Title: Hybrid Deep Question Answering

Author: Ahmad Aghaebrahimian

Institute: Institute of Formal and Applied Linguistics

Supervisor: RNDr. Martin Holub, Ph.D., Institute of Formal and Applied Linguistics

Abstract: As one of the oldest tasks of Natural Language Processing, Question Answering is one of the most exciting and challenging research areas with lots of scientific and commercial applications. Question Answering as a discipline in the conjunction of computer science, statistics, linguistics, and cognitive science is concerned with building systems that automatically retrieve answers to questions posed by humans in a natural language. This doctoral dissertation presents the author's research carried out in this discipline. It highlights his studies and research toward a hybrid Question Answering system consisting of two engines for Question Answering over structured and unstructured data. The structured engine comprises a state-of-the-art Question Answering system based on knowledge graphs. The unstructured engine consists of a state-of-the-art sentence-level Question Answering system and a word-level Question Answering system with results near to human performance. This work introduces a new Question Answering dataset for answering word- and sentence-level questions as well. Starting from a simple Logistic Regression model, the author presents more advanced models including multi-layer Neural Network, Convolution Neural Network and Recurrent Neural Network solutions for different Question Answering systems mentioned above. Although the core of all developed modules in this work are new Deep Neural Network architectures, several linguistic intuitions such as phrase structure theory and constituency relations are integrated into them to improve their performance.

Keywords: Question Answering, Word-level Question Answering, Sentence-level Question Answering, Neural Network, Knowledge Graph, Convolution Neural Network, Recurrent Neural Network, Constituency Relation