

# Posudek oponenta diplomové práce

Jméno a příjmení autora posudku: RNDr. Daniel Průša, Ph.D.

Jméno a příjmení autora práce: Bc. Jan Hajič

Název práce: Matching Images to Texts

---

Vlastní text (sem prosím napište text posudku, délka textu posudku není omezena):

Cílem předložené diplomové práce bylo navrhnout a implementovat systém pro vyhledávání obrazových dat na základě plných textů dokumentů. Autor své řešení založil na neuronových sítích a hlubokém učení. Navrhnul společný pravděpodobnostní model textu a obrázků, implementoval jej a natrénoval pro data z novinových článků. Úspěšného vyhledávacího systému se mu dosáhnout nepodařilo, zejména z důvodu selhání trénování neuronových sítí na řídkých textových datech.

Studovaná úloha je obtížná a vývoj vyhledávacího systému, který by fungoval pro libovolná data, je ambiciózní projekt. Hodnocení proto nezaměřuji na celkovou funkčnost, ale na potenciál dílčích kroků, efektivitu jejich provedení a prezentaci.

Použití neuronových sítí pro daný problém lze považovat za vhodnou volbu opodstatněnou jejich úspěšností a rozmachem v posledních letech. Přehled v úvodní části práce spolu s množstvím citovaných článků svědčí o tom, že diplomant dobře zmapoval stav vědění. S úspěchem využil i metody z jiných oblastí, například techniky pro zpracování přirozeného jazyka k odstranění řídkosti textových dat.

Pro reprezentaci obrazových dat zvolil autor stav poslední vrstvy existující konvoluční sítě pro kategorizaci obrázků. Zde je otázkou, nakolik je tato reprezentace vhodná. Sít' je natrénovaná pro určitou skupinu kategorií, která se jen částečně překrývá s klíčovými slovy z novinových článků. Zároveň kategorie nezastupují složitější vztahy mezi klíčovými slovy. V práci by mělo být lépe vysvětleno rozhodnutí pro tento krok a diskutovány i jiné možnosti.

Větší pozornost by zasloužila i analýza samotného zadaného problému. Je otázkou, pro jaké typy textů může natrénovaná síť z principiálního hlediska dobře fungovat. Možnosti generalizace systému pro různé texty budou patrně limitované. O vhodnosti či nevhodnosti přiřazení obrázku k textu mohou rozhodovat drobné niance. Trénovací data stěží pokryjí variabilitu všech možných vstupů. Pro testování systému může být prospěšné zúžit zaměření textů na určitou oblast (např. technické popisy).

Kromě výše uvedeného nemám ale žádných dalších výhrad. Jako hlavní přínosy práce vyzvihuji zejména následující:

- Návrh modelu.
- Pořízení databáze textů a obrázků, návrh metodologie pro anotování.
- Rozsáhlou implementaci knihovny pro podporu experimentů.
- Podrobnou analýzu možných příčin neúspěchu při trénování neuronové sítě.

Toho vše přispívá k posunu při řešení náročného úkolu. Diplomant získal cenné zkušenosti – identifikoval a analyzoval problematické části. Toho jistě zúročí, pokud bude ve výzkumu pokračovat. Nápomocná mu bude vytvořená infrastruktura (anotovaná data, knihovna). Knihovna SAFIRE zároveň může dobře posloužit i pro vývoj jiných aplikací založených na neuronových sítích.

O samotném textu práce mohu konstatovat, že je dobře strukturovaný, obsahově vyvážený, srozumitelný, bez zjevných chyb. Nechybí ani řada ilustrativních schémat. Připomínku mám pouze k typografické podobě matematických funkcí ( $\log$ ,  $\max$ ,  $\tanh$ ,  $\text{softmax}$ ) – nebylo využito příkazů LaTeXu ( $\log$ ,  $\max$ ) a jsou proto zobrazeny kurzívou, stejně jako proměnné, což působí nezvykle.

## Otázka k obhajobě:

S přihlédnutím k získaným zkušenostem a experimentálním poznatkům, vidíte nyní způsob, jak postupovat a jaké změny v modelu a použitém aparátu učinit, aby byla větší jistota, že ve stejném čase (odpovídajícímu předchozímu vývoji) dosáhnete systému funkčního alespoň pro jednoduché typy vstupních dat (i za cenu, že systém nepůjde zobecnit pro složitější data)? Popište příslušné možnosti.

## Nalezené drobné chyby/překlepy v textu:

- str. 11, řádek 10: L+1-th  $\rightarrow$  L+1-st
- str. 16, Figure 2.1:  $x^{\hat{}}$   $\rightarrow$  x
- str. 21, řádek 4: opposit  $\rightarrow$  opposite
- str. 22, řádek -7: 2x written
- str. 29, řádek 16: chybná reference [?]
- str. 36, řádek -7: +inf  $\rightarrow$   $+\infty$
- str. 36, řádek -4: This has works  $\rightarrow$  This works
- str. 39, Figure 5.2: osamocené "The" na konci popisu
- str. 45, řádek 2-3: smlaler  $\rightarrow$  smaller
- str. 45, řádek 10: in Theano. 8.2  $\rightarrow$  in Theano, see section 8.2.
- str. 45, poznámka pod čarou: chybná reference [?]
- str. 49: chybí Figure 7.5, je uvedený pouze popis

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	ANO <input type="checkbox"/>
---	------------------------------

Seznam soutěží studentských prací, viz <http://www.mff.cuni.cz/studium/bcmgr/prace/>

Pokud jste výše zaškrtnli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

--

V Praze dne: 1.9.2014

Podpis:\*\*

\* *nehodící se škrtněte (vymažte)*

\*\* *do SISu vkládejte formulář nepodepsaný (ve formátu PDF), podpis je potřeba doplnit až na vytištěný posudek.*