

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Autor/ka: David Bakrlík
Název práce: Modelování dat charakterizujících virtuální server pomocí Support Vector Machine
Vedoucí práce: prof. Ing. Vladimír Olej, CSc.
Oponent: doc. Ing. Pavel Petr, Ph.D.

Hodnocení práce:

(hodnocení: 1 je výborný, ... , 4 je nevyhovující)

	1	1-	2	2-	3	4
1. Náročnost tématu			X			
2. Volba vhodné metodiky zpracování			X			
3. Splnění cílů práce		X				
4. Odborný přínos autora/rky			X			
5. Logický postup řešení			X			
6. Využití teoretických znalostí		X				
7. Návrh modelu řešení			X			
8. Sběr dat, materiálu pro řešení				X		
9. Výstižný souhrn práce a vhodná klíč. sl.			X			
10. Průběžná citace použité literatury					X	
11. Formální úprava textu				X		
12. Formální úprava grafická (obrázky, ...)			X			
13. Celkové hodnocení práce			X			

Další připomínky, vyjádření, doporučení nebo nedoporučení k obhajobě:

Definovaný cíl práce byl splněn. Jedná se o standardní úlohu klasifikace s využitím SVM.

K práci mám celou řadu připomínek:

- V první kapitole autor na cca 25 stranách popisuje problematiku virtualizace, i když to přímo nesouvisí s řešenou problematikou.
- Kapitola 4 a 5 je věnována návrhu klasifikace a analýze modelu. Bohužel zde není ani ukázka analyzovaných dat a tudíž nelze posoudit vhodnost nebo dostatečnost předzpracování. Konstatování, že data jsou ve čtyřech různých souborech a je třeba je sloučit, je nedostatečné. Stejně tak je v této části celá řada nepřesností. Je uvedeno, že výsledný soubor obsahuje cca 42000 záznamů. Následně se ale pro zpracování používá 206. Proč? Postup segmentace na str. 55 je popsán jen velmi vágně. Pokud chceme omezit počet analyzovaných záznamů, je třeba vybrat reprezentativní vzorek, který bude mít stejnou popisnou statistiku jako původní soubor. A to takto nejde.
- Dále zde autor popisuje využití segmentace na rozdělení záznamů na trénovací a testovací množinu s využitím segmentace. Následně ale využije standardní náhodně rozdělení s využitím nástroje Clementine.
- Po formální stránce práce obsahuje celou řadu nedostatků (vyberu jen některé):

- nejednotné titulky u obrázků a tabulek (malá i velká počáteční písmena);
- chybějící zdroje u některých obrázků a tabulek;
- velice nedůsledná citace použité literatury v textu (např. u Cloudu skoro žádná);
- velké množství stylistických a gramatických chyb.

Práci doporučuji k obhajobě

Hodnocení práce známkou: VELMI DOBŘE

Otázky k obhajobě:

1. Objasněte důvod pro určení počtu segmentu 10 v části předzpracování.
2. Vysvětlete, proč jste nepoužil dělení dat na trénovací, testovací a validační množinu.
3. Vysvětlete redukci datové matice z 42000 záznamů na 206.
4. Popište, co jsou parametry C, N, L v tabulce 5.

V Pardubicích, dne 29. 5. 2013