelů služeb se nejprve zaměřilo na omezené množství nízkonákladových oblastí. Dokonce se ukázalo, že tradiční Indické společnosti se zaměřily na Indii jakožto své jádro a poté rozšiřovaly best practices do satelitních center v dalších zemích. (Salonen 2010) nazývá tento model jako Hub and Spoke Regional Delivery Model.



Obrázek 10: Hub and Spoke Regional Delivery Model, (Salonen 2010)

I tento model však přinášel řadu komplikací a trend multisourcingu se projevil i na těchto modelech. V důsledku toho "mnoho poskytovatelů služeb vyvíjí globální delivery modely a významně investují do vyvážené kombinace jak lokálních služeb, tak služeb poskytovaných z globálních center, umístění jednotlivých lokalit a implementace metodologií, aby mohli nabídnout svým zákazníkům vyšší kvalitu a konzistenci". Tento model je nazýván Full Global Delivery Model (Salonen 2010).



Obrázek 11: Full Global Delivery Model, (Salonen 2010)

Lokální versus globální služby

Vytváření velkých globálních delivery center však neznamená, že již není třeba lokálních dodávek služeb. I když jsou služby poskytovány z globálního centra, zákazníci například požadují lokální rozhraní pro přesun a poskytování služeb, v praxi to znamená, že komunikace, koordinace či pravidelné hodnocení služeb je prováděno lokálně, tj. v místě i jazyce zákazníka. Mohou existovat právní či politické důvody proč není možné využít globálního centra nebo kulturní rozdíly mohou bránit ve využívání globálního centra. Je proto velmi důležité mít jasnou definici toho, jaké služby a jak budou dodávány z lokálního centra a globálního centra.

Základním prvkem je fungovat v rámci mezinárodní sítě globálních a lokálních produkčních center, kde jsou používány standardizované a zautomatizované procesy, nástroje a techniky, které jsou neustále zlepšovány a inovovány. Globální centra poskytují tedy standardizované služby kde ‘standardní služby’ znamená, že tyto služby se řídí stejnými procesy, nástroji a technikami bez ohledu na to, kde a kam jsou poskytovány. ‘Globalizovatelný’ znamená, že tyto standardní služby mohou být dodávány do mnoha zemí. Globální centrum je zodpovědné za návrh globálního řešení, project management, service delivery management a delivery portfolio management. Lokální centrum je zodpovědné za návrh řešení, project management a service delivery management ve vztahu ke svým lokálním zákazníkům a service delivery management včetně služeb globálního centra.

Zvýšené využívání globálních zdrojů vede k diskuzi o využívání lokálních zdrojů a snah o lokální konsolidaci. Nové modely mění charakter služeb. Tradičně, outsourcingové služby jsou považovány za "zralé", ale dopad hlavních trendů, jako jsou globální dodávky, standardizace, virtualizace, cloud computing i nadále přehodnocují složení trhu a nabídky poskytovatelů.

Zaměstnanci v globálních modelech IT

Vývoj IT profesionála

Priority požadavků na IT profesionála se měnily chronologicky s vývojem IT od vysoké technické znalosti a zároveň IT univerzálnosti k perfektní znalosti technologií v kombinaci jejich využitím pro potřeby podniku společně s efektivní komunikací, jazykovou vybaveností a schopností multikulturní spolupráce.

Metodika Českého statistického úřadu (Český statistický úřad uvádí nedatováno), že „IT odborníci se podle mezinárodní definice dělí na dvě hlavní skupiny. Základem pro toto členění je klasifikace ISCO 88 (v ČR odpovídající klasifikaci zaměstnání KZAM-R):
KZAM-R 213 – Vědci a odborníci v oblasti výpočetní techniky

2131 Projektanti a analytici výpočetních systémů
2132 Programátoři
2139 Ostatní odborníci zabývající se výpočetní technikou
KZAM-R 312 – Techničtí pracovníci v oblasti výpočetní techniky
3121 Poradenství v ICT
3122 Operátoři a obsluha výpočetní techniky
3123 Operátoři průmyslových strojů, NC strojů
3129 Ostatní technici v ICT
Informační a komunikační technologie

V oblasti poskytování zboží a služeb spojených s informačními a komunikačními technologiemi (ICT) již byla klasifikace KZAM-R zastaralá. Proto byly v klasifikaci CZ-ISCO vytvořeny třídy 25 Specialisté v oblasti informačních a komunikačních technologií a 35 Technici v oblasti informačních a komunikačních technologií. Zaměstnání v oblasti informačních a komunikačních technologií jsou však zahrnuta i v podskupinách jiných tříd (např. 2356 Lektoři a učitelé informačních technologií na ostatních školách, 7422 Mechanici a opraváři informačních a komunikačních technologií).“

S určitým zjednodušením lze říci, že vědečtí pracovníci se podílí na samotném vývoji nových technologií a souvisejících konceptů, jde především o analytiky a vývojáře softwaru a počítačových aplikací (programátory) a specialisty na databáze a počítačové sítě, zatímco techničtí pracovníci spíše na provozu a podpoře těchto systémů, jde především o techniky uživatelské podpory informačních technologií či správce webu. Stinnou stránkou vymezení IT odborníků podle klasifikace KZAM – R (resp. ISCO 88) bohužel zůstává fakt, že neodráží rychlé změny, které v IT sektoru probíhají. Nezahrnuje např. pozice typu analytik business procesů v oblasti IT, obchodník s ICT (který musí mít velmi dobrou znalost produktů, které prodává) a další.

Globální modely poskytování služeb kladou na zaměstnanci další rozměr – a to především schopnosti a dovednosti interkulturní, kde nejen jazyková vybavenost ale schopnost spolupráce v mezinárodním prostředí s překonáváním kulturních rozdílností je nedílnou součástí práce v IT.

Globální společnosti se snaží řídit a implementovat rovnou personální politiku, politiku, která tyto globální modely podporuje. V ideálním světě by měla být personální politika v rámci celého globálního IT týmu konzistentní, spravedlivá a citlivá. Globální konzistence musí umožnit a sladit se s místními zákony a kulturními zvyklostmi, což není snadný úkol. Kromě toho, že životní náklady značně liší podle regionu. Takže z hlediska zaměstnanců, model, založený na konceptu jedna velikost pro všechny, je neproveditelný. Vedle boje o konzistenci, musí CIO najít způsoby, jak pro vzdálené týmy zůstat ve spojení s jádrem podnikání.

Jedním z doporučených postupů je mít svého HR specialistu v každé lokalitě a najímání, propouštění pracovníků stejně jako kariérní růst musí sice podléhat celkové koncepci a HR metodice společnosti, ale zároveň respektovat místní právní předpisy a postupy stejně jako kulturní zvyklosti. Pro usnadnění řízení a porovnání finanční náročnosti v jednotlivých zemích globální modely používají jednotnou metodiku pracovních pozic a rolí, která snadné porovnání hodinových cen, ale také řízení a hodnocení výkonu. Důležitou a nedílnou součástí je neustálá motivace a zvyšování zainteresovanosti jednotlivých zaměstnanců, čemuž mimo jiné přispívá vysoká informovanost.

Podpůrnými nástroji pro řízení v rámci globálních modelů jsou také nástroje podporující spolupráci.

Výkonná rada CIO (Pastore 2008) provedla výzkum, jaké jsou výzvy globalizace z hlediska jejich relativního významu a rozsahu. Tento žebříček je založen na procentech respondentů.

CHALLENGE	VERY IMPORTANT	WORLDWIDE SCOPE
1. Managing virtual teams	70%	70%
2. Consolidation	61%	70%
3. Centralized/decentralized system decisions	61%	65%
4. Organizational structure	43%	70%
5. Leadership/ownership/ governance	48%	61%
6. Global vendor partner selection	35%	65%
7. Privacy regulations	35%	48%
8. Outsourcing versus insourcing	35%	48%
9. Intellectual property and knowledge management	30%	52%
10. Cultural issues and appropriate behavior	30%	43%

Source: CIO Executive Council poll, May 2007

Obrázek 12: Globální výzvy, (Pastore 2008)

Výsledky tohoto výzkumu ukazují, že více než polovina výzev se týká lidských zdrojů, počínaje řízením virtuálních zdrojů, organizační strukturou či kulturní otázky a vhodné chování.

Požadavky na zaměstnance

Pracovní prostředí IT má progresivní vývoj a požadavky trhu práce na zaměstnance jsou, jak uvádí (Zelený 2010), nejenom perfektní znalost technologií v kombinaci s jejich využitím pro potřeby klienta, rozšířené vzdělání, nutné k pochopení řešení technických problémů v širších souvislostech – v globálním, ekonomickém, ekologickém a sociálním kontextu, ale také multikulturní povědomí, znalost cizích jazyků, schopnost práce v mezinárodním prostředí a schopnost práce v heterogenních týmech (s obchodníky, právníky atd.) s efektivní komunikací s technickou i netechnickou veřejností.

Pokud budeme používat globální model a přesuneme činnosti dříve prováděné lokálně do offshoringového centra - globálního produkčního centra změní se podmínky pro lokální zaměstnance. Management služeb se stane klíčovou součástí poskytování služeb, a proto je nezbytné posílit tuto část řízení, a zejména úlohu projektového řízení a service delivery managementu. Zaměstnanci v těchto rolích musí být schopni mezinárodní spolupráce a řízení projektů/služeb v různých kulturních prostředích. Požadavky na tyto zaměstnance jsou obvykle v následujících oblastech:

a) Technická (Technologie/Metodologie/ Profesní znalost)

Požadavky na znalosti technologií, metodologií a IT jsou velmi podobné jako při poskytování služeb na lokální úrovni a změna modelu nemá významný vliv na požadavky.

b) Jazyková

Základním požadavkem je obvykle nativní znalost místního jazyka a vysokou znalost jazyka používaného v oficiální korespondenci - zpravidla angličtina. Změna modelu má významný vliv na tyto požadavky.

c) Interkulturní

Zkušenosti v interkulturním prostředí, zkušenosti s mezinárodním nasazením a/nebo zkušenosti spolupráce na mezinárodních projektech jsou žádoucí. Změna modelu má významný vliv na tyto požadavky a nikoli pouze jazykové požadavky, ale na stejné úrovni, je požadována schopnost interkulturní komunikace a empatie. Kultura (multikultura) ovlivňuje pracovní návyky, časovou zónu, místní kulturní pozadí, různé chápání odpovědnosti atd. Změna modelu má významný vliv na tyto požadavky.

d) Měkké dovednosti

Požadavky na efektivní komunikaci, prezentační dovednosti. Změna modelu nemá významný vliv na tyto požadavky.

e) Leadership

Zkušenosti s řízením virtuálních týmů a skupin. Změna modelu má dopad na tyto požadavky.

Dopady implementace globálního modelu poskytování služeb

Je nemožné ignorovat trend globalizace IT služeb obecně a také zcela jistě obchodní trendy využití globálních modelů poskytování IT služeb formou outsourcingu. Potenciál, poskytovat lepší produkty a služby za nižších nákladů, může však být využit předavším za předpokladu zvýšené pozornosti tomu, co se děje lokálně.

Globální společnosti řeší touto formou také snahu o převzetí kontroly nad celkovým děním v dané oblasti či přenášení zaměstnanosti do strategických lokalit nezávisle na ceně a celkovém výsledku. Na straně lokálního produkčního centra je potřeba si uvědomit, že kroky, které mohou být vnímány jako omezení uživatelského komfortu, zvýšení byrokracie či celkově implementaci horšího řešení jsou obvykle kroky nezbytně nutné pro úspěšnou implementaci globálního projektu a globální cíle se nutně nemusí shodovat s cíli lokálními, zkrátka je potřeba vzít na vědomí, že to, co není na první pohled výhodné pro lokální pobočku je výhodné pro celou korporaci.

Dopady z hlediska lokálního produkčního centra

I když poskytování služeb přesuneme do globálního centra, zákazník zůstává na místě a očekává nejenom poskytování služeb, ale také řízení služeb a komunikaci na lokální úrovni. Proto je nezbytně nutné věnovat zvýšenou pozornost projektovému vedení a service delivery managementu a připravit lokální řízení na změnu přístupu a myšlení, tak aby zákazník byl stále na prvním místě, nicméně nabídka služeb se nutně musí změnit z nestandardních, specializovaných řešení na řešení postaveném na standardu či komoditě.

Nejvíce zasaženi jsou zaměstnanci lokální organizace. I když nejsou požadováni v současném rozsahu poskytování služeb na lokální úrovni, jsou i nadále základem, na kterém je postavena spolupráce na lokální úrovni, a to především v úrovni vnímání koncového zákazníka. Nastává změna lokálních profesí - obvykle snížení počtu či ztráta přesunovaných profesí jako např. administrátor, vývojář, technik a zároveň zvýšená potřeba profesí Project manager, Service delivery manager, Relationship manager.

Neméně významným dopadem při lokálním snížení počtu zaměstnanců je zároveň ztráta kritické masy pro poskytování vybraných služeb nebo ztráta lokálního know how.

Dopady z hlediska globálního produkčního centra

Tendence industrializace části služeb informačních technologií má vliv na požadavky lidských zdrojů. V globálních centrech, kde se poskytují především industrializované služby je potřeba získat levné zdroje, vykonávající především rutinní činnosti bez mimořádné vlastní iniciativy, kde řízení pracovníků je velmi obdobné jako ve výrobě, tzn., měříme, vyhodnocujeme, soutěžíme. Vzhledem k centralizaci je potřeba velké množství takovýchto profesí v jedné lokalitě a společnosti přistupují k používání levných pracovních zdrojů nabíraných mimo zemi umístění, mimo kontinent. To s sebou nese celou řadu oblastí k řešení, od náboru ke změně leadershipu.

Design všech služeb je striktně centralizován a je navrhován a schvalován úzce specializovanou skupinou odborníků, kde celkový návrh řešení je postaven tak, aby vyhovoval jak technologické strategii tak především s ohledem na klíčové lokality.

Spolupráce lokálních a globálních produkčních center

I přes profesionální školení a tréninky na interkulturní spolupráci, bariéry spolupráce založené na neschopnosti empatie, odlišných kulturních i životních zvyklostí vede k rozlišení na MY a ONI a strategiím, jejichž důsledkem je negativní vnímání koncových zákazníků (nikoli vedení společností) k použití těchto modelů. Jsem přesvědčena, že změnit tuto situaci bude vyžadovat změnu našeho uvažování a pokud budeme chtít uspět jako IT profesionálové, tyto bariéry budou muset být zbořeny jak v týmech, které řídíme, tak v nás samotných.

Závěry

Závěry vyplývající z kapitol výše můžeme shrnout v následujících bodech:

Dopady z hlediska lokálního produkčního centra:

- Lokální snížení počtu zaměstnanců, zároveň ztráta kritické masy pro poskytování vybraných služeb, ztráta lokálního know how.
- Změnu lokálních profesí obvykle snížení počtu či ztráta profesí jako administrátor, vývojář, technik na zvýšenou potřebu profesí Project manager, Service delivery manager, Relationship manager.
- Změnu řízení poskytování služeb.
- Změnu nabídky služeb, z nestandardních, specializovaných řešení na řešení postaveném na standardu či komoditě.

Dopady z hlediska globálního produkčního centra:

- Tendence industrializace části služeb informačních technologií má vliv na požadavky lidských zdrojů. V globálních centrech, kde se poskytují především komodizované služby je potřeba získat levné zdroje, vykonávající především rutinní činnosti bez mimořádné vlastní iniciativy.
- Vzhledem k centralizaci je potřeba velké množství takovýchto profesí v jedné lokalitě a společnosti přistupují k používání levných pracovních zdrojů nabíraných mimo zemi umístění, mimo kontinent. To s sebou nese celou řadu oblastí k řešení, od náboru ke změně leadershipu.
- Řízení pracovníků je velmi obdobné jako ve výrobě, tzn. měříme, vyhodnocujeme, soutěžíme…
- Design všech služeb je striktně centralizován a je navrhován a schvalován úzce specializovanou skupinou.

Spolupráce lokálních a globálních produkčních center:

- Na poskytování služeb i po překonání jazykových bariér mají vliv specifika multikulturních vlivů a schopnosti mezinárodní spolupráce.

IT jako komodita nebo nervová soustava společnosti:

- Část IT, především provozní a uživatelská, se stává komoditou a při poskytování z globálních center je třeba se vyhnout rizikům, podobně jako při standardním projektu.

Na přelomu století, kdy většina IT služeb byla prováděna lokálně požadavky na v IT zcela upřednostňovaly technické znalosti, event. certifikace. Od IT odborníka se také očekávala jakási univerzálnost a kreativita pro tvorbu specifických řešení.

Centralizací služeb IT služeb a vývoji směrem ke standardním řešením docházelo ke změnám a to především v tlaku na technické specializace a nasazení procesů a znalost metodologií. Komodizace vybraných IT služeb i rychlé změny technologií však vedly ke zvýšené potřebě umět zákazníkovi/uživateli vysvětlit, umět komunikovat a porozumět jeho potřebám.

Můžeme dále bez obav část IT nazvat komoditním trhem, kde nemohou konkurovat jen IT služby samy, úspory nákladů, nejnovější technologie a dodržování SLA. Zákazníci očekávají dlouhodobou tvorbu hodnoty a chtějí mít přístup k obchodní dovednosti, intelektuálnímu kapitálu a odborné znalosti odvětví a samotného businessu.

Globalizace a s ní spojené nové přístupy k řízení služeb a IT jako celku s sebou přináší samozřejmě mnohá rizika a je proto nezbytně nutné přistupovat k obdobným projektům jako ke každému jinému projektu, ať se ocitneme na jakékoli straně – lokální produkční centrum, globální produkční centrum, zákazník a zároveň se nebránit těmto novým přístupům a naopak využít výhod i znalostí, které nám mohou poskytnout.

Seznam literatury

Český statistický úřad. Klasifikace zaměstnání (CZ-ISCO). [online]. Dostupné z: https://www.czso.cz/csu/czso/klasifikace_zamestnani_-cz_isco-

Noska, Martin. 2013. Zájem o outsourcing roste pomaleji než se předpokládalo. [online]. 2013. [17.07.2013]. Dostupné z: http://www.ictmanazer.cz/2013/07/zajem-o-outsourcing-roste-pomaleji-nez-se-predpokladalo/

Pastore, R. 2008. Global Team Management: It's a Small World After All; CIO News; [online]. [23.1.2008]. Dostupné z: http://www.cio.com/article/174750/Global_Team_Management_It_s_a_Small_World_After_All

Pastore, R. 2008. Models for Global IT Governance; CIO News; Dostupné z: http://www.cio.com/article/191700/Models_for_Global_IT_Governance

Salonen S., 2010. Tieto Delivery Excelence. [online]. Dostupné z http://webcast.tieto.com/cmd2010/PDF/04_SSalonen_CMD2010_Delivery%20excellence.pdf

Sokolovsky, Z. 2010. Vliv trendů v podnikovém nasazení IT na nutné znalosti; Liberecké Informatické Fórum 2010. ISBN 978-80-7372-656-0.

Tříska, Dušan. Nedokonalosti trhu a jejich řešení. In: seminář Asymetrické informace - nová cesta ke zdůvodnění státních zásahů, Žofín 13. 11. 2001 sborník č. 19 Investiční pobídky [online]. Dostupné z http://www.cepin.cz/cze/prednaska.php?ID=227

Voříšek, J. 2003. Strategické řízení informačního systému a systémová integrace; Praha; Management Press 2003; ISBN: 8085943409

Zelený, J. 2010. Multidisciplinarita – požadavek výuky technických disciplin; Sborník Liberecké informatické fórum 2010; ISBN 978-80-7372-656-0

JEL Classification: L15, M15

Summary

Employees in the global models of IT service outsourcing

Multinational companies providing information technology services in the form outsourcing constantly evolving global models to provide these services so that they are able to offer both industrialized IT services from global production centers, as well as specific services requiring high customer understanding. Provision of IT services from global outsourcing centers has implications for the employee as a global and a local production center both on the requirements of individual IT professionals and the skills and knowledge required by IT professional and the profession.

Given the trend towards the use of new IT service delivery model is the main objectives of the article describe global delivery models for IT services, their development from the provision of services from different locations, off shoring, the full global model, define the basic rules for services provided locally and globally, characterize the development requirements employee during the past 20 years, the requirements in the global models and requirements for IT service as a commodity, to describe the impact on employees in moving services into the global production center.

Keywords: outsourcing, global model, employees, methodology, ITIL, commodity, industrialization.

Speciální asistenční místnosti pro pacienty s mentálním a motorickým postižením

Jiří Zumr

jiri.zumr@gmail.com

Doktorand oboru Aplikovaná informatika

Školitel: prof. Ing. Petr Berka, CSc., berka@vse.cz

Abstrakt: Prezentujeme výsledky pilotní studie jedné části projektu chráněné místnosti, která bude bez přímého vizuálního dohledu zabezpečovat alarm při nastalém zdravotním riziku (například při poruše kardiálního nebo dechového rytmu či epileptickém záchvatu) psychomotoricky retardované či jinak handicapované osoby.

Jedná se o aplikaci piezoměničů ve smyslu senzorů, schopných evidovat širší škálu hodnot a tím získat větší množství dat k dalšímu zpracování a event. „učení“ systému. Běžně dostupné senzory, které jsou schopny evidovat pouze dvě hodnoty (zátěž/bez zatížení, tedy – 0 a 1), tomuto projetu nevyhovují.

Klíčová slova: asistivní technologie, monitorování polohy, monitorování pohybu, piezosenzor, Arduino

Úvod

Každodenní péče o psychomotoricky retardované pacienty je stresujícím faktorem pro všechny členy rodiny. Proto na základě zkušeností s problematikou „chytrých domů“ a z popudu dětských neurologů i rodičů handicapovaných dětí vznikl projekt speciální asistenční místnosti pro pacienty s mentálním a motorickým postižením, poskytující alarm při náznaku zdravotního rizika (například při poruše srdečního nebo dechového rytmu či epileptickém záchvatu) i bez přímého vizuálního dohledu pečující osoby. Na základě rešerší tuzemské i světové literatury bylo zjištěno, se jedná o zcela inovativní pohled na tuto problematiku, který by mohl být využit i jako patentově chráněný projekt.

