Master thesis proposal

Question answering in context

Natural language understanding and question answering are important topics within natural language processing (NLP) that have received attention in the last years. The generality of the problems (many other natural language processing problems can be expressed as special cases of these problems) are part of the reason for this interest, along with its difficulty. The tasks require a system to capture just the right amount of meaning from the provided texts (data of other modalities are sometimes incorporated), along with non-trivial reasoning. In some cases the agents also need to keep track of a dialog with state and progression. Deep learning, and specifically recurrent neural networks (RNN) have been crucial components to many state-of-the-art solutions (see Reading below).

**Question Answering in Context** is a dataset for modeling, understanding, and participating in information seeking dialog. Data instances consist of an interactive dialog between two crowd workers: (1) a student who poses a sequence of freeform questions to learn as much as possible about a hidden Wikipedia text, and (2) a teacher who answers the questions by providing short excerpts (spans) from the text. QuAC introduces challenges not found in existing machine comprehension datasets: its questions are often more open-ended, unanswerable, or only meaningful within the dialog context. (See quac.ai for more details. Description reproduced from web page).

In this master thesis, you will implement machine learning models for question answering in context. The work requires students that are skilled in machine learning. Familiarity with natural language processing is beneficial, and statistical inference is a corner stone in the techniques that will be employed. You will start with some simpler, existing models that were developed for similar tasks, and eventually try out some more advanced solutions. The work is suitable for two students. Working without a partner is also a possibility for a motivated student. The work may include handling large datasets and perform the neccessary processing on the data, prior to model development.

The work will be performed in collaboration with RISE AI at Research institute of Sweden (see below).
Related courses:

- Machine learning (TDA231)
- Deep machine learning (SSY340)

Required skills:

- Experience of implementing machine learning models
- Experience of natural language processing (NLP)
- Experience of working with large data
- Programming skills (and preferably with some experience of relevant frameworks such as Tensorflow, Pytorch, or Keras. You are free to choose programming language, but a substantial part of related work is implemented in Python, so working with other languages requires some good reasons and perhaps extra work on your part.

Keywords: Machine learning, natural language processing, dialog systems.

Potential supervisors: Olof Mogren, RISE AI, olof.mogren@ri.se

RISE AI is part of Research institutes of Sweden (and formerly SICS). We are an organization of around 60 researchers working on AI-related tasks within different fields including natural language processing, computer vision, and network analysis. Our Gothenburg office was started in May 2018 and comprises a (so far) small team of researchers focusing on machine learning.

For more info, visit http://ri.se or email olof.mogren@ri.se.

Reading and resources:

1. quac.ai (http://quac.ai/)
3. SQuAD (https://rajpurkar.github.io/SQuAD-explorer/)