
Abstrakt

Detektory malvéru založené na strojovom učení môžu dosahovať vysokú predikčnú výkonnosť, samotná presnosť však nerobí ich rozhodnutia zrozumiteľnými ani obhájitelnými. Vysvetlenia by mali zodpovedať rozhodovaniu detektora, odhaľovať rozhodovaciu štruktúru a prepájať dôkazy so zmysluplnými konceptmi kybernetickej bezpečnosti. Táto práca postupuje od post-hoc analýzy čiernoskrinkových klasifikátorov k relačným detektorom, ktorých predikcie možno interpretovať prostredníctvom ich vlastného výpočtu.

Práca rozširuje post-hoc atribúciu o analýzu interakcií a viacpríznakové symbolické pravidlá, vyvíja Tailored Logic Explained Networks pre kompaktné a verné vysvetlenia a predstavuje PaTEx, ktorý používa regionálne náhradné stromy na odhalenie ciest, interakcií, alternatív a kontrafaktuálov. Následne skúma znalostné grafy odvodené z ontológií pre relačnú charakterizáciu malvéru. Hlavným príspevkom je QuinEx, vnútorne rozložiteľné vnorenie znalostného grafu typu Matrix Product State, ktorého skóre sa presne rozkladá na aditívne kanály. To umožňuje presnú atribúciu kanálov, úplné priradenie rozhodovacieho rozdielu správaniu a konceptové pravidlá zosúladené s modelom, pričom koherencia taktík MITRE ATT&CK nezávisle meria zosúladenie s doménou kybernetickej bezpečnosti.

Experimenty so statickými a dynamickými reprezentáciami ukazujú konkurencieschopnú výkonnosť spolu s vysvetľovacími schopnosťami, ktoré silným referenčným modelom chýbajú. Na primárnom grafe MAECO-Lite dosiahol QuinEx ROC-AUC $0,908 \pm 0,027$ a F1 $0,839 \pm 0,027$ v desiatich behoch. Jeho konceptové pravidlá dosiahli F1 $0,832 \pm 0,017$ a koherenciu taktík $0,813 \pm 0,112$, s vyššou nameranou koherenciou než vysvetlenia LIME, SHAP a rozhodovacieho stromu. Sekundárny experiment so statickou ontológiou EMBER potvrdil konkurencieschopnú výkonnosť.

Práca rozlišuje predikčnú výkonnosť, vernosť, granularitu, kompaktnosť a doménové zosúladenie ako odlišné, ale spoločne dôležité vlastnosti, a posúva vysvetľovanie malvéru k štruktúrovanej a výpočtovo podloženej podpore rozhodovania.