Posudek oponenta diplomové práce

Student: Hošták Viliam Samuel, Bc.
Téma: Učení se automatů pro rychlou detekci anomálií v síťovém provozu (id 23093)
Oponent: Matoušek Petr, Ing., Ph.D., M.A., UIFS FIT VUT

1. Náročnost zadání **značně obtížné zadání**

Téma DP je výzkumné a zaměřuje se na vytváření pravděpodobnostních konečných automatů ze vzorků síťové komunikace a následné využití natrénovaných automatů na detekci anomálií v průmyslové řídící komunikaci pomocí protokolu IEC 104. Zadání považuji za značně obtížné.

2. Splnění požadavků zadání **zadání splněno**

Zadání práce bylo splněno ve všech bodech.

3. Rozsah technické zprávy **přesahuje obvyklé rozmezí**

Rozsah práce přesahuje obvyklé rozmezí. Je to dáno zejména tím, že dle zadání měl student prostudovat algoritmy na učení KA L*, Lambda* či Alergia. Pro vlastní výzkum a experimenty byl vhodný pouze poslední přístup, takže některé části se mohou zdát zbytečně podrobné. Nicméně pro přehled a zhodnocení možných přístupů jsou užitečné.

4. Prezentační úroveň předložené práce **90 b. (A)**

Struktura technické zprávy je dobrá, třetinu tvoří teorie popisující KA a metody jejich učení na základě vstupní množiny řetězců, zbytek práce popisuje návrh systému učení, zpracování dat a experimenty. Práce obsahuje velké množství definic, vzorců a vět, které jsou dobře popsány a odkazovány. Našel jsem pouze drobné nepřesnosti, např. v definici deterministického KA na str. 12 či chybějící test hodnoty t_0 v algoritmu Alergia na str. 35. Implementace je však správně.

5. Formální úprava technické zprávy **85 b. (B)**

Typografická a jazyková stránka práce je vyhovující. Doporučil bych číslovat vzorečky (pro odkazování na ně je to výhodné). Místo slova "výrok" se v odborné terminologii používat výraz věta, lemma či teorém. Některé hodnoty v grafech v kapitole 7.3 nejsou dobře rozlišitelné a bylo by vhodnější uvést je například formou tabulky, která by srovnala tyto hodnotu u všech experimentů.

6. Práce s literaturou **100 b. (A)**

Student používal pro řešení odbornou literaturu a vědecké články týkající se teorie jazyků a konečných automatů. Všechny převzaté části dobře ocitoval. Z popisu algoritmů a teorií je vidět, že se student v oboru velmi dobře orientuje a je schopen prostudovat i teoretické odborné články. Zároveň je schopen přidat vlastní hodnocení a použitelnost pro danou oblast aplikace.

7. Realizační výstup **95 b. (A)**

Výstupem je sada prototypových nástrojů, které zpracovávají vstupní řetězce a vytvářejí model ve formě pravděpodobnostních automatů (trénovací fáze). Student dále implentoval skripty pro detekci anomálií. Zdrojové kódy jsou dobře označené a komentované. Součástí odevzdaných souborů je i podrobný popis instalace a spouštění s příklady (soubor README).

8. Využitelnost výsledků

Téma práce vychází z výzkumu bezpečnosti průmyslových protokolů, které se řeší v projektu MV Bonnet. Vytvořené výsledky jsou užitečné zejména pro vyhodnocení různých metod vytváření modelů komunikace na základě vzorků a jejich využití pro detekci anomálií. Zejména zajímavé jsou části týkající se paketového či konverzačního módu či porovnání různých metod slučování stavů vytvářených konečných automatů.

9. Otázky k obhajobě

- Mohl byste shrnout, které typy testovaných anomálií (útoků) je možné odhalit pomocí pravděpodobnostních automatů a pro které je nutné využít statistickou analýzu? Uveďte výhody či omezení těchto přístupů.
- Jak by vypadala detekce anomálií v případě, že místo pravděpodobnostního modelu vytvoříme pouze frekvenční prefixový strom? Zhodnoťte fázi učení a detekce z pohledu rychlosti a přesnosti při odhalování anomálií.

10. Souhrnné hodnocení **95 b. výborně (A)**

Odevzdanou práci považuji za jednu z nejlepších, kterou jsem měl možnost hodnotit, jak z pohledu teoretického zpracování, tak i vzhledem k provedeným experimentům. Student nejenže nastudoval a pochopil netriviální látku, ale byl schopen navrhnout vlastní vylepšení, například různé režimy zpracování (paketový, konverzační, využití tříd ekvivalence) či různé heuristiky pro efektivní slučování stavů. Navrhuji hodnocení A, výborně.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 7. srpna 2020

Matoušek Petr, Ing., Ph.D., M.A.
oponent