Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Mostafa Abdou
Název práce	An investigation of the task of Universal Semantic Tagging using Neural Architectures, Multi-task Learning, and Multi-lingual Learning
Rok odevzdání	2017
Studijní program	Informatika **Studijní obor** Matematická lingvistika

Autor posudku	Jind¿ich Libovick **Role** oponent
Pracoviště	stav form ln a aplikovan lingvistiky

Text posudku:

Mostafa's master thesis presents a solid study of multi-task learning in the context of semantic tagging which is currently a research topic of a high interest of the NLP community. The thesis presents several successful applications of multi-task learning with semantic tagging as an auxiliary task for less complex NLP tasks and cross-lingual model augmentation.

The thesis clearly shows that its author is familiar with the newest development of deep learning and its applications for NLP. He mastered the experimental methodology and is able to tackle research problems in a very systematic way. All the experiments were described in great detail. The experiment results are convincing and well explained.

The contribution of the excellent experimental results is stressed by the fact that a paper summarizing some of the results was accepted for publication at EMNLP, one of the most prestigious conferences in the field NLP.

On the other hand, the thesis could have been better organized. The theoretical introduction suffers from many inaccurate and misleading formulations. Most of the following critical remarks address this point.

Section 1.1 mentions the distributional hypothesis and mentions several techniques that explicitly rely on this language property. The section makes an impression the mentioned techniques, some of them not related to the topic of the thesis, are the only ones that exploit distributional hypothesis. The truth is that we do not rely on distributional hypothesis only while working with distributional semantics. All ML approaches to NLP and deep learning, in particular, utilize almost exclusively word cooccurrence and thus exclusively rely on the distributional hypothesis in the same way as vector-space word models or IR models.

Section 1.2.5 mentions the vanishing gradient problem in RNNs. The main reason for the vanishing gradient problem is not that activation functions are limited, but that they almost

constant in most of their domains and thus their derivatives are close to zero. It makes virtually impossible to consider the long-distance relationship between input and output because the gets repetitively multiplied by numbers close to zero during the back-propagation. I also miss an explanation why LSTMs and GRUs do not suffer from this problem.

Explanation of residual connections and highway networks (Section 2.1.4) would fit better in the introductory chapter, such that all the techniques used in the thesis are explained at one place not mixed with the experimental setup (literature overview of residual networks is placed immediately after technical details on word-embedding pre-training).

The conclusions state that approach taken in the thesis "presents a promising range of possibilities for moving away from narrow artificial intelligence towards a more general and human-like artificial intelligence." This is, in my opinion, a vastly exaggerating claim. Compared to current NLP models, human language acquisition happens with much fewer data and for most of the data without any supervision. Most of the experiments presented in the thesis relied on demanding expert annotation of a large body of text. Moreover, results from Chapter 1 shows that it is much easier to take advantage of semantic annotation in solving a syntax and morphology problems and not the other way round as you would expect while simulating human-like language acquisition.

Question for the author:

1. From a theoretical perspective, this is an interesting result because it suggests neural networks learns conceptualization that is similar to levels of linguistic analysis. From the engineering point of the results are however less encouraging. Semantic tags are quite hard to get, require complicated manual annotation, annotation rules become unclear in case of noisy user-generated data. This makes semantic tagging and other tasks requiring semantic annotation a rather inconvenient auxiliary task for end-to-end tasks in less resourced languages. Can you think of another task that can be as useful as semantic tagging, but the training data are easier to obtain?

2. Both German and Dutch belong to the same language family English. The successful knowledge transfer in the cross-lingual experiments might be also due to this relatedness. The Parallel Meaning Bank also contains Italian? Why did you exclude it from your experiments?

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 30. 8. 2018

Podpis: