Identification of Sympathy in Free Conversation

Tomotaka Fukuoka and Kiyoaki Shirai

Japan Advanced Institute of Science and Technology 1-1, Asahidai, Nomi, Ishikawa, 923-1292, Japan {s1320010, kshirai}@jaist.ac.jp

Abstract

Dialog systems are generally categorized into two types: task oriented and non task oriented systems. Recently, the study of non task oriented dialog systems or chat systems becomes more important since robotic pets or nursing care robots are paid much attention in our daily life. In this paper, as a fundamental technique in a chat system, we propose a method to identify if a speaker displays sympathy in his/her utterance. Our method is based on supervised machine learning. New features are proposed to train a classifier for identifying the sympathy in user's utterance. Results of our experiments show that the proposed features improve the F-measure by 3-4% over a baseline.

1 Introduction

Dialog systems could be broadly divided into two categories. One is a task oriented dialog system. It focuses on a specific task such as guidance on sightseeing, hotel reservation or promotion of products, and communicates with a user to achieve a goal of the task. The other is a non task oriented dialog system or chat system. It does not suppose any specific tasks but can handle a wide variety of topics to freely chat with the user. Most of the past researches focus on task oriented dialog systems become more important since robotic pets or nursing care robots are paid much attention (Libin and Libin, 2004).

One of important characteristics in free conversation is sympathy of a speaker for the topics in

the conversation (Anderson and Keltner, 2002; Higashinaka et al., 2008). The topics in free conversation are not fixed but could be changed by the speakers at any time. To make the conversation natural and smooth, however, a non task oriented dialog system can not arbitrarily change the topics. It is uncomfortable for the user if the system would suddenly change the topic when the user wants to continue to talk on the current topic, or if the system would keep the same topic when the user is bored and does not want to talk on the topic any more. If the system fails to shift the topic at appropriate time, the user may break the conversation. The sympathy of the user is one of the useful clues to guess good timing for changing the topic. If the user shows the sympathy for the current topic, the system should continue the conversation with the same topic. On the other hand, if the user does not display the sympathy, the system should provide other topics. Therefore, it is essential for the chat system to guess the sympathy of the user.

This paper proposes a method to automatically judge whether the user displays the sympathy in his/her utterance as a fundamental technique in a non task oriented dialog system. In this paper, we define 'sympathetic utterance' as the utterance where the speaker expresses the sympathy or approval especially when he/she replies to subjective utterance of the other participant. Note that the utterance just showing agreement is not defined as sympathetic. Various kinds of clues could be applicable for identification of the sympathy, such as facial expressions, gesture or the contents of the utterance. Since we focus on a text based chat system, our

29th Pacific Asia Conference on Language, Information and Computation: Posters, pages 1 - 9 Shanghai, China, October 30 - November 1, 2015 Copyright 2015 by Tomotaka Fukuoka and Kiyoaki Shirai

1

method only considers the content and detects the user's sympathy in a transcript of the utterance. In addition to ordinary n-gram features, new features for the sympathy identification are introduced. The effectiveness of our proposed features will be proved via empirical evaluation.

The remaining parts of this paper are organized as follows. Section 2 discusses related work for the sympathy identification. Section 3 presents our proposed method. Section 4 reports results of evaluation experiments. Finally, Section 5 concludes the paper.

2 Related work

A considerable number of studies have been made on an automatic tagging of utterance in a dialog corpus. That is, each utterance in the dialog is automatically annotated with some useful information such as dialog acts. Hereafter we call it 'dialog tag'. Supervised machine learning is often used for automatic identification of dialog tags. Since the sympathy of the speaker is also regarded as a kind of dialog tags, we introduce several related work automatically classifying utterance into dialog tags including the sympathy¹.

Xioa et al. (2012) proposed a method to estimate the sympathy speech using the language model learning tool SRILM (Stolcke, 2002). In their method, n-gram of words were used as the features to classify if the utterance indicated the sympathy of the speaker. They reported that bi-gram was the most effective feature and the accuracy of the sympathy identification was around 60%.

A set of 29 dialog acts including 'empathy' was proposed toward an open-ended dialog system (Minami et al., 2012). They performed the automatic recognition of them using a weighted finite-state transducer with the words in the utterance.

Sekino et al. (2010) tried to identify the dialog acts using Conditional Random Fields (CRF). SWBD-DAMSL tag set (Jurafsky et al., 1997) were used as a set of dialog acts. Note that the tag 'sympathy' is included in SWBD-DAMSL. The features used for training CRF were the tag of the previous utterance, the number of content words in the utterance, the length of the utterance and so on.

To identify the dialog acts of the sentences in microblogging, semantic category patterns were introduced as the feature of Support Vector Machine (SVM) classifier (Meguro et al., 2013). The words in the utterance were converted into their semantic categories (or abstract concepts) using a thesaurus, then n-gram of not words but semantic categories is used as the feature. Results of this study showed that n-gram of the semantic categories was more effective than word n-gram.

This study also applies supervised learning for automatic identification of the sympathy. Especially, we investigate what are the useful features to infer the sympathy in the utterance. Therefore, we focus on identification of the sympathy only, although many previous work handled the sympathy as one of the dialog acts. Several studies reported that characteristics of the sympathy could be found in an expression at the end of the utterance (Itoh and Nagata, 2007; Huifang, 2009). In addition, there might be more linguistic features indicating the sympathy of the speaker. The main contribution of the paper is that new features for the sympathy identification are proposed through manual analysis of a free conversation corpus. Furthermore, the effectiveness of these features is empirically evaluated by experiments. Note that the target language in this study is Japanese.

3 Proposed method

Our system accepts a text of utterance in free conversation as an input, then guesses whether it indicates the speaker's sympathy. Support Vector Machine (SVM) (Chih-Chung and Chih-Jen, 2001) is applied to train a binary classifier to judge if the given utterance is sympathetic 2 .

3.1 Feature

We design the following 9 features for sympathy identification. Note that all features are binary, that is, the weight in the feature vector is 1 if it is present in the utterance, 0 otherwise.

F_{ng} : Word n-gram

¹Since we focus on the methods that handle Japanese utterance, some of the related papers are written in Japanese.

²Memory-based learning (TiMBL) (Daelemanset al., 2010) is also applied in our preliminary experiment, but SVM slightly outperformed TiMBL.

The word n-gram (n=1,2,3) is used as the feature, since it represents the content of the utterance. This is the basic feature widely used for identification of the dialog tags in the previous work. Since the content of the previous utterance is also important, we use the word n-gram of both the current and previous utterance.

*Flen***:** Length of utterance

Since the sympathetic utterance tends to be short, the length of the utterance (the number of characters) is considered. In the simple approach, the length feature is defined according to intervals, such as ' $1 \sim 5$ ', ' $6 \sim 10$ ' and 'more than 10'. However, it is rather difficult to determine the optimum intervals. In this study, the length features are defined as in (1) and (2)

$$f_{len}^{(i)}$$
: if l_u is in $[i-2, i+2]$ (1)

$$f_{len}^{(long)}: \quad \text{if } l_u \ge 20 \tag{2}$$

, where l_u stands for the length of the utterance. We use 17 length features $f_{len}^{(i)} \ (3 \le i \le 19)$ as well as an extra feature $f_{len}^{(long)}$ indicating the utterance is long. This approach enables us to incorporate the information of the length of utterance into SVM more flexibly.

F_{tu}: Turn taking

In our conversation corpus, the speakers may give two or more utterance in one turn. This feature indicates the presence of turn taking, i.e. whether the speaker of the current and previous utterance is the same.

F_{rw1} : Repetition of word (1)

The speakers often show their sympathy by repeating a word in a previous utterance of the other. For example, in the simple conversation below ³, the speaker B repeats the word '傑作 (fine work)' to agree with A's comment.

- A: あの/ 映画 /は/ 傑作 / だ (that) (movie) (fine work) (be) (That movie is a fine work.)
- B: 傑作 / だ /ね (fine work) (be) (It is a fine work.)

We introduce a feature indicating if the same word appears in the current and previous utterance.

F_{rw2} : Repetition of word (2)

Repetition of the words does not always indicates the sympathy. Let us consider the following example.

- A: 海草類 / 嫌い /な/の/? (seaweed) (dislike) (Do you dislike seaweed?) B: そう/で/も/ない/よ/、/ 海草
- (so) (not) (seaweed.)

The speaker B repeats the word '海草 (seaweed)', but his/her utterance does not show the sympathy.

This feature is similar to F_{rw1} , but more strictly checks the presence of repetition of the content words. The feature F_{rw2} is activated if either condition below is fulfilled:

- The last predicative word in the previous utterance is also found in the current utterance.
- There is only one content word in the current utterance and it also appears in the previous utterance.

F_{rc1} : Repetition of semantic class (1)

Repetition of not words but semantic classes is considered in this feature. In the following example, no content word is overlapped in two utterance, but the speaker B express his/her sympathy by saying '楽しかっ(fun)' whose meaning is similar to '面白かっ(interesting)' in the speaker A's utterance.

A: あの/ 映画 /は/ 面白かっ /た (that) (movie) (interesting) (That movie was interesting.)

This feature is activated if the same semantic class appears in both current and previous utterance. A Japanese thesaurus 'Bunruigoihyo' (National Institute for Japanese Language

³Note that '/' stands for the word segmentation, and a word or a sentence in parentheses is an English translation. The words without translations are function words that have no meaning.

and Linguistics, 2004) is used to obtain the semantic classes of the words. If one word has two or more semantic classes in the thesaurus, all of them are used to check repetition in two utterance. That is, we build the lists of all possible semantic classes of all content words in the current and previous utterance, and check if there is an overlap between them.

F_{rc2} : Repetition of semantic class (2)

Similar to F_{rw2} , repetition of the semantic classes are strictly checked as follows:

- The semantic class of the last predicative word in the previous utterance is also found in the current utterance.
- There is only one content word in the current utterance and its semantic class also appears in the previous utterance.

*F*_{da}: Dialog act

Dialog act is also a useful feature to identify the sympathy. When we hear the other's assertion or opinion, we sometimes show our sympathy with it. However, we seldom express the sympathy for a simple yes-no question. In this study, we define a set of dialog acts in free conversation as in Figure 1.



Figure 1: Dialog act

We manually annotate the conversation corpus with the dialog acts and use them as the features. In future, the dialog acts should be automatically identified to derive this feature.

Fend: End of sentence

The speakers often show their sympathy in an expression at the end of their utterance. For example, in Japanese, " $\mathcal{E}[da] / \mathcal{I}_{\alpha}[ne]$ " or " $\mathcal{L}[yo] / \mathcal{I}_{\alpha}[ne]$ "⁴ at the end of the sentence strongly indicates the sympathetic mood of the

speaker. Based on the above observation, the expression at the end is introduced as the feature. In this paper, it is represented by a sequence of function words at the end of each sentence in the utterance.

3.2 Combination features

In the preliminary experiment, we investigated several types of kernels of the SVM classifier: linear kernel, polynomial kernel, radial basis function and so on ⁵. We found that the kernels except for the linear kernel performed very poorly on our data set. Therefore, we chose the linear kernel. However, the individual features are regarded as independent each other in the SVM with the liner kernel, although the dependency between the features should be considered.

To tackle this problem, we introduce a feature composed by combination of the existing features. When a feature set $F = \{\dots, f_i \dots\}$ is derived from one utterance, where f_i is one of the features described in Subsection 3.1, all possible pairs of features $[f_i, f_j]$ $(i \neq j)$ are also added to the feature set. Hereafter, $[f_i, f_j]$ is referred to as a combination feature. The combination features enable the classifier to consider the dependency between two features. Since the number of this feature are increased combinatorially, feature selection is applied as described in the next subsection.

3.3 Feature selection

A simple feature selection procedure is introduced. We apply the feature selection only for the word n-gram feature (F_{ng}) and the combination feature, since the numbers of these features are extremely high.

The correlation between a sympathy class and a feature f_i is measured by χ^2 value. The features are discarded when χ^2 value is less than a threshold. We denote the threshold of χ^2 value for the n-gram and combination feature as T_{ng} and T_{comb} , respectively. In the experiment in Section 4, these thresholds will be optimized with a development data.

3.4 Filtering of negative samples

In supervised machine learning, it is inappropriate that the numbers of positive and negative samples in

⁴Parentheses show pronunciation of each word. '/' stands for word segmentation. Note that these words are particles and have no meaning.

⁵We used LIBSVM (Chih-Chung and Chih-Jen, 2001).

the training data are extremely imbalanced, since the trained classifier may display strong bias for the majority class. In general, however, the sympathetic utterance does not frequently appear in free conversation. Actually, the ratio of the sympathetic utterance is 1.1% in our conversation corpus as will be shown in Table 1. To tackle this problem, a filtering process to remove the negative samples is introduced to correct imbalance of the training data.

The basic idea of our filtering method is that we try to remove redundant negative samples. Here 'redundant' sample stands for a sample that is similar to other samples in the training data. Similar negative samples might be redundant and could be removed from the training data without any significant loss of the classification performance. The similarity between two samples (utterance) is measured by cosine similarity of the vector consisting of the word n-gram feature only.

It is time consuming to calculate the similarity between all possible pairs of the utterance in the training data. Instead, we reduce the computational cost by constructing clusters of the utterance as the prepossessing. First, the clusters are constructed from the set of the negative samples. A fast clustering algorithm 'Repeated Bisections' is used, where the number of the cluster is set to 1000^6 .

For each cluster, the redundant negative samples are detected by Algorithm 1. Given a set of utterance in a cluster U, the algorithm divides the utterance into a set of non-redundant utterance U_k to be kept and redundant utterance U_d to be deleted. For each utterance u_i , if the similarity between u_i and the rest of the utterance u_j is greater than the threshold S_{fil} , u_j is added to the set U_d . Then u_i is added to U_k . Intuitively, if several similar utterance are found, only the first appeared one is remained in the training data. Note that the threshold S_{fil} controls the number of the removed negative samples. It is optimized on a development data.

4 Evaluation

This section reports experiments to evaluate our proposed method. In this experiment, the systems are evaluated and compared by the precision, recall and

Algorithm 1: Search for redundant negative samples

F-measure of the identification of sympathetic utterance.

4.1 Data

Meidai conversation corpus ⁷ is used to train and evaluate our proposed method. It is a collection of transcription of actual conversation or chat in Japanese. Two to four participants joined free conversation. Dialogs where the number of the participants is two are chosen from the corpus, then each utterance is manually annotated with 'sympathy tag' indicating whether it expresses the sympathy of the speaker or not⁸.

We randomly divide the conversation corpus into three sets: 80% training, 10% development and 10% test set. Table 1 shows the number of the dialogs, sympathetic utterance (sym) and non-sympathetic utterance (non-sym) in each data. The ratio of the positive and negative samples stands at 1 to 86, that is, the number of sympathetic utterance is much fewer than non-sympathetic. A balanced data in-

⁶We used the clustering tool CLUTO. http://glaros.dtc.umn.edu/gkhome/views/cluto

⁷https://dbms.ninjal.ac.jp/nuc/index.php ?mode=viewnuc

⁸Each dialog in the corpus is annotated by one person. To measure inter-annotator agreement, another annotator put sympathy tags to only three dialogs. Cohen's kappa was 0.27. It indicates the difficulty of the sympathy identification task. In future, the definition of sympathetic utterance should be more clarified to make a better annotation guideline for consistent annotation.

Table 1: Statistics in the conversation corpus

data	dialog	sym	non-sym
training	77	861	73378
development	10	103	8882
test	10	99	8598

cluding the same number of the positive and negative samples is also used for evaluation. It is made by keeping all positive samples and randomly choosing the equal number of the negative samples in the training, development and test data. We repeat to construct the balanced data five times, and evaluate the systems in these five data sets. Note that the results on the balanced data shown below are the average precision, recall and F-measure of five trials.

4.2 Results and discussion

4.2.1 Parameter optimization

First, the parameter T_{ng} for selection of n-gram feature was optimized on the development data. Figure 2 shows a change in precision(P), recall(R) and F-measure(F) on the development data. We chose $T_{ng} = 0.9$ as the best parameter where the precision, recall and F-measure were the highest. In this case, 4378 features, which are 1% of all n-gram features, were selected.



Figure 2: Optimization of T_{ng}

Another parameters T_{comb} and S_{fil} were also optimized. The details will be reported later.

4.2.2 Results

We define the baseline as the classifier with the word n-gram feature only. Table 2 reveals the performance of the baseline and our proposed method on the test data, while Table 3 shows the results on the balanced test data. In these tables, the filtering of the negative samples is not applied. Our proposed method outperformed the baseline on the whole, although the precision was comparable on the balanced data. In the imbalanced test data, the F-measure was not so high. This is because the sympathetic utterance does not frequently appear in the conversation corpus. Since the participants of some dialogs were strangers, they might hesitate to express their sympathy. On the other hand, in the balanced test data, the results were reasonably high. If our method is applied for the conversation between close friends where they frequently show their sympathy, it will achieve better performance than the results in Table 2.

	Table 2:	Results	on t	he	imbal	lanced	test	data
--	----------	---------	------	----	-------	--------	------	------

	Р	R	F			
Baseline (F_{ng}) 0.23 0.11 0.15						
Proposed method 0.28 0.13 0.18						
Table 3: Results on the balanced test data						

	Р	R	F
Baseline (F_{ng})	0.80	0.73	0.76
Proposed method	0.81	0.76	0.80

4.2.3 Effectiveness of features

Next, to investigate the effectiveness of our proposed features, the models with several feature sets are compared. We train the classifiers with the basic word n-gram feature and one of the other features (denoted as $F_{ng} + F_*$), and compared it with the baseline model (F_{ng}). We also compare the classifier with all features (denoted as F_{ALL}). Table 4 and 5 show the results on the imbalanced and balanced test data. Note that the combination features are not used in this experiment.

On the imbalanced test data, adding the feature F_{len} , F_{rc2} , F_{da} and F_{end} caused a decline of the Fmeasure. Furthermore, the classifier using all features were comparable with the baseline. However, on the balanced data, almost all types of the features contributed to gain the F-measure. In addition, precision, recall and F-measure of F_{ALL} were better than the baseline.

From the results in Table 4 and 5, turn taking (F_{tu})

Feature set	Р	R	F
F_{ng}	0.23	0.11	0.15
$F_{ng} + F_{len}$	0.18	0.08	0.11
$F_{ng} + F_{tu}$	0.25	0.12	0.16
$F_{ng} + F_{rw1}$	0.25	0.11	0.15
$F_{ng} + F_{rw2}$	0.26	0.11	0.15
$F_{ng} + F_{rc1}$	0.23	0.11	0.15
$F_{ng} + F_{rc2}$	0.21	0.10	0.14
$F_{ng} + F_{da}$	0.19	0.08	0.11
$F_{ng} + F_{end}$	0.19	0.10	0.13
F_{ALL}	0.24	0.11	0.15

 Table 4: Effectiveness of the features on the imbalanced test data

Table 5: Effectiveness of the features on the balanced test data

Feature set	Р	R	F
F_{ng}	0.80	0.73	0.76
$F_{ng} + F_{len}$	0.81	0.73	0.77
$F_{ng} + F_{tu}$	0.81	0.75	0.78
$F_{ng} + F_{rw1}$	0.81	0.73	0.77
$F_{ng} + F_{rw2}$	0.81	0.73	0.77
$F_{ng} + F_{rc1}$	0.81	0.72	0.76
$F_{ng} + F_{rc2}$	0.81	0.73	0.77
$F_{ng} + F_{da}$	0.81	0.73	0.77
$F_{ng} + F_{end}$	0.82	0.74	0.78
F_{ALL}	0.83	0.77	0.80

and repetition of word (F_{rw1} and F_{rw2}) seem the most effective features. Since the increase or decrease caused by adding one feature is inconsistent for several features on the imbalanced and balanced data, however, the effectiveness of them are rather unclear.

