

# Tomáš Musil: Exploring natural language principles with respect to algorithms of deep neural networks

## Oponentský posudek

Hlavním cílem předložené práce je zjistit, jak neuronové sítě, natrénované pro některou úlohu zpracování přirozeného jazyka, zachycují jazykový význam. Práce přesahuje do oblasti filozofie jazyka, ve které je třeba najít vhodnou definici toho, co je to vlastně význam.

Celkové hodnocení:

- Práce je dobře motivovaná a rozumně strukturovaná, u návrhu disertace bych však očekával více prostoru věnovaného plánům do budoucna (v předkládaném textu se tomu věnuje vlastně jen druhá polovina desáté strany).
- Práce je psána srozumitelně a dobrou angličtinou.

Komentáře a otázky k obhajobě:

- Strana 4: „*Because we are interested in representations of meaning, we will restrict this section to embeddings of words.*“ ... Toto je zvláštní formulace. Nechce snad autor říct, že jednotky větší nebo naopak menší než slovo nejsou součástí jazykového významu? Rozumím tomu, že experimenty s reprezentací celých vět by byly výpočetně příliš náročné, jak ostatně autor v následujícím textu upozorňuje, ale příčina uvedená v této větě je podle mne problematická.
- Strana 6: „... *into derivational trees. For each such pair ...*“ ... Jaký pár? Předcházející text o žádných párech nemluví.
- Strana 6: Semantic and syntactic category ... Následuje příklad množiny českých slov a autor o nich uvádí, že jde o slovesa pohybu v minulém čase a v mužském rodě. To ale zjevně platí pouze o některých slovech z uvedené množiny. V jednom případě jde o ženský rod, to lze přičíst nedokonalosti natrénovaného neuronového modelu. Proč se ale infinitivy považují za syntakticky blízké minulému času?
- Autor výslovně nezmiňuje, že by chtěl zkoumat význam pouze v jednom konkrétním jazyce, nicméně všechny dosud provedené experimenty jsou zaměřené na češtinu. Podle mého názoru by u takto obecně zaměřeného tématu bylo velmi vhodné ověřit navržené postupy alespoň na jednom dalším jazyku, pokud možno typologicky odlišném od češtiny. Je něco takového v plánu?
- V oddíle 7.2 se od pojmu „umělá inteligence“ odvíjí diskuse o tom, kde je hranice mezi „přirozeným“ a „umělým“, vůbec se však neřeší, co je a není „inteligence“. Neuronové sítě sice možná jsou „přirozené“ v tom smyslu, že řeší danou úlohu a neumíme vysvětlit, jak to dělají, přesto lze těžko hovořit o inteligenci. Neuronový model je postavený na velkém

množství jazykových dat, případně i dalších dat, např. obrazových, přesto se ale nenaučí totéž co člověk. Když člověku někdo řekne, že „kamna pálí“, a dotyčný člověk si na ta kamna sáhne, tak ihned pozná význam slova „pálí“. Většina současných neuronových modelů k této zkušenosti přístup nemá. Stejně tak necítí spoustu dalších věcí, třeba emoce.

- Nejdůležitější otázka: Co tedy mají být další kroky a co by mělo být nakonec obsahem disertační práce? Téma nastíněné v aktuálním textu je velmi obecné a ambiciózní, některé formulace naznačují cíle, které podle mne nejsou v rámci času vymezeného na doktorské studium reálné („*we will find a theory of language that ...*“). Na druhou stranu, nadpis práce („*Exploring natural language principles...*“) zní mnohem přiměřeněji, je ale příliš obecný, než aby bylo možné si udělat představu, jakého cíle má být dosaženo. Bude se celá práce točit pouze kolem vektorové reprezentace a významu jednotlivých slov, tak jako dosavadní experimenty? Nebo půjde i o větší jednotky, ustálená slovní spojení, klauze, věty, diskurz, metaforu, postoj mluvčího, ironii, modalitu, dosah kvantifikátorů? Jak by zhruba měly vypadat další experimenty, které k vytčeným cílům povedou?

V Jenštejně 27. února 2020

RNDr. Daniel Zeman, Ph.D.