Discourse Relation Recognition by Comparing Various Units of Sentence Expression with Recursive Neural Network

Atsushi Otsuka, Toru Hirano, Chiaki Miyazaki, Ryo Masumura,
Ryuichiro Higashinaka, Toshiro Makino and **Yoshihiro Matsuo**
NTT Media Intelligence Laboratories
NTT Corporation
{otsuka.atsushi, hirano.tohru, miyazaki.chiaki, masumura.ryo, higashinaka.ryuichiro, makino.toshiro, matsuo.yoshihiro}@lab.ntt.co.jp

Abstract

We propose a method for implicit discourse relation recognition using a recursive neural network (RNN). Many previous studies have used the word-pair feature to compare the meaning of two sentences for implicit discourse relation recognition. Our proposed method differs in that we use various-sized sentence expression units and compare the meaning of the expressions between two sentences by converting the expressions into vectors using the RNN. Experiments showed that our method significantly improves the accuracy of identifying implicit discourse relations compared with the word-pair method.

1 Introduction

Discourse relation recognition is a technique to identify the type of discourse relation between two sentences. Because discourse relation contributes to the coherence of sentences, it has potential applications in many natural language processing (NLP) tasks. For example, in text summarization, it makes summary documents more consistent by using discourse relations and structures (Gerani et al., 2014). Similarly, in conversational systems (Higashinaka et al., 2014), discourse relations can help the system select contextually appropriate system utterances.

Discourse relations are categorized into explicit and implicit relations. Explicit relations have a discourse marker such as a connective, making them easy to identify with a high degree of accuracy (Pitler and Nenkova, 2009). Implicit discourse relations, in contrast, have no discourse marker between sentences. Previous studies have proposed many methods for implicit discourse recognition, among them reasoning-based (Sugiura et al., 2013) and pattern-based (Saito et al., 2006) methods. Many of these earlier studies (Marcu and Echihabi, 2002; Lin et al., 2009; Pitler et al., 2009; Wang et al., 2012; Lan et al., 2013; Biran and McKeown, 2013; Rutherford and Xue, 2014) focused on using word pairs or their derivative features. For example, take the two following sentences:

$A1$: *I like summer.*
$B1$: *I prefer winter.*

In this case, we can easily identify the relation as "comparison" by focusing on the word pair "*summer - winter*". However, there is emerging evidence that word pairs may no longer have a role to play in implicit discourse relation recognition (Park and Cardie, 2012). This is because identification is not always possible by using just word pairs. When we consider the following sentences,

$A2$: *I got soaked by the sudden rain yesterday.*
$B2_1$: *Did you forget your umbrella at the office?*
$B2_2$: *The rain was so heavy that my umbrella was useless.*

discourse $A2 - B2_1$ and $A2 - B2_2$ have different relations. discourse $A2 - B2_1$ is causal relation: $B2_1$ explains the reason for $A1$, and $A2 - B2_2$ is expansion relation: $B2_2$ is a supplemental explanation about the "*sudden rain*" in $A2$. Nevertheless, the same word pair "*soaked - umbrella*" can be ex-

29th Pacific Asia Conference on Language, Information and Computation pages 63 - 72
Shanghai, China, October 30 - November 1, 2015

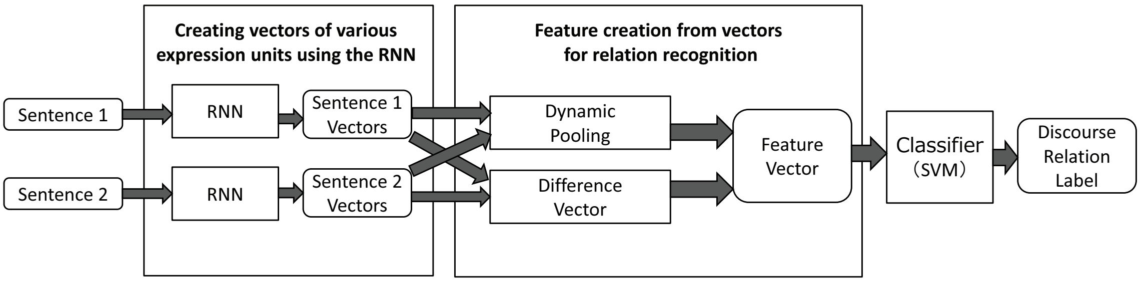

Figure 1: Overview of proposed discourse relation recognition.

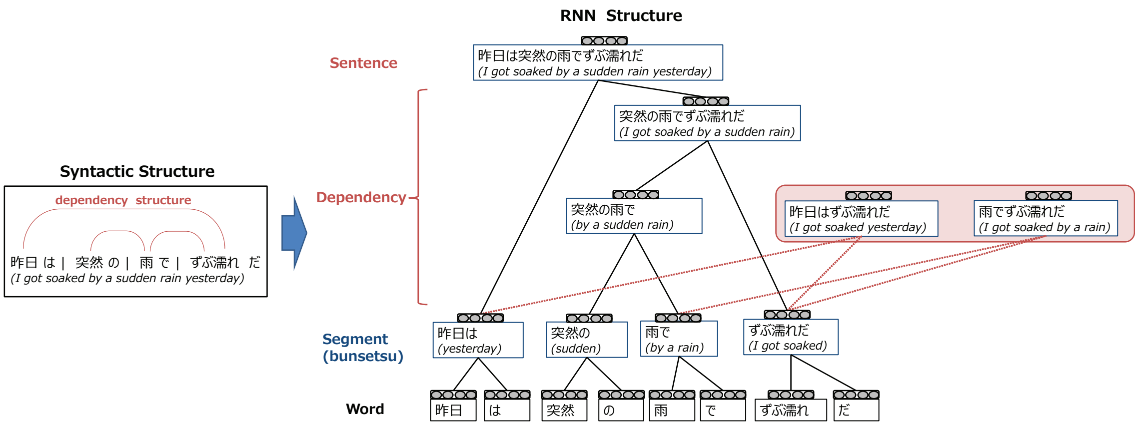

Figure 2: RNN structure in Japanese dependency structure.

tracted for both cases, making little contribution to relation recognition. If we can use pairs of longer expressions such as "*I got soaked* - *forget your umbrella*" and "*I got soaked by the sudden rain* - *so heavy that my umbrella was useless* ", it will be easier to perform relation recognition because the units employed are more specific and distinguishing of discourse relations.

This paper proposes a novel method for implicit discourse relation recognition that compares various expression units between two sentences. The smallest units of a sentence expression are words, and the largest are the entire sentence. To consider various expression units, we turn to a recursive neural network (RNN) based approach. The RNN is the neural network based method to create vectors of various expression units on the basis of the syntactic structure of a sentence and has been applied to various NLP tasks (Socher et al., 2011; Li et al., 2014; Liu et al., 2014). Here, we employ the RNN based approach for implicit discourse recognition and show that our proposed method significantly outperforms the word pair based approach.

In this paper, we demonstrate through experiments using Japanese conversational data that our method can improve the estimation performance of implicit discourse relation recognition more than the conventional word pair method. In the following sections, we first describe our proposed method using the RNN with Japanese sentences in Section 2. Section 3 explains the experiments we performed on implicit discourse recognition in Japanese dialogue, and we discuss the results in Section 4. Finally, we conclude in Section 5.

2 Discourse relation recognition by comparing various units of sentence expressions

Figure 1 shows an overview of the proposed method using various units of expressions in a sentence to identify implicit discourse relations. First, we input sentences to the RNN. The RNN then creates vectors of various expression units on the basis of the input syntactic structures in a bottom-up fashion. Next, we create a feature vector by comparing vectors of various units of expression. The discourse relation is identified by a discriminative classifier such as a support vector machine (SVM). In this section, we explain how the RNN works, describe how the vectors are created by the RNN, and show how to create the feature for the classifier from vectors.

2.1 Recursive neural network

The RNN is a kind of deep neural network created by applying the same set of weights recursively over a structure. The RNN has a binary tree structure, and its framework computes the representation for each parent iteratively in a bottom-up fashion on the basis of its children. We assume that word vectors c_1, c_2, and c_3 have N dimensions. Each word is given vectors in advance by word embeddings (e.g., *word2vec* (Mikolov et al., 2013a)). Segment vectors are created by combining word vectors from left to right in each segment. The c_1 and c_2's parent representation vector p_1 is computed as

$$p_1 = f(W_e[c_1; c_2] + b_e) \tag{1}$$

where $[c_1; c_2]$ is the $2N$-dimension concatenation vector of c_1 and c_2, W_e is the $N \times 2N$ encoding matrix, b_e is the N-dimension encode bias vector, and f denotes an element-wise activation function (we use tanh). The next parent representation vector p_2, which has children p_1 and c_3, is computed in the same way by an input concatenation vector $[p_1; c_3]$ and encoding parameters W_e and b_e.

2.2 Creating vectors of various expression units using the RNN

The RNN creates vectors of various expression units during the process of creating a sentence vector. Our approach compares the meaning of two sentences by using these interim vectors. In this subsection, we introduce a method for extracting vectors of various expression units by the RNN for Japanese sentences.

Figure 2 shows the RNN structure based on Japanese dependency structure. Japanese sentences have dependency structures made up of bunsetsu segments (bunsetsu is a Japanese expression unit comprising one or more content words with zero or more function words). We obtain the syntactic structures of sentences by Japanese dependency parsing. Refer to (Kudo and Matsumoto, 2003) for how Japanese dependency parsing works in general.

We create segment vectors by combining word vectors. The sentence vector is the root vector of the RNN created at the end of the combining process. In this paper, we construct an RNN tree structure on top of the Japanese dependency structure. In Japanese, dependency relationships are generally directed from left to right, so we constantly combine segment vectors from the right-most segment to obtain the segment vector, as in the example shown in Fig. 2.

Because Japanese dependency structures are not a binary tree, there are some vectors that are not used in the process of creating the sentence vector. For example, the vectors of the expressions "*I got soaked yesterday*" and "*I got soaked by rain*" are not created in the process of creating the sentence vector in Fig. 2. Since these vectors have an independent meaning and can be useful, in our proposed method, we use all the vectors (including ones that do not lead to the sentence vector) in the RNN structure for discourse relation recognition as we describe in the following section.

2.3 Feature creation from vectors for discourse relation recognition

If sentences 1 and 2 have n and m vectors, respectively, we have to create a feature vector considering $n \times m$ patterns. However, the feature vector for the classifier must be fixed-length although the number of vectors extracted from a sentence changes dynamically depending on the number of words and on the syntactic structure. Therefore, we need to create a fixed-length feature vector without dependence on the number of vectors. The simplest approach to do this is to use a concatenation of sentence vectors as the feature vector. However, this way does not allow us to directly compare the meaning of intermediate expression units. Here, we create fixed-length feature vectors by dynamic pooling and difference vectors as follows:

Dynamic Pooling

Dynamic Pooling (DP) (Socher et al., 2011) is a method to create fixed-length features using the similarity between two vectors (Fig. 3). First, we create a similarity matrix between the vectors within the two sentences. The similarity between two vectors is computed with cosine

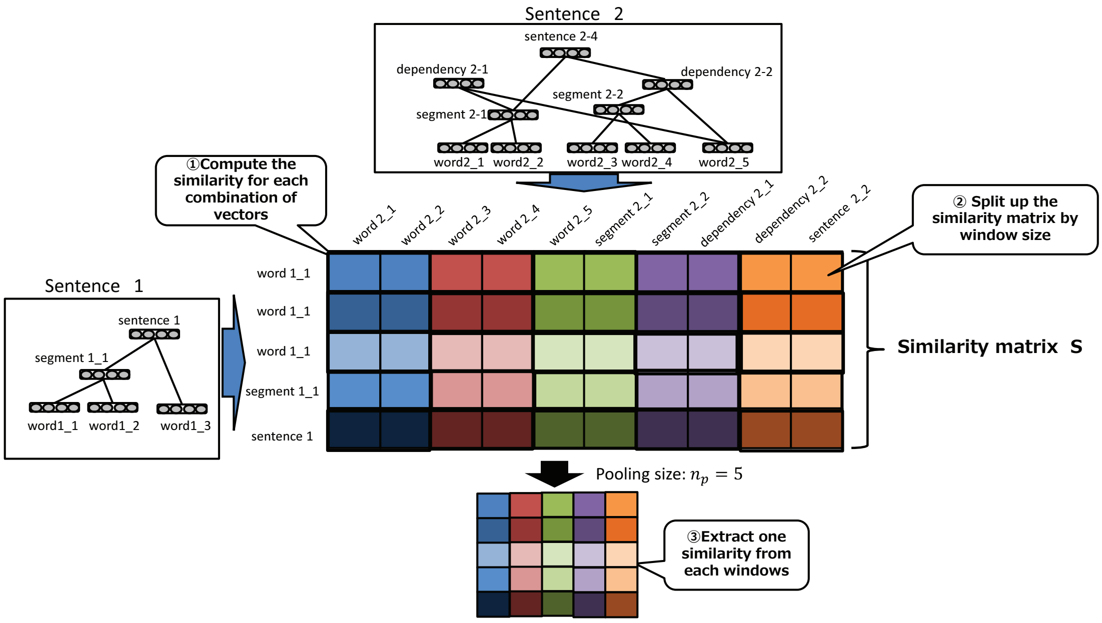

Figure 3: Overview of Dynamic Pooling.

similarity, as follows:

$$sim(v_1, v_2) = \frac{v_1 \cdot v_2}{|v_1||v_2|} \quad (2)$$

where v_1 and v_2 denote vectors extracted from sentences 1 and 2, respectively. The row and column order of the matrix is placed depth-first in the RNN tree, right-to-left. Specifically, matrix element s_{00}, which is the first element of similarity matrix S, is the degree of similarity between the left-most word vectors in each sentence.

In DP, the similarity matrix is split up into a sub-matrix by a grid window. The size of the grid window is computed depending on pooling size n_p. If sentences 1 and 2 have N and M vectors, respectively, the grid window size is $[\frac{N}{n_p}] \times [\frac{M}{n_p}]$. We extract a maximum similarity value element in each sub-matrix to create a pooled matrix. This pooled matrix is consistently fixed-length because the grid window size dynamically varies depending on sentence length. Similarity information between two sentences is consolidated into a fixed-length feature by the DP.

Difference vectors

Recent studies of word embeddings such as word2vec (Mikolov et al., 2013b) have revealed that difference vectors are meaningful. In the well-known word2vec example, the vector operation "*king* - *man* + *woman* = *queen*" holds. That is, the difference vector "*king* - *man* " represents the information of kingship. Following this insight, we use the difference vector in the hope that it can capture some relations between sentences. The difference vector is computed by subtracting two vectors, v_1 and v_2,

$$diff(v_1, v_2) = \frac{v_1 - v_2}{|v_1 - v_2|} \quad (3)$$

where vectors v_1 and v_2 denote vectors created by the RNN. In this paper, we utilize the mean vector of all difference vectors created by a combination of all the vectors (i.e., vectors that correspond to all the cells in the matrix S in Fig. 3) of two sentences as a feature vector.

3 Experiment

We performed experiments using a Japanese conversational corpus. First, we explain the dataset used

```
<s line = "1" speaker= "A">
  普段はお酒を飲まれますか？  ( Do you drink alcoholic beverages in your daily life? )
</s>

<s line = "2" speaker = "B">
  <connective category="Implicit" class="EXPANSION"  type="Instantiation"  rel="1" marker="たとえば(For example)" />
  スミノフアイスをよく飲みます。 ( I often drink Smirnoff-ice. )
</s>

<s line = "3" speaker = "A">
 <connective category="Implicit"  class="CONTINGENCY"  type="Cause"  rel="2" marker ="なぜなら(Because)" />
  スッキリしますよね。(It has a refreshing taste. )
</s>

<s line = "4" speaker = "A">
  日本酒は飲みますか？( Do you drink Japanese Sake? )
</s>

<s line = "5" speaker = "B">
 <connective category="Implicit"  class="COMPARISON"  type="Contrast"  rel="2" marker ="でも (But)" />
  日本酒はあまり飲みません。( I rarely drink Japanese Sake. )
</s>

<s line = "6" speaker = "B">
  独特の味がする (its taste is so unique.)
   <connective category="Explicit"  class="CONTINGENCY"  type="Cause"  rel="5" >
     ので。(Because )
   </connective>
</s>
```

Figure 4: Discourse relation corpus from Japanese dialogue.

Utterance 1	Utterance 2	Relation	Connective
普段はお酒を飲まれますか？ (*Do you drink alcoholic beverages in your daily life?*))	スミノフアイスをよく飲みます。(*I often drink Smirnoff Ice.*)	Implicit EXPANSION Instantiation	例えば (*For example*)
スミノフアイスをよく飲みます。(*I often drink Smirnoff Ice.*)	スッキリしますよね。(*It has a refreshing taste.*)	Implicit CONTINGENCY Cause	なぜなら (*Because*)
スミノフアイスをよく飲みます。(*I often drink Smirnoff Ice.*)	日本酒はあまり飲みません。(*I rarely drink Japanese sake.*)	Implicit COMPARISON Contrast	でも (*But*)
日本酒はあまり飲みません。(*I rarely drink Japanese sake.*)	独特の味がするので。(*Because I think its taste is so unique.*)	Explicit CONTINGENCY Cause	ので (*Because*)

Table 1: Examples of utterance pairs and discourse relations extracted from Fig. 4.

for the experiment. Next, we describe our experimental methodology and comparative methods. Finally, we present the experimental results.

3.1 Dataset

In this paper, we focus on conversational dialogue because we want sophistication of dialogue analysis by using discourse relations.

The annotation framework follows the Penn Discourse Treebank (PDTB), a corpus of English texts from the Wall Street Journal in which the relations between abstract objects in discourse are annotated (Prasad et al., 2008). The PDTB has four classes (CONTINGENCY, COMPARISON, EXPANSION, and TEMPORAL) and 16 types of discourse relation within its hierarchical structure. In the PDTB, the discourse relations are decided with connectives: "*because*" , "*and*", "*but*", and so on. If a discourse marker (e.g., a connective) is written clearly in either target sentence, the discourse relation is categorized as *Explicit*. Discourse relations without any discourse marker are called *Implicit*.

We annotated PDTB-style discourse relations to the Japanese conversational dialogue corpus created by Higashinaka et al. (2014). Figure 4 shows the annotated Japanese conversational dialogue corpus. We provide connective tags to each utterance if they have a connective. Connective elements have five attributes: *category*, which denotes discourse relation category and can be either explicit or implicit; *class*, which includes the four discourse relations; *type*, which denotes detailed relation types; *rel*, which denotes an utterance line number that has a discourse relation; and *marker*, which denotes the connective appropriate for discourse relation if the relation is *Implicit*. Table 1 gives a tabular view of the utterance pairs from Fig. 4.

Note that there is another dialogue corpus annotated with PDTB-style discourse relations (Tonelli et al., 2010); however, they focus on the design of the corpus and do not tackle the problem of discourse relation recognition.

3.2 Experimental method and results

We evaluate our proposed approach using the annotated conversational dialogue corpus. We created an implicit discourse relation classifier using an SVM with training data consisting of utterance pairs that have an explicit discourse relation. Explicit relations are more certain than implicit relations, so explicit relational data have been used as training data (Pitler and Nenkova, 2009).

We performed the evaluation by classifying three discourse relations (CONTINGENCY, COMPARISON , and EXPANSION) using classifiers. Here, we do not use the TEMPORAL relation class because far fewer utterance pairs have a relation to TEMPORAL than the other relations. Training data consisted of 5,000 utterance pairs for each relation. Test data were utterance pairs that have an implicit discourse relation, with each relation containing 500 utterance pairs by random sampling.

We evaluated our proposed method along with several comparative methods. All the methods derive features for two sentences to be classified by the SVM. The features used by the methods are described as below.

- Comparative methods

  Word pair

  The word pair feature is a basic feature for discourse recognition. Input sentences are split into words by a morphological analyzer MeCab[1] (we used this analyzer throughout the paper). We create word pair tokens from the combination of words between two sentences. Finally, the word pair feature is created by creating word-pair appearance frequency vectors.