Původní, čistě informačně-technologická (IT) představa projektu zahrnovala implementaci nejnovějších technologií. Na základě pilotních přednášek o studované problematice a plánovaném řešení pro zainteresované skupiny lékařů i rodin pacientů bylo nutné přehodnotit výchozí parametry – a to zejména ve smyslu pořizovací ceny výsledného produktu. Je nutné vycházet z faktu, že, bohužel, socioekonomická situace rodin s postiženými dětmi neumožňuje vybavit byt finančně nákladnými monitorovacími prostředky.

Základním cílem práce se tedy stal finančně běžně dostupný produkt, který umožní členům rodiny či jiným pečujícím spolehnout se na monitoraci handicapovaného jedince tak, aby nebyly ohroženy jeho životní funkce.

Ve světě sice existuje mnoho výzkumných projektů, které se zabývají teorií přenosu dat mezi jednotlivými senzory a místem, kde jsou vyhodnocována. Jedná se však většinou o přenos dat ze senzorů umístěných přímo na pacientovi a průběžně získávaná data jsou odesílána do pracovny ošetřujícího lékaře. Lze shrnout, že prioritami těchto teoretických prací je problematika přenosu dat na větší vzdálenost a s tím související jejich ochrana plus řešení způsobu shromažďování a ukládání. To však není prioritou předkládaného projektu, protože získaná data nebudou odesílána mimo prostory monitorovaného prostředí, ale budou ukládána a zpracovávána systémem, který bude umístěn v dosažitelné blízkosti aplikovaných senzorů. Díky tomuto uspořádání nebudou získávaná data vystavena nebezpečí zneužití třetí osobou. Také není třeba dlouhodobě shromažďovat větší objem starších dat k dalším studiím – nejedná se o problém diagnostický, ale monitorovací s akcentem na okamžitý alarm.

V literatuře lze dohledat přípravné vědecké studie rozebírající chování pacientů - např. ve stylu práce „The challenges in monitoring and preventing patient safety incidents for people with intellectual disabilities in NHS acute hospitals: evidence from a mixed-methods study“(TUFFREY-WIJNE, 2014) na základě téměř dvouletého sledování rozebírala rizikové situace, do kterých se ve zdravotnickém zařízení mohou dostat mentálně postižení pacienti. Na základě podkladů, poskytnutých dětskými neuro-

logy i rodiči handicapovaných pacientů však byly propozice pro chráněnou místnost natolik jasné, že nebylo třeba předběžných studií dalších expertů.

Z hlediska projektu je velmi důležitá rozsáhlá teoretická souhrnná práce „A Survey of Body Sensor Networks“ (LAI, 2013) čínských autorů z roku 2013. Podrobně se zabývá nejen nejpoužívanějšími senzory, ale i jejich zapojením, kdy je třeba vzít v úvahu a splnit řadu požadavků například na nízkou spotřebu energie, nízkou opotřebovatelnost, diagnostiku poruch apod. Podrobně je rozebírána otázka energetické kontroly, kontroly chyb a možnost co největší redukce senzorických virtuálních uzlů. Samostatná kapitola je věnována možnostem zpracování pořízených dat a v návaznosti na tento proces přehled užívaných tělesných senzorových sítí (Body Sensor Networks; BSNs). Upozorňují na současné problémy s aplikací senzorů, které by měly být co nejmenší a s minimální spotřebou energie. Některé dříve používané měly i nežádoucí účinky na tělesnou tkáň a mohlo docházet k popálení. Upozorňují i na etické a právní problémy vyplývající z aplikací a zdůrazňují nutnost multidisciplinárního přístupu k danému problému.

Nezanedbatelným momentem - finanční náročností - se zabývá stať „Virtualization of Wireless Sensor Network: Smart House perspective“ (ISLAM, 2013) korejských autorů Md. Motaharul Islam a Eui-Nam Huh, ve které je brán zřetel na využití virtuálních senzorů pro inteligentní či asistivní domy podle dostupnosti, tedy i pro běžného uživatele.

Základem projektované místnosti je detekce polohy/pohybu po podložce. Proto byla v tomto směru provedena rozsáhlá rešerše, ze které vyplynulo, že na trhu jsou již sice běžně dostupné senzory poskytující vysokou rozpoznávací schopnost v oblasti, pro kterou jsou určeny (od polohy, pohybu, teploty, přes vlhkost, frekvenci pulsu, dýchání …), ale jejich cena bývá stále příliš vysoká. Pro účely sledování handicapovaného dítěte by bylo teoreticky možné využít mnohem finančně dostupnější a tedy jednodušší varianty. Například pro účely zjištění polohy osoby v místnosti lze použít několik různých technik – snímání kamerou, laserovým paprskem (Kinect), nebo podlahovým senzorem. Nejjednodušší z výše uvedených možností je snímání pomocí podlahových senzorů. Zde se lze poučit z projektů, které jsou primárně určeny starším lidem a mají za úkol především monitorovat pády. Jako velmi užitečná se z tohoto pohledu jeví práce „Smart carpet using differential piezoresistive pressure sensors for elderly fall detection”(CHACCOUR, 2015) libanonských autorů Kabalana Chaccoura a kol. z loňského roku, která využívá piezoelektrické tlakové senzory. Tuto technologii jsme nezávisle na publikované práci vyhodnotili i my jako nejperspektivnější a naše zkušenosti prezentujeme.

Piezoelektrický jev byl objeven Jacquem a Pierrem Curie (PIEZO SYSTEMS, INC., 2016) již na konci předminulého století (1980). Jedná se o schopnost krystalu generovat elektrické napětí během jeho deformace. Piezoelektrický efekt známe jako přímý a nepřímý. Nepřímý efekt znamená, že pod vlivem elektrického pole dochází k deformaci krystalu, kdežto přímý piezoelektrický efekt je schopnost některých materiálů vytvořit elektrický potenciál v závislosti na aplikovaném mechanickém namáhání. To mění hustotu polarizace v materiálu a vede k pozorované změně potenciálu. Pro účely sledování pohybu pacienta po podložce využíváme tedy efektu přímého.

Piezoelektrické snímače jsou univerzálními nástroji pro měření řady různých procesů s vynikající spolehlivostí. Jsou používány pro zajištění kvality a řízení procesů i pro výzkum a vývoj v mnoha odvětvích, například v medicíně, letectví, automobilovém průmyslu, ale třeba i jako snímač tlaku na tlačítkách mobilních telefonů. Piezoelektrický jev vykazují pouze materiály, které nemají centrálně symetrickou krystalovou strukturu. Některé z běžně používaných a známých piezoelektrických materiálů jsou křemen (SiO2), oxid zinečnatý (ZnO), polyvinylidenefluorid (PVDF) a zirkoničitan-titaničitanu olova, (PZT nebo Pb (Zr, Ti) O_3) (SINGH, 2015).

Cíl a metodika

V rámci celkového cíle projektu asistenční místnosti bylo nutné stanovit dílčí cíle. Prvním dílčím cílem bylo vytvoření programu, který bude schopen zaznamenávat data z podlahového senzoru a reagovat na ně přednastaveným způsobem.

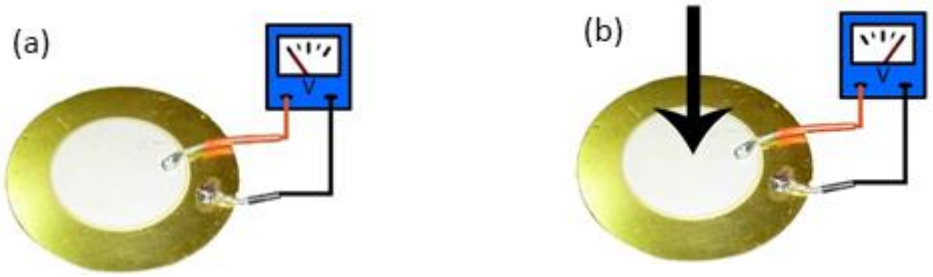
Původně byla jedním z námi zkoušených systémů senzorů taneční podložka známa jako Dancemat, kterou německá společnost v době nákupu přes internet vydávala za dotykový senzor. Technologie

v těchto podložkách se však ukázala pro naše potřeby jako nedostatečná, protože poskytuje pouze výstup v hodnotách 1 nebo 0, v případě uvažovaného podlahového senzoru by tyto hodnoty byly vyhodnoceny jako „sešlápnuto“ nebo „nesešlápnuto“. Ačkoliv by tyto hodnoty stačily pro bazální vyhodnocení, je mnohem vhodnější, aby senzor byl schopen vyhodnotit také míru zatížení/sešlápnutí. Kromě toho jsou senzory v taneční podložce uloženy ve velkém odstupu – na ploše 1m x 1m je jich umístěno pouze 9, což neumožňuje určit přesnou polohu a směr pohybu sledované osoby v pokoji.

S ohledem na finanční dostupnost jsme využili běžného piezoelektrického senzoru, „piezoměniče“, který je dostupný v nabídkách obchodů s běžnou elektronikou. Dle různých velikostí se cena pohybuje kolem 10,-Kč.

Vycházeli jsme z toho, že piezoelektrický senzor je zařízení, které využívá piezoelektrický jev k měření změn tlaku, zrychlení, nebo napínání převáděním do elektrického náboje. Opakovaná aplikace tlaku na piezoelektrický materiál může vést ke stavu, kdy takto vytvořený elektrický proud může být použit například pro žárovku (diodu), nebo pro nabíjení mobilního zařízení za chůze uživatele (čehož není možné dosáhnout v klidu). Díky přímému piezoelektrickému efektu je lze využít jako senzory tlaku, kdy na jedné straně je senzor a na druhé snímač, který převádí analogový signál (proud) na digitální.

Na obrázku 1a,b je názorně ukázán piezoelektrický materiál před stiskem, kdy není produkován žádný elektrický náboj a je tentýž materiál při mechanickém namáhání.



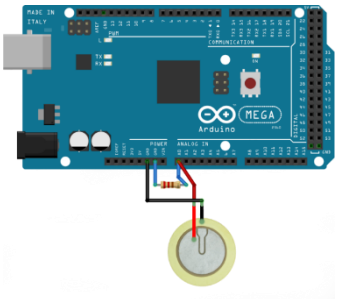
Obrázek 1a,b: Piezoelektrický materiál před stiskem (a) a po stisku (b)

Testování piezoelektrického senzoru (senzoru)

Pro otestování piezoelektrických vlastností piezoměniče bylo nutné převést analogový signál (náboj), který piezoměnič během deformace vytváří, na digitální signál neboli hodnotu. K tomuto účelu jsme využili vývojovou platformu s názvem Arduino (Arduino, 2016).
Pro získání údajů z piezoměniče bylo použito:

- Arduino Mega 2560
- Megaohmový rezistor
- Piezoměnič
- Rovná plocha představující podlahu

Vyjmenované součásti byly zapojeny podle schématu znázorněného na obrázku 2.



Obrázek 2: Zapojení Arduino a piezoměniče

Piezoměniče jsou polarizované – to znamená, že jimi (nebo z nich) prochází napětí jen v určitém směru. Aby piezoměnič reagoval na stlačení, je nutné připojit černý vodič (nižší napětí) k nulovému potenciálu (GND) u Arduina a červený vodič (vyšší napětí) na analogový vstup, tedy pin 0 (ANALOG IN 0). Každý piezoměnič generuje během deformace různé napětí, a proto je nutné připojit 1 mega-ohmový rezistor paralelně s piezoměničem. Tento omezí napětí a proud vyvolaný piezoměničem a chrání tak analogový vstup u vývojové desky Arduino.

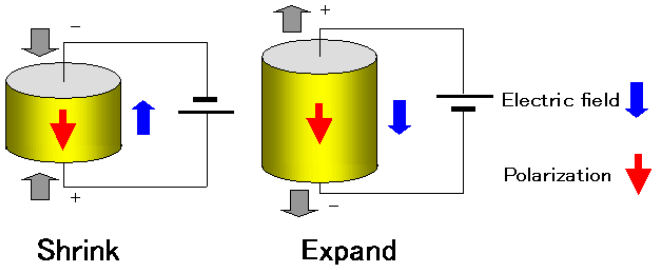
Provedené experimenty

Pro načítání dat pomocí Ardunina byl sestaven následující kód – obrázek 3:

```
/*
 Tlakovy senzor
 Tento program nacita napeti generovane piezomenicem a vypisuje na obrazovku.
 Z analogoveho pinu jsou nacitany hodnoty a prevadeny do hodnot digitálních.
*/

// definice promennych
const int SenzorTlaku = A0; // nastaveni pinu, kde je pripojen piezomenic (Analog 0)
int hodnota = 0;            // promenna pro ukladani hodnoty ze senzoru

void setup() {
  Serial.begin(9600);       // aktivace sériového portu
}

void loop() {
  hodnota = analogRead(SenzorTlaku);
  Serial.println(hodnota);
  delay(100);  // zpozdeni proti pretizeni vyrovnavaci pameti serioveho portu
}
```

Obrázek 3: Kód pro Arduino

Během pokusů vykonávaných s piezoměničem bylo ověřeno, že při stlačení (stisku) senzoru dochází ke generování kladného náboje a při jeho uvolnění dochází naopak ke generování záporného náboje - obrázek 4.



Obrázek 4: Generování náboje piezoměničem Zdroj: SINGH, 2015

Díky tomuto efektu je možné pomocí piezoměniče detekovat nejen zatížení (šlápnutí na podložku), ale také uvolnění (odchod z podložky). Vytvořený program zaznamenává informace z prototypové podložky a zapisuje (sčítá) množství elektrického pole generovaného po stisku/uvolnění.

Při stlačení senzoru vyhodnocuje Arduino zvýšení napětí v rozmezí 0-1024 hodnot (rozsah použitého AD převodníku). Zásadním problémem Arduina je neschopnost zaznamenávat na analogových vstupech i záporné hodnoty. Proto v případě uvolnění senzoru po předchozím sešlápnutí jsou hodnoty po určitou dobu vyhodnocovány jako rovné 0 (přestože jsou ve skutečnosti záporné). Množství takto zachycených záporných hodnot je úměrné nastavené frekvenci záznamu hodnot Arduinem.

V případě tohoto experimentu, kdy není nutné zaznamenávat sílu sešlápnutí, je možné tento efekt využít k detekci sešlápnutí/odšlápnutí uvedeným programovým kódem. V případě kardiopulmonálního selhání pacienta při běžné každodenní činnosti/hře se nebude pravděpodobně jednat o okamžitou zástavu motorických funkcí (nulová hodnota změny zátěže podložky ve všech senzorech), ale o řadu dyskoordinovaných záškubů s nepravidelnou frekvencí. Odlišit tyto záškuby od běžné hry je obtížné, avšak na základě algoritmu „učení“ v programu, či v počátečních fázích na základě nastavení sledování nepravidelného zatěžování podložky jen v určitém časovém intervalu s následným „silent“ intervalem, lze i tyto stavy rozlišit.

Dalším předpokládaným vzorcem je rytmické zatěžování (cca 10x a méně/sekundu) podložky při záškubech během klonické fáze epileptického záchvatu, většinou předcházených tonickou, tedy

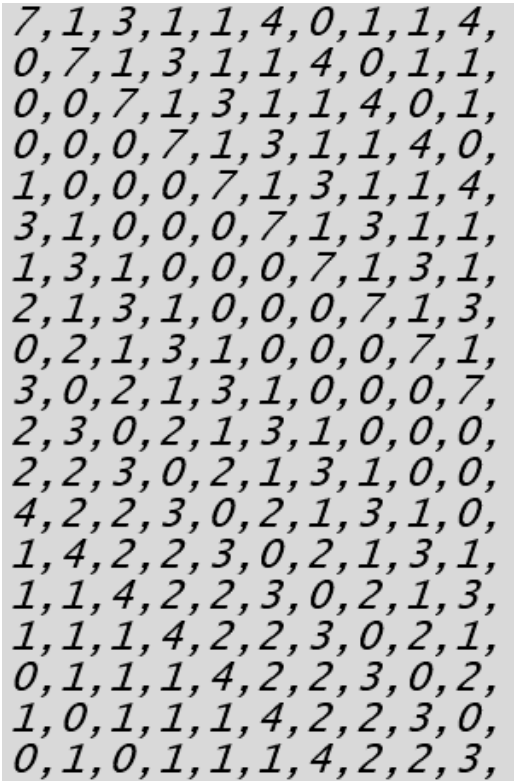
z hlediska podložky „klidovou“, fází. V tomto případě je nutné spolupracovat s rodinou a diferencovat charakter záchvatů u sledovaného jedince.

Jiné typy záchvatů, například „klasické“ absence, tedy několikasekundovou zástavu motorické činnosti pacienta, nelze tou metodou snadno odlišit. Nejedná se však o stavy ohrožující život, pouze je nutné poukázat na skutečnost, že tuto metodu nelze použít jako monitorování frekvence specifického typu epileptických záchvatů. Zkušenost ukázala nemožnost vyhodnocovat přímo jednotlivé aktuálně snímané hodnoty. Problémem piezosnímače spočívá v neposkytování stálé nulové hodnoty i bez působení tlaku. Z tohoto důvodu je nutné zavést snímání delšího časového úseku (posloupnost více hodnot). Z pohledu obslužného programu je tento problém vyřešen přidáním pole hodnot do zdrojového kódu a tedy uchovávající určité množství hodnot.

Pokud z důvodu rychlosti programu požadujeme pole co nejkratší, musí být jeho délka přizpůsobena rychlosti záznamu hodnot Arduinem. Během testování byla zjištěna vhodná velikost pole o 10 prvcích při záznamu 20 hodnot za sekundu ze senzoru. Těchto deset hodnot bude zaznamenáváno a postupně ukládáno do pole následujícím způsobem (princip posuvného registru / paměti):

Při založení pole jsou všechny jeho prvky rovny 0 (0,0,0,0,0,0,0,0,0,0). V první iteraci je po načtení první hodnoty z piezoměniče (označme jako a) uložena tato do pozice 0 v poli - (a,0,0,0,0,0,0,0,0,0). V každé iteraci je načítána hodnota na pozici 0 v poli a musíme tedy hodnoty v poli postupně posouvat, aby nedocházelo k jejímu přepsání. Po prvním posunutí bude v poli na prvním místě hodnota 0 a původní hodnota „a“ bude uvedena na druhém místě - (0,a,0,0,0,0,0,0,0,0). V druhé iteraci dojde opět k načtení aktuální hodnoty z piezoměniče (označme jako b) a hodnota je uložena do pole (b,a,0,0,0,0,0,0,0,0) a pole posunuto (0,b,a,0,0,0,0,0,0,0).

Během pokusů byly zaznamenány například následující hodnoty – obrázky 5-7.



Obrázek 5: Hodnoty zaznamenané bez zatížení senzoru

```
39,3,0,0,0,0,0,0,0,0,
0,39,3,0,0,0,0,0,0,0,
207,0,39,3,0,0,0,0,0,0,
44,207,0,39,3,0,0,0,0,0,
280,44,207,0,39,3,0,0,0,0,
429,280,44,207,0,39,3,0,0,0,
65,429,280,44,207,0,39,3,0,0,
80,65,429,280,44,207,0,39,3,0,
17,80,65,429,280,44,207,0,39,3,
8,17,80,65,429,280,44,207,0,39,
1,8,17,80,65,429,280,44,207,0,
14,1,8,17,80,65,429,280,44,207,
17,14,1,8,17,80,65,429,280,44,
11,17,14,1,8,17,80,65,429,280,
0,11,17,14,1,8,17,80,65,429,
0,0,11,17,14,1,8,17,80,65,
0,0,0,11,17,14,1,8,17,80,
0,0,0,0,11,17,14,1,8,17,
0,0,0,0,0,11,17,14,1,8,
0,0,0,0,0,0,11,17,14,1,
0,0,0,0,0,0,0,11,17,14,
0,0,0,0,0,0,0,0,11,17,
0,0,0,0,0,0,0,0,0,11,
0,0,0,0,0,0,0,0,0,0,
16,0,0,0,0,0,0,0,0,0,
```

Obrázek 6: Hodnoty zaznamenané s pomalým sešlápnutím senzoru

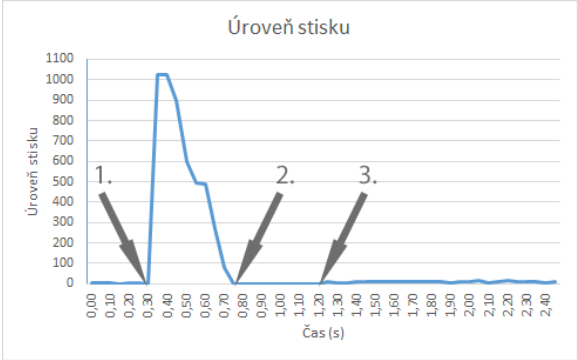
```
1,4,3,7,2,6,2,4,3,5,
7,1,4,3,7,2,6,2,4,3,
1023,7,1,4,3,7,2,6,2,4,
1023,1023,7,1,4,3,7,2,6,2,
621,1023,1023,7,1,4,3,7,2,6,
242,621,1023,1023,7,1,4,3,7,2,
51,242,621,1023,1023,7,1,4,3,7,
5,51,242,621,1023,1023,7,1,4,3,
25,5,51,242,621,1023,1023,7,1,4,
80,25,5,51,242,621,1023,1023,7,1,
0,80,25,5,51,242,621,1023,1023,7,
0,0,80,25,5,51,242,621,1023,1023,
0,0,0,80,25,5,51,242,621,1023,
0,0,0,0,80,25,5,51,242,621,
0,0,0,0,0,80,25,5,51,242,
0,0,0,0,0,0,80,25,5,51,
0,0,0,0,0,0,0,80,25,5,
0,0,0,0,0,0,0,0,80,25,
0,0,0,0,0,0,0,0,0,80,
0,0,0,0,0,0,0,0,0,0,
16,0,0,0,0,0,0,0,0,0,
```

Obrázek 7: Hodnoty zaznamenané s rychlým sešlápnutím senzoru

Pokud není na senzor aplikován žádný tlak, jsou získané hodnoty rovny 0, nebo jsou velice blízké této hodnotě (0-25). V případě pomalého sešlapávání je zaznamenáno větší množství hodnot v rozmezí 100-300. V případě rychlého sešlápnutí je zaznamenáno menší množství hodnot >1000. Ve všech případech mají hodnoty společnou jednu vlastnost – jsou následovány sadou záporných hodnot, vyjádřených v Arduinu jako sada 0.

