Chia-Hung, Huang ; Hsin-Wei, Yau ; Pei - Hsuan, Huang
Mentor: Prof. Su, Keh-Yih
August 27, 2017

SSQA Project Proposal

Social Studies Question Answering based on DNN

Abstraction

We apply NLP and Deep Learning technology to solve Chinese Social Studies Question Answering, comparing the accuracy between only using NLP and some domain knowledge. In order to make the machine learn Entailment Recognition between evidence and hypothesis. Using the third year to fifth grade primary school text and exam questions to be our corpus which have done the evidence search and tokenization, we apply not only Decomposable Attention Model, which is composed of three part of feed forward neural network, attend, compare and compare but also Doc2Vec model, which is an advanced application using word2vec on it. In some parameters tuning and punctuation adjusting, we achieved higher performance on Entailment Recognition than some feature extracting. In Decomposition Attention Model, we get 10% performance accuracy higher than the same subject project before.

Introduction

Since the beginning of 1956, with the progress of computer, artificial intelligence has become the trend of modern technology and the future trend. Nature Language Process(NLP) technology is an important development in a branch of artificial intelligence. Using NLP technology to answer questions such as Question Answering(QA) is the expected dream of much technical staff.

The recognition of textual entailment is one of QA's challenge. This topic wants to build a relation between hypothesis and evidence about Binary Class(BC), Multi Class(MC). To do that, there are two main steps needed to be executed: Evidence Search, Entailment Recognition; Evidence Search, based on the current question and search the most similar evidence from the corpus, can be applied to many domains, such as Medical(Asma et al. 2017). Entailment Recognition, determine the question's answer based on evidence search's result, can be used on the testing machine, such as AI-MATHS, Torobo-Kun. Our project is focused on the latter. For examples, consider the following sentences:

- 臺灣島南北總長度約四百公里，東西最寬約一百四十公里，是個南北狹長的島嶼。
- 臺灣島的南北總長度比東西寬度長，是個南北較狹長的島嶼。

We knew that it is fairly easy to conclude that the second sentence follows from the first one, by simple aligning 台灣島 with 台灣島 and 南北狹長 with 南北較狹長 and recognizing that they are synonyms.

Researches done in the past are mainly used machine learning(ML) technique and statistical-based model to solve this entailment recognition problem. Probabilistic Entailment Model(Glickman and Dagan, 2005) judges the answer by calculating hypothesis probability. Support Vector Machine(SVM) which is ML method applied on deciding which option to choose(Prodromos Malakasiotis, 2007).

However, Deep Learning is booming on each domain in recent year, and also have much better accuracy performance than before. Therefore, we reference two models try to improve the entailment.