



V Praze 19.2.2023

Posudek školitele k obhajobě disertační práce

Název: General Framework for Classification at the Top

Autor: Ing. Václav Mácha

Téma:

Předložená disertační práce se zabývá problémem binární klasifikace s různou cenou chyby klasifikace do výsledných tříd. Toto téma je často opomíjené v běžných aplikacích, je však velice důležité v citlivých oblastech jako je bezpečnost. I proto jsou metody testovány na příkladech z počítačové bezpečnosti.

Díky specifčnosti tématu je v literatuře popsáno mnoho metod, které jsou speciálními případy tohoto problému. Jsou však vytvořené pro různé aplikace v různém kontextu a proto bylo součástí práce i vytvoření ucelené formulace a jednotícího rámce v němž jsou tyto problémy studovány.

V práci se rozlišují lineární a nelineární modely především z důvodu jejich rozdílných možností jejich analýzy. Lineární modely mohou vést na konvexní problémy, které je možno studovat a srovnávat jejich vlastnosti teoreticky. Pro nelineární modely takové možnosti nemáme a jejich srovnání je tedy převážně experimentální. Nelineární modely však mají zásadní výhodu v praktických aplikacích.

Obecně je téma velice důležité a dosažené výsledky najdou uplatnění ve zmíněných kritických aplikacích.

Přínos:

Základním přínosem práce je jednotná formulace problému jako optimalizační úlohy s omezením a kategorizace stávajících metod jako jejich speciálních případů nebo aproximací. Důsledkem této systematizace je široká paleta metod, které vznikají kombinací různých známých variant formulace a aproximace. Zcela nová je formulace metody Pat&Mat a Pat&Mat-NP která je alternativou formulací TopMeanK a τ -FPL, avšak s lepšími teoretickými vlastnostmi.

Analýza vlastností vybraných formulací lineárních klasifikátorů z hlediska robustnosti a konvexity je také nová a přináší vysvětlení experimentálních výsledků známých i navržených metod.

Významnou částí práce jsou metody využívající duální formulaci problému. Ačkoliv duální formulaci využívají i známé metody, autorovým přínosem jsou významné pokroky týkající se inicializace a řešení problému modifikacemi metod typu gradient descent.

Z praktického hlediska je velmi významná metoda nelineární formulace typu DeepTopPush protože představuje pouze relativně jednoduchou modifikaci stávajících metod trénování. K aplikaci této metody v praxi stačí uchovávat hodnotu prahu jako součást trénovacího minibatche. Tato úprava minimalizuje vychýlenost odhadu klasifikačního prahu od skutečné hodnoty. Díky její jednoduchosti má tato metoda potenciál pro širší uplatnění v praxi.

Autorův podíl

Autor na tématu pracoval velmi samostatně. Analytické výsledky vznikly ve spolupráci se školitelem specialistou, na jejich základě odvodil autor práce své vlastní výsledky především v oblasti duálních formulací. Veškeré experimentální srovnání studovaných metod je samostatná práce autora.

Celkové hodnocení studia

Václav Mácha prokázal své mimořádné kvality již za magisterského studia, kdy publikoval výsledky svých absolventských prací. V doktorském studiu pokračoval v samostatné a pilné práci. Je autorem dvou článků přijatých do časopisů s impakt faktorem. Další dva články jsou připraveny ve formě preprintu a jsou dokončovány k odeslání do časopisů. Významným přínosem je též publikování metod pro zpracování dat jako otevřeného kódu na platformě github. Díky vysoké kvalitě kódu a aktivní podpory ze strany autora jsou tyto balíčky aktivně používány výzkumnou komunitou.

V publikační aktivitě je nižší podíl konferenčních příspěvků než je obvyklé. Toto bylo negativně ovlivněno pandemií koronaviru, při níž klesl počet pořádaných konferencí.

Závěr

V celkovém hodnocení považuji vědecký přínos autora za originální a dostatečně prokázaný významnými publikacemi. Předložená práce splňuje nároky na disertační práci a jednoznačně ji doporučuji k obhajobě.

doc. Ing. Václav Šmídl, Ph.D.
Ústav teorie informace a automatizace, AV ČR