V poli tedy můžeme následně vyhledávat tyto „záporné“ hodnoty a tím identifikovat „odšlápnutí“ ze senzoru. Během testování bylo zjištěno, že pro záznam 20ti hodnot za sekundu je kontrola sedmi za sebou jdoucích 0 možné identifikovat jako stav „odšlápnuto“.

Na následujícím grafu (obrázek 8) je možné lépe demonstrovat činnost senzoru během sešlápnutí a uvolnění:



Obrázek 8: Průběh činnosti senzoru
(1 - začátek tlaku na senzor, 2 - konec tlaku na senzor, 3 - senzor v normálním/klidovém stavu)

Závěr

Pilotní studie tedy prokázala vhodnost použití piezoměničů jako detektorů polohy či pohybu sledované osoby.. Jejich výhoda oproti jiným řešením, spočívá v možnosti libovolného uspořádání pod podložkou a tím tedy monitorování handicapovaných pacientů v chráněném prostředí.

Vytvořený program sice umožnil lokalizovat polohu a při větším množství zabudovaných senzorů i pohyb osoby v místnosti, na druhou stranu - kdybychom zátěž plosky nohy monitorovali s větším rozlišením, byli bychom schopni v časných stadiích detekovat i takové podrobnosti, jako je počínající plantární fasciitida (zánět šlach umístěných na plosce nohy). V případě zachycení rytmických pohybů je systém možné využít k signalizaci nebezpečí – např. epileptického záchvatu.

Dalším plánovaným krokem je aplikovat do stávajícího obvodu senzoru zesilovač měřeného signálu a tak otestovat senzory využívající jiné technologie plus data ze senzorů monitorující puls, dech, event. i z kamerového systému. Projekt předpokládá, že data, získaná podlahovými a dalšími senzory jsou ukládána po převedení z analogového signálu na digitální do databáze a dále analyzována virtuálním senzorem. Ukládání dat z podlahových senzorů předpokládá databázi, ve které budou data rozdělena do řádků, reprezentující jednotlivé záznamy, a sloupců, identifikujících zapsaná data. Nejdůležitějším atributem, který je v této databázi uveden, je čas, díky kterému bude celý mechanismus synchronizován. Pokud uvažujeme tlakový senzor, bude ještě nutné definovat, zda je nutné zaznamenávat všechny údaje, nebo jen ty, které označí virtuální senzor za důležité. Pokud bychom zaznamenávali všechny údaje, bude množství ukládaných dat enormní. Uvažujeme-li pokoj o rozměrech 5x5 metrů a v případě testované piezoelektrické tlakové podložky budou senzory přibližně 10 centimetrů od sebe, vzniká matice o 2500 senzorech. Pokud budou měřit 10x za sekundu, vyšlou každou sekundu 25000 potencionálních řádků databáze (nebo, pokud bychom nastavili sloupce jako hodnoty jednotlivých senzorů, bude do databáze přibývat každou sekundu 10 řádků o minimálně 2502 sloupcích). Lze předpokládat, že data z dalších senzorů budou vzhledem ke svému charakteru mnohem méně objemná a tím jednodušší ke zpracování.

Na úrovni připravovaného virtuálního senzoru probíhá v první úrovni analýza okamžitě po příchodu dat tak, aby bylo možné ihned zareagovat na situaci, která by mohla ohrožovat pacienta na životě (např. rytmické údery do podložky mohou signalizovat křeče pacienta při epileptickém záchvatu). V tomto případě nejsou data analyzována z již uložených dat, ani s nimi srovnávána, ale vyhodnocena ještě před uložením do databáze.

V druhé úrovni virtuální senzor kontroluje krátkodobá data, jejichž změny by mohly naznačovat nadcházející akutní problém (např. nižší prokrvení končetin detekované oxymetrem začátek

epileptického záchvatu nebo vdechnutí cizího tělesa). V takovém případě jsou analyzována data, která jsou již uložena do databáze, a kontrola probíhá v rámci posledních zjištěných 12 -24 hodin.

Třetí úrovní je kontrola dlouhodobých dat, uložených za 5-10 dnů, které mohou naznačovat chronické změny v organismu pacienta (hyperaktivita nebo naopak hypoaktivita pacienta může signalizovat počínající nemoc, nebo zvýšení frekvence záchvatů bez křečové složky, tzv. „absencí“, atd.).

Výsledkem vyhodnocení situace, která ohrožuje pacientův stav, bude spuštění zvukového a/nebo světelného alarmu. Po synchronizaci a ověření funkčnosti jednotlivých typů senzorů bude probíhat optimalizace navrženého virtuálního senzoru a odzkoušení celého systému v praxi (předpokládaná spolupráce s Centrem asistivních technologií ČVUT a rodinami pacientů).

Literatura

TUFFREY-WIJNE, Irene, Lucy GOULDING, Vanessa GORDON, Elisabeth ABRAHAM, Nikoletta GIATRAS, Christine EDWARDS, Steve GILLARD a Sheila HOLLINS. The challenges in monitoring and preventing patient safety incidents for people with intellectual disabilities in NHS acute hospitals: evidence from a mixed-methods study. BMC Health Services Research [online]. 2014, 14(1), 432- [cit. 2016-01-19]. DOI: 10.1186/1472-6963-14-432. ISSN 1472-6963. Dostupné z: http://www.biomedcentral.com/1472-6963/14/432

LAI, Xiaochen, Quanli LIU, Xin WEI, Wei WANG, Guoqiao ZHOU a Guangyi HAN. A Survey of Body Sensor Networks. Sensors [online]. 2013, 13(5), 5406-5447 [cit. 2016-01-19]. DOI: 10.3390/s130505406. ISSN 1424-8220. Dostupné z: http://www.mdpi.com/1424-8220/13/5/5406/

ISLAM, Md. Motaharul, Jun Hyuk LEE a Eui-Nam HUH. An Efficient Model for Smart Home by the Virtualization of Wireless Sensor Network. International Journal of Distributed Sensor Networks [online]. 2013, 2013, 1-10 [cit. 2016-01-19]. DOI: 10.1155/2013/168735. ISSN 1550-1329. Dostupné z: http://www.hindawi.com/journals/ijdsn/2013/168735/

CHACCOUR, Kabalan, Rony DARAZI, Amir HAJJAM EL HASSANS a Emmanuel ANDRES. Smart carpet using differential piezoresistive pressure sensors for elderly fall detection. In: 2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) [online]. IEEE, 2015, s. 225-229 [cit. 2016-01-18]. DOI: 10.1109/WiMOB.2015.7347965. ISBN 978-1-4673-7701-0. Dostupné z: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7347965

HISTORY OF PIEZOELECTRICITY. PIEZO SYSTEMS, INC. [online]. 65 Tower Office Park Woburn, MA 01801 USA, 2016 [cit. 2016-01-30]. Dostupné z: http://www.piezo.com/tech4history.html

SINGH, S. Materials with unique combination of properties. [online]. 2015, 2015-02-21 [cit. 2015-12-04]. Dostupné z: http://www.startinn.com/materials-with-unique-combination-of-properties/

What is Arduino? *Arduino* [online]. USA, 2016 [cit. 2016-01-30]. Dostupné z: https://www.arduino.cc/en/Guide/Introduction

JEL Classification: L86.

Summary

Special assistance rooms for patients with mental and motor disabilities

We present the results of a pilot study within the scope of the protected room project that will activate the alarm without direct visual supervision of a caring person when the patient is at heath risk (such as a disruption of cardiac or respiratory rhythm or epileptic seizure).

It is about an application of piezoelemets within the meaning of sensors that are able to record a wider range of values and thus obtain more data for further processing and "learning" of the system. Commercially available sensors which are able to register only two values (load / no load - 0 and 1), are inadequate for this project.

Keywords: assistive technology, position monitoring, movement monitoring, piezoelectric sensors, Arduino

Rovnomerné maticové rozvrhy v intuitívnej fuzzy logike

Roman Hajtmanek

roman.hajtmanek@fri.uniza.sk

Doktorand študijného odboru 9.2.9 Aplikovaná informatika

Školiteľ: doc. RNDr. Štefan Peško, CSc., (stefan.pesko@fri.uniza.sk)

Abstrakt: V praxi vzniká potreba vytvárať rozvrhy s rovnomernými výkonmi. V prípade neistých údajov je jedným z riešení intuitívna fuzzy logika. V príspevku sa rieši problém rozvrhu, ktorý vznikol pri zrovnomerňovaní týždenných výkonov vodičov modelovaných maticou. Cieľom je minimalizovať vhodnú mieru nerovnomernosti vektora riadkových súčtov matice pomocou permutácií stĺpcových vektorov matice. Tento NP-ťažký problém je riešiteľný v polynomickom čase pre dvojstĺpcovú reálnu maticu. Ukážeme, ako možno tento problém riešiť v prípade intuitívnej fuzzy matice priradením zoraďovacej funkcie k jej prvkom. Prezentujeme pravdepodobnostný algoritmus pre všeobecný prípad intuitívnej fuzzy matice, ktorý je založený na agregačnom dvojstĺpcovom princípe známeho z riešenia v reálnej aritmetike.

Kľúčové slová: rovnomerné rozvrhy, problém stĺpcových permutácií matice, intuitívne fuzzy čísla, pravdepodobnostný algoritmus.

Úvod

Prvá zmienka o rovnomerných rozvrhoch bola uvedená v článku (Coffman, Yannakakis, 1984) v reálnej aritmetike. Úvod do intuitívnych fuzzy čísiel sme čerpali z článku (Singh., Yadav, 2014). Tu sa obmedzíme na alternatívnu metódu výpočtu výkonov inšpirovanou nasledujúcim rozvrhovacím problémom:

Majme m vozidiel (smetiarske autá) a n dní. V každý deň sa musí vykonať m úloh. Predpokladajme, že prvkami danej matice A^I sú zisky z vykonanej práce dané intuitívnymi fuzzy číslami: $\tilde{a}_{ij}^I = (a_{ij1}, a_{ij2}, a_{ij3}; a'_{ij1}, a_{ij2}, a'_{ij3})$, kde i-te vozidlo vykonalo úlohu v j-ty deň. Celý zisk z i-teho vozidla je $\tilde{s}_i^I = \sum_{j=1}^{n} \tilde{a}_{ij}^I$. Ak je príliš veľký rozdiel medzi ziskami $\tilde{s}_1^I, \tilde{s}_2^I, \dots \tilde{s}_m^I$, vodiči vozidiel budú proti tomu protestovať. Preto zamestnávateľ permutuje prvky v jednotlivých stĺpcoch matice $\tilde{A}^I$, aby dosiahol čo najrovnomernejšie súčty riadkov.

Deterministická interpretácia (s reálnou maticou) tohoto problému bola riešená v (Kluvánek, 1983) a formulovaná v (Tegze, Vlach, 1984) ako MPP (Matrix Permutation Problem). Rozbor nerovnomernosti v MPP sa uvádza v (Černý, 1985). Grafové zovšeobecnenie MPP, kde vrcholy sú prvkami reálnej matice bolo preštudované v prácach (Czimmermann, Peško, 2003) a (Czimmermann P., 2003). Zovšeobecnenie MPP s pozitívnymi váhami riadkov bolo uvedené v (Černý, Peško, nedatované). Podľa množinových mier nerovnomernosti klasifikoval Dell'Olmo a kol. 32 typov rovnomerných k-partition problémov a určil ich výpočtovú zložitosť v (Dell'Olmo, Hansen, Pallottino, Storchi, 2005). MPP v intervalovej aritmetike bolo uvedené v (Peško, 2006). Riešenie MPP uvádza (Boudt, Vanduffel, Verbeken, 2015). Možnosť použitia MPP vo férovom rozvrhovaní uvádza (Peško, Hajtmanek, 2014).

Základy teórie intuitívnych fuzzy množín

V tomto odstavci sa obmedzíme len na tie základné pojmy a tvrdenia, ktoré ďalej využijeme.

Nech X je univerzum. Potom fuzzy množina $\tilde{A}$ v X je definovaná ako usporiadaná dvojica:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) : x \in X\}, \quad (1)$$

kde $\mu_{\tilde{A}}: X \to [0,1]$.

Nech X je univerzum. Potom intuitívna fuzzy množina IFS (Intuitionistic Fuzuzy Set) $\tilde{A}^I$ v X je definovaná ako usporiadaná trojica:

$$\tilde{A}^I = \left\{\left(x, \mu_{\tilde{A}^I}(x), \nu_{\tilde{A}^I}(x)\right) : x \in X\right\},$$

kde $\mu_{\tilde{A}^I}(x), \nu_{\tilde{A}^I}(x): X \to \langle 0,1\rangle$ sú také funkcie, že $0 \leq \mu_{\tilde{A}^I}(x) + \nu_{\tilde{A}^I}(x) \leq 1, \forall x \in X$. Funkcia $\mu_{\tilde{A}^I}(x)$ reprezentuje stupeň príslušnosti (degree of membership) a funkcia $\nu_{\tilde{A}^I}(x)$ stupeň nepríslušnosti (degree of nonmembership) prvku $x \in X$ do $\tilde{A}^I$. $h(x) = 1 - \mu_{\tilde{A}^I}(x) - \nu_{\tilde{A}^I}(x), \forall x \in X$ sa nazýva stupeň pochybnosti (degree of hesitation) prvku $x \in X$ do $\tilde{A}^I$.

Intuitívnu podmnožina $\tilde{A}^I = \left\{\left(x, \mu_{\tilde{A}^I}(x), \nu_{\tilde{A}^I}(x)\right) : x \in X\right\}$reálnej osi $\Re$, nazývame intuitívne fuzzy číslo (IFN) ak spĺňa nasledovné:

1. Existuje $r \in \Re$také že, $\mu_{\tilde{A}^I}(r) = 1$ a $\nu_{\tilde{A}^I}(r) = 0$ (r sa nazýva stredná hodnota $\tilde{A}^I$).
2. $\mu_{\tilde{A}^I}(x)$ a $\nu_{\tilde{A}^I}(x)$ sú po častiach spojito mapované z $\mathbb{R}$ do uzavretého intervalu $\langle 0,1\rangle$ a relácia $0 \leq \mu_{\tilde{A}^I}(x) + \nu_{\tilde{A}^I}(x) \leq 1$ platí pre všetky $x \in \Re$.

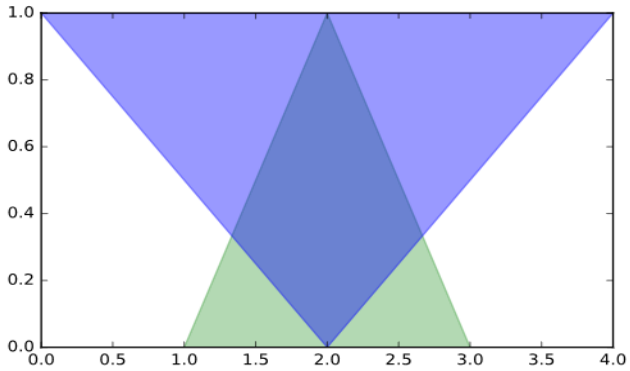
Funkcia príslušnosti (membership function)$\mu_{\tilde{A}^I}(x)$ a funkcia nepríslušnosti (non-membership function) $\nu_{\tilde{A}^I}(x)$ majú nasledovný tvar:

$$\mu_{\tilde{A}^I}(x) = \begin{cases} g_1(x), & r - \alpha \leq x \leq r, \\ 1, & x = r, \\ h_1(x), & r \leq x \leq r + \beta, \\ 0, & \text{inak}, \end{cases}$$

kde $g_1(x)$ a $h_1(x)$ sú funkcie po častiach spojité, rastúce respektíve klesajúce na intervaloch $\langle r - \alpha, r\rangle$ a $\langle r, r + \beta\rangle$.

$$\nu_{\tilde{A}^I}(x) = \begin{cases} g_2(x), & r - \alpha' \leq x \leq r; 0 \leq g_1(x) + g_2(x) \leq 1, \\ 0, & x = r, \\ h_2(x), & r \leq x \leq r + \beta', 0 \leq h_1(x) + h_2(x) \leq 1, \\ 1, & \text{inak}, \end{cases}$$

kde r je stredná hodnota $\tilde{A}^I$; α a β sú ľavý a pravý rozsah funkcie príslušnosti$\mu_{\tilde{A}^I}(x)$ respektívne α' a β' sú ľavý a pravý rozsah funkcie nepríslušnosti $\nu_{\tilde{A}^I}(x)$. Potom intuitívne fuzzy číslo IFN (Intuitionistic Fuzzy Number)$\tilde{A}^I$ je reprezentované výrazom $\tilde{A}^I = (r; \alpha, \beta'; \alpha')$.



Obrázok 13: Trojuholníkové intuitívne fuzzy číslo (0,2,4;1,2,3) Zdroj: Autor

Nech $\tilde{A}^I$ je intuitívna fuzzy podmnožina v $\Re$ s funkciou príslušnosti $\mu_{\tilde{A}^I}(x)$ a funkciou nepríslušnosti $\nu_{\tilde{A}^I}(x)$ definovanými výrazmi:

$$\mu_{\tilde{A}^I}(x) = \begin{cases} \dfrac{x - a_1}{a_2 - a_1}, & a_1 \leq x \leq a_2 \\ \dfrac{a_3 - x}{a_3 - a_2}, & a_2 \leq x \leq a_3 \\ 0, & \text{inak}, \end{cases} \qquad \nu_{\tilde{A}^I}(x) = \begin{cases} \dfrac{a_2 - x}{a_2 - a'_1}, & a'_1 \leq x \leq a_2 \\ \dfrac{x - a_2}{a'_3 - a_2}, & a_2 \leq x \leq a'_3 \\ 1, & \text{inak}. \end{cases}$$

kde $a'_1 \leq a_1 < a_2 < a_3 \leq a'_3$. Túto IFN budeme volať trojuholníkové intuitívne fuzzy číslo TIFN (Triangular Intuitionistic Fuzzy Number) a značiť $\tilde{a}^I = (a_1, a_2, a_3; a'_1, a_2, a'_3)$. Množina všetkých TIFN sa značí $IF(\Re)$. Grafická reprezentácia TIFN, jeho funkcií príslušnosti a nepríslušnosti je znázornená na obrázku (Obr. 1).

Nech $\tilde{a}^I, \tilde{b}^I \in IF(\Re)$ potom je operácia ich súčtu definovaná:

$$\tilde{a}^I + \tilde{b}^I = (a_1 + b_1, a_2 + b_2, a_3 + b_3; a'_1 + b'_1, a_2 + b_2, a'_3 + b'_3)$$

Poradovú funkciu (Accuracy function) $\tilde{a}^I \in IF(\Re)$ značíme $f(\tilde{a}^I)$a definujeme vzťahom:

$$f(\tilde{a}^I) = \frac{(a_1 + 2a_2 + a_3) + (a'_1 + 2a_2 + a'_3)}{8}.$$

Možno dokázať [Sujeet, Shiv 2014], že takto definovaná poradová funkcia $f(\tilde{a}^I)$je lineárna funkcia.

$$f\left(k\tilde{a}^I + \tilde{b}^I\right) = kf(\tilde{a}^I) + f\left(\tilde{b}^I\right) \qquad (2)$$

Túto vlastnosť využijeme v našom prístupe pri určovaní poradia TIFN čísiel.

Nech $\tilde{a}^I = (a_1, a_2, a_3; a'_1, a_2, a'_3)$ a $\tilde{b}^I = (b_1, b_2, b_3; b'_1, b_2, b'_3)$ sú dve TFIN čísla. Potom môžu byť relačné vzťahy medzi trojuholníkovými intuitívnymi fuzzy číslami, definované takto:

1. $\tilde{a}^I \geq \tilde{b}^I$ak $f(\tilde{a}^I) \geq f\left(\tilde{b}^I\right)$,
2. $\tilde{a}^I \leq \tilde{b}^I$ak $f(\tilde{a}^I) \leq f\left(\tilde{b}^I\right)$,
3. $\tilde{a}^I = \tilde{b}^I$ak $f(\tilde{a}^I) = f\left(\tilde{b}^I\right)$,
4. $min\left(\tilde{a}^I, \tilde{b}^I\right) = \tilde{a}^I$ak $\tilde{a}^I \leq \tilde{b}^I$ alebo $\tilde{b}^I \geq \tilde{a}^I$,
5. $max\left(\tilde{a}^I, \tilde{b}^I\right) = \tilde{a}^I$ak $\tilde{a}^I \geq \tilde{b}^I$ alebo $\tilde{b}^I \leq \tilde{a}^I$.

