FAKULTA
APLIKOVANÝCH VĚD
ZÁPADOČESKÉ
UNIVERZITY
V PLZNI

Posudek oponenta disertační práce

Název práce: **Integration of Voice of Customer into Customer Experience Measurement**

Autorka: **Ing. Lucie Šperková**

Práce se zabývá měřením zákaznické zkušenosti z převážně nestrukturovaných zákaznických dat (Voice of Customer). Cílem je integrace takových dat se strukturovanými daty, např. ze CRM. Téma je určitě aktuální, zasahuje částečně do výzkumu zpracování přirozeného jazyka, ale její hodnotu spatřuji především v praktické oblasti Customer Experience.

V rámci disertační práce vzniklo 17 publikací. U některých z nich je řešitelka hlavním autorem. U ostatních by bylo dobré vědět roli autorky v týmu. Neznám osobně uvedené časopisy nebo konference a impaktní faktor nemusí být vždy vypovídající, zajímalo by mě tedy, co považuje za největší osobní publikační úspěch. V každém případě je publikační činnost dostačující pro obhajobu disertační práce.

Jako přínosy práce vidím zjištění stavu oblasti Customer experience v českých organizacích, navržení datového modelu a metodiky zaintegrování nestruktovaných dat a následnou validaci s několika experty. V oblasti zpracování přirozeného jazyka posun state-of-the-art nevidím, ale takové cíle si disertace ani nakladla.

Rozsah práce je velmi široký, zaobírá se oblastí customer care z technického i byznysového pohledu. Oceňuji kapitolu 3, zorganizovat a vyhodnotit tolik rozhovorů s relevantními zástupci českých firem (navíc z managementu), co mají na starost customer experience, muselo dát hodně práce. Dále se mi líbí aplikování analýzy na opravdu reálná data a v zásadě validace od lidí, kteří se danému byznysu věnují. Použité NLP je založené na slovnících a pravidlech. Přístup je jednoduchý a výhodou je, že nepotřebuje anotovaná data. Ale většinou trpí nižší úplností. Zajímavé by bylo ukázat na některých experimentech posun v případě použití pokročilejších metod, např. embeddingů a neuronových sítí. Dalo by se také potom ověřit, jak

Katedra informatiky a výpočetní techniky **IČ:** 49777513 **Telefon:** (+420) 377632401
Univerzitní 8, 306 14 Plzeň **DIČ:** CZ 49777513 **Fax:** (+420) 377632402
E-mail: cse@kiv.zcu.cz

zlepšená kvalita NLP ovlivňuje byznysovou stránku customer care. V případě některých experimentů mi chybí formálnější ověření výstupu, srovnání více metod, nebo analýza chyb.

Práce je velmi dobře a jasně strukturována. Je také napsána velmi pěknou angličtinou, tabulky a obrázky vhodně dokreslují text. Množství překlepů je minimální. Našel jsem jen:
- LSA jako Latent Semantic Association místo Analysis,
- world embeddings.

Konkrétní přípomínky/náměty pro diskuzi:
- V kapitole 2 jsem očekával skóre měření jako NPS, potom jsem je ale nalezl v kapitole 5. Bylo by zajímavé zjístit při rozhovorech, jaká skóre v praxi převládají. V poslední době často shledávám NPS skóre + feedback volným textem.
- Analýza byla provedena nad jedním typem dat a jedním jazykem. Jaký by byl problém to samé udělat pro jiné industry/jazyk?
- Zákaznická data (obváště emaily) jsou velmi citlivá a firmy mají problém je poskytovat. Je to určitě jeden z aspektů, který velmi s prací souvisí.
- V práci jsem nenašel seznam firem, které dodávají NLP nad zákaznickými daty. Např. v oblasti customer feedback analysis je takových firem hodně, ovšem uznávám, že množina se zúži pokud půjde o češtinu.
- Existuje mnoho use casů, jak zpracování textových zákaznických použít. Jejich zmapování by určité nebylo jednoduché, ale mohlo by práci posunout zas dále.
- Detekce aspektů 4.5: Zajímalo by mě více informací – na jakých datech byla přesnost 0,7 dosažena, srovnání s baseline, pojmenování clusterů v obrázku 4.2.
- V obrázku 4.5, „frequent terms“ vidím stopslova. Je to záměrné? Také se tam vyskytují stemy, které nemusí být vždy čitelné. Možná by se hodilo je nahradit nejčastější formou slova nebo lemmatem, pokud je to možné.
- Crawlery: byla data autory sbírána? Crawlování komentářů a jejich metadat z webu je často složité. Ukázka indikuje spíše menší množství textů.
- Kapitola 5, metriky: Největší problém je propojit ID uživatele mezi různnými zdroji dat. Např. komentáře z webu, návštěvy webových stránek, stížnosti. Jak bylo v tomhle postupováno? Vlastně jde o spojení *Internal systems* a *External systems* v obrázku 5.1.
- Jazyk v sociálních médií je většinou velmi neformální. Nástroje parsingu jsou většinou trénovány na „hezkých“ textech. Jak dobře parsing fungoval? (str. 105)
- Obrázek 5.5: Datum je spojen k názoru a ne ke komentáři, je to záměrem?

Katedra informatiky a výpočetní techniky **IČ:** 49777513 **Telefon:** (+420) 377632401
Univerzitní 8, 306 14 Plzeň **DIČ:** CZ 49777513 **Fax:** (+420) 377632402
E-mail: cse@kiv.zcu.cz

- Nástroj Korektor (str 126): Co zvládá nástroj opravovat? Jaká je jeho úspěšnost?
- Anotace aspektů (str. 127): Pravděpodobně to chtělo nadefinovat striktnější guidelines.
- Syntaktická pravidla: Jsou nadefinovaná pravidla dostatečná?
- Úspěšnost klasifikace sentimentu (str. 135) se zdá být spíše nízká.
- LDA: Kolik dokumentů bylo pro trénování použito? Bojím se že alespoň 100k a 1000 iterací je třeba. Pojmenování témat (str. 140) mi přijde hodně subjektivní (hlavně témata 7-10).
- Úspěšnost detekce aspektů 98% si zasluhuje větší vysvětlení. Pokud se neuvažují *false positives* tak nemusí být vypovídající.
- Str. 145: Chápu správně, že klasifikace emocí (8 tříd) měla *accuracy* 94%? Zdá se mi velmi vysoká.
- Chápu, že práce se zaobírala pouze češinou. Ale zajímavé by bylo srovnání navržených metod na úlohách SemEvalu (např. ABSA). Zjistilo by se, jak si navržený přístup stojí oproti ostatním.
- Je nějaká představa o přesnosti detekce personality?
- Recenze jsou pravděpodobně z pohledu NLP nejjednodušší zdroj. Emaily, obsah na Facebooku nebo přepisy hovorů přináší větší výzvy.
- Z bysnysového pohledu mi trochů chybí diskuze o odchodu zákazníků (churn prediction).

Celkově se mi práce velmi líbila. Záběr je obrovský a další výzkum může práci rozšířit do mnoha směrů.

Práci doporučuji k obhajobě před příslušnou zkušební komisí pro obhajobu disertačních prací.

V Plzni, 17. ledna 2020

……

doc. Ing. Josef Steinberger, Ph.D.

Katedra informatiky a výpočetní techniky **IČ:** 49777513 **Telefon:** (+420) 377632401
Univerzitní 8, 306 14 Plzeň **DIČ:** CZ 49777513 **Fax:** (+420) 377632402
E-mail: cse@kiv.zcu.cz