

Posudek oponenta diplomové práce

Název práce: Decision Forests in the Task of Semi-Supervised Learning

Autor diplomové práce: Bc. Jan Brabec

Posudek vypracoval: Ing. Vojtěch Franc, Ph.D.

Předložená práce studuje použití rozhodovacích stromů pro klasifikaci v případě, kdy jsou trénovací data navyvážená a pouze pozitivní příklady jsou jednoznačně označeny. Jádrem práce je experimentálně ověřit funkčnost existujících učících algoritmů a jejich modifikací na popsaném typu trénovacích dat. Metody jsou testovány na standardních veřejných sadách a na problému detekce malware v síťové komunikaci.

Práce je psaná dobrou angličtinou. Text je logicky uspořádaný do osmi kapitol. Obzvláště zdařilé jsou kapitoly 2 až 5, které velmi jasně popisují základní principy, testované algoritmy a jejich modifikace. Výsledky experimentů a jejich analýza mají potenciál být použitelné v dalším vývoji metod pro detekci malwaru.

K práci mám jen málo drobných výhrad:

- Metody se porovnávají na základě středního odhadu metrik (precision, recall, atd.), který se počítá z několika náhodně generovaných vzorků dat. Konfidence odhadů ale chybí. Není tedy jasné, jak jsou rozdíly v přesnosti jednotlivých metod statisticky významné. Přitom by k středním hodnotám stačilo uvést například varianci odhadu.
- V grafu 6.2 porovnávaným algoritmy pro “semi-supervised learning” (odstavec 6.1.1) není jasné, zdali se velikost trénovací množiny vztahuje k označeným (labeled) a nebo neoznačeným (unlabeled) příkladům. Popřípadě jaký byl jejich poměr. Lepší možnost by byla počet označených příkladů zafixovat a v grafu ukázat závislost přesnosti na rostoucím počtu neoznačených příkladů.
- Grafy 6.7, 6.8, 6.9 jsou poměrně těžko čitelné. Pomohlo by je zvětšit a použít snadněji rozlišitelné barvy.
- Rovnost $(1 - 1/n)^n = 1/e$ (vzorec (4.1)) platí pouze přibližně pro velká n .

Celkově se jedná o nadprůměrně dobrou práci. Doporučuji ji k obhajobě a hodnotím ji známkou **A-výborně**.

V Praze 27. ledna 2017

Ing. Vojtěch Franc, Ph.D.
Centrum strojového vnímání
Katedra kybernetiky, FEL ČVUT
Karlovo nám. 13, 121 35 Praha 2