

Analysis of semantic parsers on downstream applications

Vanda Balogh

II. year Computer Science, MSc

Supervisor: Richárd Farkas

SZTE TTIK Department of Computer Algorithms and Artificial Intelligence

Semantic representations are getting more and more attention in Natural Language Processing (NLP) lately, but only a little progress has been made to estimate how effective these representations can be on downstream tasks. Semantic parsing is the task of converting natural language into formal meaning representation. Semantic representations may differ in their target meaning component, including predicate-argument relations, discourse relations, semantic roles and logical structures.

In terms of natural language parsing, a downstream application means an NLP task or component on which the parser is applied, making this downstream component dependent on the parser, itself. In this context, a downstream application can be e.g. machine translation, information extraction or question answering. An intrinsic evaluation of the parser refers to the analysis of the results returned by the parser as an independent, stand-alone system, while an extrinsic evaluation examines the impact of the parser on an NLP application. This way, the comparison of different parsers is possible. The first shared task on the extrinsic evaluation of parsers is the Extrinsic Parser Evaluation (EPE 2017) task. A shared task is organized to solve specific, challenging problems by groups of researchers. A key element of shared tasks is the availability of data.

In this work, I present the notion of syntactic and semantic parsers in detail, then briefly introduce a semantic annotation called Universal Conceptual Cognitive Annotation, UCCA for short, and its sole parser called Transition-based UCCA Parser, TUPA as an acronym. The ultimate goal is to assess the competency of semantic parsers by competitively evaluating syntactic and semantic parsers using an extrinsic evaluation system.

Keywords: NLP, semantic parsing, UCCA, EPE

Szemantikus mondatelemzők vizsgálata célfeladatokon

Balogh Vanda

II évf. programtervező informatikus, MSc

Témavezető: Farkas Richárd

SZTE TTIK Számítógépes Algoritmusok és Mesterséges Intelligencia Tanszék

A szemantikus reprezentációk az utóbbi időben egyre jobban elterjedtek a természetes nyelvfeldolgozás területén. Ennek ellenére eddig kevés figyelmet kapott, hogy mennyire hatékonyak ezek a reprezentációk különféle célfeladatokon. A szemantikus elemzők természetes nyelvi szövegeket konvertálnak gép által kezelhető formális jelentés-reprezentációkká. A szemantikus jelöléseket megkülönböztethetjük jelentéssel bíró elemeik alapján. Ilyen lehet például a predikátum-argumentum reláció, diszkurzus reláció, szemantikai szerep vagy logikai struktúra.

A természetes nyelvfeldolgozásban többféle célfeladat létezik, amelyekben a mondatelemzők hasznosnak bizonyulhatnak. Ilyen célalkalmazás lehet például gépi fordítás, információ kinyerés vagy kérdés megválaszolás. Az elemzők külső kiértékelése azt vizsgálja, hogy hogyan befolyásolja egy adott elemző a célalkalmazást. Ez lehetővé teszi, hogy különféle elemzők eredményeit hasonlítsuk össze ugyanazon a célfeladaton. Az elemzők külső kiértékeléséről szóló első nemzetközi verseny az Extrinsic Parser Evaluation (EPE 2017). Egy nemzetközi verseny vagy angol nevén „shared task” célja, hogy sok kutatót egy közös feladat megoldására sarkalljon, melyhez az adatok rendelkezésre állnak.

Munkámban részletesen bemutatom napjaink szintaktikai és szemantikus elemzőit, majd szó lesz egy szemantikus reprezentációról, melynek megnevezése Universal Conceptual Cognitive Annotation vagy röviden UCCA, és az eddigi egyetlen ismert hozzá készült elemzőről, ami a Transition-based UCCA Parser nevet viseli, rövidítése TUPA. A dolgozat célja az, hogy felmérjük a szemantikus elemzők képességét azáltal, hogy szintaktikai és szemantikus elemzőket vetünk össze és értékelünk ki az EPE 2017 verseny adatbázisain.

Kulcsszavak: NLP, szemantikus elemző, UCCA, EPE