Sinica Semantic Parser for ESWC'14 Concept-Level Semantic Analysis Challenge

Shafqat Mumtaz Virk, Yann-Huei Lee, and Lun-Wei Ku

Institute of Information Science (IIS)
Academia Sinica Taiwan
{virk.shafqat,andycyrus}@gmail.com,lwku@iis.sinica.edu.tw

Abstract. We present a semantic parsing system to decompose a sentence into *semantic-expressions/concepts* for ESWC'14 semantic analysis challenge. The proposed system has a pipeline architecture, and is based on syntactic parsing and semantic role labeling of the candidate sentence. For the former task, we use Stanford English parser; and for the later task, we use an in-house developed semantic role labeling system. From the syntactically and semantically annotated sentence, the *concepts* are formulated using a set of hand-build concept-formulation patterns. We compare the proposed system's performance to SenticNet with the help of few examples.

Keywords: Syntactic Parsing, Semantic Parsing, Semantic Role Labeling, Concept Formulation Templates

1 Introduction

Natural languages are both complex and ambiguous. Unless machines are capable of handling these issues in an intelligent way, building smart natural language processing (NLP) applications is a tough and challenging task. Maybe, one way to ease this toughness is to try to make computers understand natural language text. For the same purpose, the trend in NLP is shifting from exploring 'What it is?' to 'What it means?' (i.e. from syntax to semantics).

During the last couple of decades, a number of sub-fields have emerged under the umbrella term *computational semantics* including but not limited to sentiment analysis, textual entailment, question answering, and semantic parsing. These are among rapidly growing areas of NLP, and the research community has recognized their worth in recent times. This can be realized by the fact that their are many workshops and/or special tracks/challenges in conferences dedicated to these tasks. The ESWC'14 challenge on semantic analysis is one among those special challenges, and is scheduled to be held together with the 11^{th} European Semantic Web Conference 2014 (ESWC'14). The challenge has a number of advanced tasks in addition to an elementary task on polarity detection. The advanced task#2 is titled "Semantic Parsing", and refers to the task of de-constructing natural language text into a number of *semantic-expressions/concepts*. Though the term *semantic-expression/concept* is very gen-

eral in itself, and is hard to define clearly, we take it to be a single-word/multi-word expression for which we have semantics.

In this paper, we propose a system for semantic parsing in the context of task#2 of the challenge. The system has a pipeline architecture and relies on syntactic and semantic analysis of a candidate sentence. We use Stanford English parser [3] for syntactic parsing, and an in-house built semantic role labeling system for semantic interpretations. To formulate the concepts into desired format, we propose a set of hand-build concept formulation templates.

2 Proposed System

The architecture of the proposed system is shown in Fig. 1. It has three major components: a syntactic parsing, a semantic role labeling, and a concept formulation component. The purpose and importance of each component is explained in the following paragraphs.

Syntactic Parsing: As a preliminary step, the input sentence is syntactically

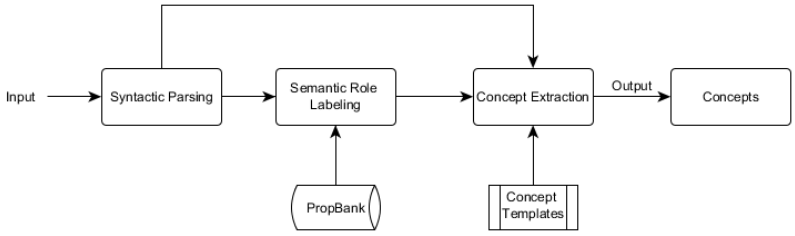

Fig. 1: System Architecture

analyzed to get a syntactic parse tree. This step is necessary for the major reason that almost all automatic semantic role labeling system rely on a preliminary syntactic parsing step [10].

Semantic Role Labeling: Semantic role labeling (SRL), also known as shallow semantic parsing, is the task of semantically annotating natural language text. Conventionally, a syntactically parsed sentence is taken as input, and semantic arguments associated with predicate of the sentence are identified and classified to a particular semantic class. The first automatic SRL systems was reported by Gildea and Jurafsky in 2002 [5], and since then, their ideas have been dominating the field. In their approach, they emphasized on selection of appropriate lexical and syntactical features for SRL, use of statistical classifiers and their combinations, and ways to handle the data sparseness issue. Researchers have tried to build on that by augmenting and/or altering the feature set [9], by experimenting with various classification approaches [11, 6], and by attempting different ways

to handle data sparseness [1]. For this challenge, we developed a SRL system, which is based largely on previously explored features and maximum entropy classifiers. The classifiers were trained using English Penn Treebank [8], and Propbank [7] data. However, we have proposed a number of additional features to enhance its performance. The details of the SRL system are beyond the scope of this paper, and are supposed to be covered in another planned article.

Concept Formulation: Once the sentence has been annotated syntactically and semantically, the concepts can be formulated using a set of hand-build concept templates. Table 1 lists few of the templates used in our experiments. Here,

#	Concept Template	#	Concept Template
1	ARG0_Pred	10	Pred_in_the_direction_ARGM-DIR
2	Pred_ARG1	11	Pred_because_ARGM-ARGM-CAU
3	Pred_ARG1_ARG2	12	Pred_when_ARGM-TMP
4	Pred_ARG1_ARG2_ARG3	13	Pred_ARGM-GOL
5	Pred_ARG1_ARG2_ARG3_ARG4	14	Pred_by_ARGM-EXT
6	Pred_ARG1_ARG2_ARG3_ARG4_ARG5	15	Pred_ARGM-MNR
7	Pred_with_ARGM-COM	16	Pred_ARGM-NEG
8	Pred_in_ARGM-LOC	17	ARGX's
9	Pred_in_order_to_ARGM-PRP	18	ARGM's

Table 1: Concept Templates

Pred and ARG1, ARG2, ARGM-LOC, ARGM-GOL, etc. refer to the predicate, and to the semantic role classes used in the prop-bank labeling scheme (see [2] for details on these classes).

3 An Example

To explain how our proposed system works at different levels, lets take an example sentence: *This film served as great entertainment for young people.*, and go through all the steps that the proposed system will perform to extract *concepts*. As a first step the sentence is syntactically parsed, which is semantically annotated by the SRL system as the second step. The resulting syntactically and semantically annotated tree is shown in Fig 2. From the semantically annotated tree, the extractable predicate-argument information is given in Table 2. Using this information and the templates given in Table 1, the following concepts can be formulated:

```
(1) This_film_serve (2) serve_as_great_entertainment
(3) serve_for_young_people (4) great_entertainment
(5) This_film (6) young_people
```

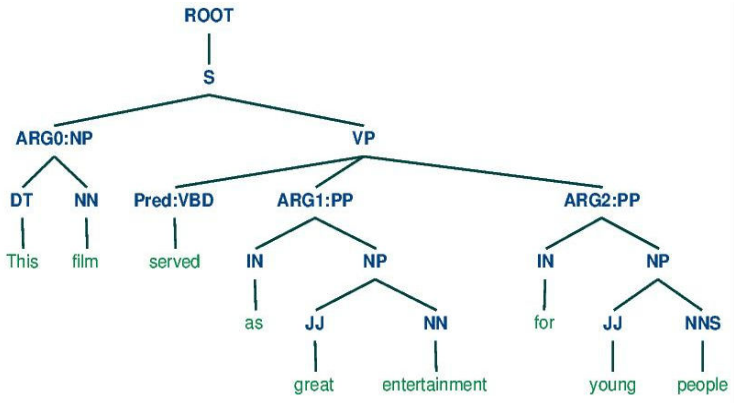

Fig. 2: Syntactically and Semantically Annotated Parse Tree

Predicate	Arguments
serve	Arg0: This film
	ARG1: as great entertainment
	ARG2: for young people

Table 2: Predicate-Argument Information

4 Comparison to SenticNet

At the current stage of our experiments, we did not perform any automatic comparison or performance measurement leaving it to the official evaluation during the challenge days. However to give an idea to the reviewers, Table 3 lists a couple of example sentences together with the extracted[1] concepts both by the proposed system[2], and SenticNet [4]. We leave it to the reviewers to compare the outputs.

Acknowledgment

We would like to acknowledge that this work was partially supported by National Science Council, Taiwan, under the contract NSC 102-2221-E-001-026.

[1] The SenticNet concepts were extracted using its web-demo version available at (http://sentic.net/demo/)

[2] A web-demo of the proposed system is available at (http://andycyrus.github.io/ESCW2014-challenge)

Sentence	**Proposed System's Output**	**SenticNet's Output**
I went to the market, bought fresh fruits and vegetables and came back	(1)bought_fresh_fruits (2)I_went (3)I_bought (4)vegetables (5)bought_vegetables (6)the_market (7)fresh_fruits (8)went_to_the_market (9)came_{in_the_direction}_ back (10)came_I	(1)go_to_market (2)market (3)buy_fruit (4)buy_vegetable (5)fresh_fruit (6)back_come
We also ordered the bedding and got the pillow	(1)got_the_pillow (2)the_pillow (3)We_got (4)the_bedding (5)We_order (6)order_also (7)order_the_bedding	(1)also_order (2)order_bed (3)bed (4)get_pillow (5)pillow

Table 3: Example Sentences and Extracted Concepts

References

1. Beńat Zapirain and Eneko Agirre and Lluís Màrquez. Ubc-upc: Sequential srl using selectional preferences: An aproach with maximum entropy markov models. In Proceedings of the 4th International Workshop on Semantic Evaluations, pages 354-357, Stroudsburg, PA, USA. Association for Computational Linguistics. 2007.
2. Claire Bonial, Jena Hwang, Julia Bonn, Kathryn Conger, Olga Babko-Malaya and Martha Palmer, English PropBank Annotation Guidelines, Center for Computational Language and Education Research Institute of Cognitive Science University of Colorado at Boulder, Nov, 2012.
3. Dan Klein and Christopher D. Manning. Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430, 2003.
4. Erik Cambria, Daniel Olsher, Dheeraj Rajagopal. SenticNet 3: A Common and Common-Sense Knowledge Base for Cognition-Driven Sentiment Analysis, Association for the Advancement of Artificial Intelligence. 2014.
5. Gildea Daniel and Daniel Jurafsky. Automatic labeling of semantic roles. Computational Linguistics, 28(3): 245-288, 2002.
6. Kyung-Mi Park and Hae-Chang Rim. Maximum entropy based semantic role labeling. In Proceedings of the Ninth Conference on Computational Natural Language Learning, CONLL ?05, pages 209-212, Stroudsburg, PA, USA. Association for Computational Linguistics, 2005.
7. Martha Palmer, Dan Gildea, Paul Kingsbury, The Proposition Bank: A Corpus Annotated with Semantic Roles, Computational Linguistics Journal, 31:1, 2005.
8. Mitchell P., Beatrice Santorini and Mary Ann Marcinkiewicz, Building a Large Annotated Corpus of English: The Penn Treebank: COMPUTATIONAL LINGUISTICS, 10(2), pages 313-330, 1993.
9. Nianwen Xue. Calibrating features for semantic role labeling. In Proceedings of EMNLP 2004, pages 88-94. 2004.
10. Richard Johansson and Pierre Nuges, The Effect of Syntactic Representation on Semantic Role Labeling, Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), pages 393-400, Manchester, August 2008.
11. Sammer Pradhan, Wayne Ward, Kadri Haciuglu, James Martin, and Dan Jurafsky. 2004. Shallow semantic parsing using support vector machines. In Proceedings of HLT/NAACL-2004.