



Hodnocení vedoucího závěrečné práce

Student: Bc. Michal Buchovecký
Vedoucí práce: Mgr. Martin Jureček
Název práce: Semi-supervised learning pro detekci malware
Obor: Počítačová bezpečnost

Datum vytvoření: 14. 5. 2019

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.	
Komentář: Diplomová práce mala za cieľ využiť okrem označených i neoznačené vzorky, ktoré denne prichádzajú antivírusovým spoločnostiam a pomocou nich zvýšiť presnosť detekcie malware. Všetky body popísané v pokynoch pre vypracovanie považujem za splnené. Menšiu výhradu mám ku kapitole 3, ktorá by mala byť ťažiskom práce. Kapitola obsahuje zjednodušené popisy metód, v ktorých sa vyskytuje pár chýb (napr. v Alg. 3 nie je uvedená iterácia) a nie je zjednotená notácia (napr. $k=c$, $X=D$).	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	80 (B)
Popis kritéria: Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.	
Komentář: Rozsah predloženej DP je v súlade s požadovaným rozsahom podľa príslušnej fakultnej smernice. Práca študenta je dobre čitateľná, avšak v niekoľkých častiach textu sa študent miesto vysvetlenia danej metódy (napr. kalibrácia modelu alebo hashing trik) len odkázal na literatúru. Zoznam použitej literatúry práce obsahuje veľké množstvo relevantných odkazov, celkom 55. Práca ďalej obsahuje nasledujúce okrajové chyby: - text obsahuje pár desiatok chýb, ide hlavne o diakritiku a interpunkciu, ktoré ale pri čítaní nepôsobia rušivo - číslovanie tabuliek nejde postupne a u niektorých obrázkov chýbajú popisky os - výskyt neformálneho vyjadrovania ako napr. "o čosi rýchlejší" alebo "akási istota predikcie".	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
3. Nepísemná část, přílohy	90 (A)
Popis kritéria: Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů	
Komentář: Implementácia algoritmov bola vykonaná v jazyku Python a bola využitá známa knižnica scikit-learn obsahujúca potrebné machine learning algoritmy. Všetky použité technológie sú adekvátne.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

4. Hodnocení výsledků, jejich využitelnost

87 (B)

Popis kritéria:

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Komentář:

Študentove výsledky nepreukázali, že by využitie vybraných semi-supervised learning algoritmov zvýšilo presnosť detekcie malware. To môže byť spôsobené tým, že bolo použitých pomerne málo príznakov (40) alebo bolo uvažovaných len pár klasifikátorov. Na druhej strane študentove výsledky by mohli pomôcť každému, kto by sa tejto problematike chcel venovať. Všetky dosiahnuté experimentálne výsledky je možné overiť.

Hodnotící kritérium:

Způsob hodnocení – následující škálou 1 až 5:

5. Aktivita a samostatnost studenta

5a:

1=výborná aktivita,

2=velmi dobrá aktivita,

3=průměrná aktivita,

4=slabší, ale ještě dostatečná aktivita,

5=nedostatečná aktivita

5b:

1=výborná samostatnost,

2=velmi dobrá samostatnost,

3=průměrná samostatnost,

4=slabší, ale ještě dostatečná samostatnost,

5=nedostatečná samostatnost

Popis kritéria:

V souvislosti s průběhem a výsledkem práce posudte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (5a). Posudte schopnost studenta samostatně tvůrčí práce (5b).

Komentář:

So študentom sme absolvovali niekoľko konzultácií a udržovali emailovú komunikáciu. Študent pracoval na diplomovej práci samostatne.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

6. Celkové hodnocení

87 (B)

Popis kritéria:

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

Text hodnocení:

Celkové hodnotenie študenta je na hranici medzi A a B, bližšie k B. Ak by študent mal ambície na lepšie hodnotenie, tak na obhajobe by mal preukázať, že rozumie danej problematike (napr. v prezentácii by mal lepšie vysvetliť vzťah (3.1) na str. 29, ďalej kalibráciu modelu a čo znamená podmienená nezávislosť spomenutá u Co-training a či bola splnená v experimentoch).

Podpis vedoucího práce: