

JAIST: A two-phase machine learning approach for identifying discourse relations in newswire texts

Nguyen Truong Son
University of Science, VNU
Ho Chi Minh City
Viet Nam
ntson@fit.hcmus.edu.vn

Ho Bao Quoc
University of Science, VNU
Ho Chi Minh City
Viet Nam
hbquoc@fit.hcmus.edu.vn

Nguyen Le Minh
Japan Advanced Institute of
Science and Technology
Ishikawa, 923-1292
Japan
nguyenml@jaist.ac.jp

Abstract

In this paper, we present a machine learning approach for identifying shallow discourse relations in news wire text. Our approach has 2 phases. The arguments detection phase will identify arguments and explicit connectives by using the Conditional Random Fields (CRFs) learning algorithm with a set of features such as words, parts of speech (POS) and features extracted from the parsing tree of sentences. The second phase, the sense classification phase, will classify arguments and explicit connectives into one of fifteen types of senses by using the SMO classifier with a simple feature set. The performance of system was evaluated three different data sets given by the CoNLL 2015 Shared Task. The parser of our system was ranked 4 of 16 participating systems on F-measure when evaluating on the blind data set (strict matching).

1 Introduction

The shallow discourse parsing task given by the CoNLL 2015 Shared Task proposed by Xue et al. (2015) aims to extract discourse relations in newswire texts. Each discourse relation is a set of four: two arguments, connective words and senses. However, the connective words may not be available in case of implicit discourses. Identifying discourse relations is clearly an important part of natural language understanding that benefits a wide range of natural language applications. A number of applications of discourse information have been proposed for recent years. For example, in the task of identifying paraphrase texts, Bach et al. (2014) has used discourse information to compute the similarity

score between two sentences or Somasundaran et al. (2009) has used discourse relations to improve the performance of the opinion polarity classification task.

In the past, this task is solved at different levels. Lin et al. (2009) have used supervised learning method to build a maximum entropy classifier to identify implicit relations. Ghosh et al. (2011, 2012) have used CRFs with a set of local and global features to recognize arguments of discourses from texts. However, in contrast to the CoNLL 2015 SDP Shared Task, Ghosh et al. (2011, 2012) just considered explicit relations with explicit connectives have been provided.

Our team approach for this shared task composes two phases. In the first phase, we use CRFs and a set of features such as words, POS and pattern features based on parsing tree of sentences to build models for recognizing arguments and connective words. In the second phase, we use the SMO algorithm, an optimization of SVM, to build a classifier to predict the senses of discourse relations.

The remainder of this paper is structured as follows: Section 2 describes the details of the proposed system for solving the task of identifying shallow discourse relations given by CoNLL 2015 Shared Task. We also describe the experimental results and some analysis in Section 3. Finally, Section 4 presents our conclusions and future works.

2 System description

Our parser system is divided into 2 phases. Firstly, documents without discourse information will be passed through the argument detection phase to recognize components of discourse relations such as both of arguments and explicit connectives if it is possible. Secondly, the sense classification phase will identify the sense of discourse relation by using a SVM classifier then format

the results according to the expected output of evaluate system.

2.1 Phase 1: detection of arguments

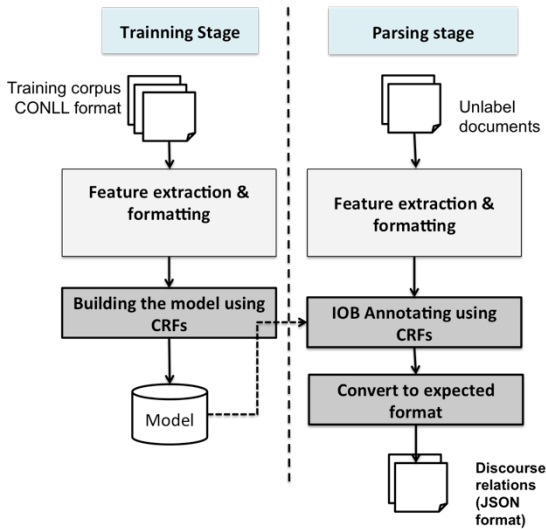


Figure 1. Workflow of the arguments detection phase

The workflow of the first phase consists of two stages. In the training stage, we use machine-learning algorithms to build models, which will be used to identify boundaries of components in the parsing stage. In order to learn models by using machine learning (ML) algorithms, we use some popular features in such as words, parts of speech. Besides, we extract a set of pattern features based on the parsing tree of sentences.

