Thesis advisors
Prof. Dietrich Klakow
Dr. Volha Petukhova
A/Prof Tim Baldwin

Author
Desmond Darma Putra

Abstract

Question answering is a Natural Language Processing field that focuses on providing answer to user's question, posed in natural language. In the project that we work, we aim to build a Question Answering Dialogue System that creates an interactive game between the user and the system. In the game scenario, the system will hold a famous person's biographical facts and the user has to guess the name of this famous person. To achieve this goal, the user may ask up to ten question. However, the system prevents the user from asking direct questions about the person's name or alias.

This system consists of four modules named Interpretation Module, Dialogue Manager, Answer Extraction Module, and Utterance Generation Module. The focus of this thesis is to build Answer Extraction module (AEM) so it can automatically extract these biographical facts from raw texts. The texts are mainly gathered from Wikipedia. In order for the system to understand the texts, 53 semantic relations are defined to extract most important facts from them. Texts are annotated with defined semantic relations adding IOB (Inside, Outside, Beginning) prefixes to capture expected answer's boundaries.

Our Answer Extraction Module is comprised of three components: sequence classifiers, pattern matching tool and post processing pipeline. For sequence classifiers, we employ two well-known machine-learning algorithms, which are SVMs and CRFs. Features such as word token, lemma, Part-of-Speech tag, Named Entity tag, chunk information, capitalization and keyword are used to train the classifiers. The learning model which is the outcome of this training is later used to predict the relations' label. For pattern matching, handcrafted regular expressions are applied with the same task as the classification one. The output from the classifiers and pattern matching are then validated through the post processing pipeline.

The performance of our AEM is investigated in two experiment sets. In the first experiment, we use five-fold cross validation and apply three systems i.e. the Baseline, System 1 and System 2. Our results show that System 2 outperforms the baseline and System 1. The reasons are threefold. First, developing the classifiers to predict each relation separately gives more robust performance than the one that predict all relations together. Second, for several relations, pattern matching works better than the classifiers. Third, post processing pipeline is proven to filter and remove the incorrect output. In the second experiment, different sizes of training set (20%, 40%, 60 and 80%) are used to see how the classifiers work. The results demonstrate that the more training data is used, the better the classifiers' performance.