MPP v intuitívnej fuzzy logike

V tomto odseku ukážeme, ako možno modelovať neistotu pomocou TIFN čísiel v MPP:

Majme maticu typu $m \times n$ TFIN čísiel $\tilde{\mathbf{A}}^I = (\tilde{a}^I_{i,j})$. Majme n-ticu permutácií $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ množiny $\{1,2, \dots, m\}$. Nech $\tilde{\mathbf{A}}^{I\pi}$ je TIFN matica s permutovanými stĺpcovými vektormi, ktorej prvkom v i-tom riadku a j-tom stĺpci je $\tilde{a}^I_{\pi_j(i),j}$

$$\tilde{\mathbf{A}}^{I\pi} = \begin{pmatrix} \tilde{a}^I_{\pi_1(1),1} & \tilde{a}^I_{\pi_2(1),2} & \cdots & \tilde{a}^I_{\pi_n(1),n} \\ \tilde{a}^I_{\pi_1(2),1} & \tilde{a}^I_{\pi_2(2),2} & \cdots & \tilde{a}^I_{\pi_n(2),n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{a}^I_{\pi_1(m),1} & \tilde{a}^I_{\pi_2(m),2} & \cdots & \tilde{a}^I_{\pi_n(m),n} \end{pmatrix}.$$

Označme $\tilde{\mathbf{S}}^{I\pi} = (\tilde{s}^{I\pi})$ vektor m TFIN čísiel, kde i-ty prvok $\tilde{s}^{I\pi}$ je i-ty riadkový súčet matice $\tilde{\mathbf{A}}^{I\pi}$:

$$\tilde{\mathbf{S}}^{I\pi} = \begin{pmatrix} \sum_{j=1}^{n} \tilde{a}^I_{\pi_j(1),j} \\ \sum_{j=1}^{n} \tilde{a}^I_{\pi_j(2),j} \\ \vdots \\ \sum_{j=1}^{n} \tilde{a}^I_{\pi_j(m),j} \end{pmatrix} = \begin{pmatrix} \tilde{s}_1^{I\pi} \\ \tilde{s}_2^{I\pi} \\ \vdots \\ \tilde{s}_m^{I\pi} \end{pmatrix}.$$

Po aplikácii poradovej funkcie na prvky matice $\tilde{\mathbf{A}}^{I\pi}$ dostaneme reálnu maticu $\mathbf{A}^{\pi} = \left(f\left(\tilde{a}^{I\pi}_{ij}\right)\right)$poradových hodnôt:

$$\mathbf{A}^{\pi} = f\left(\tilde{\mathbf{A}}^{I\pi}\right) = \begin{pmatrix} f\left(\tilde{a}^{I}_{\pi_1(1),1}\right) & f\left(\tilde{a}^{I}_{\pi_2(1),2}\right) & \cdots & f\left(\tilde{a}^{I}_{\pi_n(1),n}\right) \\ f\left(\tilde{a}^{I}_{\pi_1(2),1}\right) & f\left(\tilde{a}^{I}_{\pi_2(2),2}\right) & \cdots & f\left(\tilde{a}^{I}_{\pi_n(2),n}\right) \\ \vdots & \vdots & \ddots & \vdots \\ f\left(\tilde{a}^{I}_{\pi_1(m),1}\right) & f\left(\tilde{a}^{I}_{\pi_2(m),2}\right) & \cdots & f\left(\tilde{a}^{I}_{\pi_n(m),n}\right) \end{pmatrix}.$$

V prípade že stĺpce matice $\mathbf{A}^{\pi}$ sú usporiadané od najmenšej hodnoty po najväčšiu, budeme hovoriť o matici hodnôt v normovanom tvare. Takáto matica zodpovedá najhoršiemu možnému usporiadaniu matice.

Vektor $\mathbf{S}^{\pi} = \left(f\left(\tilde{s}_i^{I\pi}\right)\right)$ je vektor hodnôt poradovej funkcie riadkových súčtov matice $\tilde{\mathbf{A}}^{I\pi}$:

$$\mathbf{S}^{\pi} = \begin{pmatrix} f(\tilde{s}_1^{I\pi}) \\ f(\tilde{s}_2^{I\pi}) \\ \vdots \\ f(\tilde{s}_m^{I\pi}) \end{pmatrix} = \begin{pmatrix} s_1^{\pi} \\ s_2^{\pi} \\ \vdots \\ s_m^{\pi} \end{pmatrix}.$$

V prípade normovanej matice $\mathbf{A}^{\pi}$ bude rozdiel medzi prvkami s_m^{π} a s_1^{π} maximálny.

Potom pre maticu intuitívnych fuzzy čísiel $\tilde{\mathbf{A}}^{I}$ a reálnu funkciu $g: \Re^m \to \Re$ uvažujme problém MPPIF:

$$\min\left\{ g(s_1^{\pi}, s_2^{\pi}, \ldots, s_m^{\pi}):\ \pi = (\pi_1, \pi_2, \ldots, \pi_n) \in \Pi;\ s_i^{\pi} = f\left(\sum_{j=1}^{n}\left(\tilde{a}^{I}_{\pi_j(i),j}\right)\right)\right\} \qquad (3)$$

kde g je vhodná miera nerovnomernosti vektora v reálnej aritmetike a Π je množina všetkých n-tíc permutácií.

Vzhľadom na vlastnosť (2) môžeme úlohu (3) preformulovať takto:

$$\min\left\{ g(s_1^{\pi}, s_2^{\pi}, \ldots, s_m^{\pi}):\ \pi = (\pi_1, \pi_2, \ldots, \pi_n) \in \Pi;\ s_i^{\pi} = \sum_{j=1}^{n} f\left(\tilde{a}^{I}_{\pi_j(i),j}\right)\right\}.$$

V prácach (Kluvánek, 1983) a (Tegze, Vlach, 1984), kde boli študované prípady MPP s tzv. Schur-konvexnými mierami nerovnomernosti. My sa obmedzíme len na dva prípady cieľových funkcií. Označme $I = 1,2,\ldots,m$. Nasledujúce funkcie sú príkladmi praktických mier nerovnomernosti, ktoré sú Schur-konvexné:

$$g_{dif}(x_1.\ldots,x_m) = \max_{i\in I}\{x_i\} - \min_{i\in I}\{x_i\},$$

$$g_{ssq}(x_1.\ldots,x_m) = \sqrt{\sum_{i\in I}(x_i - s)^2}, \text{kde } s = \frac{1}{m}\sum_{i\in I} x_i\,.$$

Miera nerovnomernosti g_{dif} je pomerne hrubá a preto bola zovšeobecnená v (Kmeťová, 2005) do citlivejšej miery g_{ssq}. Hodnota s predstavuje ideálny riadkový súčet matice $\mathbf{A}^{\pi}$ vo vektore $\mathbf{S}^{\pi}$.

Dvojstĺpcová inštancia

Dvojstĺpcová matica MPP je najjednoduchším prípadom, ktorý je naviac polynomiálne riešiteľný v reálnej aritmetike. Jej riešenie je založené na nasledujúcom tvrdení.

Veta 1 (Tegze,Vlach 1984) *Matica* $B = (b_{ij}) \in \Re^{m\times 2}$ *je optimálne permutovaná matica MPP pre kritériá* g_{dif} *aj* g_{ssq} *ak platí:*

$$b_{11} \le b_{21} \le \cdots \le b_{m-1,1} \le b_{m1}, \quad b_{m2} \le b_{m-1,2} \le \cdots \le b_{22} \le b_{12}. \qquad (4)$$

V prípade TIFN matice nás vzťah (2) oprávňuje k nasledujúcemu tvrdeniu:

Veta 2 *Matica* $\widetilde{\mathbf{A}}^I = (\tilde{a}_{ij}^I) \in IF(\Re)^{m\times 2}$ *je optimálne permutovaná matica IMPP pre kritériá* g_{dif} *aj* g_{ssq} *ak*

$$\tilde{a}_{11}^I \preccurlyeq a_{21} \preccurlyeq \cdots \preccurlyeq a_{m-1,1} \preccurlyeq a_{m1}, \quad a_{m2} \preccurlyeq a_{m-1,2} \preccurlyeq \cdots \preccurlyeq a_{22} \preccurlyeq a_{12}. \qquad (5)$$

Dôkaz: Ak pre maticu $\widetilde{\mathbf{A}}^I$ platí vzťah (5), potom v dôsledku (3) platí vzťah (4). V zmysle zvolenej poradovej funkcie f je matica $\widetilde{\mathbf{A}}^I$ usporiadaná optimálne.■

V prípade dvojstĺpcovej matice MPPIF $\widetilde{\mathbf{A}}^I$ teda stačí prejsť k reálnej matici poradových hodnôt **A**. Podrobne, majme $m \times 2$ maticu $\widetilde{\mathbf{A}}^I = (\tilde{a}_{i,j}^I)$. Cieľom je nájsť stĺpcové permutácie $\pi = (\pi_1, \pi_2)$ permutácií množiny $I = \{1,2,\dots,m\}$, tj. permutovanej matice

$$\widetilde{\mathbf{A}}^{I\pi} = \begin{pmatrix} \tilde{a}_{\pi_1(1),1}^I & \tilde{a}_{\pi_2(1),2}^I \\ \tilde{a}_{\pi_1(2),1}^I & \tilde{a}_{\pi_2(2),2}^I \\ \vdots & \vdots \\ \tilde{a}_{\pi_1(m),1}^I & \tilde{a}_{\pi_2(m),2}^I \end{pmatrix},$$

zodpovedá matica poradových hodnôt

$$\mathbf{A}^{\pi} = \begin{pmatrix} \tilde{a}_{\pi_1(1),1}^I & \tilde{a}_{\pi_2(1),2}^I \\ \tilde{a}_{\pi_1(2),1}^I & \tilde{a}_{\pi_2(2),2}^I \\ \vdots & \vdots \\ \tilde{a}_{\pi_1(m),1}^I & \tilde{a}_{\pi_2(m),2}^I \end{pmatrix},$$

ktorá minimalizuje hodnotu miery nerovnomernosti:

$$g_{ssqr}(s_1^{\pi}, \dots, s_m^{\pi}) = \sum_{i\in I}^{\cdot} (s_i^{\pi} - s)^2$$

resp.

$$g_{dif}(s_1^{\pi}, \dots, s_m^{\pi}) = \max_{i\in I}\{s_i^{\pi}\} - \min_{i\in I}\{s_i^{\pi}\}$$

kde $s_i^{\pi} = f\left(\tilde{a}_{\pi_1(1),1}^I\right) + f\left(\tilde{a}_{\pi_2(1),2}^I\right)$ pre $i \in I$, $s = \frac{1}{m}\sum_{i\in I} s_i^{\pi}$.

Nasledujúci príklad demonštruje vyššie uvedený postup. Majme:

$$\widetilde{\mathbf{A}}^I = \begin{pmatrix} [0,0,2,5,8] & [0,3,5,6,12] \\ [0,2,6,6,6] & [5,6,7,9,9] \\ [1,5,5,7,8] & [0,1,2,5,47] \\ [2,4,6,6,9] & [3,4,5,7,30] \\ [6,7,8,8,8] & [1,1,2,9,62] \end{pmatrix} \quad \widetilde{\mathbf{S}}^I = \begin{pmatrix} [0,3,7,11,20] \\ [5,8,13,15,15] \\ [1,6,7,12,55] \\ [5,8,11,13,39] \\ [7,8,10,17,70] \end{pmatrix}$$

$$f\left(\widetilde{\mathbf{A}}^I\right) = \begin{pmatrix} 21 & 41 \\ 38 & 57 \\ 41 & 61 \\ 45 & 64 \\ 61 & 81 \end{pmatrix} \quad f\left(\widetilde{\mathbf{S}}^I\right) = \begin{pmatrix} 62 \\ 95 \\ 102 \\ 109 \\ 142 \end{pmatrix}.$$

Miery nerovnomernosti vektora poradových hodnôt TFIN riadkových súčtov $f\left(\widetilde{\mathbf{S}}^{\mathbf{I}}\right)$majú hodnoty $g_{dif} = 80$ a $g_{ssqr} = 25{,}68$.

Stĺpce matice poradových hodnôt funkcií TFIN usporiadame vo vzájomne opačnom poradí, čím dostaneme optimálne usporiadanie matice $\widetilde{\mathbf{A}}^I$.

$$\widetilde{\mathbf{A}}^{I\pi} = \begin{pmatrix} [0,0,2,5,8] & [1,1,2,9,62] \\ [0,2,6,6,6] & [3,4,5,7,30] \\ [1,5,5,7,8] & [0,1,2,5,47] \\ [2,4,6,6,9] & [5,6,7,9,9] \\ [6,7,8,8,8]] & [0,3,5,6,12] \end{pmatrix} \quad \tilde{\mathbf{S}}^{I\pi} = \begin{pmatrix} [1,1,4,14,70] \\ [3,6,11,13,36] \\ [1,6,7,12,55] \\ [7,10,13,15,18] \\ [6,10,12,14,20] \end{pmatrix}$$

$$\boldsymbol{A}^{\pi} = f\left(\widetilde{\mathbf{A}}^{I\pi}\right) = \begin{pmatrix} 21 & 81 \\ 38 & 64 \\ 41 & 61 \\ 45 & 57 \\ 61 & 41 \end{pmatrix} \quad \boldsymbol{S}^{\pi} = f\left(\tilde{\mathbf{S}}^{I\pi}\right) = \begin{pmatrix} 102 \\ 102 \\ 102 \\ 102 \\ 102 \end{pmatrix} \quad \pi = \begin{pmatrix} 1 & 5 \\ 2 & 4 \\ 3 & 3 \\ 4 & 2 \\ 2 & 1 \end{pmatrix}.$$

Miery nerovnomernosti vektora riadkových súčtov $f\left(\tilde{\mathbf{S}}^{I\pi}\right)$ hodnôt poradových funkcií TFIN majú hodnoty $g_{dif} = 0$ a $g_{ssqr} = 0$.

Všeobecná inštancia

Riešenie MPPIF s väčším počtom stĺpcov ako 2 sa dá nájsť podobne ako v prípade MPP pomocou nasledujúceho agregačného heuristického algoritmu:

Krok 1: Náhodne zvolenej matici $\widetilde{\mathbf{A}}^{I}$ typu $m \times n$ priradíme maticu poradových hodnôt **A** a n-ticu identických stĺpcových permutácií $\pi = (\pi_1, \pi_2, \dots, \pi_n)$.

Krok 2: Náhodne vyberieme neprázdnu podmnožinu $J \neq \emptyset$ množiny stĺpcov $J \subset \{1,2,\dots,n\}$ a definujeme $K = \{1,2,\dots,n\} - J$.

Krok 3: Riešime MPP pre $m \times 2$ v reálnej aritmetike v dvojstĺpcovej matici B^{π} s prvkami:

$$b_{i1}^{\pi} = \sum_{j \in J} a_{\pi_j(i),j} \quad b_{i,2}^{\pi} = \sum_{j \in J} j \in K a_{\pi_j(i),j}$$

a vzájomne opačným usporiadaním oboch stĺpcov podľa hodnôt poradových funkcií ich prvkov $b_{i,1}^{\psi}$, získame permutačnú maticu $\psi = (\psi_1, \psi_2)$

Krok 4: Aplikujeme permutáciu ψ_1 na všetky permutácie π_j pre $j \in J$ a tiež ψ_2 na všetky permutácie π_j pre $j \in K$ – t.j. aktualizujeme permutačnú maticu π

Krok 5: Ak je splnené nejaké z kritérií zastavenia (počet krokov, čas výpočtu), tak **STOP,** inak **GOTO Krok 2**.

V prípade že vstupná matica $\widetilde{\mathbf{A}}^{I}$ bola celočíselná, bude aj $8 * \mathbf{A}$ celočíselná vzhľadom na linearitu poradovej funkcie f podľa (2) a môžeme požiť kritériá $g_{dif} \leq 1$ resp. $g_{sqr} \leq 1$. Počítačové experimenty (Peško, Kaukič, 2015) so známymi optimálnymi riešeniami ukazujú, že táto heuristika je veľmi dobrá. Ako ukazujú nasledovné počítačové experimenty, prekvapujúco dobré výsledky sme dosiahli aj so vstupnou normovanou maticou, ktorá predstavuje najhoršiu možnú inštanciu daného typu matice.

Vykonané počítačové experimenty

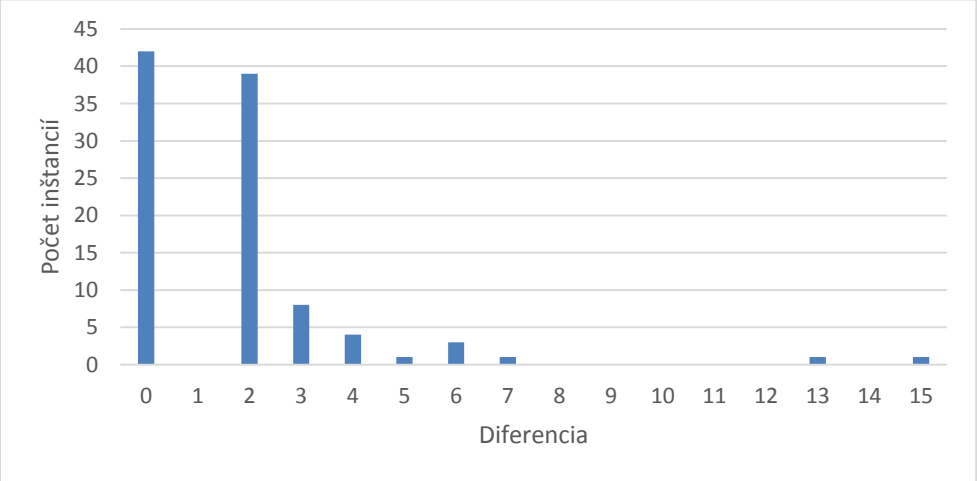
Naše experimenty a simulácie sme vykonali na HP ProBook 6550b (Intel(R) Core(TM) i5 CPU M450 @ 2,40GHz, RAM 4GB s OS Windows 10/OS Linux (Debian/Jessie)). Použili sme Pythonovské softwarové nástroje, menovite Python 3,0 (Rossum, 2016), spolu s knižnicami Numpy, SciPy (Jones, Oliphant, Peterson, 2015) a Matplotlib (Hunter, 2007) v prostrediach Jupyter notebbok (Anon 2015) a IPython (Perez, Granger, 2007).

Ako sme už spomenuli, testovali sme heuristiku na celočíselných inštanciách. Údaje reálneho sveta môžeme skonvertovať do celočíselných hodnôt prenásobením konštantou podľa požadovanej presnosti (napr. 100 pre dve desatinné miesta) a zaokrúhlením. Vykonali sme niekoľko experimentov líšiacich sa typom vstupnej matice $\widetilde{\mathbf{A}}^{I}$ aj počtom inštancií daného typu. Zo všetkých sme pre tento príspevok vybrali experiment matice typu 5x5, kde bol vykonaný výpočet pre 100 rôznych inštancií s maticou v

normovanom tvare s výsledkami spracovanými v grafe na obrázku 2. V tabuľke 1 je pre prehľadnosť výpis skrátený.

Tabuľka 5: Výsledky experimentu s vybranými maticami typu 5x5 s 2000 iteráciami Zdroj: Autor

Inštancia	Počet iterácií	Trvanie iterácie [s]	Počiatočné hodnoty		Koncové hodnoty	
			g_{dif}	g_{ssqr}	g_{dif}	g_{ssqr}
1	1999	0,600	227	76,929	2	0,632
2	1999	0,540	275	93,815	2	0,632
3	1999	0,548	345	117,769	2	0,632
4	1999	0,546	149	52,828	2	0,632
5	210	0,057	315	110,104	0	0,000
6	56	0,017	337	113,583	0	0,000
7	1999	0,549	241	85,091	2	0,632
8	1999	0,535	247	89,342	2	0,632
⋮	⋮	⋮	⋮	⋮	⋮	⋮
46	1999	0,539	254	90,078	2	0,632
47	1999	0,549	171	57,428	2	0,632
48	1999	0,542	478	173,489	13	4,648
49	1999	0,542	354	123,147	6	2,098
50	1999	0,541	172	59,890	2	0,632
51	26	0,007	373	136,820	0	0,000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
96	1999	0,548	369	124,732	6	2,280
97	1999	0,565	193	69,294	2	0,894
98	1999	0,566	221	73,149	2	0,632
99	1999	0,569	237	85,339	2	0,632
100	1999	0,566	182	61,002	2	0,894

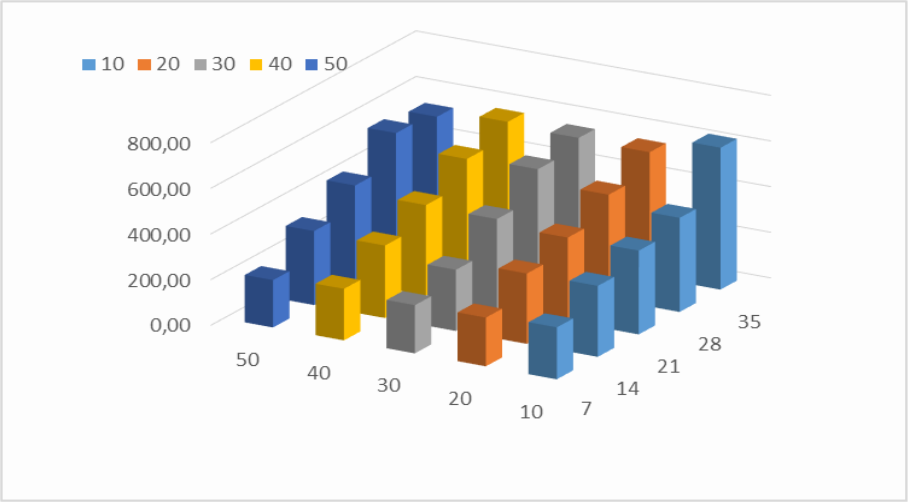


Obrázok 14: Histogram experimentu s maticou typu 5x5, 100 inštancií s 2000 iteráciami
Zdroj: Autor

Z tabuľky 1 a obrázku 2 vidieť úspešnosť algoritmu v prípade riešenia matice, pre ktorú je známe optimálne riešenie. Možno pozorovať absenciu diferencie 1, ktorá je spôsobená skutočnosťou, že sme nepoužili matice s optimálnou diferenciou 1. Pri rozsiahlejších maticiach bola už úspešnosť algoritmu 100%. V ďalších experimentoch sme preto sledovali počet iterácií pri ktorých bolo dosiahnuté optimálne riešenie v zmysle poradovej funkcie s takýmito výsledkami.

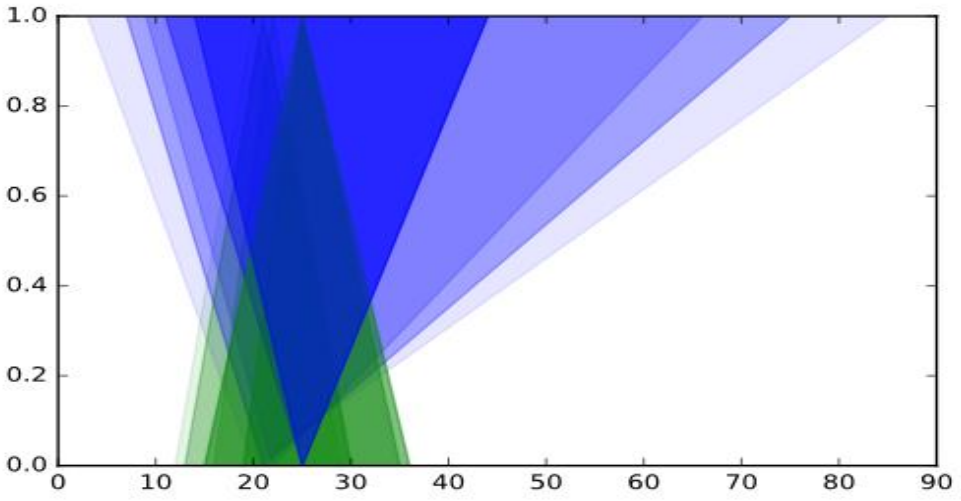
Tabuľka 6: Priemerný počet iterácií Zdroj: Autor

m\n	7	14	21	28	35
10	227,26	310,74	366,88	412,32	621,72
20	213,48	308,04	367,52	457,31	545,41
30	211,07	267,75	391,85	512,13	551,56
40	225,94	317,35	396,65	500,66	565,37
50	207,37	324,23	425,74	557,31	530,87



Obrázok 15: Počet iterácií v závislosti na rozsahu matice Zdroj: Autor

Aj keď sa výsledky experimentov javia ako mimoriadne kvalitné, skutočnosť je iná. Hoci je výsledná funkcia nerovnomernosti optimálna v zmysle poradovej funkcie, z hľadiska prvkov vektora riadkových súčtov $\tilde{\mathbf{S}}^{I\pi}$ sme optimum nedosiahli, pretože hodnoty jednotlivých zložiek intuitívnych čísiel táto metóda nevyrovnáva. Vidieť to na vizualizácii tohoto vektora (Obrázok 4), pričom sú hodnoty mier nerovnomernosti jemu priradeného vektora $\boldsymbol{S}^{\pi}$ rovné nule.



Obrázok 16: Vektor $\tilde{\mathbf{S}}^{I\pi}$ optimálnej matice $\tilde{\mathbf{A}}^{I\pi}$ Zdroj: Autor

Záver

Aplikácie MPPIF predstavujú jednu z ciest ako modelovať neurčitosť v jednom netradičnom type rovnomerných rozvrhov. V tomto príspevku je ukázané ako možno využiť poznatky z riešenia MPP v deterministickej (reálnej) aritmetike pri riešení úloh v intuitívnej fuzzy aritmetike. Umožnila nám to definícia poradovej funkcie, ktorá je súčasťou teórie intuitívnych fuzzy množín. Na základe počítačových experimentov však konštatujeme, že tento prístup len čiastočne rieši náš problém.

V ďalšom výskume sa preto zameriame na vhodnejšiu definíciu poradovej funkcie intuitívneho fuzzy čísla. Budeme vychádzať jednak z doterajších experimentov, kde boli inštancie generované náhodne s rovnomerným rozdelením fuzzy čísiel v matici, ale aj z ďalších experimentov s odlišným rozdelením pravdepodobnosti intuitívnych fuzzy čísiel. Domnievame sa totiž, že ani tak úspešný agregačný algoritmus v reálnej aritmetike, nebude pravdepodobne dosahovať uspokojivé výsledky s fuzzy číslami.

Ďalší výskum preto zameriame na tvorbu vhodného MILP (Mixed Integer Linear Programming) modelu, ktorý sa tiež ukázal úspešný (Peško Kaukič, 2014) v riešení MPP v celočíselnej aritmetike. Tento prístup naviac umožňuje zapracovanie aj takých obmedzujúcich podmienok, ktoré nie sú riešiteľné agregačným algoritmom (bezpečnostná prestávka v mesačných rozvrhoch, technické údržby vozidiel, zdravotné prehliadky vodičov, resp. školenia vodičov atď.)

V záverečnej fáze výskumu sa sústredíme na porovnanie prezentovaného fuzzy prístupu s modelmi založenými na teórii pravdepodobnosti a teórii hrubých množín, ktoré tiež umožňujú modelovať neistotu údajov v MPP rozvrhoch.

Literatúra

PEŠKO, Š., HAJTMANEK, R., 2014. *Matrix permutation problem for fair workload distribution.* In Mathematical methods in economics, MME 2014, 32nd international conference, pages 783–788, Olomouc, Palacky University in Olomouc.

ANON., 2015. Jupyter Documentation. [online] © Copyright 2015, Jupyter Team, dostupné z: http://jupyter.readthedocs.org/en/latest/index.html#user-docs.

BOUDT, K., VANDUFFEL, S., VERBEKEN, K., 2015. *Block rearranging elements within matrix columns to minimize the variability of row sums.* http://ssrn.com/abstract=2634669.

COFFMAN, E., YANNAKAKIS, M., 1984. *Permuting elements within columns of a matrix in order to minimize maximum row sum.* Mathematics of Operations Research 9 (3), 384-390.

CZIMMERMANN, P., 2003a. *The most regular perfect matchings in graph.* Studies of the University of Zilina, Mathematical Series, 16(1):15–18.

CZIMMERMANN, P., 2003b. *A note on using graphs in regular scheduling problems.* Communications, (4):47–48.

CZIMMERMANN, P., PEŠKO, Š., 2003. *The Regular Permutation Scheduling on Graph.* Journal of Information Control and Management System, Vol. 1, pp. 15-22.

ČERNÝ, J., 1985. *How to minimize irregularity?* Pokroky matematiky, fyziky a astronomie, 30(3):145–150.

ČERNÝ, J., PEŠKO, Š., 2006. *Uniform splitting in managerial decision making.* E+M, Economics and Management, 9(4):67–71.

ČERNÝ, J., PEŠKO, Š., Nedatované. *A Note to Weighted Matrix Permutation Problem.* prepared for publication

DELL'OLMO, P., HANSEN, P., PALLOTTINO, S., STORCHI, G., 2005. *On uniform kpartition problems.* Discrete Applied Mathematics, 150(1-3):121–139.

HUNTER, J. D., 2007. *Matplotlib: A 2d graphics environment.* Computing In Science and Engineering, 9(3):90–95.

JONES, E., OLIPHANT, T., PETERSON, P. et al, 2015. *SciPy: Open source scientific tools for Python.* http://www.scipy.org/.

KLUVÁNEK, P. et al., 1983. *Optimalizácia riadenia dopravných systémov*. Report III-9-6/09, Žilina.

PEŠKO, Š., 2006. *The Matrix Permutation Problem in Interval Arithmetic.* Journal of Information, Control and Management Systems.

PEŠKO, Š., KAUKIČ, M., 2014. *Stochastic algorithm for uniform distribution problems.* Research gran APVV-0760-11 Designing of Fair Service Systems on Transportation networks.

PEREZ, F., GRANGER, B. E., 2007. *IPython: a system for interactive scientific computing.* Computing in Science and Engineering, 9(3):21–29, http://ipython.org.

ROSSUM, G.,2016. *Python 3.5.1 documentation.* https://docs.python.org/3/tutorial/http://docs.python.org/2/.

SINGH S. K., YADAV S.P., 2014. *A new approach for solving intuitionistic fuzzy transportation problem of type-2,* Ann Oper Res DOI 10.1007/s10479-014-1724-1,© Springer Science+Business Media New York.

TEGZE, M., VLACH, M., 1984. *The Matrix Permutation Problem.* Report 84-54, Institute fur mathematic, University¨at Graz.

TEGZE, M., VLACH, M., 1984. *On the matrix permutation problem*. Zeitschrift fr Operations Research, 30(3):A155–A159.

JEL Classification: C02, C44, C61, C63, L91

Dedikace: Príspevok je spracovaný ako súčasť výskumného grantu APVV-14-0658 “Optimalizácia mestskej a regionálnej dopravy”.

Summary

Uniform matrix schedules in intuitive fuzzy logic

In this paper the scheduling problem was solved, which was modeled by a matrix.

Our goal is to minimize a suitable non-uniformity measure of row sum vector of the matrix by permuting the column vectors of the matrix.

This NP-hard problem is polynomial solvable for a two column real matrix. There is a probability aggregate algorithm for general matrix. Ordering of columns is the key operation to optimize the matrix.

When uncertain data is given, we get a matrix, whose elements are not simple real numbers, but have more components. Mostly it is not possible to get properly ordered matrix with such elements.

In case of uncertain data, one of the ways is to use the theory of intuitive fuzzy sets, which part is a definition of order function, for which we decided to investigate its suitability for given problem solving.

In this paper we showed how to transform intuitive fuzzy matrix into real matrix by assigning the order function to its elements and used it to optimize the original intuitive fuzzy matrix.

We presented a probability algorithm for the general case of intuitive fuzzy matrix, which is based on aggregate two column principle of known from solving in real arithmetic.

Key words: uniform schedules, matrix permutation problem, intuitive fuzzy numbers, probability

Propensity score matching a alternativní přístup k párování osob pro účely dopadové evaluace

Michal Švarc

xsvam00@vse.cz

Doktorand oboru Ekonometrie a operační výzkum

Školitel: doc. Ing. Tomáš Cahlík, CSc., (tomas.cahlik@vse.cz)

Abstrakt: Tento článek si klade za cíl položit základy rozšíření přiřazovacího problému pro účely dopadové evaluace. V první části tohoto článku jsou shrnuty klasické možnosti párování osob, které využívají metodu propensity score matching (PSM) a zároveň jsou standardně dostupné ve statistických softwarech určených pro kontafaktuální evaluaci. Druhá část článku je věnována zejména přiřazovacímu problému a možnostem jeho rozšíření o dodatečné omezující podmínky, které mají zajistit shodnost profilů párovaných skupin osob.

Klíčová slova: kontrafaktuální dopadová evaluace, propensity score matching, přiřazovací problém

Úvod

Pro vyhodnocování vlivu intervence, jako je například nasazení experimentální léčby nebo poskytnutí dodatečného vzdělání, se požívají metody kontrafaktuální evaluace. Samotné slovo *kontrafaktuál* představuje hypotetickou situaci, ve které by bylo možné zkoumat změnu určité charakteristiky u jedné osoby v případě, že by vlivu intervence vystavena současně nebyla i byla. Samozřejmě takováto skutečnost nemůže v realitě nikdy nastat. Proto v případě čistě experimentálního designu vyhodnocování jsou osoby náhodně rozděleny do dvou skupin, přičemž jedna skupina osob je vystavena vlivu intervence a druhá skupina osob nikoliv. Posléze jsou porovnávány průměrné změny obou skupin a na základě tohoto porovnání je učiněn závěr o vlivu intervence. Bohužel náhodné rozdělení osob není zejména v oblasti sociální politiky zcela obvyklé a naráží zejména na praktické, politické či morální bariéry. Kvůli těmto bariérám byly vyvinuty speciální metody pro odhad průměrného vlivu intervence v anglické literatuře označovaného jako *average treatment effect* (**ATE**) či *average treatment effect on treated* (**ATET**), což je pojem označující průměrný vliv intervence na osoby, které byly intervenci vystaveny. [WOOLDRIDGE, 2010]

Metody odhadu ATE případně ATET lze rozdělit do tří skupin. První skupinu tvoří regresní metody. Patří sem například postup označovaný jako *Difference in Differences* (**DiD**) nebo *First Differences Estimator* (**FD**) a další. Skupinu druhou tvoří metody založené na párování a využívají zejména *propensity score matching*. Poslední skupina metod využívá instrumentálních proměnných k získání odhadu vlivu intervence. [CAMERON, 2005]

Tento příspěvek je věnován právě druhé skupině metod, které využívají párování pro odhad vlivu intervence. První část článku ukazuje dnes již standardní metody PSM jako je *Nearest-Neighbor Matching, Kernel Matching Method, Calliper / Radius Matching a Stratification or Interval Matching*. Druhá část článku je potom zaměřena na využití a modifikaci přiřazovacího problému pro účely párování osob na základě jejich charakteristik. [ROSENBAUM, 1983].

Propensity score matching

Ještě před vysvětlením pojmu *propensity score*, budou vysvětleny některé symboly a označení, které jsou dále využity v textu. Proměnná *w* je indikátorovou binární proměnnou, jež může nabývat hodnot 0 (v případě, že daná osoba nebyla vystavena intervenci) nebo 1, pokud osoba patřila do intervenované skupiny. Písmenem **x** je označen charakteristický vektor osob. Dobrým příkladem pro charakteristický vektor osoby muže být například následující vektor **x** = (1,35,17), který označuje muže (první prvek vektoru **x**), ve věku 35 let (druhý prvek vektoru **x**), jenž věnoval studiu 17 let (poslední prvek vektoru **x**).

Základním předpokladem pro využití konceptu PSM je skutečnost, že pro libovolný charakteristický vektor **x** existuje nenulová pravděpodobnost, že osoba s tímto vektorem patří do intervenované skupiny osob. Stejně tak ovšem musí existovat nenulová pravděpodobnost, že osoba je součástí kontrolní skupiny, tedy skupiny osob bez vlivu intervence. V následující rovnici je tento slovně popsaný vztah ještě zapsán matematicky.

$$0 < \mathrm{P}(w = 1|\mathbf{x}) < 1$$

Jinými slovy lze říct, že pro odhad vlivu intervence ATE je nezbytné, aby existovaly alespoň dvě osoby se stejným nebo obdobným charakteristickým vektorem **x**, přičemž alespoň jedna bude patřit do kontrolní skupiny a alespoň jedna bude ze skupiny osob vystavených intervenci. Pojem *Propensity Score* označuje tedy podmíněnou pravděpodobnost intervence na základě charakteristického vektoru **x.**

$$\mathrm{p}(\mathbf{x}) = \mathrm{P}(\mathrm{w} = 1|\mathbf{X} = \mathbf{x})$$

Tuto podmíněnou pravděpodobnost lze získat například za využití modelů jako je logit nebo probit. [CAMERON, 2005]

Právě pro kvazi-experimentální výzkumný design, kdy rozdělení osob do skupin není náhodné, je PSM vhodnou metodou k odhadu ATET. Bohužel, i propensity score matching s sebou nese určitá úskalí. Prvním problémem je samotná volba charakteristik, na základě kterých budou jednotlivé osoby párovány. Čím vyšší počet charakteristik je zvolen, tím větší naděje je, že osoby si budou opravdu svou charakteristikou podobné, a že nebude opomenut žádný klíčový faktor ovlivňující zkoumaný výsledek. Na druhou stranu ovšem s rostoucím počtem charakteristik pro párování klesá pravděpodobnost nalezení přesné shody, která muže být kompenzována pouze zvětšováním počtu osob ve vzorku.

Následující volba, kterou je nutno učinit, je rovněž spojena s možností určitého zkreslení. Volba spočívá v rozhodnutí, zda v případě, že byl nalezen shodný pár osob na základě charakteristik, jsou spárované osoby vyloučeny z dalšího párování nebo ne. Toto rozhodnutí většinou souvisí s velikostí obou skupin osob. Pokud jsou skupiny dostatečně velké, zejména pak kontrolní skupina, lze rozhodnout, že spárované osoby již nemohou být využity pro další párování. Posledním problémem spojeným s párováním je volba párovací techniky, přičemž pro různé druhy párování mohou vycházet odhady ATET velice odlišně.

Ještě před uvedením jednotlivých technik pro párování je odvozen v následující rovnici výpočet ATET pro případ, kdy ve vektoru **x** jsou pouze diskrétní proměnné.

$$\begin{aligned} E\left(y_{1,i} - y_{0,i}|w_i = 1\right) &= E\left[\left\{E\left(y_{1,i}|\mathbf{x}_i, w_i = 1\right) - E\left(y_{0,i}|\mathbf{x}_i, w_i = 0\right)\right\}|w_i = 1\right] \\ &= E(\Delta_{\mathbf{x}}|w_i = 1) \\ &= \sum_{\mathbf{x}} \Delta_{\mathbf{x}} P(\mathbf{x}_i = \mathbf{x}|w_i = 1) \end{aligned}$$

První způsob matchingu je takzvané přesné párování. Tento způsob párování je vhodné využívat v případě, že je párováno na základě relativně malého počtu proměnných, které nabývají nízkého počtu hodnot. V případě, kdy je párováno na základě mnoha charakteristik nebo v případě matchingu spojitých proměnných již tento způsob párování přestává být výhodný a dobrou alternativou k přesnému párování jsou metody založené na přibližném párování.

Mezi nejčastěji používané metody přibližného párování patří tato čtveřice:

- párování s nejbližším sousedem (*Nearest-Neighbor Matching*),
- párování na základě kernelu (*Kernel matching method*),
- kaliper / párování dle poloměru (*Calliper / Radius matching*),
- stratifikační či intervalové párování (*Stratification or interval matching*)

přičemž první dvě metody lze použít jak v případě párování na základě charakteristického vektoru, tak i v případě, že párování jednotek již probíhá na základě získaného propensity score nebo jiné agregace charakteristického vektoru. Zbylé dvě techniky párování lze použít pouze v případě, kdy jsou k dispozici již agregované hodnoty vektoru **x.**

Ještě před popisem jednotlivých technik přibližného párování je zde zadefinován obecný vzorec pro výpočet odhadu ATET včetně značení, které je dále použito i pro jednotlivé párovací techniky.

Symbolem $A_j(\mathbf{x})$ je značena množina osob z kontrolní skupiny, ve které jsou osoby, jež mají charakteristický vektor přibližně odpovídající charakteristickému vektoru *i*-tého jedince z intervenované skupiny osob. Takto definovanou množina lze zaspat: $A_j(\mathbf{x}) = \left[j \middle| \mathbf{x_j} \in c(\mathbf{x_i})\right]$, přičemž právě $c(\mathbf{x_i})$ je okolím charakteristického vektoru *i*-tého jednice. Počet osob v kontrolní skupině je značen jako N_c a počet osob ve skupině s intervencí je označen N_T. Jako $w_{(i,j)}$ je značená váha *j*-tého jedince z kontrolní skupiny v porovnání s *i*-tým jedincem z intervenované skupiny osob, přičemž platí, že $\sum_j w(i,j) = 1$. Obecný vzorec pro výpočet ATET je uveden v následující rovnici:

$$ATET = \frac{1}{N_T} \sum_{i \in (w=1)} [y_{1,i} - \sum_j w(i,j) y_{0,j}].$$

Nearest-Neighbor Matching

Způsob přibližného párování, který ke každé osobě ze skupiny intervenovaných jedinců hledá takového jedince, který je svých charakterem nejpodobnější, lze zapsat takto:

$$A_i(\mathbf{x}) = \left\{j \middle| min_j \parallel \mathbf{x}_i - \mathbf{x}_j \parallel\right\}.$$

Výraz ve složených závorkách $\parallel \mathbf{x}_i - \mathbf{x}_j \parallel$ reprezentuje eukleidovskou vzdálenost obou charakteristických vektorů. Tento výraz může být případně nahrazen rozdílem hodnot propensity score. V závislosti na počtu nalezených nejbližších sousedů je odvozena i hodnota váhy *w(i,j)*. Hodnotu váhy lze univerzálně zapsat jako *1/n*, kde *n* je právě počet nalezených nejbližších sousedů pro *i*-tého jedince z intervenované skupiny osob. [BECKER, 2002]

Kernel matching

Stejně jak metoda Nearest-Neighbor Matching je aplikovatelná jak v případě párování na základě propensity score, tak i v případě dle charakteristických vektorů. Tato párovací technika využívá k získání váhy *w(i,j)* funkci kernel:

$$w(i,j) = \frac{\mathrm{K}(\mathbf{x}_j - \mathbf{x}_i)}{\sum_{j=1}^{N_{C,i}} \mathrm{K}\,(\mathbf{x}_j - \mathbf{x}_i)}$$

Funkce kernel neporovnává vždy pouze dva charakteristické vektory, nýbrž pro jednotlivé profily osob z intervenované skupiny osob je počítán na základě kernel funkce vážený průměr ze všech profilů a výsledků, které jsou dostupné v kontrolní skupině. [BECKER, 2002]

Calliper / radius matching

Metoda, která je podmíněna tím, že jejím vstupem je skalár jako již aproximovaný charakteristický vektor. Pro každou jednotku ze skupiny osob zapojených do intervence je na základě propensity score vybrána skupina osob z kontrolní skupiny, které mají propensity score v okolí propensity score jednotky z intervenované skupiny. Okolí je definováno určitým poloměrem, který představuje hodnotu dále označovanou jako *r*.

$$A_i(p(\mathbf{x})) = (p_j || \mathrm{p_i} - \mathrm{p_j} || < r)$$

Párování na základě stanovení určité velikosti okolí propensity score vychází z myšlenky Nearest-Neighbor Matching, přičemž obě metody lze kombinovat. Před aplikací metody párování na základě nejbližšího souseda se doporučuje určit velikost okolí, ve kterém jsou nejbližší sousedé hledáni. Pokud v takovém okolí žádný protějšek z kontrolní skupiny není nalezen, pak je takovýto profil reprezentovaný vektorem **x** z analýzy efektů intervence vyloučen.

