

## Posudek diplomové práce

### Peter Lacký: Teoretické způsoby modelování uživatelského rozhodování

Jméno oponenta: Marta Vomlelová

Práce se zabývá několika jazyky pro práci s nejistotou. S tématem modelování uživatelského rozhodování je spojena příkladem v šesté kapitole. Práce plně v souladu se zadáním neobsahuje implementaci algoritmů ani modelů.

Práci se zdráhám nazvat teoretickou, neboť její zpracování je velmi povrchní, spíše odpovídající semestrální práci či snesitelné jako úvod k praktické implementační části.

#### Zásadní výhrady:

1. Dle zadání měl řešitel sestavit přehled stávajících teoretických modelů uživatelských preferencí, porovnat je a pokud možno zahrnout i pohled psychologů. Kapitola 3 o preferencích je příliš obecná na to, aby bylo zřejmé, co se má modelovat, čili neposkytuje dostatečný základ pro následující výběr modelovacího jazyka. Ač je Mgr. A. Eckhardt uveden jako konzultant diplomové práce, řešitel necituje žádný z jeho článků o uživatelských preferencích a kapitola 3 zůstává hluboko pod úrovní těchto článků (např. A. Eckhardt: *Inductive Models of User Preferences for Semantic Web, CEUR Workshop Proceedings 2007*).
2. Vzhledem k bodu 1. nejsem přesvědčena o správnosti volby modelů v kapitole 4. Je zřejmé, že z širokého spektra modelovacích jazyků musí řešitel vybrat jen některé, tato volba by se ale neměla jevit náhodná.
3. Klíčové myšlenky by měla obsahovat kapitola 5, Porovnání modelů. Autor se téměř omezuje na jediné kritérium, a to vyjadřovací sílu. Uvítala bych širší spektrum kritérií, např. blízkost vyjadřování experta (pokud očekáváme modely tvořené experty), snadnost strojového učení (pokud očekáváme učení modelu z dat), nároky na paměť a časové nároky na vyhodnocení nejčastějších dotazů (které by se také měly ujasnit v prvních kapitolách).
4. I samotné transformace uváděné v Kapitole 5 jsou povrchní, bez-hlubší znalosti věci. Autor několikrát zdůrazňuje, že bayesovské sítě nepřipouští cykly, proto nejsou schopny reprezentovat formule typu obr. 5.1. Skutečnost je však taková, že libovolná bayesovská síť s uzly dle obr. 5.1. a třemi hranami je schopna reprezentovat libovolné pravděpodobnostní rozložení, tedy i to odpovídající formuli z obrázku („Pokud jsou A, B přátelé, tak kouří buď oba, nebo nikdo“).  
Převod BLP do MLN zcela pomíjí agregující operátory v BLP, obecně jsou pomíjeny číselné informace, takže vůbec není zřejmé, zda ze stejných kvantitativních hodnot v různých modelech dojdeme ke stejnému výsledku. Kapitola neobsahuje žádné precizně definované tvrzení, které by šlo dokázat či vyvrátit, je to jen volný text myšlenek, z nichž některé jsou pravdivé.
5. Příklad v kapitole 6 částečně, ale naprosto nedostatečně, nahrazuje analýzu problému, která měla být provedena v úvodu. Jako absolutní minimum definice problému bych přijala několikánásobně rozsáhlejší příklad, který by se co nejvíce blížil v realitě použitelnému modelu, bez uzlů nazvaných A, B, C. Zdroj inspirace může řešitel vzít např. z publikací o *collaborative filtering*.

Autor neprokázal schopnost zpracovat zadané téma na úrovni magisterské práce, práci nedoporučuji uznat jako diplomovou.

Mgr. Marta Vomlelová PhD.  
KTIML

V Praze 21.1.2009

Následující dvě strany obsahují soupis drobnějších nedostatků, není třeba je v případě mé nepřítomnosti na obhajobě číst, pouze na výslovnou žádost členů komise či autora.

V kapitole 5 je třeba uvádět odkazy na publikace, přesně rozlišit, co je přínosem autora. Je nutné přesně formulovat věty (tvrzení) a dokázat je. Vůbec není zřejmá číselná ekvivalence výsledků v různých modelech, linearita převodů apod.

Návrhy na rozšíření v kapitole 7 jsou jednoduché úvahy bez hlubší znalosti problematiky. Bayesovské sítě umí pracovat se spojitými veličinami (např. pomocí MTE – aproximace mixture of truncated exponentials, po částech exponenciální funkcí). Poslední odstavec zase pokrývají dynamické bayesovské sítě. Zbývá pouze úvaha o fuzzy FLP, kterou nemohu zhodnotit, ale přinejmenším by zasloužila více analyzovat.

Volba klíčových slov mi nepřipadá adekvátní – BN a MN jsou v práci zmíněny spíš okrajově, zato BLP, MLN, FLP v klíčových slovech chybí.

Úvod do logiky (2.1) je třeba psát i formálně logicky správně, protože o korektním formalizmu logika je.

