

Designing a Discourse Parser for the Evaluative Text Genre

Hammad Ali, Giuseppe Carenini, Gabriel Murray, and Raymond Ng
Department of Computer Science
University of British Columbia
Vancouver, B.C. Canada V6T 1Z4
{hammad, carenini, gabrielm, rng} @cs.ubc.ca

Abstract

We propose designing a discourse parser specifically for the evaluative text genre. We aim to see whether focusing on a certain genre and relations specific to that genre offers performance gain beyond more generic discourse parsers. In this extended abstract we describe the approach we intend to take, and how this differs from what has been done previously.

The problem of discourse parsing

It has often been pointed out that in any text sentences are not understood in isolation but in relation to each other. Several theories postulate that all discourse has a hierarchical structure of coherence relations which reflect the author's intended effect on the reader. Discourse structure is often based on semantic and pragmatic relationships, and results in a tree structure whose leaves are the elementary discourse units (EDUs) of the text. The goal of discourse parsing is to build this tree structure and to identify the rhetorical relations labeling the internal nodes of the tree. Discourse parsing consists of three co-dependent subtasks: i) identifying the EDUs, ii) determining which discourse segments are related to each other, and iii) identifying the rhetorical relation between these segments. Discourse parsing has applications in other NLP tasks such as sentence compression, summarization and natural language generation (NLG). For instance, knowing that two EDUs are in an *Elaboration* relation, and one of them is the nucleus, would enable us to prune the satellite EDU when generating a summary. In NLG, knowing what features usually signal a *Concession* relation can enable us to generate such instances ourselves.

Our task

There are several different theories about the taxonomy of rhetorical relations, with each proposing its own set of fine-grained relations. However, for most purposes we can identify a common core of relations, possibly at the cost of losing some subtle differences. For our study we are interested in extracting information from evaluative text, for instance product and service reviews. Based on an initial annotation phase, we have narrowed our relation set down to 8 key relations we believe to be most informative for our genre, compared to the 18 relations used by duVerle and Predinger (2009) for generic discourse parsing, and the 26 relations used by Subba and Di Eugenio (2009) on discourse parsing for instruction manuals. Among the relations we are interested in identifying is *Evidence*, where one span presents a claim, and the other presents a justification for this claim. Another relation is *Concession*, where both spans are claims about an entity, with one span qualifying the other, possibly through the use of contrasting opinions. These relations offer information about what claims are being made about an entity, whether these claims are equivocal, and the justifications for these claims. It is important to emphasize here that we are not claiming that relations beyond this subset do not occur in evaluative text, just that they do not offer enough useful information to merit inclusion as separate relations. Instead we merge them all into a generic, catch-all relation which we refer to as *Adjacent*.

We believe that focusing on genre-specific knowledge and genre-specific relations, can offer performance gains over parsers that use generic corpora and relation sets. We are greatly motivated by the promising results achieved by Subba and Di Eugenio (2009). They focus on the genre of instruction manuals, and identify a small subset of relations believed to be most informative for this genre. While most previous work on rhetorical parsing has used lexical and syntactic features, they also use more genre-specific features and demonstrate that this offers performance gain. Since they are dealing with instruction manuals, they introduce verb semantics as an additional feature. Their results give us reason to believe that developing genre-specific parsers is a promising avenue of research. As a contribution beyond the use of lexical and syntactic features, we propose to use detection of subjectivity and polarity as an additional feature. Since we are dealing with evaluative text, it will be interesting to investigate how much information such features offer in addition to the basic feature set. During the initial annotation phase, we

noticed that knowledge of polarity aids relation identification. We would like to explore further and see how much, if any, performance gain these features offer.

While taking the supervised approach creates the need for an annotated corpus, we believe that development of such a corpus will benefit the NLP research community. Further, the unsupervised approach to discourse parsing has faced some criticism. Marcu and Echihiabi (2002) tried to generate training data in an unsupervised manner, through the use of explicit discourse relations as an approximation to implicit discourse relations. Later, Sasha Blair-Goldensohn *et al.* (2007) extends their work in breadth and depth. However, the results from this unsupervised approach have been difficult to reproduce. In particular, Sporleder (2008) claim that there is some innate difference between “organic” and “synthetic” instances that is lost in unsupervised data generation.

Preliminary results

High-Level Discourse Analyzer (HILDA) is a generic document-level discourse parser. HILDA uses the 18 relation set proposed by Rhetorical Structure Theory (RST), and was trained on the RST-DT corpus, which corresponds to a subset of the Penn Treebank (PTB).

We wanted to see the extent to which restricting the number of possible relations, as opposed to using the entire RST relation set, offers an advantage in parsing. Thus we obtained HILDA and ran some preliminary tests on our corpus. For comparison we compute F-scores at the unit, span, nuclearity and relation level. We had a small subset of our corpus manually annotated in order to measure HILDA’s performance on it. HILDA performs much worse on this corpus than on RST-DT, with a 4.55% F-score at the relation level compared to 55.3% reported in their paper. While genre-switching is one cause for this sharp decline, we believe that using a smaller relation set with the same feature set as HILDA can offer better performance. Further, we also want to see the effect of using subjectivity and polarity as additional features.

After a manual analysis of the output trees, we notice that most of the disagreements at the relation level are due to ambiguous relations. For instance, relations like *Elaboration* are too generic to be reliably distinguished from more specific, informative relations like *Evidence* or *Restatement*. This observation strengthens our belief that developing a classifier that focuses on a more customized relation set can address this issue. We also intend to present the relations along with their order of specificity, hoping this will lead to the annotators favoring relations that are more informative over ones that are less so.

Current work

As a starting point for our study, we want to develop an annotated corpus of evaluative text. We have identified a small subset of relations, obtained by identifying relations of interest, merging some of them into broader classes, and treating all other relations as a generic label. There is a clear mapping from the 18 label RST relation set to our 8 label relation set, which enables us to carry out a fair comparison.

We have prepared a rigorous annotation manual in which we present our relation set divided into levels from more to less specific, in order to facilitate the annotation process. We also provide comprehensive guidelines to help distinguish a relation from all the other ones that share characteristics with it.

The next task is to hire and train annotators, and complete the annotation process. Following this, we propose to use the same lexical and syntactic features as duVerle and Predinger (2009) to see if this already yields better performance on our corpus due to training and testing on the same genre as well as to the smaller subset of relations. We then propose to use detection of subjectivity and polarity as an additional feature, with the hope that this will offer further performance gain specifically for this genre.

References

1. Building a Discourse-tagged corpus in the framework of Rhetorical Structure Theory – Lynn Carlson, Daniel Marcu, Mary Ellen Okurowski, ACL 2001
2. An unsupervised approach to recognizing discourse relations – Daniel Marchu, Abdessamad Echihiabi, ACL 2002
3. Building and refining rhetorical-semantic relation models – Sasha Blair-Goldensohn, Kathleen R. McKeown, Owen C. Rambow, NAACL 2007
4. Manually vs. Automatically labeled data in Discourse Relation Classification – Caroline Sporleder, LDV Forum, Volume 14, Issue 3, July 2008
5. An effective Discourse Parser that uses Rich Linguistic information – Rajen Subba, Barbara Di Eugenio, NAACL 2009A Novel Discourse Parser based on Support Vector Machine Classification – David A. duVerle, Helmut Predinger, ACL 2009
6. A Novel Discourse Parser based on Support Vector Machine Classification – David A. duVerle, Helmut Predinger, ACL 2009.