Stratication or interval matching

Tato metoda vychází z myšlenky, že lze interval *(0,1)*, ve kterém se propensity score nalézá, rozdělit do jednotlivých bloků (stratas) a v rámci jednotlivých stratas porovnávat výsledky osob z obou skupin. Lze uvažovat, že v jednotlivých blocích jsou osoby s přibližně podobným profilem. Nevýhodou tohoto přístupu je skutečnost, že pokud v rámci nějakého bloku nejsou zastoupeny osoby v obou skupinách, nelze uvažovat o odhadu efektů intervence.

Vzorec pro výpočet ATET pro jednotlivé bloky je takovýto:

$$ATET_b^S = \frac{\sum_{i \in I(b)} Y_{1,i}}{N_b^T} - \frac{\sum_{j \in I(b)} Y_{0,j}}{N_b^C}$$

Množina jednotek v rámci jednoho bloku *b* je označena jako *I(b)*. Počet jednotek v bloku *b*, které prošly intervencí, je definován jako N_b^T a počet jednotek z kontrolní skupiny je značen N_b^C. Výpočet celkového ATET je možné provést na základě následujícího vzorce:

$$ATET^{total} = \sum_{b=1}^{B} ATET_b^S \times \frac{\sum_{i \in I(b)} w_i}{\sum w_i},$$

který je ve své podstatě váženým průměrem. Dílčí ATET za jednotlivé bloky jsou váženy počtem jednotek v daném bloku. [CAMERON, 2005]

Z výše uvedených metod pro párování osob z kontrolní a intervenované skupiny lze doporučit ty, které jsou založeny na párování na základě charakteristického vektoru, protože vektor **x** s sebou nese větší informace ve srovnání s propensity score.

Způsob výpočtu odhadu ATE a ATET je poslední věcí, která je v rámci představení metody odhadu efektů intervence na základě užití propensity score. Po odhadu propensity score lze získat odhad ATE dle vztahu:

$$\hat{ATE} = \frac{1}{N} \sum_{i=1}^{N} \frac{[w_i - \hat{p}(\mathbf{x}_i) y_1]}{\hat{p}(\mathbf{x}_i)[1 - \hat{p}(\mathbf{x}_i)]},$$

A odhad ATET dle vztahu:

$$\hat{ATET} = \frac{\sum_{i=1}^{N} \frac{1}{N} \left[\frac{(w_i - \hat{p}(\mathbf{x}_i) y_1)}{\hat{p}(\mathbf{x}_i)(1 - \hat{p}(\mathbf{x}_i))}\right]}{\frac{1}{N} \sum_{i=1}^{N} w_i}.$$

Přiřazovací problém a jeho rozšíření pro účely párování

Alternativní způsob přiřazování k již zmiňovaným technikám matchingu z předchozí části je využití a modifikace úlohy známé z operačního výzkumu jako přiřazovacího problému.

Úloha přiřazovacího problému hledá vzájemně jednoznačné přiřazení prvků ze dvou předem definovaných skupin tak, aby přiřazení splňovalo stanovené podmínky a vzhledem ke stanovenému cíli přinášelo optimální řešení. [LAGOVÁ, 2005]

Matematický model přiřazovacího problému je ve svojí základní podobě následující:

Účelová funkce:

$$z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} x_{i,j}$$

Omezující podmínky:

$$\sum_{i=1}^{m} x_{i,j} = 1 \qquad j = 1,2,\dots,n\,,$$

$$\sum_{j=1}^{n} x_{i,j} = 1 \qquad i = 1,2,\dots,m\,,$$

$$x_{i,j} = 0/1 \qquad i = 1,2,\dots,m,\ j = 1,2,\dots,n$$

Proměnná $x_{i,j}$ je binární proměnnou, která nabývá hodnoty jedna v případě, že je *i*-tý prvek z množiny ***A*** přiřazen *j*-tému prvku z množiny ***B***. V opačném případě je hodnota této proměnné 0. Koeficient $c_{i,j}$ označuje ohodnocení vzniklého párů. První dvě omezující podmínky společně vytvářejí požadavek na to, že každý prvek z množiny ***A*** musí být přiřazen právě jednomu prvku z množiny ***B*** a stejně tak každý prvek z množiny ***B*** musí být asociován k jednomu z prvků množiny ***A***. [LAGOVÁ, 2005]

Modifikace přiřazovacího problému pro účely matchingu

Matematický model přiřazovacího problému ovšem dovoluje vyšší kontrolu při matchingu, zvláště co se týká požadavku vyváženosti, oproti technikám přibližného párování. Tato vlastnost je dána možností přidávat různá omezení do podmínek optimalizace. Před uvedením možností modifikací přiřazovacího problému je diskutováno řešení problémů zjišťování shody profilů jednotlivých osob a následně je zde nově odvozena modifikace účelové funkce přiřazovacího problému.

Shodu profilů osob je možné definovat dvěma možnými způsoby. První možností, jak získat cenový koeficient $c_{i,j}$ pro každý pár osob tvořený jednou osobou z kontrolní skupiny a druhou osobou z intervenované skupiny osob, je absolutní hodnota rozdílu odhadnutých propensity score.

$$c_{i,j} = |\hat{y}_i - \hat{y}_j|$$

Druhým možným způsobem, jak vymezit cenový koeficient $c_{i,j}$, je jeho zadefinování jako prostý nebo vážený součet normalizovaných diferencí jednotlivých charakteristik:

$$c_{i,j} = \sum_{k=1}^{o} v_k e_{i,j}^k.$$

V případě, že se jedná o prostý součet, pak koeficient v_k je pro všechny charakteristiky roven jedné. Pokud jde ovšem o vážený součet, pak koeficient v_k může nabývat různých nezáporných hodnot, avšak je vhodné si stanovit podmínku, že $\sum_{k=1}^{o} v_k = 1$. V rovnici uvedené výše je zmíněna normalizovaná diference profilů a je označována jako $e_{i,j}^k$, avšak pro její výpočet je nutné uvažovat rozdílnost typů proměnných vyjadřujících jednotlivé charakteristiky v profilech osob. Proměnné je možné roztřídit do čtyř kategorií:

- binární/dichotomická proměnná,
- nominální proměnná,
- ordinální proměnná,
- kvantitativní proměnná.

Výpočet diference profilů u binárních proměnných

V případě, že *k*-tá charakteristika může nabývat pouze dvou hodnot, pak se pro její zakódování do datové matice používají tzv. binární nebo též dichotomické proměnné. Pro *k*-tou charakteristiku i-té intervenované osoby je zvoleno označení d_i^k. V případě osoby z kontrolní skupiny se mění u označení pouze dolní index: d_j^k a množina hodnot, kterých takováto proměnná může nabývat, je pouze dvouprvková, čítající hodnoty 0 a jedna. Jako příklad charakteristiky, která je vhodná pro dichotomickou proměnnou, je pohlaví. Hodnotu normalizované diference lze získat na základě vztahu:

$$e_{i,j}^k = |d_i^k - d_j^k| \leftrightarrow \begin{matrix} e_{i,j}^k = 0 \leftrightarrow d_i^k \neq d_j^k \\ e_{i,j}^k = 1 \leftrightarrow d_i^k = d_j^k \end{matrix}$$
$$\max_{i,j}(e_{i,j}^k) = 1$$
$$\min_{i,j}(e_{i,j}^k) = 0$$

Výpočet diference profilů u nominálních proměnných

Charakteristiky profilů, které mohou nabývat více než dvou hodnot a zároveň není možno tyto hodnoty nikterak uspořádat, pak jsou proměnné označující takovéto charakteristiky nazývány jako nominální. Mezi nominální proměnné lze začlenit proměnné označující barvu pleti nebo kraj trvalého bydliště.

Pro *k*-tou charakteristiku *i*-té intervenované osoby je zvoleno označení d_i^k. V případě osoby z kontrolní skupiny se mění u označení pouze dolní index: d_j^k a množina hodnot, kterých takováto proměnná může nabývat, je totožná s množinou přirozených čísel. Výpočet normované diference lze zapsat obdobně, jako tomu bylo v případě binomické proměnné:

$$e_{i,j}^k = |d_i^k - d_j^k| \leftrightarrow \begin{matrix} f(d_i^k, d_j^k) = 0 \leftrightarrow d_i^k \neq d_j^k \\ f(d_i^k, d_j^k) = 1 \leftrightarrow d_i^k = d_j^k \end{matrix}$$
$$\max_{i,j}(e_{i,j}^k) = 1$$
$$\min_{i,j}(e_{i,j}^k) = 0$$

Výpočet diference profilů v případě ordinálních proměnných

Je-li určitá vlastnost nebo charakteristika, jejíž hodnoty lze nějak logicky uspořádat, pak proměnnou, která zastupuje tuto charakteristiku, značíme jako proměnnou ordinální. Ideálním příkladem pro ordinální proměnnou je pak nejvyšší dosažené vzdělání.

V tomto případě ovšem není určení diference profilů úplně triviální. Obtíže jsou způsobeny dodatečnou informací o kolik kategorií se jednotlivé profily liší. Pro získání normalizované diference je v prvním kroku nutno jednotlivé kategorie uspořádat a přiřadit jim číselné hodnoty. Rovněž v případě ordinální proměnné je množina přirozených čísel dostačující pro všechny možné hodnoty, kterých ordinální proměnná může nabývat. Dalším krokem pro získání $e_{i,j}^k$ je určení o kolik kategorií se profily osob odlišují. Jelikož odlišnost může čítat i několik kategorií, je nutno tuto skutečnost eliminovat na základě normalizace na jednotku. Normalizací je myšlen podíl počtu kategorií, o které se profily odlišují a maximálního možného počtu kategorií, o který se mohou profily teoreticky odlišovat. Pokud počet kategorií dané ordinální proměnné je *k*, pak maximální možný počet kategorií, o které se mohou profily odlišovat je *k-1*. Výpočet normované diference lze matematicky zapsat následujícím vztahem:

$$e_{i,j}^k = \frac{|d_i^k - d_j^k|}{k-1}$$
$$\max_{i,j}(e_{i,j}^k) = 1$$
$$\min_{i,j}(e_{i,j}^k) = 0$$

V této rovnici je ve jmenovateli uvedena absolutní hodnota rozdílu pořadových čísel kategorií dané charakteristiky, které byly získány na základě uspořádání. Jmenovatel zlomku je počet kategorií dané charakteristiky snížený o jednotku. [HEBÁK, 2013]

Výpočet diference profilů u ordinálních proměnných

Kvantitativní proměnné mohou nabývat jakýchkoliv hodnot v množině reálných čísel. Typickou charakteristikou pro kvantitativní proměnnou je například příjem z výdělečné činnosti. Normalizovaná diference pro kvantitativní proměnné je definována jako:

$$e_{i,j}^{k} = \frac{|d_i^k - d_j^k|}{max[\underset{i}{max}(d^k), \underset{j}{max}(d^k)] - min[\underset{i}{min}(d^k), \underset{j}{min}(d^k)]}$$

$$\underset{i,j}{max}(e_{i,j}^{k}) = 1$$

$$\underset{i,j}{min}(e_{i,j}^{k}) = 0$$

přičemž v čitateli zlomku je absolutní hodnota rozdílu kvantitativních proměnných a jmenovatel má normalizační funkci. Hodnota jmenovatele je rozdílem maximální a minimální hodnoty kvantitativní proměnné.

Po definování normované diference lze sestavit obecný vztah pro výpočet cenového koeficientu $c_{i,j}$. Je-li množina ***D*** tvořena podmnožinami ***R***, ***S***, ***T*** a ***U***, které jsou popořadě množinami binárních, nominálních, ordinálních a kvantitativních proměnných, přičemž množina ***R*** obsahuje celkem *r* proměnných binárního charakteru, množina ***S*** obsahuje celkem *s* nominálních proměnných, množina ***T*** obsahuje celkem *t* ordinálních proměnných a množina ***U*** obsahuje celkem *u* proměnných kvantitativních, lze vztah pro výpočet $c_{i,j}$ zapsat takto:

$$c_{i,j} = \sum_{k\in\mathbf{R}} v_k\left[\left|d_i^k - d_j^k\right|\right] + \sum_{k\in\mathbf{S}} v_k f(d_i^k, d_j^k) + \sum_{k\in\mathbf{T}} v_k\left[\frac{|d_i^k - d_j^k|}{(n-1)}\right] + \sum_{k\in\mathbf{U}} v_k\left\{\frac{|d_i^k - d_j^k|}{max[\underset{i}{max}(d^k), \underset{j}{max}(d^k)] - min[\underset{i}{min}(d^k), \underset{j}{min}(d^k)]}\right\}$$

Rozšířený matematický model přiřazovacího problému

Účelová funkce:

$$z = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}x_{i,j}$$

Omezující podmínky:

$$\sum_{i=1}^{m} x_{i,j} = b_j \qquad j = 1,2,\dots,n\,,$$

$$\sum_{j=1}^{n} x_{i,j} = a_i \qquad i = 1,2,\dots,m\,,$$

$$\sum_{i=1}^{m} a_i \geq f$$

$$\sum_{j=1}^{n} b_j \geq f$$

$$f \geq g$$

$$x_{i,j} = 0/1 \qquad i = 1,2,\dots,m,\ j = 1,2,\dots,n$$

Výhodou tohoto rozšíření je parametr g. Tento parametr umožňuje nastavovat minimální počet párovaných osob. Další odchylkou od základního modelu je užití parametrů a_i a b_j. Pokud je parametrům dovoleno nabývat hodnot vyšších než jedna, pak je možné využít jednu osobu pro více párů.

Další možné rozšíření přiřazovacího problému je zaměřeno na vyváženost spárovaných osob. Vyváženost může být hlídána vzhledem k charakteristikám celé populace, charakteristikám osob ze skupiny s intervencí, případně vzhledem k vyváženosti párovaných osob z kontrolní a intervenované skupiny. Následující text nově odvozuje omezující podmínky pro přiřazovací problém a je členěn dle toho, zda pojednává o podmínkách vyváženosti vzhledem ke známým hodnotám z populace nebo zda pojednává o vyváženosti párovaných skupin. Další rozčlenění v rámci již vzniklých částí je učiněno na základě typu proměnné, u níž je hlídaná vyváženost.

Pro rozšiřující podmínky přiřazovacího problému, které jsou do matematického modelu začleněny z důvodů kontroly vyváženosti vzhledem k charakteristikám populace, předpokládejme, že je známo, jakých hodnot tyto charakteristiky nabývají. Rovněž se předpokládá, že je znám požadovaný počet spárovaných osob.

Vyváženost charakteristik binárních proměnných při známých charakteristikách populace:

U binomické proměnné je kontrola vyváženosti postavena na základě testu o jedné relativní četnosti. Pokud je k dispozici dostatečně velký vzorek a zároveň známe relativní četnosti z populace, lze provést test o jedné relativní četnosti. (Specifikace testu v příloze 2.)

Na základě dostupných informací o počtu párů a známé relativní četnosti je možné testovací kritérium přepsat pro účely rozšíření podmínek přiřazovacího problému takto:

$$Y < n\pi_0 + u_{1-\frac{\alpha}{2}}\sqrt{n\pi_0(1-\pi_0)},$$

$$Y > n\pi_0 - u_{1-\frac{\alpha}{2}}\sqrt{n\pi_0(1-\pi_0)}.$$

Ještě před uvedením rozšiřujících podmínek pro matematický model je uveden vztah mezi Y a počtem spárovaných osob s danou vlastností dle matematického modelu. Pro osoby z kontrolní skupiny platí, že $Y = \sum_{j=1}^{n} d_j^k \sum_{i=1}^{m} x_{i,j}$ V případě osob ze skupiny, která se zapojila do programu spojeného s intervencí je $Y = \sum_{i=1}^{m} d_i^k \sum_{j=1}^{n} x_{i,j}$.

Pro každou binární proměnnou, která je pro párování využitá a má být u ní hlídána vyváženost vzhledem k známým hodnotám z populace, je do matematického modelu doplněna následující sada čtyř omezujících podmínek.

$$\sum_{j=1}^{n} d_j^k \sum_{i=1}^{m} x_{i,j} < f\pi_0 + u_{1-\frac{\alpha}{2}}\sqrt{f\pi_0(1-\pi_0)},$$

$$\sum_{j=1}^{n} d_j^k \sum_{i=1}^{m} x_{i,j} > f\pi_0 - u_{1-\frac{\alpha}{2}}\sqrt{f\pi_0(1-\pi_0)},$$

$$\sum_{i=1}^{m} d_i^k \sum_{j=1}^{n} x_{i,j} < f\pi_0 + u_{1-\frac{\alpha}{2}}\sqrt{f\pi_0(1-\pi_0)},$$

$$\sum_{i=1}^{m} d_i^k \sum_{j=1}^{n} x_{i,j} > f\pi_0 - u_{1-\frac{\alpha}{2}}\sqrt{f\pi_0(1-\pi_0)}.$$

Jelikož je známý počet požadovaných párů a zároveň je známá i hodnota π_0, je možné považovat výrazy na pravé straně za konstanty. Tato skutečnost zachovává linearitu přidaných omezujících odpovědí a nemění typ optimalizační úlohy.

Vyváženost charakteristiky nominálních a ordinálních proměnných při známých charakteristikách populace:

Jak pro proměnné nominální, tak i pro proměnné ordinální, lze využit pro zajištění vyváženosti relativních četností vzhledem k již známým relativním četnostem z populace Chí-kvadrát testu dobré shody.(Specifikace testu v příloze 2) [PECÁKOVÁ, 2011]

Aby mohlo být usuzováno, že spárované osoby z obou skupin odpovídají svými charakteristikami známému rozdělení z populace, je nutné, aby byla splněna následující podmínka:

$$\sum_{l=1}^{L} \frac{(n_l - n\pi_{l,0})^2}{n\pi_{l,0}} < \chi^2_{1-\alpha}(L-1).$$

Jelikož je počet kategorií dané nominální nebo ordinální proměnné předem známý a rovněž výraz $n\pi_{l,0}$ lze považovat za konstantu, je možné tuto nerovnici zahrnout mezi omezující podmínky matematického modelu. Ještě před zapsáním omezujících podmínek je zde naznačena transformace vektoru obsahujícího sledovanou charakteristiku pro všechny osoby dané skupiny. Vektor je transformován do matice obsahující pouze nuly a jedničky. Počet sloupců matice odpovídá počtu kategorií dané proměnné a počet řádku matice je totožný s počtem osob v dané skupině.

$$\begin{matrix} & & & & \begin{bmatrix} p & q & \cdots & r \end{bmatrix} \\ \mathbf{d^k} = \begin{bmatrix} d_1^k \\ d_2^k \\ d_3^k \\ \vdots \\ d_i^k \\ \vdots \\ d_l^k \end{bmatrix} = \begin{bmatrix} p \\ q \\ p \\ \vdots \\ p \\ \vdots \\ r \end{bmatrix} & \rightarrow \mathbf{W} = & \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} & \begin{bmatrix} i=1 \\ i=2 \\ i=3 \\ \vdots \\ i \\ \vdots \\ i=m \end{bmatrix} \end{matrix}$$

Matice **W** se pro osoby z kontrolní a intervenované skupiny liší pouze počtem řádků, pokud i počty osob v obou skupinách jsou odlišné. Matice **W** pro osoby z kontrolní skupiny a *k*-tou proměnnou bude dále označována jako $\mathbf{W^{C,k}}$. Pro osoby z intervenované skupiny je tato matice označována jako $\mathbf{W^{T,k}}$. S využitím nově zavedených matic **W** lze omezující podmínky pro matematický model zapsat takto:

$$\text{pro intervenovanou skupinu: } \sum_{l=1}^{L} \frac{(\sum_{j=1}^{n} w_{j,l}^{Tk} \sum_{i=1}^{m} x_{i,j} - f\pi_{l,0})^2}{f\pi_{l,0}} < \chi^2_{1-\alpha}(L-1),$$

$$\text{pro kontrolní skupinu: } \sum_{l=1}^{L} \frac{(\sum_{i=1}^{m} w_{i,l}^{Tk} \sum_{j=1}^{n} x_{i,j} - f\pi_{l,0})^2}{f\pi_{l,0}} < \chi^2_{1-\alpha}(L-1).$$

Díky druhé mocnině ve výše uvedených rovnicích jsou ovšem nerovnice nelineární a optimalizační postup nezaručuje nalezení globálního optima. Řešením tohoto problému lze nalézt v nahrazení Chí-kvadrát testu dobré shody testem o jedné relativní četnosti. Pro každou kategorii dané nominální nebo ordinální proměnné by byly do matematického modelu přiřazeny další čtyři nerovnice odvozené v části pojednávající o využití testu o jedné relativní četnosti. Dalším omezením Chí-kvadrát testu dobré shody je podmínka, že četnost v jednotlivých kategoriích musí být větší než 5.

Vyváženost charakteristik kvantitativních proměnných při známých charakteristikách populace:

U kvantitativních proměnných lze vyváženost sledovat na základě testu o shodě dvou středních hodnot. Bohužel tento test vyžaduje znalost výběrových rozptylů, což pro účely lineárního programování není vhodné. Jelikož je předpokládáno, že jsou známy charakteristiky dané proměnné z populace, lze přistoupit k sledování vyváženosti alternativním způsobem.

