

Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

posudek vedoucího

posudek oponenta

Autor/ka: Dominik Franěk
Název práce: Klastrovací algoritmy
Studijní program a obor: Obecná informatika
Rok odevzdání: 2008

Jméno a tituly vedoucího/opponenta: Doc. RNDr. Iveta Mrázová, CSc.
Pracoviště: KTIML MFF UK

| | excelentní | odpovídající | slabší | nevyhovující |
|-------------------------------------|------------|--------------|--------|--------------|
| Náročnost zadaného tématu | X | | | |
| Míra splnění zadání | | X | | |
| Rozsah práce | | X | | |
| Struktura textové části práce | | X | | |
| Jazyková a typografická úroveň | | | X | |
| Analýza | | X | | |
| Vývojová dokumentace | | X | | |
| Uživatelská dokumentace | | X | | |
| Návrh a design implementace | X | | | |
| Kvalita zpracování softwarové části | X | | | |
| Stabilita aplikace | X | | | |

Nejvýznamnější klady:

- Předkládaná práce je věnována popisu a analýze vlastností celé řady známých klastrovacích algoritmů – k-means, fuzzy k-means, hierarchické klastrování, Kohonenovy mapy, rostoucí Kohonenovy mapy a LVQ.
- Analýza klastrovacích algoritmů zahrnuje i porovnání výpočetní složitosti studovaných algoritmů a otázku volby odpovídajícího počtu shluků pomocí validačních kritérií.
- Součástí práce je velmi pěkně navržená a plně funkční implementace prostředí pro vývoj a testování (adaptivních) klastrovacích algoritmů. Systém podporuje grafické znázornění detekovaných shluků i dat.
- Práce obsahuje i jednoduché porovnání vlastností, resp. výkonnosti vyvinutého systému s dalšími dostupnými systémy (Weka a MATLAB).

Nejzávažnější nedostatky:

- Textová část práce obsahuje velké množství překlepů (např. index i namísto j na 4. ř. sh. na str. 18, 15. ř. zd. na str. 22, chybějící k na 7. ř. sh. na str. 33), typografických (např. 10. ř. sh. na str. 12, 1. ř. sh. na str. 42), gramatických (např. 6. ř. sh. na str. 8, 8. ř. zd. na str. 17) a stylistických chyb (např. posl. věta na str. 8, první věta na str. 14). Větší výhrady lze mít k práci i z věcného hlediska – viz např. výpočet vzdálenosti na str. 9, analýza konvergence algoritmu k-means (algoritmus využívá diskretní členskou funkci), opačné pořadí kroků 2 a 3 v algoritmu na str. 19 (v uvedeném případě budou výstupem algoritmu nové, aktualizované středy shluků, kterým ovšem neodpovídá dříve přepočítaná členská funkce).
- V práci je poměrně často použito nevhodné značení – např. v odstavci Dohoda na str. 9 (zřejmě se předpokládá $k=2$), resp. při aktualizaci těžiště na 5. ř. sh. na str. 16 (n by mělo odpovídat počtu vzorů v daném shluku), podobně i ve vztahu (2.12) na str. 32.
- Text na mnoha místech neobsahuje původní zdroj informací.

Další poznámky:

- Celkově má práce dobrou úroveň. Především vyvinutá aplikace představuje kvalitní a dobře navržený nástroj pro práci s celou řadou známých klastrovacích algoritmů a dává zároveň i výborný základ pro vývoj dalších, pokročilejších shlukovacích technik. Proto doporučuji uznat práci Dominika Fraňka jako bakalářskou práci.
- Modul pro grafické znázornění výsledků klastrování by měl podporovat volbu atributů pro zobrazování a testování implementovaného systému by mohlo být rozsáhlejší.

| | výborně | velmi dobře | dobře | něprospěš/a |
|--------------|---------|-------------|-------|-------------|
| Návrh známky | | X | | |

Datum: 29.8.2008

Podpis: