

Posudek práce

předložené na Ústavu aplikované informatiky Přírodovědecké fakulty JU

- | | |
|--|--|
| <input type="checkbox"/> posudek vedoucího | <input checked="" type="checkbox"/> posudek oponenta |
| <input type="checkbox"/> bakalářské práce | <input checked="" type="checkbox"/> diplomové práce |

Autor/ka: Bc. Miloslav Thon

Název práce: Aplikace algoritmů umělé inteligence pro data mining v prostředí reálného trhu

Studijní obor: Aplikovaná informatika

Datum odevzdání: 12. 12. 2014

Jméno a tituly vedoucího/opponenta: Ing. Jiří Jelínek, CSc.

Pracoviště: Ústav aplikované informatiky

Kontaktní e-mail: jjelinek@prf.jcu.cz

Odborná úroveň práce:

- vynikající velmi dobrá průměrná podprůměrná nevyhovující

Věcné chyby:

- téměř žádné vzhledem k rozsahu přiměřený počet méně podstatné četné závažné

Výsledky:

- originální původní i převzaté netriviální kompilace citované z literatury opsané

Rozsah práce:

- veliký standardní dostatečný nedostatečný

Grafická, jazyková a formální úroveň:

- vynikající velmi dobrá průměrná podprůměrná nevyhovující

Tiskové chyby:

- téměř žádné vzhledem k rozsahu a tématu přiměřený počet četné

Celková úroveň práce:

- vynikající velmi dobrá průměrná podprůměrná nevyhovující

Slovní vyjádření, komentáře a připomínky vedoucího/oponenta:

Autor se v práci zaměřuje na uplatnění standardních algoritmů z oblasti zpracování dat a data miningu na problematiku realitního trhu. Primární zaměření je na textová inzertní data.

Práce je rozdělena do několika kapitol, v rámci kterých autor (tam, kde to lze) vhodně dodržuje jednotící formu popisu obsahu. Není oddělena teoretická a praktická část, je použito tématické členění (více méně dle metodiky zpracování dat), které zlepšuje orientaci v celé problematice.

Celé téma je zpracováno na velmi dobré úrovni, významná pozornost je (v souladu s obecným stavem v data miningu) věnována předzpracování dat. Jednotlivé postupy jsou dobře vysvětleny. Pozitivně lze hodnotit i vytvoření UI frontendu pro konzolové programy.

I přesto je možné k obsahu uvést několik připomínek. Značná část práce je zaměřena na popis obecných metod dataminingu, doplněna by však měla být rešerše zaměřená na uplatnění těchto metod na inzertní data a v prostředí realitního trhu. Jasně zdůvodněn by měl být i výběr použitých metod.

Mezi drobnější nedostatky lze zařadit následující:

- U metody k -Nearest Neighbor by bylo vhodné doplnit podrobnější analýzu volby hodnoty k .
- U kapitoly 5.3 by bylo potřeba doplnit vyhodnocení výsledků a specifikaci koncové podmínky.
- Experimenty v kap. 6.3 a 6.4 by měly být provedeny na stejném zadání, autor tento nedostatek však odstraňuje zanesením porovnatelných výsledků do tab. 6.2.

Množství použité literatury je omezené, autor však vycházel z významných souhrnných zdrojů pro oblasti Information Retrieval a Data miningu.

Práce obsahuje několik drobných textových chyb, její kvalita by mohla být ještě zvýšena dodržením typografických pravidel. Obrázky, tabulky a vzorce obsah vhodně doplňují.

Přiložené CD obsahuje kódy jednotlivých programů a další v práci zmiňované materiály. Použitá struktura je odpovídající.

Případné otázky při obhajobě a náměty do diskuze:

1. Jaký vliv bude mít skladba trénovací množiny na top-down clustering?
2. Jakou sofistikovanější podmínku by bylo možné použít v kap. 5.2 pro ukončení algoritmu?
3. Jaké další metody by bylo možno použít pro zpracování inzertních dat?

Práci

doporučuji

nedoporučuji

uznat jako diplomovou/bakalářskou.

Navrhuji hodnocení stupněm:

výborně velmi dobře dobře neprospěl/a

Místo, datum a podpis vedoucího/oponenta: České Budějovice, 22. 12. 2014



Posudek práce

předložené na Ústavu Aplikované Informatiky Přírodovědecké fakulty JU

posudek oponenta diplomové práce

Autor: Bc. Miloslav Thon

Název práce: Aplikace algoritmů umělé inteligence pro data mining v prostředí reálného trhu

Studijní obor: Informatika

Datum odevzdání: 22.12.2014

Jméno a tituly oponenta: Ing. Marta Vohnoutová

Pracoviště: Katedra informatiky

Kontaktní e-mail: marta.vohnoutova@hp.com

Odborná úroveň práce:	vynikající
Věcné chyby:	téměř žádné
Výsledky:	originální
Rozsah práce:	standardní
Grafická, jazyková a formální úroveň:	vynikající
Tiskové chyby:	téměř žádné
Celková úroveň práce:	vynikající

Slovní vyjádření, komentáře a připomínky vedoucího práce:

Předložená diplomová práce je nadprůměrná. Má logickou strukturu, vhodné členění. Způsob popisu problematiky i řešení je jasný a srozumitelný. Vyjadřování autora je vyspělé. V diplomové práci je poznat, že byla autorem skutečně odpracována a že autor tématice skutečně rozumí.

Připomínky:

1. I když je v diplomové práci množství příkladů, uvítala bych, kdyby jich bylo ještě více. Již proto, že ne každý čtenář rozumí zcela všem popisovaným metodám a je důležité, aby byly pochopeny správně.
2. V diplomové práci mi chyběla možnost používání synonym a podobných výrazů, které by mohly zpřesnit hledání.
3. Na trhu existuje množství různě sofistikovaných produktů. Uvítala bych tabulku srovnání, i když je jasné, že komerční produkty musí mít více možností než výstup diplomové práce.

Případné otázky při obhajobě a náměty do diskuze:

Náměty do diskuse

- kap.7 Detekce duplicitních inzerátů - zde by mohlo pomoci porovnání fotografií, které budou stejné nebo hodně podobné a u většiny inzerátů asi budou. Uvažovat autor i o tomto?

- kap. 8.6 Prezentace výsledků - Uživatel si může zadat, co hledá, ale bylo by dobré, aby systém umožňoval evidovat synonyma. Např. Uživatel zadá "zemědělská usedlost" a systém najde i "prodám stodolu, stáj a seník".

Otázky

- v čem tkví specifičnost na oblast realitních dat a naopak co by bylo nutno předělat, aby bylo možné použít Práci na jinou oblast dat?

- proč se použila shluková analýza, když pracuje na principu, že si systém vytváření shluků vytváří sama? V případě realitních inzerátů je mnoho možností, jak upřesnit pravidla na vytváření shluků - v našem případě víme, co hledáme. Jinými slovy, připadá mi, že autor použil metodu, která znovuobjevuje to, co již od počátku víme. Na druhé straně však použití této metody dělá asi celé řešení univerzálněji použitelné. Může se k tomuto autor vyjádřit?

- kap.6 Vyhledávání podobných inzerátů - jak autor správně napsal, co pro jednoho může být podobné, pro jiného nemusí. Umožňuje program, aby uživatel sám dával nějaké omezující podmínky tomuto hledání?

Práci **doporučuji** uznat jako **diplomovou**.

Navrhuji hodnocení stupněm: **výborně**



Místo, datum a podpis vedoucího: České Budějovice, 2.1.2015