Detailní výhrady: str. 7:

- P není prvotní formule.
- Na více místech v práci není jasné kde končí definice, např. užití definice jazyka  $L_p$  se objevuje další definice DNF, CNF.
- Ve formulaci str. 7 poslední řádek přebývají dvě čárky.  
str. 8:
  - *Goal clause* je něco jiného než fakt.
  - *Definite clause* je něco jiného než uspořádaná klauzule.
  - LD a SLD rezoluce jsou podivně definované, pochybuji, že korektně.
  - Není jasné, jakou rezoluci nakonec autor používá a zda se jedná o rezoluci úplnou.  
str. 10
- Není jasné, kde začínají a končí definice ohodnocení a interpretace termů, obsah definice interpretace termů bych spíše nazvala rozšíření ohodnocení  $e$  na termy.  
str. 11
- Poslední slovo strany „neplatí“ není logický termín. Není pravdivá? Není splnitelná?  
str. 12
- Definice substituce není správně, neboť pomíjí (ne)substituovatelnost termů.
- Bylo by záhodno zmínit vztah logického programu a prologovského programu, místo příkladů uvést definici faktu, pravidla, dotazu.
- Strana končí v půlce věty, chybí tečka a především možnost, že program nedá ani pozitivní, ani negativní odpověď.  
str. 14
- Kde končí definice ekvivalentní relace? (Chybí tečka.)  
str. 15
- Množina Omega není množina jistých jevů.
- Autor by si měl ujasnit, jestli v pravděpodobnosti několika jevů bude psát průnik, čárky, nebo nic, ale nelze to neustále střídat.
- Ve větě o celkové pravděpodobnosti vypadl znak pravděpodobnosti  $P$  z  $P(B_n)$
- Definice nezávislých jevů by měla být uvozena slovem Definice.  
str. 17
- Definice rodiče uzlu je špatně.
- Ve formulích se najednou objevuje pro násobení znak hvězdička, jinde bez znaku.
- Učení parametrů BN z úplných dat je lineární, není třeba „jednoduchší učení sítě“, propagace v BN není obecně v lineárním čase vzhledem k počtu hran.
- Je třeba zmínit, že  $A_i$  jsou z množiny  $U$ .
- Slovo poznámka uvozuje definici, čímž vzniká zmatečný odstavec.
- Chybí tečka na konci věty.

- Autor si musí ujasnit, zda používat desetinnou tečku nebo čárku, nelze to střídat.  
str. 18
- Pojem „potencionálních“ je nejspíš špatně, resp. česky „potenciálních funkcí“ by bylo na daném místě špatně (lze napsat např. množiny (pravděpodobnostních) potenciálů).
- $x\{k\}$  převed'te do matematického modu.  
str. 19
- *Ontologie* není závislost na jazycích.  
str. 22
- Překvapuje mě, že ze seznamu možných základů pro modelování preferencí zcela vypadly expertní znalosti.  
str. 26
- Obrázek 4.1. je velmi užitečný, bylo by vhodné vysvětlit význam šipek a doplnit ty vztahy, kterým se věnuje následující kapitola.  
str. 28
- Definice výpočetního kroku obsahuje nedefinované symboly a je nesrozumitelná.  
str. 29
- Indexy L a Z jsou nedefinované (ať si čtenář domyslí), přitom na nich výsledné číslo záleží.
- Definice Bayesovy klauzule je neúplná, postrádá CPT.  
str. 31
- tabulky ukázek CPT jsou špatně popsány (třetí sloupec je podmíněná pravděpodobnost)  
str. 32
- Pořadí třídy a atributu A.X, X.A nelze měnit v průběhu práce.  
str. 35
- Formule v dolní třetině strany není správně (vypadlo  $1/Z$ ) a zavádění funkce  $\varphi$  považují za zbytečné, neboť její jednodušší výraz se nachází v textu. O (ne)jednoznačnosti převodu viz výše.  
str. 37
- Kde je slibovaný graf obsahující všechny modely spolu s existujícími transformacemi?  
str. 38
- Je D(...) pro MLN definované?
- V TeXu doporučuji používat dynamickou velikost závorek  $\backslash( a \backslash)$ , formule by vypadala lépe.
- V BLP existuje agregační funkce více klauzulí se stejnou hlavou, převod jí nemůže pomíjet.
- Podivné je přetěžování konstant v příkladu, je třeba rozebrat, sníží prostorové nároky za cenu zavedení zmatku pro uživatele modelu.  
str. 43
- Je PRM model schopen predikce pro nové turisty? Pokud ne, proč ho vlastně vytváříme?
- Kde se vzala čísla v dolní části strany? Bylo by vhodné ukázat číselnou ekvivalenci informací odvozených v jednotlivých modelech.  
str. 46
- Pro srovnání modelů je důležité specifikovat nejčastější dotazy. Teprve pak můžeme posuzovat, zda je lepší mít výsledky „předpočítané“ (např. v BN), nebo je vždy odvozovat znovu z plné databáze.

Práce občas obsahuje gramatické chyby, občas chybí slova nebo věta nedává smysl. Např. str. 5: „Zisťovanie závislosti a ich spracovanie sú tvz.“ závislosti čeho? Závislostí z dat? Str. 13: „viachonotovú“, str. 16 „... pravdepodobnostné premennými“. Především je třeba doplnit tečky na konec vět.

*Vom!*

