

Semantic parsing for Vietnamese Question Answering System

January 17, 2014

Abstract

We apply a learning mechanism in to semantic parsing for Vietnamese. This kind of learning reduces the burden of providing supervision. The training data does not need to be carefully annotated. We only need some kind of "Question - Answer" data to train our model.

1 Introduction

Semantic parsing is a very important process in Natural Language Processing, helping computers to "understand" natural languages. Generally, semantic parsing converts a natural language sentence into a special representation so that computers can read and process it. We call that special representation as "Computer-understandable language". An effective way to do this task is using learning methods to build a model of relationship between natural language structure and "Computer-understandable language" (CUL) structure. Unfortunately, most of the current approaches require very large amounts of fully annotated data in order to obtain a good model. Annotated data means that for each input sentence of natural language, we need the corresponding result sentence of "Computer-understandable language". The later is almost prepared manually by human, which takes a lot of efforts.

Recently, James et al. implemented a new learning paradigm aimed at alleviating the supervision burden. The algorithm is able to predict complex structures which only rely on a binary feedback. Borrowing the idea from these authors, we developed "Vietnamese geometric question answering system" (VGQAS). The core of the system is semantic parser.

2 Computer-understandable language

The representation of meaning (the result of semantic parsing) depends on each system and need to be unified. In our system, the format of "Computer-understandable language" is composition of pre-defined functions. Each function has only one argument. For example, `capital(stateid("alaska"))` is a CUL sentence, being a result of semantic parsing for "Th ph ca alaska l g?" ("What is the capital of alaska?").

3 VGQAS overview

4 Semantic parsing model

As mentioned, each parser system has a model for translating a natural language sentence to CUL sentence. In VGQAS, model consists of three parts: Word-Function feature calculator, Function-Function feature calculator and learning parameters.

4.1 Word-Function feature

References

- [1] Clarke James, Goldwasser Dan, Chang Ming-Wei, and Roth Dan. Driving semantic parsing from the world's response. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, CoNLL '10, pages 18–27, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. ISBN 978-1-932432-83-1. URL <http://dl.acm.org/citation.cfm?id=1870568.1870571>.