

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Aktivní učení pro klasifikaci textů
Jméno autora:	Marko Sahan
Typ práce:	diplomová
Fakulta/ústav:	Fakulta jaderná a fyzikálně inženýrská (FJFI)
Katedra/ústav:	
Oponent práce:	Ing. Jan Drchal, Ph.D.
Pracoviště oponenta práce:	FEL, Katedra počítačů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Výzkum aktivního učení vyžaduje dobré pochopení pravděpodobnostních principů strojového učení.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Splněno bez výhrad.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Zvolený postup považuji za v zásadě správný. Všechny kroky i hodnotící metodika jsou vesměs dobře motivovány a vysvětleny. Výtky a dotazy mám k následujícímu: <ol style="list-style-type: none"> Není mi jasné, proč byl v rámci aktivního učení testovací dataset zmenšován (sekce 6.2.1) – předpokládal bych, že bude zvolen pevný dataset a ten bude používán pro všechny experimenty, včetně pasivního učení. Pokud to dobře chápu, zdrojových dat, ze kterých byly vzorkovány finální datasety bylo dost. Výsledky pasivního učení jsou prezentovány (kapitola 5) pouze na datech „Sports and Comedy“. Představoval bych si tabulku výsledků, pro všechny datové množiny, která by umožnila porovnání s výsledky aktivního učení. V závěrečné kapitole postrádám lepší vzájemné porovnání jednotlivých metod – pouze grafy s rozdílným škálováním osy y jsou nedostatečné. Nejlepší by bylo přidat shrnující tabulku. Chybí pořádná diskuse srovnávající výsledky aktivního a pasivního přístupu (viz bod 3). Demonstrace metod na two-moon problému považuji za velmi přínosné. Není však například hladkost zmiňovaná v sekci 6.1.3 způsobena spíše malým počtem neuronů/vrstev prezentované neuronové architektury? 	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Z textu je zřejmé, že se student v problematice dobře orientuje. Práce má vysokou odbornou úroveň. Moc se mi líbí formalismus zavedený v úvodních kapitolách – pěkně je sjednocen přístup k různým typům modelů strojového učení.	

Formální a jazyková úroveň, rozsah práce	C - dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Text je psán kvalitní, čtivou angličtinou s minimem překlepů. Struktura práce je vhodně zvolena. Menší výhrady mám k: <ol style="list-style-type: none"> Popis SGLD by mohl být detailnější. Na obrázku 6.2 a podobných je použit příliš malý font. Nesrovnané osy y (AUC) u obrázku 6.8 a podobných. Chybí popis křivek spolehlivosti. 	

Výběr zdrojů, korektnost citací	A - výborně
--	--------------------

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Zdroje jsou citovány správně.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Výsledky práce mají po úpravách (viz výše) publikační potenciál.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Objem odvedené práce je velký. Hodnocení pro mě snižuje neúplně porovnání jednotlivých metod a příliš stručná diskuze rozdílů výsledků pasivního a aktivního učení (viz výše). Celkově jsem ale s předloženou závěrečnou prací spokojen a hodnotím ji klasifikačním stupněm **C - dobře**.

Datum: 17.7.2020

Podpis: