FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	Ing. Matouš Kozák
Student:	Bc. Vojtěch Skalák
Název práce:	Klasifikace malwaru na základě samoorganizačních map
Obor / specializace:	Počítačová bezpečnost
Vytvořeno dne:	31. května 2024

Hodnotící kritéria

1. Splnění zadání

[1] zadání splněno
▸ **[2] zadání splněno s menšími výhradami**
[3] zadání splněno s většími výhradami
[4] zadání nesplněno

Všechny body zadání se vyskytují v textu. Nicméně bod 1. (rešerše použití self-organizing maps (SOM)) a bod 4. (diskuze nad výsledky a porovnání s jinými metodami) mohli být z mého pohledu podrobněji zpracovány (viz. komentáře níže).

2. Písemná část práce 55 /100 (E)

Úvod práce má neobvyklou strukturu, kde místo úvodu čtenáře do tématu a nastínění struktury práce, obsahuje podrobný popis detekčních technik a popis PE file formátu. Kapitola popisující SOM je velmi pěkná a je škoda, že zbytek práce se nenese v podobném duchu (jak obsahově tak formátově). Kapitola "Literature review" se pouze okrajově věnuje využití SOM např. v oblasti anomaly detection nebo malware clustering. Podle zadání bych očekával, že toto bude jedním z hlavních tématu práce. Popis experimentů je napsán dobře. Kapitola "Results" se především věnuje laděním hyperparametrů SOM, ale téměř vůbec se nevěnuje vyhodnocení na testovacích datech a porovnání s jinými modely. Dále popis KNN a MLP nepatří do prezentace výsledků a mohl by být zmíněny v některé z předešlých kapitol. Závěr práce (kapitola "Discussion") by se pravděpodobně neměla nazývat diskuzí, když zde je minimum textu, který by se za diskuzi mohl považovat. Očekával bych zde např. úvahu nad tím proč výsledky SOM byly tak nízké v porovnání s jinými modely, nebo proč KNN a MLP dosáhly horších výsledků na datasetu "malware families" než na "malware and benign files". Naopak si myslím, že by kapitola nemela obsahovat analýzu výsledných clusterů.

Z pohledu jazyka mi práce přišla napsána dobrou angličtinou s občasnými chybami, které ale nebránily pochopení textu. Co ale z mého pohledu kazí dojem z práce je občasné

neexistují napojení odstavců, kde předešlý odstavec měl minimální souvislost s nadcházejícím. Z hlediska typografie a vzhledu je práce nedodělaná, práce obsahuje velké množství tabulek, které "přetékají" text. Dále odstavce, které mají různé odsazení, které nemá logické opodstatnění. Interpunkce na začátku řádku, apod.

Zdroje mi přijdou relevantní a jsou většinou dobře citovány (občas se citace nachází před tečkou občas za). Seznam literatury obsahuje povětšinou všechny potřebné údaje pro identifikaci zdroje, výjimku tvoří reference 16, 20, 24, 27, které neobsahují unikátní identifikátor.

3. Nepísemná část, přílohy 90 /100 (A)

Zdrojové kódy jsou v jazyce C++ a Python a jsou přiloženy k práci. Velmi oceňuji vlastnoruční implementaci SOM v jazyce C++, kde je vidět, že na ní bylo stráveno hodně času.

4. Hodnocení výsledků, jejich využitelnost 55 /100 (E)

Bohužel naměřené výsledky (přesnost) nejsou vysoké a bez chybějící dalších experimentů není dost možné určit proč tomu tak je. Jedním z důvodu by mohlo být, že jednotlivé hyperparametry SOM byly nastaveny zvlášť (s výjimkou radius a neighborhood functions). To může mít za důsledek, že i když se individuální hodnota hyperparametrů tváří jako optimální, tak to nemusí platit o kombinaci s dalšími hyperparametry. Pro budoucí práce bych doporučoval kombinace hyperparametrů testovat společně (tzv. grid search). Dále porovnání s jinými metodami je velmi stručné (jak teoreticky tak experimentálně) a spolu se špatnými výsledky si nemyslím, že by práce byla využitelná v praxi.

V kapitole "Discussion" je zmíněno, že natrénovaná SOM byla otestována na EMBER test datasetu, ale v kapitole "Methodology" je zmíněno, že EMBER dataset byl rozdělen do 3 podskupin (trénovací/validační/testovací), tak aby byly jednotlivé rodiny malware rovnoměrně zastoupeny. Není mi tedy jasné, zda byla SOM otestována na EMBER test datasetu (fixní dataset vybrán autory EMBERu) nebo na nějakém jinem subsetu EMBERu. Toto rozlišení může být pro čtenáře velmi důležité, protože EMBER dataset slouží k porovnání různých detekčních modelů a pro zaručení férovosti se většinou používá jednotná testovací sada. Samozřejmě je možné si zvolit i vlastní testovací sadu, ale poté bych doporučil nepoužívat název "EMBER test dataset", protože by to mohlo vést ke zmatení čtenáře.

Celkové hodnocení 60 /100 (D)

Práci hodnotím známkou D, kvůli nedostatkům v textové a experimentální části. Autor velmi kvalitně zpracoval popis a implementaci SOM, kde ukázal, že je schopen kvalitní práce. Nicméně zbytek práce se mi zdá nedodělaný/uspěchaný což považuji za škodu. Za nedostatky považuji: stručnou rešerši využití SOM, uspěchaná experimentální část, nedostatečná diskuze nad výsledky a špatná typografie/vzhled práce.

Otázky k obhajobě

1. Byla natrénovaná SOM (a ostatní modely) otestovány na EMBER test datasetu nebo na nějaké jiné části datasetu EMBER? Viz. komentář k bodu (4):
2. Jak si vysvětlujete, že KNN a MLP dosáhly horších výsledků na datasetu "malware families" než na "malware and benign files"?
3. Jak si vysvětlujete horší výsledky SOM v porovnání s jinými modely?

Instrukce

Splnění zadání

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.

Písemná část práce

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. 52/2021, 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.

Nepísemná část, přílohy

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.

Hodnocení výsledků, jejich využitelnost

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.

Celkové hodnocení

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.