Improving discourse sense classification by the form of discourse relations

Frances Yung   Yuji Matsumoto
Nara Institute of Science and Technology
{pikyufrances-y,matsu}@is.naist.jp

1 Introduction

Discourse relations are relations betwen unit of texts that make a document coherent. These relations can be marked explicitly or implied implicitly.

For example, the word *but* in Example (1) marks a *Contrast* relation. We call *but* an explicit discourse connective (DC). On the other hand, a *Result* relation can be inferred between the two sentences in Example (2) although there is not an explicit markers. We say the two sentences (called *arguments*) are connected by an implicit DC.

1. It is late, *but* he is still awake.
2. It is late. I go to bed.

1.1 Penn discourse treebank

The Penn Discourse Treebank (PDTB) [5] is the largest available discourse-annotated resource in English. The raw text are collected from news articles of the Wall Street Journals. On the PDTB, all explicit DCs are annotated with a discourse sense, while implicit discourse senses are annotated between two adjacent sentences. Each discourse relation is labelled with 1 to 2 senses. Following existing works, we split multi-sense samples into multiple samples, each labeled with one of the senses.

The discourse senses defined in PDTB are arranged in a hierarchy of 3 levels, resulting in a total of 42 distinct sense labels, as shown in Figure 1.

1.2 Shallow discourse parsing

The task of shallow discourse parsing, also known as PDTB-style discourse parsing, is to retrieve a list of discourse relations from an input text, where each discourse relation is a tuple of 1) the explicit DC in the text or 'implicit DC'; 2) the positions and spans of the arguments; and 3) the sense of the discourse relation.

Since some senses in the PDTB occur at very low frequency, similar and related senses are practically grouped together. For example, in the CONLL shared task of shallow discourse parsing [6], the 42 senses are mapped to 15 senses (as shown in the first column of Table 1), making the task a managable 15-way classification task.

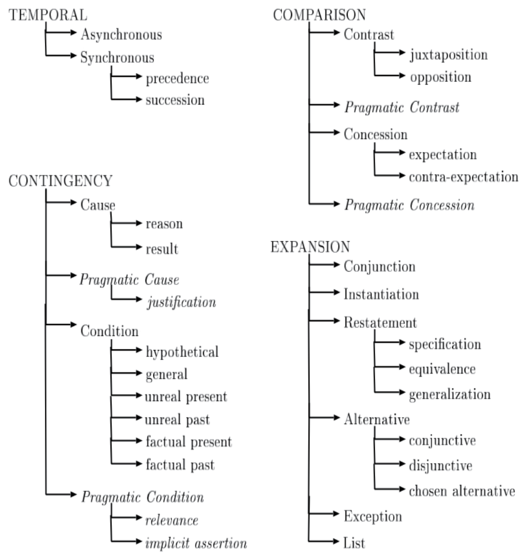

Figure 1: Sense hierarchy of PDTB. (from Prasad et al., 2008)

A pipeline approach, which is first proposed by [1], is generally adopted to identified the elements of discourse relations one after another in separated models . Figure 2 shows the pipeline of the discourse parser of [7], which was the winning parser of the CONLL shared task.

In a conventional discourse parsing pipeline, explicit DCs are first identified, followed by identification of the location and spans of the arguments for these explicit DCs. The senses of the explicit DCs are classified, based on features extracted from the identified arugments. Next, adjacent sentences that are not connected by any explicit relations are selected as arguments for implicit DCs. Similarly, the

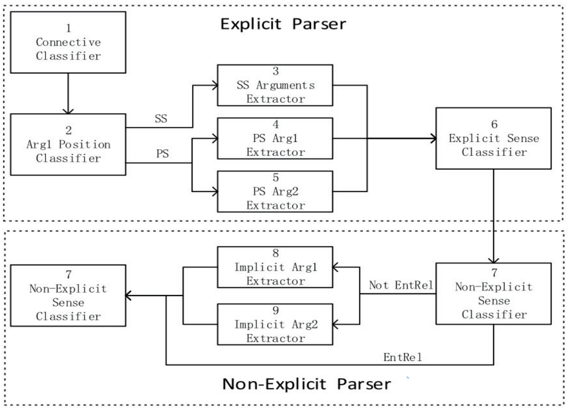

Figure 2: Pipepine of discourse parser. (from Wang and Lang, 2015)

senses of the implicit DCs are classified, based on features extracted from the identified arugments.

Since explicit and implicit relations are first separated in the parsing pipeline, and separated classification models are built to classify explicit and implicit discourse senses respectively, the form of the discourse relation is essentially used for sense classification.

In this work, we propose to directly make use of 'the fact that a particular relation is expressed explicitly/implicitly'. Our approach is driven by a cognitive motivation, which is explained in Section 2.

2 Motivation

The presentation and interpretation of discourse relations can be viewed as a kind of communication between speakers and listeners (or authors and readers). Recent psycholinguistic studies proposed that when human communicate, the speaker attempts to be informative while the listener use Bayesian inference to reason the speaker's intended message [2, 3, 4].

Specifically, the speaker selects an utterance that, s/he thinks, is unambiguous for the listener. This is based on the speaker's prediction on how likely the listener can figure out his/her intended message from the utterance s/he uses.

On the other hand, the listener interprets the utterance taking into account how likely, s/he thinks, the speaker chooses that utterance. These inferences between speakers and listeners are formally defined in the *Rational Speech Acts* (RSA) model [2].