According to our analysis of discourse relations, two arguments of each discourse relation may be appeared at different positions: in the same sentence, in two consecutive sentences or in far apart sentences. Based on the statistic of discourse relations in the training dataset, we see that the number of discourse relations which two arguments located in the same sentence or two consecutive sentences is in a large quantity (92.5%). Therefore, our system focus on identifying these kinds of discourse relations by building two models: one for recognizing discourse relations in the same sentence (**SS**) and another model for recognizing discourse relations in two consecutive sentences (**2CS**).

To build learning models using ML algorithms, we need to extract the features from the data set for the input of the ML algorithm. Each type of discourse relation (SS-type or 2CS-type) has some common features and some reserved features. Table 1 describes all features that are used for machine learning approaches in our experiments.

Table 1. List of all features using for identifying arguments and connectives

#	Feature description
<i>Common features for both SS-type and 2CS-type</i>	
A	Word
B	POS
C	Stem
D	Belongs to connective list
E	Brown cluster
F	Noun phrase / verb phrase
G	CLAUSES from S
<i>Pattern features of SS discourse relations</i>	
H	S_CC_S
I	SBAR_CC_SBAR
K	SBAR_IN_S
<i>Pattern features of 2CS discourse relations</i>	
L	1 st sentence: RIGHTMOST_S 2 nd sentence: S_begin_with_CC
M	1 st sentence: RIGHTMOST_S 2 nd sentence: NP_ADVP_VP
N	1 st sentence: RIGHTMOST_S 2 nd sentence: S_begin_with_ADVP
O	1 st sentence: RIGHTMOST_S 2 nd sentence: S_begin_with_PP

After all required features are extracted, the training data and these extracted features will be formatted as the input format of the machine learning algorithm tool in which words of discourse relations are marked labels using IOB notations. We use CRF++ (Taku Kudo, 2005), an implementation of the Conditional Random Fields (John Lafferty et al, 2001) to train models from the training data sets.

After models are built, they were used to predict the discourse labels of new documents (in the parsing stage) then the result will be converted into expected format.

Section 2.1.1 and 2.1.2 will describe the details of all features we used in our experiments.

2.1.1 Common features:

- *Popular language features (A-C)*: including words, their parts of speeches and their stems.
- *Connective features (D)*: The features show whether or not the words belong to a predefined connective list. Predefined connective lists are constructed from connective words in the training data set. Then we use these lists to extract this feature for building the model.
- *Brown clusters features (E)*: Brown clusters, introduced and prepared by Turian (2010), were successfully applied in some

named entity recognition tasks. In Brown clusters, the semantic similarities of words in the same cluster are higher than of words in different clusters. We use the Brown cluster index of words as a feature for the ML process.

- Noun phrases, verb phrases and clauses features (F, G): all words of a noun phrase, verb phrase or clauses are often located entirely in arguments. Moreover, the beginning of arguments is often the same with the beginning of noun phrases, verb phrases or clauses. We extract noun phrases, verb phrases and clause based on the syntactic parse tree of sentences.

2.1.2 Pattern based features based on syntactic parse trees

Our analysis on the training corpus shows that the syntactic information based on the syntactic parse trees is very important for identify discourse relations. According to our analysis, sentences that express discourse relations are usually follow some special syntax. Therefore, if we can extract features based on these special syntaxes, the system will recognize arguments of discourse relations more exactly.

Due to the linguist characteristic of discourses in sentences, each kind of discourse relations (SS-type or 2CS-type) has different pattern feature sets. Below are patterns based on syntactic parse trees we have used to extract features for each of type:

Pattern features for SS-type discourses recognition (H, I, K): We have three patterns that help to recognize boundaries of arguments of SS-type discourse relations. These patterns are based on the syntactic characteristic of discourse expressions using prepositions or conjunctions such as *and*, *but*, *if*, *although*, ... For example, the pattern S_CC_S (feature **H**) and $SBAR_CC_SBAR$ (feature **I**) indicate S nodes of which child nodes matched with the pattern $S(.*)CC(.*)S(.*)$ or $SBAR(.*)CC(.*)SBAR(.*)$. In this case, related S-nodes or SBAR-nodes may be the arguments of a discourse relation. Figure 2 shows an example of sentences which matches with pattern S_CC_S . In this example, the matched left S node and the right S node are arguments of a discourse relation in the training data set. Another pattern is $SBAR_IN_S$ (Feature **K**). This pattern matched with sentences of which SBAR node has an IN node (“if”, “although”, “before”, “after”, “though”) follow by

an S node. If a sentence match with this pattern, the S node is often the first arguments and the rest is often the second argument of a discourse relation. Figure 3 shows an example of sentences matched with the pattern $SBAR_IN_S$.

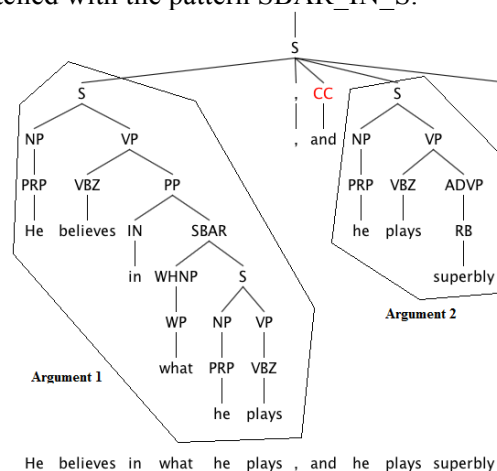


Figure 2. The matching of a discourse relation with the pattern S_CC_S

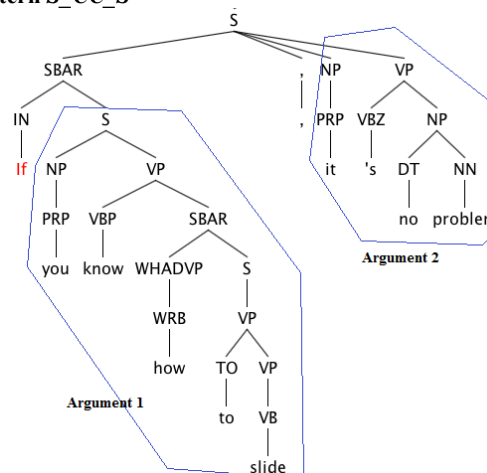


Figure 3. The matching of a discourse relation with the IN_SBAR pattern

Pattern features for 2CS-type discourses recognition (L, M, N, O): When arguments of discourse relations are not located in the same sentence the task is more difficult. To build the model for identifying 2CS discourses, we will extract pattern-based features of each pair of sentences in the training data set based on parsing tree of sentences. Our analysis on the training corpus shows that if a pair of sentence in which the second sentence begins with *a conjunction*, *an adverb*, and *a preposition* (e.g. “for example”, “by comparison” and so on) or *a noun phrase followed by an adverb* (e.g., “also”), the right most clauses of first sentence and the left most sentence in the second sentence may be arguments of a discourse relation. We use patterns from L-O to extract these features.

2.2 Phase 2: Sense classification

After arguments and explicit connectives of discourses are identified, we need to identify the sense of these discourses. These discourses without sense information are passed through a classifier with a model trained in the training stage to identify the correct senses of discourse relations.

