

Školitelův posudek diplomové práce  
„Stochastické metody propojující neuronové sítě s náhodnými lesy a jejich využití v analýze medicínských dat“

studenta  
**Bc. Martina Oharka**

Student se v rámci předložené diplomové práce zabýval klasifikačními algoritmy, přesněji propojením neuronových sítí a rozhodovacích lesů, na kterých pracoval již během svého předchozího studia. Jak je v práci podrobně vysvětleno, obě metody mají své výhody i nedostatky, a tak je teoreticky možné vhodným propojením algoritmů získat kvalitnější klasifikátor. Přestože je takovýto postup přirozený, v odborné literatuře není tato kombinace dostatečně prozkoumána. Výchozím bodem práce je článek ilustrující možnost prostého převedení libovolného rozhodovacího stromu na neuronovou síť, což ale samo o sobě nepřináší významné benefity. Student samostatně navrhl velké množství úprav takto vzniklé sítě, kterými byly odstraněny klasifikační omezení náhodných lesů a zároveň zachoval jednoduchou kostru algoritmu, čímž udržel relativně malé množství hran i vrcholů neuronové sítě. Výsledné algoritmy nakonec splnily počáteční pozitivní očekávání, jak je doloženo na velkém množství testů na veřejných data setech, speciálně připravených studií i na lékařských datech, jejichž komplexnost nás k této práci inspirovala.

Práce psaná v anglickém jazyce obsahuje 108 stran textu a ilustrací rozdělených do osmi kapitol a dvou dodatků. Po formální i jazykové stránce je práce v pořádku, čemuž pomohla i jazyková korektura. Za drobný nedostatek považuji chyby v diakritice matematických výrazů, které přečkaly všechny kola revizí.

První polovina (kapitoly 1 - 4) je zaměřena na popis použitých klasifikačních metod. Vyzvedl bych snahu studenta srozumitelně a do detailu vysvětlit používané postupy, což je v kontrastu s většinou dnešních publikací v této oblasti, kde se určité části algoritmu schovávají za „známe balíčky kódu“. V této práci je diskutován efekt všech částí algoritmu, včetně matematického popisu.

Druhá polovina je už zacílena na konstrukci vlastního algoritmu a jeho testování. Hybridní model na většině dat vykazuje z pohledu různých klasifikačních metrik lepší výsledky než obě výchozí metody, v hraničních případech „remizuje“. Pozitivní je i dobrá stabilita parametrů, jejichž efekty jsou taktéž dobře popsány.

„K dokonalosti“ práci chybí širší zakotvení výchozích algoritmů do palety použitelných metod, nicméně i současný rozsah řešerše je dostatečný. Tento nedostatek je více než kompenzován velkým potenciálem vyvinutého algoritmu, který bychom rádi využili nejen na zmíněnou analýzu medicínských záznamů, ale i na klasifikaci egyptologických dat, se kterými se student setkal v předchozích projektech.

S ohledem na veškeré výše zmíněné skutečnosti hodnotím tuto diplomovou práci stupněm **A** (výborně).

V Havlíčkově Brodě dne 11. 7. 2020

Ing. Marek Bukáček