  Vector centroid

  We create a sentence vector by computing the centroid of all word vectors in the sentence and use the vector as a feature. Here, word vectors are given by the word2vec model created using Japanese Wikipedia data. Note that the word centroid vector reflects the whole meaning of the sentence without syntactic structure or word order.

  RNN sentence

  The RNN sentence feature is the root node vector of the RNN structure. Parameters of the RNN are trained with data consisting of 100,000 utterances from the aforementioned dialogue corpus. The sentence vector differs from the word centroid vector in that it includes the information of syntactic structure.

- Proposed methods

  RNN + DP

  The RNN+DP feature is a concatenation vector with the RNN sentence vector and Dynamic Pooling vector (window size: 5).

  RNN + DP + diff

  The RNN+DP+diff feature is a concatenation vector with the RNN sentence vector, Dynamic Pooling, and a difference vector.

Figure 5 shows the results of the overall classification accuracy and McNemar's testing, and Table 2 shows the implicit discourse classification performance for each discourse relation by using precision, recall, and F-score. As can be seen in Fig. 5, our proposed method **(RNN + DP + diff)** had the

[1] `http://taku910.github.io/mecab/`

	CONTINGENCY			COMPARISON			EXPANSION		
	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score
Word pair	0.38	**0.60**	**0.46**	0.41	0.32	0.36	0.37	0.22	0.28
Vector centroid	0.39	0.38	0.38	0.40	**0.40**	**0.40**	**0.41**	0.43	0.42
RNN sentence	0.42	0.26	0.32	0.47	0.26	0.34	0.36	**0.66**	0.47
RNN + DP	0.41	0.41	0.41	0.46	0.29	0.36	0.39	0.53	0.45
RNN + DP + diff	**0.45**	0.41	0.43	**0.48**	0.30	0.37	**0.41**	0.60	**0.49**

Table 2: Implicit discourse classification scores.

Utterance 1	Utterance 2	Predicted relation
Example of correct classification by all methods		
私もスキー得意です！(*I'm good at skiing too!*)	ボードは難しくて (*Snowboarding is hard for me.*)	COMPARISON
好きな番組のジャンルありますか (*What type of TV programs do you like?*)	バラエティですかね (*I like variety shows.*)	EXPANSION
Example of correct classification by **RNN + DP + diff**		
昨日遊園地に行きました (*I went to an amusement park yesterday.*)	好きなバンドのライブがそこであったんです (*My favorite band performed played a live show there.*)	CONTINGENCY
メイクの仕方ってどこで学ぶんでしょね？(*Where do you learn your makeup techniques?*)	私は雑誌を見ながらですね (*I learn them by reading magazines.*)	EXPANSION

Table 3: Examples of discourse relation recognition between two utterances.

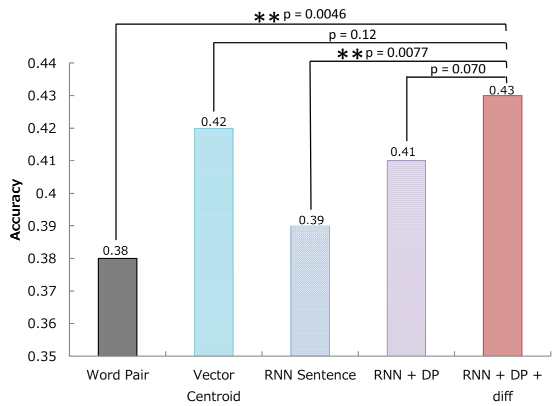

Figure 5: Comparison of classification accuracy.

highest accuracy ($accuracy = 0.43$), and the results of McNemar's testing reveal a significant difference between the **(Word pair)** and **(RNN + DP + diff)** methods ($p = 0.0046$, $p < 0.001$) and between the **(RNN sentence)** and **(RNN + DP + diff)** methods ($p = 0.0077$, $p < 0.001$). In contrast, the difference between the **(Vector centroid)** and **(RNN + DP + diff)** methods was only marginally significant ($p = 0.12$).