Pro *k*-tou kvantitativní proměnnou je určen interval spolehlivosti pro její střední hodnotu:

$$(\overline{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) \leq E(x) \leq (\overline{x} + u_{1+\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}),$$

který se stane hranicemi pro střední hodnotu dané proměnné u spárovaných osob. Do matematického modelu přibývají proto následující čtyři omezení:

$$\text{pro intervenovanou skupinu: } \frac{\sum_{i=1}^{m} d_i^k \sum_{j=1}^{n} x_{i,j}}{f} > \overline{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}),$$

$$\frac{\sum_{i=1}^{m} d_i^k \sum_{j=1}^{n} x_{i,j}}{f} < \overline{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}),$$

pro kontrolní skupinu:
$$\frac{\sum_{j=1}^{n} d_j^k \sum_{i=1}^{m} x_{i,j}}{f} > \overline{x} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}),$$
$$\frac{\sum_{j=1}^{n} d_j^k \sum_{i=1}^{m} x_{i,j}}{f} < \overline{x} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}).$$

Pravé strany těchto omezení lze považovat za konstanty, protože jsou předem známy, stejně jako hodnotu *f*, která reprezentuje počet vzniklých párů. Tato omezení jsou lineární, a tudíž jejich přidáním do matematického modelu se nemění typ úlohy. Tato úloha zůstává úlohou lineárního programování a nalezené řešení je globálním optimem.

Test vzájemné vyváženosti párovaných skupin:

Pokud má být testována vzájemná vyváženost kontrolní skupiny vzhledem ke skupině intervenovaných osob, lze využít statistické testy určené pro dva závislé vzorky. Jako první je uveden párový T-test. Další možný test, který je dobře aplikovatelný v matematickém modelu, je neparametrický znaménkový test.

Párový test je vhodný pro porovnávání shodné úrovně sledované kvantitativní proměnné. Předpokladem testu je normalita testované proměnné a alespoň 50 párů. Párový test je specifikován v příloze 2. [PECÁKOVÁ, 2011]

Použití párového T-testu přidává do modelu dvě omezující podmínky, přičemž tyto podmínky jsou nelineární, takže při optimalizačním postupu nelze zaručit globální optimum. Pro přehlednost matematického modelu je nejdříve zavedena proměnná *avDiff* vyjadřující průměrnou diferenci a proměnná *sdDiff* označující výběrovou směrodatnou odchylku.

$$avDiff = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left(d_i^k - d_j^k\right) x_{i,j}}{f},$$
$$sdDiff = \frac{\{\sum_{i=1}^{m} \sum_{j=1}^{n} [(d_i^k - d_j^k) x_{i,j}] - x(i,j) avDiff\}^2}{f-1}.$$

S využitím předem připravených proměnných *avDiff* a *sdDiff* je zápis omezujících podmínek, které zaručují negativní výsledek párového T-testu následující:

$$\frac{avDiff}{sdDiff} < t_{n-1}\left(1-\frac{\alpha}{2}\right)\sqrt{f},$$
$$\frac{avDiff}{sdDiff} > -t_{n-1}\left(1-\frac{\alpha}{2}\right)\sqrt{f}.$$

Pravé strany nerovnic lze opět na základě známého požadovaného počtu vytvářených párů považovat za předem známé konstanty.

V pořadí druhým uváděným testem je znaménkový test. Tento test je používán v případě, že je malý rozsah výběru, stejně tak jako v případě, pokud sledovaná veličina je ordinální. Test spočívá ve sledování, zda hodnota první jednotky z dvojice je větší (+), nebo zda jsou hodnoty obou jednotek z dvojice stejné (0), případně, zda hodnota první jednotky z dvojice je menší(-). Celkový počet případů, ve kterých je první jednotka z dvojice větší než druhá, je označen jako $n_{(+)}$. Pro opačný případ je užit symbol $n_{(-)}$. K ověření hypotézy o stejné úrovni se využívá binomický test, avšak v případě dostatečné velikosti parametru $n_{()}$ velkého vzorku a dvoustranné hypotézy lze využít McNemarovu statistiku. Znaménkový test v případě užití McNemarovy statistiky je specifikován v příloze 2. [PECÁKOVÁ, 2011]

Omezující podmínky, které zajišťují, aby McNemarův test nezamítal nulovou hypotézu o shodné úrovni sledované veličiny, vyžadují zavedení dalších proměnných do modelu. První dvě proměnné ${}_{diff}^{+}d_{i,j}^{k}$ a ${}_{diff}^{-}d_{i,j}^{k}$ jsou nezáporné a označují popořadě kladnou a zápornou odchylku *k*-té charakteris-

tiky v případě páru tvořeného *i*-tou osobou z intervenované skupiny osob a *j*-tou osobou z kontrolní skupiny. Další dvě proměnné ${}_{ind}^{+}d_{i,j}^{k}$ a ${}_{ind}^{-}d_{i,j}^{k}$ jsou binární a slouží jako indikátory nenulovosti příslušné odchylky:

$${}_{ind}^{+}d_{i,j}^{k} = \begin{cases} 1 \Leftrightarrow {}_{diff}^{+}d_{i,j}^{k} > 0 \\ 0 \Leftrightarrow {}_{diff}^{+}d_{i,j}^{k} = 0 \end{cases}$$

$${}_{ind}^{-}d_{i,j}^{k} = \begin{cases} 1 \Leftrightarrow {}_{diff}^{-}d_{i,j}^{k} > 0 \\ 0 \Leftrightarrow {}_{diff}^{-}d_{i,j}^{k} = 0 \end{cases}$$

Aby všechny čtyři nově přidané proměnné zaručily nezamítnutí nulové hypotézy McNemarova testu, je nutné do modelu kromě proměnných samotných přidat i dostatečně velkou konstantu značenou jako Q a následující sadu omezení:

$$(d_i^k - d_j^k)x_{i,j} = {}_{diff}^{+}d_{i,j}^{k} - {}_{diff}^{-}d_{i,j}^{k} \quad i = 1,2,\dots,m, \quad j = 1,2,\dots,n$$

$${}_{ind}^{+}d_{i,j}^{k} + {}_{ind}^{-}d_{i,j}^{k} \leq 1 \quad i = 1,2,\dots,m, \quad j = 1,2,\dots,n$$

$${}_{diff}^{+}d_{i,j}^{k} \leq ({}_{ind}^{+}d_{i,j}^{k})Q \quad i = 1,2,\dots,m, \quad j = 1,2,\dots,n$$

$${}_{diff}^{+}d_{i,j}^{k} \geq {}_{ind}^{+}d_{i,j}^{k} \quad i = 1,2,\dots,m, \quad j = 1,2,\dots,n$$

$${}_{diff}^{-}d_{i,j}^{k} \leq ({}_{ind}^{-}d_{i,j}^{k})Q \quad i = 1,2,\dots,m, \quad j = 1,2,\dots,n$$

$${}_{diff}^{-}d_{i,j}^{k} \geq {}_{ind}^{-}d_{i,j}^{k} \quad i = 1,2,\dots,m, \quad j = 1,2,\dots,n$$

Tato sada omezení využívá obdobný postup jako v případě linearizace absolutní hodnoty, kterou uvádí [JABLONSKÝ, 2011]. Omezující podmínka zaručující nevýznamnost výsledku McNemarova testu je ovšem opět nelineární, což ovlivňuje vlastnost výsledného řešení, které je pouze lokálním optimem. Omezující podmínka je uvedena zde:

$$\frac{[(\sum_{i=1}^{m}\ \sum_{j=1}^{n}\ {}_{ind}^{-}d_{i,j}^{k}) - (\sum_{i=1}^{m}\ \sum_{j=1}^{n}\ {}_{ind}^{+}d_{i,j}^{k})]^2}{(\sum_{i=1}^{m}\ \sum_{j=1}^{n}\ {}_{ind}^{-}d_{i,j}^{k}) + (\sum_{i=1}^{m}\ \sum_{j=1}^{n}\ {}_{ind}^{+}d_{i,j}^{k})} < \chi_{1-\alpha}^{2}(1).$$

Všechny zmiňované omezující podmínky si kladou za cíl zajistit, aby párované skupiny osob si byly co nejvíce podobné, případně se shodovaly i s charakteristikami populace, ze které byly vybrány. Jelikož některé omezující podmínky činí optimalizační úlohu nelineární, lze doporučit tyto podmínky v prvním kroku hledání optimálního spárování vynechat nebo nahradit lineárními a nalézt optimální řešení. Dalším krokem je nalezení optimálního řešení v případě, kdy v modelu jsou všechny (i nelineární) omezující podmínky. Jelikož o nalezeném řešení lze tvrdit pouze to, že je optimální vzhledem ke svému okolí, je vhodné jej konfrontovat s řešením nalezeným v případě lineárního modelu. Pokud jsou hodnoty optimalizované funkce totožné, je nalezené řešení optimální.

Závěr

Způsob párování osob s využitím modifikovaného přiřazovacího problému se zdá býti výhodnější než všechny výše zmiňované způsoby přibližného párování, protože hledá optimální řešení vzhledem ke stanoveným kritériím a zároveň umožnuje zohledňovat i požadavky na vyváženost vzorků párovaných osob vzhledem k jejich charakteristikám z populace, případně obou skupin. Určitou nevýhodou je výpočetní náročnost tohoto způsobu přiřazování při rozsáhlých počtech osob v jednotlivých skupinách a relativně rychle rostoucí počet jak omezujících podmínek, tak i proměnných při zařazování jednotlivých rozšíření do původního matematického modelu. Další nevýhodou tohoto postupu je oproti standardním metodám PSM nutnost využití optimalizačního softwaru. Výše uvedený způsob párovaní vy-

užívající optimalizační model lze proto doporučit zejména při potřebách párovat desítky osob podle nevelkého počtu jejich atributů.

Za hlavní přínos tohoto příspěvku lze považovat odvození omezujících podmínek pro zaručení vyváženosti párovaných osob na základě statistických testů. Dále byla v tomto příspěvku nově odvozena forma modifikace účelové funkce přiřazovacího problému, která umožňuje zohlednění diference profilů párovaných osob.

Literatura

BECKER, Sascha O. a Andrea ICHINO. Estimation of average treatment eects based on propensity scores. *The Stata Journal*. 2002, č. 4, 358{377. Dostupné z: http://www.stata-journal.com/sjpdf.html?articlenum=st0026

CAMERON, Adrian Colin a P TRIVEDI. Microeconometrics: methods and applications. 1st ed. Cambridge: Cambridge University Press, 2005, xxii, 1034 s. ISBN 0-521-84805-9.

HEBÁK, Petr. Statistické myšlení a nástroje analýzy dat. Vyd. 1. Praha: Informatorium, 2013, 877 s. ISBN 978-80-7333-105-4.

JABLONSKÝ, Josef. *Programy pro matematické modelování*. Vyd. 2., preprac. Praha: Oeconomica, 2011, 258 s. ISBN 978-80-245-1810-7.

LAGOVÁ, Milada a Josef JABLONSKÝ. Lineární modely. Vyd. 2.,preprac. Praha: Oeconomica, 2009, 300 s. ISBN 978-80-245-1511-3.

PECÁKOVÁ, Iva. *Statistika v terénich pruzkumech*. 2. dopl. vyd. Praha: Professional Publishing, 2011, 236 s. ISBN 978-80-7431-039-3.

ROSENBAUM, Paul R. a Donald B. RUBIN. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983, vol. 70, issue 1, s. 41-55. DOI:10.1093/biomet/70.1.41.

WOOLDRIDGE, Jerey M. Econometric analysis of cross section and panel data. 2nd ed. Cambridge,Mass.: MIT Press, c2010, xxvii, 1064 p. ISBN 02-622-3258-8.

Příloha 1 – Seznam užitých symbolů a jejich vysvětlení

$\boldsymbol{A^T}$ je množina reprezentující osoby zapojené do programu spojeného s intervencí.

m je počet osob v množině reprezentující osoby zapojené do programu spojeného s intervencí.

i je index označující konkrétní osobu v rámci skupiny osob zapojených do programu spojeného s intervencí.

$\boldsymbol{B^C}$ je množina reprezentující osoby z kontrolní skupiny.

n je počet osob v množině reprezentující osoby z kontrolní skupiny.

j je index označující konkrétní osobu v rámci kontrolní skupiny.

$\boldsymbol{D}$ je množina charakteristik osob, která je společná jak pro osoby z množiny $\boldsymbol{A^T}$, tak i z množiny $\boldsymbol{B^C}$.

o je velikost množiny charakteristik osob, která je společná jak pro osoby z množiny $\boldsymbol{A^T}$, tak i z množiny $\boldsymbol{B^C}$.

d_i^k, d_j^k je hodnota k-té charakteristiky v profilu i-té nebo j-té osoby.

$\boldsymbol{W^{T,k}}$ matice vzniklá transformací vektoru charakteristik d^k pro k-tou proměnnou a intervenovanou skupinu osob.

$\boldsymbol{W^{C,k}}$ matice vzniklá transformací vektoru charakteristik d^k pro *k*-tou proměnnou a kontrolní skupinu osob.

k index označující pořadí charakteristiky v množině charakteristik $\boldsymbol{D}$.

v_k váha *k*-té charakteristicky.

$c_{i,j}$ cenový koeficient ohodnocení shody profilu *i*-té osoby z množiny $\boldsymbol{A^T}$ a *j*-té osoby z množiny $\boldsymbol{B^C}$.

$x_{i,j}$ binární proměnná indikující hodnotou 1 přiřazení *i*-té osoby z intervenované skupiny k j-té osobě z kontrolní skupiny.

$\hat{y}_i, \hat{y}_j$ hodnota odhadnuté podmíněné pravděpodobnosti přirazení do skupiny intervenovaných osob na základě modelu diskrétní volby (propensity score).

$e_{i,j}^k$ normalizovaná diference profilu *i*-té a *j*-té osoby v případě *k*-té charakteristiky.

a_i proměnná označující počet osob z množiny $\boldsymbol{A^T}$, který muže být přiřazen *i*-té osobě.

b_j proměnná označující počet osob z množiny $\boldsymbol{B^C}$, který muže být přiřazen *j*-té osobě.

f celkový počet vytvořených páru.

g minimální počet vytvářených páru.

Příloha 2 – Statistické testy

Test o shodě jedné relativní četnosti

Hypotézy:

$$H_0: p = \pi_0$$
$$H_1: p \neq \pi_0$$

Testovací kritérium:

$$U = \frac{Y - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}}$$

Rozhodnutí:

$$\text{zamítnutí:} H_0 \Leftrightarrow |U| > u_{1-\frac{\alpha}{2}}$$

Chí-kvadrát test dobré shody

Hypotézy:

$$H_0: n_l = n\pi_{l,0}, l = 1,2,\dots,L$$
$$H_1: nonH_0 \neq \pi_0$$

Testovací kritérium:

$$\chi^2 = \sum_{l=1}^{L} \frac{(n_l - n\pi_{l,0})^2}{n\pi_{l,0}}$$

Rozhodnutí:

$$\text{zamítnutí:} H_0 \Leftrightarrow \chi^2 > \chi^2_{1-\alpha}(L-1)$$

Párový test

Hypotézy:

$$H_0: \mu_X - \mu_Y \Leftrightarrow \mu_Z = 0, \mu_Z = \mu_X - \mu_Y$$
$$H_1: \mu_X \neq \mu_Y \Leftrightarrow \mu_Z \neq 0$$

Testovací kritérium:

$$T_n = \sqrt{n}\frac{\overline{Z}}{S} = \sqrt{n}\frac{\overline{X} - \overline{Y}}{S},$$

kde: S je výběrová směrodatná odchylka diference $\overline{Z}, \overline{Z}$ je průměrná diference,n je počet párů.

Rozhodnutí:

$$\text{zamítnutí:} H_0 \Leftrightarrow |T_n| > t_{n-1}(1 - \frac{\alpha}{2})$$

Znaménkový test

Hypotézy:

$$H_0: \pi_{(-)} = \pi_{(+)}$$
$$H_1: \pi_{(-)} \neq \pi_{(+)}$$

Testovací kritérium:

$$\chi^2 = \frac{(n_{(-)} - n_{(+)})^2}{n_{(-)} + n_{(+)}}$$

Rozhodnutí:

$$\text{zamítnutí:} H_0 \Leftrightarrow \chi^2 > \chi^2_{1-\alpha}(1)$$

JEL Classification: C44, C21

Summary

Propensity score matching and alternative approach for persons matching for counterfactual evaluation purposes

First part of this paper provides brief explanation of counterfactual impact evaluation especially propensity score matching methods like Nearest-Neighbor Matching, Kernel matching, Calliper/radius matching and Stratification or interval matching. Following part is dedicated to mathematical model called in operation research assignment problem and its modification for matching purposes.

Keywords: propensity score matching, impact evaluation, assignment problem

Vliv vybraných ekonomických ukazatelů na sebevraždy v České republice

Michaela Antovová

michaela.antovova@vse.cz

Doktorandka oboru statistika

Školitel: doc. Ing. Ladislav Průša, CSc., (ladislav.prusa@vups.cz)

Abstrakt: Článek se zabývá analýzou vlivu hospodářského cyklu, míry nezaměstnanosti a průměrného příjmu na sebevražednost v České republice. Závislost sebevražednosti na vybraných ekonomických ukazatelích byla zkoumána pomocí kointegrace v časových řadách. Pro výpočty byl použit statistický software eViews.

Práce zmiňuje několik teorií významných osobností týkajících se vztahu míry sebevražednosti, hospodářského cyklu, míry nezaměstnanosti a průměrného příjmu.

Dále je proveden vlastní výzkum prostřednictvím časových řad příslušných ukazatelů České republiky. Vztah byl prokázán je u dvou případů ze tří. Výsledky analýzy se s některými v článku zmíněnými teoriemi shodují.

Klíčová slova: sebevražednost, HDP, míra nezaměstnanosti, průměrná mzda

Úvod

Sebevražednost je stále aktuální a diskutované téma. Týká se nejen rozvinutých zemí, ale i rozvojových. V různých kulturách je na ni pohlíženo odlišně – pozitivně či negativně. Na sebevražednost působí mnoho faktorů. Například sociologické, demografické, historické, kulturní, medicínské a v neposlední řadě rovněž ekonomické, jimiž se tato práce zabývá.

V posledních letech došlo nejen v České republice k výrazným změnám hospodářské situace. V roce 2008 proběhla hospodářská krize, která započala ve Spojených státech amerických a postupně zasáhla většinu zemí po celém světě, Českou republiku nevyjímaje. Hospodářská krize měla značný vliv na HDP a míru nezaměstnanosti.

Cílem práce je analyzovat závislost míry sebevražednosti na vybraných ekonomických ukazatelích. Konkrétně na HDP na osobu, míře nezaměstnanosti a průměrném příjmu osoby. Poté jsou výsledky porovnány s ekonomickými teoriemi. Durkheimova, Ginsbergova a Henryho a Shortova se týkají dopadu HDP na sebevražednost. Okunův zákon se zabývá vztahem sebevražednosti a nezaměstnanosti Tvrzení Hamermeshe a Sosse lze aplikovat nejen na sebevražednost a nezaměstnanost, ale i sebevražednost a průměrný příjem.

Metodologie analýzy vlivu ekonomických ukazatelů na míru sebevražednosti

Závislost míry sebevražednosti na HDP, míře nezaměstnanosti a průměrném příjmu byla zkoumána pomocí kointegrace časových řad. Kointegrační analýza se zabývá pouze analýzou vztahů nestacionárních časových řad. (Arlt, Arltová, 2007)

Rozlišujeme krátkodobé a dlouhodobé vztahy mezi časovými řadami. První typ se objevuje mezi časovými řadami jen v relativně krátkém období, časem zmizí a neobnovuje se. Tyto vztahy se objevují pouze u stacionárních časových řad, tedy časovými řadami s krátkou pamětí. Druhý typ je dlouhodobý a s postupem času nemizí. Objevuje se u nestacionárních časových řad, tedy tzv. řad s dlouhodobou pamětí. Informace z minulosti se kumulují. (Arlt, Arltová, 2007)

Velmi důležitý pojmem je ekvilibrium, tedy rovnovážný stav. Časové řady jsou vystaveny neustálým šokům, tudíž se nikdy nenachází v ekvilibriu, ale jsou v dlouhodobém ekvilibriu, což znamená stav, který k rovnovážnému stavu v čase konverguje. Pokud se ani jedna časová řada neodklání dlouhodobě od ekvilibria a existuje mez, za kterou nemůže jít, pak je možné tvrdit, že se jedná o kointegraci časo-

vých řad. Tzn. pokud se jednotlivé časové řady ubírají každá jiným směrem, pak mezi nimi není žádný vztah. (Arlt, Arltová, 2007)

Jednoduchá pravidla o lineárních kombinacích procesu typu stacionárního I(0) a nestacionárního I(1):

- jestliže platí $\{X_t\} \sim I(0)$, potom $\{a + bX_t\} \sim I(0)$ ..... stacionarita zůstává
- jestliže platí $\{X_t\} \sim I(1)$, potom $\{a + bX_t\} \sim I(1)$ ..... nestacionarita zůstává
- jestliže platí $\{X_t\} \sim I(0)$ a $\{Y_t\} \sim I(0)$, potom $\{aX_t + bY_t\} \sim I(0)$ ..... i jejich lineární kombinace je stacionární
- jestliže platí $\{X_t\} \sim I(1)$ a $\{Y_t\} \sim I(0)$, potom $\{aX_t + bY_t\} \sim I(1)$ ..... i jejich lineární kombinace je nestacionární
- v zásadě platí, že pokud $\{X_t\} \sim I(1)$ a $\{Y_t\} \sim I(1)$, potom $\{aX_t + bY_t\} \sim I(1)$ ..... i jejich lineární kombinace je nestacionární

V některých případech poslední pravidlo neplatí a lineární kombinace procesů je stacionární → $\{aX_t + bY_t\} \sim I(0)$. Tento jev je nazýván kointegrací.