This model used in the above step are built by using the Sequential Minimal Optimization algorithm (SMO), a fast algorithm for training support vector machines (John Platt, 1998), with some simple features such as: connective words; type of discourses (SS or 2CS); does the first character of connective words capital or not? The workflow of sense classification phase is shown in Figure 4.

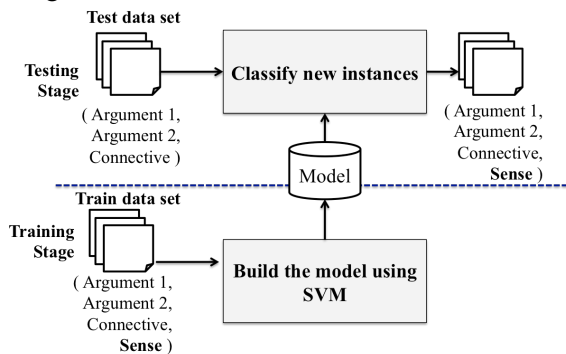


Figure 4. Workflow of the sense classification phase

We use LIBSVM (Chang and Lin, 2011) – a library that implemented SMO algorithm to build the model and classify discourses into the senses category. The trained model for sense classification task achieves an F-score of 79.8% (Precision=80.9%, R=81.6%) when evaluate using cross validation 10-fold method.

One limitation of our sense classification step is that it just takes into account discourses with explicit connectives, so the sense recognition of non-implicit discourses still has not been solved yet.

3 Experimental results

Table 2 shows the evaluation result of our system on the three data sets provided by the CONLL Shared task 2015, the rank column is the rank of our system when compare with other participating systems. In general, this task is a difficult task, so the result is not as high as our expectation. Moreover, due to the usage of special syntactic patterns extracted from parse trees, the precision scores of our system is higher than other teams. However, these patterns just cover several special cases, so the recall score of our system is low.

Table 2. The evaluation result of our system on the blind, test and dev data sets

	BLIND data set		TEST data set		DEV data set	
	score	rank	score	rank	score	rank
Arg 1 Arg2 extraction (%)						
F1	32.11	7	35.43	7	40.07	8
P	42.72	3	52.98	1	58.92	1
R	25.72	11	26.61	12	30.36	12
Arg1 extraction (%)						
F1	40.99	6	42.43	7	46.60	8
P	54.53	3	63.45	1	68.51	1
R	32.84	12	31.87	12	35.31	12
Arg2 extraction (%)						
F1	48.53	9	47.99	7	48.99	11
P	64.56	6	71.77	2	72.03	4
R	38.88	12	36.05	13	37.12	13
Explicit connective (%)						
F1	61.66	12	63.89	15	65.53	14
P	88.55	10	91.87	8	91.56	10
R	47.30	13	48.97	16	51.03	16
Overall parser performance (%)						
F1	18.28	4	20.25	8	26.10	5
P	24.31	2	30.29	2	38.38	1
R	14.64	6	15.21	10	19.78	8
Sense (%)						
F1	15.61	6	13.61	8	19.93	5
P	40.55	1	49.34	1	63.15	1
R	12.44	8	10.73	12	15.01	9

The comparison of the evaluation result between explicit discourses and non-explicit discourses are shown in Table 3. With the help of special patterns based on explicit connectives and parse trees, the result of explicit discourses recognition is higher than the result of non-explicit discourses recognition for both of precision and recall scores.