The accuracy of the baseline method **Word pair** (0.38) is very close to that of pure chance (0.33). We separately checked the inter-annotator agreement of discourse relation relation annotation and found that the accuracy of human (taking another annotator's annotation as gold standard) is 0.67. If the upper bound is 0.67, then our proposed method (0.43) achieves 64% accuracy relative to human performance, which is a lot higher than 57% accuracy (0.38) of **Word pair**, showing our contribution to implicit discourse relation recognition.

We show examples of the discourse relation recognition results between two Japanese utterances in Table 3. The upper two examples show utterance pairs that were classified correctly by all methods, while the two examples at the bottom were correctly classified by only the **(RNN + DP + diff)** method.

4 Discussion

The accuracy and McNemar's testing results indicate that our proposed approach **(RNN + DP + diff)** outperformed the word-pair and sentence vector approach, demonstrating that our approach, with its use of various units of expression, is more effective than the approach based on word pair and sentences.

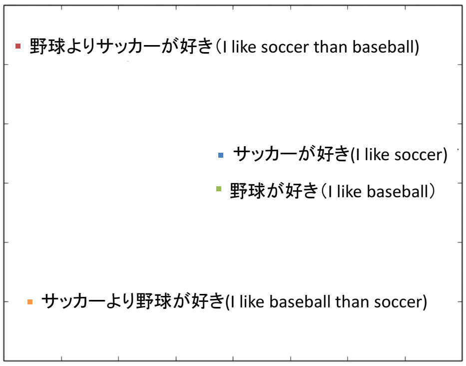

Figure 6: Visualization of RNN sentence vectors.

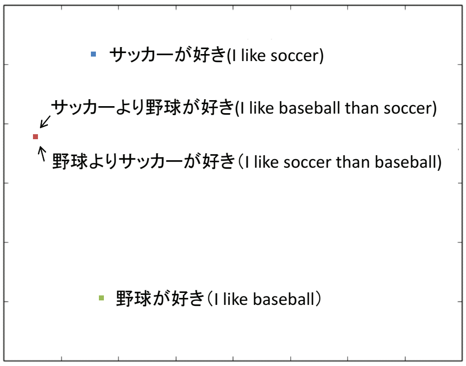

Figure 7: Visualization of word centroid vectors. Note that the vector "*I like soccer more than baseball*" overlaps with the vector "*I like baseball more than soccer.*".

In the example in Table 3, the inputs classified correctly by all methods were identified by extracting the characteristic content words from each utterance. For example, in the first example, the relation is identified as COMPARISON by extracting the pair "*skiing* - *snowboarding*". In contrast, in the last example, while the relation is difficult to identify as EXPANSION by extracting the pairs "*makeup* - *magazine*" or "*makeup* - *learn*", we can identify the relation by extracting the expression pairs "*your makeup techniques* - *by reading a magazine*". By taking advantage of the various units of expression in a sentence, our approach appropriately identifies the discourse relation between two sentences.

Our experimental results show that the RNN vectors are not always superior to word centroid vectors because there are cases where it is not necessary to consider syntax. Sometimes, word pairs are better suited for obtaining the generic topic of a sentence. However, we also found that implicit discourse relation recognition requires to detect slight differences in expressions in sentences. For example, Figs. 6 and 7 compare RNN vectors and word-centroid vectors in the visualization of vector space. The sentences "*I like baseball more than soccer.*" and "*I like soccer more than baseball.*" are in different places in Fig. 6. If the first sentence is "*I like soccer.*" and the second sentence is "*I like soccer more than baseball.*", the discourse relation between two sentences is EXPANSION (*I like soccer. Moreover, I like soccer more than baseball.*). However, if the second sentence is "*I like baseball more than soccer.*", the most appropriate discourse relation is COMPARISON (*I like soccer. But I like baseball more than soccer.*). The RNN vectors are able to capture these different structures, enabling our proposed method to recognize discourse relations more precisely.

5 Conclusion

We proposed an implicit discourse relation detection method using various units of expressions between two sentences. All expressions are converted into vectors by the RNN and then applied to Japanese dependency structures. Experimental results showed that our approach performs better than the conventional word-pair features method. This paper is the first to show that various expression units in sentences are effective for implicit discourse relation recognition.