V případě regrese časových řad může nastat několik situací. Regresní rovnice:

$Y_t = c + \beta X_t + a_t$,

kde

Y_t – vysvětlovaná proměnná (časová řada),

c – konstanta,

β – síla závislosti mezi proměnnými (časovými řadami),

X_t – vysvětlující proměnná (časová řada),

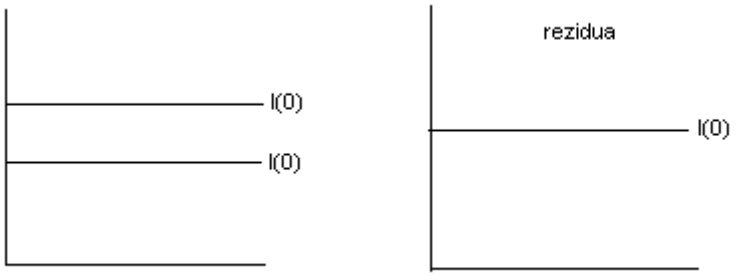
a_t – náhodná složka. (Arlt, Arltová, 2007)

Nastávají následující situace:

- Obě časové řady jsou stacionární a rezidua též. Jde o klasickou regresi, kde stacionární časová řada je vysvětlována stacionární časovou řadou.

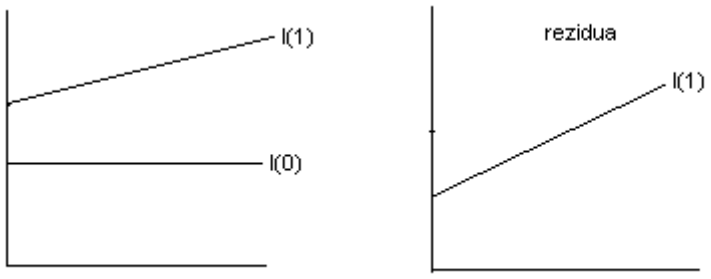
  Mohou nastat dva případy týkající se reziduí:
  a) a_t je bílý šum → statická regrese
  b) a_t je autokorelovaná = nesystematická složka není bílý šum

  Poté se do modelu přidávají zpožděné proměnné X a Y (dynamická regrese) do té doby, než je a_t vyčištěna a stane se z ní bílý šum. Používá se model ADL.



Obrázek 17: Klasická regrese v časových řadách

- Jedna časová řada je nestacionární a druhá je stacionární. Rezidua jsou nestacionární. Vysvětlující časová řada X_t nevysvětlí pohyb vysvětlované časové řady Y_t. Tedy nelze vysvětlit stacionární časovou řadu dynamikou nestacionární časové řady. Jev se nazývá nesmyslná regrese.



Obrázek 18: Nesmyslná regrese v časových řadách

- Obě časové řady a rezidua jsou nestacionární. Případ se nazývá zdánlivá regrese. I přes to, že diagnostické testy (t-testy, F test, index determinace) indikují velmi silný vztah vysvětlované a vysvětlující proměnné, mezi proměnnými není vztah. Diagnostické testy byly nadhodnoceny silnou autokorelací zbytkové složky. Problém lze řešit pomocí diferenciace. Odstraníme nestacionaritu časových řad. Dále postupujeme jako u klasické regrese. Jedná se o jiný než původní model a mezi časovými řadami je možný pouze krátkodobý vztah. Rovnice modelu po diferenciaci:

  $\Delta Y_t = c + \beta \Delta X_t + a_t$,

  kde

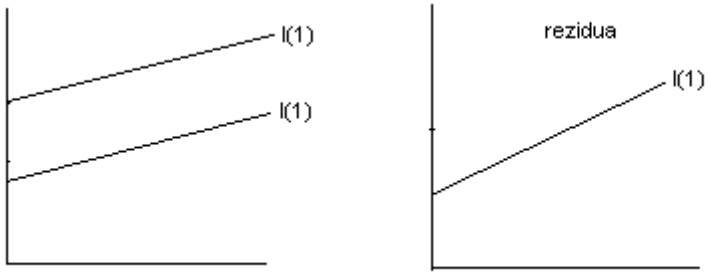
  ΔY_t – vysvětlované proměnné (Y_t, Y_{t-1}, ...),

  c – konstanta,

  β – síla závislosti mezi proměnnými,

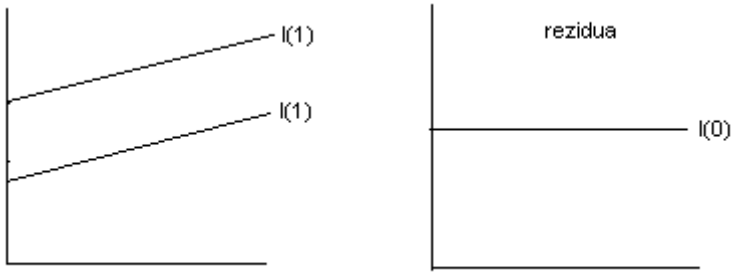
  ΔX_t – vysvětlující proměnná (X_t, X_{t-1}, ...),

  a_t – náhodná složka.



Obrázek 19: Zdánlivá regrese v časových řadách

- Poslední možná situace vzniká, když jsou obě časové řady nestacionární, ale rezidua vykazují stacionaritu. Jedná se pravou (kointegrační) regresi. V tomto případě testujeme krátkodobý i dlouhodobý vztah. (Arlt, Arltová, 2007)



Obrázek 20: Pravá (kointegrační) regrese v časových řadách

Pravá (kointegrační) regrese

V regresi mohou nastat dvě situace týkající se reziduí:

a) a_t je bílý šum → statická regrese = kointegrační regrese → časové řady jsou kointegrované, tedy provázané pouze dlouhodobým vztahem.
b) a_t je autokorelovaná = nesystematická složka není bílý šum. Pro krátkodobé vztahy využíváme modele ADL (model rozdělených zpožděných proměnných) a pro dlouhodobé model ECM (model korekce chyby).

Model krátkodobých vztahů ADL,

kde

$Y_t = c\,(1 - \rho) + \rho\, Y_{t-1} + \beta X_t - \rho\beta X_{t-1} + a_t$

Y_t – vysvětlovaná proměnná,

c – konstanta,

ρ – parametr,

X_t – vysvětlující proměnná,

β – síla závislosti mezi proměnnými,

a_t – náhodná složka.

Dynamizováním modelu ADL (přidáním zpožděných proměnných) vznikne model ECM

$\Delta Y_t = c\,(1 - \rho) + \beta \Delta X_t + (\rho - 1)\,[Y_{t-1} - \beta X_{t-1}] + a_t,$

krátkodobý vztah dlouhodobý vztah

kde

β – dlouhodobý multiplikátor

$(\rho - 1)$ – zatížení.

Zatížení, charakterizuje sílu, kterou se prosazuje dlouhodobý vztah. Čím větší je parametr v absolutní hodnotě, tím více se prosazuje dlouhodobý vztah. Model korekce chyby odlišuje krátkodobé vztahy mezi řadami, tj. vztahy diferencovaných časových řad (skrze parametr β), od vztahů dlouhodobých, tj. vztahů mezi nediferencovanými řadami (vztahy jsou obsažené v tzv. členu korekce chyby $[Y_{t-1} - \beta X_{t-1}]$).

Dle hodnoty parametru ρ mohou nastat tři typy modelu ECM:

- $\rho = 1$, ECM: $\Delta Y_t = \beta \Delta X_t + a_t$

  Mezi časovými řadami není dlouhodobý vztah jen krátkodobý. Tzn. vztah mezi diferencovanými časovými řadami, tzn. mezi přírůstky. V tomto případě má smysl pouze regrese diferencovaných časových řad. Pokud by byla realizována regrese nediferencovaných časových řad, jednalo by se o tzv. zdánlivou regresi.

- $\rho = 0$, ECM: $Y_t = c + \beta X_t + a_t$

  Mezi časovými řadami existuje pouze dlouhodobý vztah, tzv. kointegrační regrese.

- $\rho < 1$, $\Delta Y_t = c\,(1 - \rho) + \beta \Delta X_t + (\rho - 1)\,[Y_{t-1} - \beta X_{t-1}] + a_t,$

  Kointegrační regrese, kdy mezi časovými řadami existují krátkodobé i dlouhodobé vztahy.

Z důvodu analýzy jednosměrné závislosti míry sebevražednosti na vybraných ekonomických ukazatelích byla testována kointegrace jednorovnicovým modelem. (Arlt, Arltová, 2007)

Analýza závislosti míry sebevražednosti na HDP, míře nezaměstnanosti a průměrné mzdě

Analýzu závislosti míry sebevražednosti na hospodářském cyklu byla již dříve zkoumána třemi osobnostmi. Každý z nich vyvtořil odlišnou teorii reakce sebevražednosti na hospodářském cyklu. První z nich byl Emil Durkheim, který rozčlenil sebevraždy na čtyři typy dle síly sociální integrace

a regulace. Z jeho výzkumu vyplívá, že míra sebevražd roste v období hospodářské recese i expanze. V těchto obdobích totiž dochází k nízké regulaci a dle Durkheima vede nízká regulace k vyšší míře sebevražd. Ralph Ginsberg zaznamenal, že míra sebevražednosti se zvyšuje v období ekonomické expanze. Jedinec v tomto období zažívá nespokojenost, protože odměny rostou pomaleji než jeho touhy. Dochází tzv. anomickému procesu, který vede ke zvýšení míry sebevražd. Poslední teorie je od Henryho a Shorta, kteří tvrdili přesný opak, než Ginsberg. Míra sebevražd roste v období hospodářské recese. Vypozorovali, že sebevražda je typická pro vyšší vrstvy společnosti. V případě hospodářské recese, osoba s vyšším statusem ztratí více, než osoba s nižším postavením ve společnosti. Tato změna vede k frustraci a následné agresi, která je u vyšších vrstev namířena proti sobě samému. (Durkheim, 1966), (Lester, Yang, 1997), (Mühlpachr, 2008), (Petrusek, 2011), (Viewegh, 1996)

Vztah míry sebevražednosti a míry nezaměstnanosti popisuje například Okunův zákon: Jestliže se míra nezaměstnanosti zvýší o 1 %, pak je tento jev doprovázen snížením HDP o 2–3 %. Při platnosti teorie Henryho a Shorta by platilo, že pokud se zvýší míra nezaměstnanosti, dojde ke snížení HDP a to vede ke zvýšení míry sebevražednosti. Hamermesh a Soss uvedli teorii, která vychází z celoživotního užitku. Užitek je ovlivněn věkem a permanentním důchodem. Jestliže jedinec vydělává a disponuje více finančními prostředky, pak si může dovolit vyšší spotřebu. Také jeho náklad obětované příležitosti (=příjem, který by mohl vydělat, kdyby nespáchal sebevraždu) je velmi vysoký. Riziko sebevraždy je zanedbatelné. Opačný případ je možno zaznamenat u nezaměstnaného člověka, kdy jsou očekávané příjmy nejisté a nízké. Se zvyšujícím se věkem roste riziko sebevraždy, protože mezní užitek ze spotřeby klesá s věkem. (Hamermesh, 1974), (Soukup, 2010)

Vztah míry sebevražednosti a příjmu zkoumali, jak již bylo zmíněno výše, Hamermesh a Soss. Užitek jedince je ovlivněn příjmem. Největší váhu má pak výdělek v blízké budoucnosti. Z toho vyplívá, že pokud dojde k jeho snížení, povede to k menší spotřebě, uspokojení a to vyústí ve zvýšené riziko sebevraždy. (Hamermesh, 1974)

Závislost míry sebevražednosti na vybraných ekonomických ukazatelích byla analyzována pomocí kointegrace v časových řadách (metodologie popsána v předchozí kapitole). Ke zkoumání časových řad byl použit statistický software eViews. Dickey-Fullerův test odhalil, zda jsou časové řady a jejich rezidua stacionární či nestacionární. V případě, že obě dvě časové řady vykazovaly nestacionaritu a rezidua stacionaritu, se jednalo o pravou (kointegrační) regresi. Mezi řadami pak testujeme, zda je mezi nimi krátkodobý nebo dlouhodobý vztah. Pokud se u časových řad a jejich reziduí prokázala nestacionarita, šlo o zdánlivou regresi a u časových řad byl možný pouze krátkodobý vztah.

Míra sebevražednosti a HDP

Po ověření stacionarity časových řad a reziduí bylo zjištěno, že jde o případ kointegrační regrese.

Tabulka 7: Test jednotkového kořene míry sebeveražednosti, HDP a reziduí

	t- Statistics	Prob.
míra sebevražednosti	-2,010142	0,2806
HDP	-1,561744	0,4837
rezidua	-2,204980	0,0294

Z důvodu přítomnosti autokorelace bylo nutné přidávat do modelu zpožděné proměnné. Model byl dynamizován a statisticky nevýznamné proměnné odstraněny. Diagnostická kontrola modelu probíhala třemi testy: test autokorelace, normality a podmíněné heteroskedasticity. Diagnostické testy potvrdily správnost modelu (tzn. nesystematická složka nebyla autokorelována, měla normální rozdělení a byla homoskedastická).

Tabulka 8: Dynamizovaný model krátkodobého vztahu míry sebevražednosti a HDP

Dependent Variable: SEBEVRAZEDNOST
Method: Least Squares
Date: 04/02/14 Time: 20:11
Sample (adjusted): 1991 2012
Included observations: 22 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.028955	3.758682	2.136109	0.0467
SEBEVRAZEDNOST(-1)	0.629937	0.190594	3.305118	0.0039
HDP	-6.38E-05	1.86E-05	-3.419587	0.0031
HDP(-1)	5.77E-05	1.86E-05	3.099369	0.0062

R-squared	0.861651	Mean dependent var	15.80611
Adjusted R-squared	0.838593	S.D. dependent var	1.683669
S.E. of regression	0.676422	Akaike info criterion	2.218966
Sum squared resid	8.235837	Schwarz criterion	2.417337
Log likelihood	-20.40863	Hannan-Quinn criter.	2.265696
F-statistic	37.36868	Durbin-Watson stat	2.352526
Prob(F-statistic)	0.000000		

Breusch-Godfrey Serial Correlation LM Test	1,023036	Prob. F(2,16)	0,3819
Normality Test: Jarque-Bera	0,774615	Prob	0,678882
Heteroskedasticity test: ARCH	0,121979	Prob. F(1,19)	0,7307

Byl získán model ve tvaru
$SEBEVRAZEDNOST_C$= 8,028 + 0,629 $SEBEVRAZEDNOST_{t-1}$ -6,38E-05 HDP_t + 5,77E-05 HDP_{t-1} + a_t

Rovnice poukazuje na nepřímou závislost mezi mírou sebevražedností a HDP. Pokud klesne HDP, pak se zvýší sebevražednosti. Dle indexu determinace je možno zaznamenat, že model vystihuje dynamiku časové řady sebervažednosti z 86,1 %.

Dlouhodobý vztah je zapsán pomocí modelu korekce chyby EC).
$\Delta SEBEVRAZEDNOST_t$ = 8,028 -6,38E-05 ΔHDP_t -0,370($SEBEVRAZEDNOST_{t-1}$ -3,28E-04 HDP_{t-1}) + a_t

V závorce je možno vidět, že jde o nepřímý vztah mezi zkoumanými ukazateli. Zatížení dosahuje hodnoty -0,370, což je poměrně nízké číslo. Dlouhodobý vztah se v tomto případě prosazuje jen slabě.

Míra sebevražednosti a míra nezaměstnanosti

Nestacionarita časových řad a reziduí poukazovala na zdánlivou regresi.

Tabulka 9: Test jednotkového kořene míry sebeveražednosti, nezaměstnanosti a reziduí

	t- Statistics	**Prob.**
míra sebevražednosti	-2,010142	0,2806
míra nezaměstnanosti	-1,173852	0,6655
rezidua	-1,938213	0,0520

Po diferenciaci časových řad nebyl zjištěn mezi časovými řadami krátkodobý vztah.

Tabulka 10: Diferencovaný model míry sebevražednosti a nezaměstnanosti

Dependent Variable: D(SEBEVRAZEDNOST)
Method: Least Squares
Date: 04/03/14 Time: 12:10
Sample (adjusted): 1991 2012
Included observations: 22 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.186410	0.206003	-0.904891	0.3763
D(MIRA_NEZAMESTNANOSTI)	0.057684	0.141506	0.407643	0.6879

R-squared	0.008240	Mean dependent var	-0.163596
Adjusted R-squared	-0.041348	S.D. dependent var	0.911249
S.E. of regression	0.929897	Akaike info criterion	2.779022
Sum squared resid	17.29416	Schwarz criterion	2.878208
Log likelihood	-28.56924	Hannan-Quinn criter.	2.802387
F-statistic	0.166172	Durbin-Watson stat	2.315436
Prob(F-statistic)	0.687867		

Míra sebevražednosti a průměrný příjem

Test jednotkového kořene potvrdil případ kointegrační regrese.

Tabulka 11: Test jednotkového kořene míry sebeveražednosti, průměrného příjmu a reziduí

	t- Statistics	**Prob.**
míra sebevražednosti	-2,010142	0,2806
míra nezaměstnanosti	-1,037265	0,9141
rezidua	-2,239534	0,0277

Z důvodu přítomnosti korelace bylo nutné zařazovat zpožděné proměnné do modelu a poté odstranit ty, které byly statisticky nevýznamné. Všechny tři testy diagnostické kontroly potvrdily správnost modelu.

Tabulka 12: Dynamizovaný model krátkodobého vztahu míry sebevražednosti a průměrného příjmu

Dependent Variable: SEBEVRAZEDNOST
Method: Least Squares
Date: 04/03/14 Time: 12:53
Sample (adjusted): 1994 2012
Included observations: 19 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19.93313	0.926328	21.51844	0.0000
PRUMERNA_MZDA	-0.001692	0.000546	-3.099392	0.0069
PRUMERNA_MZDA(-1)	0.001509	0.000534	2.827040	0.0121

R-squared	0.651206	Mean dependent var	15.33744
Adjusted R-squared	0.607607	S.D. dependent var	1.258557
S.E. of regression	0.788376	Akaike info criterion	2.506257
Sum squared resid	9.944596	Schwarz criterion	2.655379
Log likelihood	-20.80944	Hannan-Quinn criter.	2.531494
F-statistic	14.93617	Durbin-Watson stat	1.138845
Prob(F-statistic)	0.000219		

$SEBEVRAZEDNOST_C = 19{,}933 - 0{,}00169\ PRUMERNA\_MZDA_t + 0{,}0015\ PRUMERNA\_MZDA_{t-1} + a_t$

Z rovnice je patrné, že se jedná o nepřímou závislost míry sebevražednosti a průměrné mzdy. V případě, že průměrná mzda klesne, míra sebevražednosti se zvýší. Model vysvětluje 65,1 % dynamiky časové řady míry sebevražednosti.

Závěr

Závislost sebevražednosti na HDP na obyvatele byla v programu eViews potvrzena. Mezi ukazateli je nepřímá úměra, tzn. pokud se sníží HDP, pak se zvýší míra sebevražednosti. Výsledek se shoduje s teorií Henryho a Shorta.

Vztah mezi mírou sebevražednosti a nezaměstnanosti nebyl potvrzen. Je ovšem možné, že jiné statistické metody by závislost mezi ukazateli prokázaly. Nejpravděpodobněji by se jednalo o přímou úměru. Poté by se výsledky shodovaly s Okunovým zákonem za platnosti teorie Henryho a Shorta a s teorií Hamermeshe a Sosse.

Vztah míry sebevražednosti a průměrného příjmu byl potvrzen. Jde o nepřímou závislost. Pokud se sníží průměrný příjem, zvýší se míra sebevražednosti. V tomto případě dochází ke shodě s teorií Hamermeshe a Sosse.

Z výzkumu je patrné, že ekonomické podmínky v České republice mají nezanedbatelný vliv na společnost, její chování a tedy i na sebevražednost.

Bylo by velmi zajímavé zkoumat časové řady se zpožděním, protože na sebevraha musí nejprve nějakou dobu působit rizikové faktory (mezi ně patří i ekonomická situace), než svůj úmysl vykoná.

Literatura

ARLT, Josef, ARLTOVÁ, Markéta. *Ekonomické časové řady: vlastnosti, metody modelování, příklady a aplikace.* Praha: Grada, 2007. ISBN 978-80-247-1319-9

Český statistický úřad [online]. 2015 [cit. 2016-01-2016]. Dostupné z: https://www.czso.cz/csu/czso/obyvatelstvo_lide

DURKHEIM, Emile. *Suicide: a study in sociology.* New York: The Free Press, 1966, s. 145-277.

HAMERMESH, Daniel S. a Neal M. SOSS. *An economic theory of suicide: The Journal of Political Economy* [online]. 1974, s. 83-89 [cit. 2014-01-02].

LESTER, David a Bijou YANG. *The Economy and Suicide: Economic Perspectives on Suicide.* New York: Nova Science Publishers, 1997, 186 s. ISBN 1-56072-423-4.

MÜHLPACHR, Pavel. *Sociopatologie.* BRNO: Masarykova univerzita Brno, 2008, s. 117-129. ISBN 978-80-210-4550-7.

PETRUSEK, Miloslav. *Dejiny sociologie.* Ceské Budejovice: Grada Publishing, a.s., 2011, s. 87-177. ISBN 978-80-247-3234-3.

SOUKUP, Jindřich. *Makroekonomie. 2.*, aktualiz. vyd. Praha: Management Press, 2010, 518 s. ISBN 978-80-7261-219-2.

VIEWEGH, Josef. *Sebevražda a literatura.* Ceská republika: Nakladatelství Tomáše Janecka, 1996. ISBN 80-85880-10-5.

JEL kódy: E23, J19

Summary

The influence of selected economic indicators on suicide in the Czech Republic

Artickle analyzes the impact of the economic cycle, rate of unemployment and average income on suicide in the Czech Republic. Dependence suicide on selected economic indicators was examined by cointegration in time series. Calculation were made in statistical software eViews.

The paper contains a number of significant personalities, who made theories about relation between suicides and economic cycle, rate of unemployment and average income.

Than was realized own research with time series of relevant indicators of the Czech Republic. The relation was proved in two cases from three. Results agree with some mentioned theories, with some of them disagree.

Keywords: suicide, GDP, rate of unemployment, average income