We propose to tackle the problem of discourse relation classification from the viewpoint of a rational listener. When intepreting the sense of a discourse relation, a listener considers also whether the sense is marked or not. It is because some senses (such as *'contrast'* and *'condition'*) are more often marked explicitly, while other senses (such as *'expansion'* and *'result'*) are more often implicit.

Section 3 explains the details about the RSA models and explains how we adapt the RSA framework for discourse sense classification.

3 Related work

The RSA model [2] is a variation of the game-theoretic approach. It explains the communicative reasoning of a speaker and a listener in terms of Bayesian probabilities.

A rational speaker optimizes the informativeness of his utterance based on the listener's knowledge s/he assumes and adjusted by the cost of production. S/He chooses an utterance by soft-max optimizing the expected utility of the utterance (Equation 1). α is the decision noise parameter, which is set to 1 to represent a rational speaker. Utterances that are unconventional and surprising are less useful, thus utility is defined as the negative surprisal of the utterance with respect to the message to be conveyed, deducted by the cost ($D(w)$) to produce it.

$$P_{speaker}(w|s,C) \propto e^{\alpha \cdot U(w;s,C)} \quad (1)$$

On the other hand, a rational listener assumes the utterance s/he hears contains the optimal amount of information. S/he predicts the intended message of a speaker by Bayesian inference.

$$P_{listener}(s|w,C) \propto P_{speaker}(w|s,C)P(s) \quad (2)$$

where s is the *message* of an utterence; w is the *utterance* produced by the speaker, and C is the *context*. $P_{speaker}(w|s,C)$ represents the *listener's predicted* speaker's model, and $P(s)$ represents the *salience* of the message, which is shared knowledge between the speaker and listener.

We propose to make use of the *listener model* of RSA and incoporate the likelihood that the speaker would choose expression w to convey his/her intended discourse sense s in context C. The distribution of $P_{speaker}(w|s,C)$ can be extracted from discourse annotated corpus data.

We define context C as the contextual discourse relation senses and forms, in window sizes of 1 to 2 (previous one, next one, previous two, next two, previous one paired with next one). We hypothesize that certain patterns of discourse relation forms are preferred over others. Results based on various discourse contexts are compared in the experiment.

$P(s)$ in Equation (2) represents the *salience* of the speaker's intended discourse sense. The salience of a discourse sense depends on many factors, many of which are explored as classification features used in discourse parser. Therefore, we simply integrate the probability prediction of a pipeline discourse classifier as the salience of sense s.

In this way, we can rerank the output of the discourse parsing pipeline according to the actual discourse relation form and discourse relation context. This simulates the Bayesian inference used by the listener when interpreting a discourse relation.

4 Analysis

Before examining the applicability of the proposed method, we analyze the distribution of $P_{speaker}(w|s, C)$ under various criteria. We use Sections 2-22 as the training set, from which the $P(w|s, C)$ distribution is extracted, and Sections 0-1, 23-24 are held out of future parsing experiments.

Table 1 shows part of the likelihood distribution $P(w|s, C)$. The distributions largely differ under various counting criteria. While explicit DCs are possible utterance for all discourse senses, some senses are much more likely to be marked explicitly. The third column of Table 1 shows the sense distribution of the DC 'when', which is still ambiguous. Conditioning the likelihood by contextual discourse sense (last column) further reduces the entropy of the distribution.

s	$P(Exp\|s, C)$ C =constant	$P(\text{'when'}\|s, C)$ C =constant	$P(\text{'when'}\|s, C)$ C:prev sense =Reason
Concession	0.84	0.009	0.03
Contrast	0.65	0.001	0
Reason	0.36	0.027	0.04
Result	0.29	0.001	0
Condition	0.99	0.14	0.179
Alternative	0.95	0.01	0.077
Chosen.Alt	0.41	0	0
Conjunction	0.58	0	0
Exception	0.82	0	0
Instantiation	0.17	0	0
Restatement	0.05	0.001	0
Precedence	0.63	0.003	0.01
Succession	0.86	0.185	0.193
Synchrony	0.86	0.319	0.417

Table 1: Examples of $P(w|s, C)$ distribution

5 Conclusion

In this work, we propose to use the form of discourse relation to improve discourse sense classification. Based on the theory that Bayesian inference is used in human language understanding, we propose to modeify the prediction of discourse senses based on the likelihood that the discourse sense is presented in the form it actually occurs in the text. Analysis on corpus data provides support on the applicability of this proposed method.

References

[1] Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. A pdtb-styled end-to-end discourse parser. *Natural Language Engineering*, 20(02):151–184, 2014.

[2] Michael C. Frank and Noah D. Goodman. Predicting pragmatic reasoning in lanugage games. *Science*, 336(6084):998, 2012.

[3] Noah D Goodman and Andreas Stuhlmüller. Knowledge and implicature: modeling language understanding as social cognition. *Topics in cognitive science*, 5(1):173–184, 2013.

[4] Leon Bergen, Roger Levy, and Noah D. Goodman. Pragmatic reasoning through semantic inference, 2014.

[5] Rashmi Prasad, Nikhit Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. The penn discourse treebank 2.0. *Proceedings of the Language Resource and Evaluation Conference*, 2008.

[6] Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Rashmi PrasadO Christopher Bryant, and Attapol T Rutherford. The conll-2015 shared task on shallow discourse parsing. *CoNLL 2015*, page 1, 2015.

[7] Jianxiang Wang and Man Lan. A refined end-to-end discourse parser. *CoNLL 2015*, page 17, 2015.

[8] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkey, Steven J. Bethard, and David McClosky. The standord corenlp natural language processing toolkit. *Proceedings of the Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2014.