Our future work is to enable more feature selection using intermediate expression vectors and to consider applications for dialogue systems. Current dialogue systems have problems that they choose a contextually inappropriate utterance for the user input. Since two utterances with a discourse relation can be coherent, we expect the quality of utterance selection to be increased by selecting an utterance that has a discourse relation with the user utterance.

References

Or Biran and Kathleen McKeown. 2013. Aggregated word pair features for implicit discourse relation disambiguation. *Proc of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pages 69–73.

Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng, and Bita Nejat. 2014. Abstractive summarization of product reviews using discourse structure. *Proc of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 1602–1613.

Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. 2014. Towards an Open Domain Conversational System Fully Based on Natural Language Processing. *Proc of the 25th International Conference on Computational Linguistics (COLING 2014)*, pages 928–939.

Taku Kudo and Yuji Matsumoto. 2003. Fast methods for kernel-based text analysis. *Proc of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, pages 24–31.

Man Lan, Yu Xu, and Zhengyu Niu. 2013. Leveraging Synthetic Discourse Data via Multi-task Learning for Implicit Discourse Relation Recognition. *Proc of the 51st Annual Meeting of the Association for Computational Linguistics(ACL 2013)*, pages 476–485.

Jiwei Li, Rumeng Li, and Eduard Hovy. 2014. Recursive deep models for discourse parsing. *Proc of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 2061–2069.

Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. 2009. Recognizing Implicit Discourse Relations in the Penn Discourse Treebank. *Proc of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)*, pages 343–351.

Shujie Liu, Nan Yang, Mu Li, and Ming Zhou. 2014. A recursive recurrent neural network for statistical machine translation. *Proc of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)*, pages 1491–1500.

Daniel Marcu and Abdessamad Echihabi. 2002. An Unsupervised Approach to Recognizing Discourse Relations. *Proc of the 40th Annual Meeting on Association for Computational Linguistics (ACL 2002)*, pages 368–375.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient Estimation of Word Representations in Vector Space. *Proc of the Workshop at International Conference on Learning Representations (ICLR 2013)*.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013b. Linguistic regularities in continuous space word representations. *Proc of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013)*, pages 746–751.

Joonsuk Park and Claire Cardie. 2012. Improving implicit discourse relation recognition through feature set optimization. *Proc of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2012)*, pages 108–112.

Emily Pitler and Ani Nenkova. 2009. Using Syntax to Disambiguate Explicit Discourse Connectives in Text. *Proc of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP Short Papers (ACL-IJCNLP 2009)*, pages 13–16.

Emily Pitler, Annie Louis, and Ani Nenkova. 2009. Automatic Sense Prediction for Implicit Discourse Relations in Text. *Proc of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Asian Federation of Natural Language Processing (AFNLP 2009)*, pages 683–691.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. *Proc of the sixth international conference on Language Resources and Evaluation (LREC 2008)*.

Attapol Rutherford and Nianwen Xue. 2014. Discovering implicit discourse relations through brown cluster pair representation and coreference patterns. *Proc of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, pages 645–654, April.

Manami Saito, Kazuhide Yamamoto, and Satoshi Sekine. 2006. Using phrasal patterns to identify discourse relations. *Proc of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2006)*, pages 133–136.

Richard Socher, Eric H. Huang, Jeffrey Pennin, Christopher D Manning, and Andrew Y. Ng. 2011. Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection. *Proc of Advances in Neural Information Processing Systems (NIPS 2011)*, pages 801–809.

Jun Sugiura, Naoya Inoue, and Kentaro Inui. 2013. Recognizing Implicit Discourse Relations through Abductive Reasoning with Large-scale Lexical Knowledge. *Proc of the 1st Workshop on Natural Language and Automated Reasoning (NLPAR 2013)*, pages 76–87.

Sara Tonelli, Giuseppe Riccardi, Rashmi Prasad, and Aravind Joshi. 2010. Annotation of Discourse Relations for Conversational Spoken Dialogs. *Proc of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, pages 19–21.

Xun Wang, Sujian Li, Jiwei Li, and Wenjie Li. 2012. Implicit discourse relation recognition by selecting typical training examples. *Proc of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 2757–2772.