Table 3. Comparison result between explicit discourses and non-explicit discourses

	BLIND set			TEST data set		
	ALL	Ex- plicit	Non- Exp.	ALL	Ex- plicit	Non- Exp.
Arg 1 Arg2 extraction (%)						
F1	32.11	34.23	30.44	35.43	38.16	32.44
P	42.72	49.16	38.28	52.98	54.88	50.41
R	25.72	26.26	25.27	26.61	29.25	23.92
Arg1 extraction (%)						
F1	40.99	44.08	36.9	42.43	43.82	38.85
P	54.53	63.3	46.4	63.45	63.01	60.37
R	32.84	33.81	30.63	31.87	33.59	28.64
Arg2 extraction (%)						
F1	48.53	51.35	46.13	47.99	56.25	38.85
P	64.56	73.74	58	71.77	80.89	60.37
R	38.88	39.39	38.28	36.05	43.12	28.64
Explicit connective (%)						
F1	61.66	61.66	0	63.89	63.89	0
P	88.55	88.55	0	91.87	91.87	0
R	47.3	47.3	0	48.97	48.97	0

Overall parser performance						
F1	18.28	27.2	11.25	20.25	33.22	8.01
P	24.31	39.06	14.15	30.29	47.76	12.45
R	14.64	20.86	9.34	15.21	25.46	5.91
Sense (%)						
F1	15.61	22.89	1.61	13.61	19.14	1.23
P	40.55	42.58	84.51	49.34	48.43	86.6
R	12.44	17.88	2.54	10.73	14.66	1.97

The feature set based on the syntactic parse tree is very important for our system. Table 4 shows the comparison between two different feature set on the development data set. The FULL feature set consists of all feature including lexical, part of speeches, and pattern features based on syntactic parse trees and so on. However, in the SHORT feature set, we remove all pattern features based on syntactic parse trees to evaluate the importance of these features. The result, which just considered discourse relations in the same sentences, showed that there is a significant improvement when we use the FULL feature set instead of the SHORT feature set.

Table 4. The comparison between FULL and SHORT feature set

	FULL	SHORT	FULL	SHORT
Arg 1 Arg2 extraction				
F1	0.505	0.315	0.670	0.512
P	0.684	0.559	0.886	0.885
R	0.401	0.219	0.539	0.360
Arg1 extraction				
F1	0.567	0.382	0.452	0.270
P	0.766	0.677	0.612	0.479
R	0.449	0.266	0.359	0.188
Arg2 extraction				
F1	0.612	0.433	0.232	0.159
P	0.827	0.769	0.665	0.612
R	0.485	0.302	0.184	0.109

4 Conclusion

Our approach to the Shallow Discourse Parsing at CONLL 2015 Shared task was to create a 2-phase system that identifies discourse relations in newswire text. Results show that our approach achieves the high precision of all systems and was ranked 4th in terms of F1-measure when strict matching is used.

In the future we would like to improve the recall of our approach by exploring the use of a wider range of features.

References

- N. Xue, H.T. Ng, S. Pradhan, R. Prasad, C. Bryant, A. Rutherford. 2015. *The CoNLL-2015 Shared Task on Shallow Discourse Parsing*. In Proceedings of the Nineteenth Conference on Computational Natural Language Learning: Shared Task. Beijing, China.
- N.X. Bach, N.L. Minh, and A. Shimazu. 2014. Exploiting discourse information to identify paraphrases. *Expert Systems with Applications*, 41(6):2832–2841, May.
- J. Lafferty, A. McCallum, and F.C.N Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- S. Ghosh, R. Johansson, and S. Tonelli. 2011. Shallow discourse parsing with conditional random fields. In *In Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*. Citeseer.
- S. Ghosh, G. Riccardi, and R. Johansson. 2012. Global features for shallow discourse parsing. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 150–159. Association for Computational Linguistics.
- T. Kudo. 2005. CRF++: Yet another CRF toolkit. *Software available at <http://crffpp.sourceforge.net>*.
- J. Platt. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines.
- C.C. Chang and C.J. Lin. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.
- S. Somasundaran, G. Namata, J. Wiebe, and L. Getoor. 2009. Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 170–179.
- Z. Lin, M.Y. Kan, and H.T. Ng. 2009. Recognizing implicit discourse relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 343–351.
- J. Turian, L. Ratinov, and Y. Bengio. 2010. Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 384–394.