4.2.4 Effectiveness of combination feature

In this subsection, we evaluate the combination feature. Two sets of the features are investigated: the word n-gram feature F_{ng} and all proposed features F_{ALL} . For each feature set, the combination features are added to the feature vector of the utterance.

Recall that we introduce feature selection for the combination feature. The parameter T_{comb} was optimized on the development data. T_{comb} was set as 140 for both feature sets F_{ng} and F_{all} on the imbalanced data. While, it was set as 280 and 260 for F_{ng} and F_{all} on the balanced data, respectively.

Table 6 and 7 compare the classifiers with and without the combination feature in the imbalanced and balanced test data, respectively. In Table 6, the combination feature improves both precision and recall in F_{ALL} feature set. While, combination of the n-gram features increases the precision but decreases recall and F-measure. Therefore, the combination of our proposed features worked well, but the combination of n-gram not.

In the balanced test data (Table 7), the models with and without the combination feature are comparable.

Comparing F_{ng} +COMB and F_{ALL} +COMB in Table 6, incorporation of the proposed features improved the F-measure with a loss of the precision. In the same comparison in Table 7, all three criteria were improved by using the proposed features. Therefore, it can be concluded that our proposed features are effective for identification of the sympathy, especially when the dependency between two features is considered.

Table 6: Evaluation of the combination feature on the imbalanced test data

Feature Set	Р	R	F
F_{ng}	0.23	0.11	0.15
$F_{ng}+COMB$	0.31	0.09	0.14
F_{ALL}	0.24	0.11	0.15
$F_{ALL}+COMB$	0.28	0.13	0.18

Table 7: Evaluation of the combination feature on the balanced test data

Feature Set	Р	R	F
F_{ng}	0.80	0.73	0.76
$F_{ng}+COMB$	0.80	0.73	0.77
F _{ALL}	0.83	0.77	0.80
$F_{ALL}+COMB$	0.81	0.76	0.80

4.2.5 Evaluation of filtering of negative samples

The method of negative sample filtering was evaluated using the imbalanced data set. First, the parameter S_{fil} was optimized as 0.5 that achieved the highest F-measure on the development data.

Three methods are compared in this experiment: a model without the negative sample filtering (w/o Filtering), a model with the filtering by our proposed method (Proposed Filtering) and a model where the negative samples are randomly removed (Random Filtering). In Proposed Filtering, 25,174 negative samples were removed from the training data. In Random Filtering, the same number of the negative samples were randomly removed. We repeated the training of the classifier with random filtering five times and compared the average with the other methods.

Table 8: Evaluation of filtering methods

	Р	R	F
w/o Filtering	0.28	0.13	0.18
Proposed Filtering	0.23	0.16	0.19
Random Filtering	0.25	0.18	0.22

Table 8 reveals the results of three methods. By the filtering, the recall was improved, while the precision declined. It is natural because the classifier tends to judge the utterance as sympathetic (positive) when the number of the negative samples in the training data is reduced. Since F-measure was improved, the filtering of the negative samples seems to contribute toward improvement of the performance. However, our proposed filtering method was worse than the random sampling. It is still uncertain why the idea to remove the redundant negative samples is inappropriate in this task. In future, we will investigate the reason and refine the algorithm of the negative sample filtering.

4.3 Error Analysis

We have conducted an error analysis to find major causes of the errors. First, we found many false positives (the sympathetic utterance is wrongly classified as non-sympathetic) and false negatives (the non-sympathetic utterance is wrongly classified as sympathetic) when the previous utterance was long. In such cases, the previous utterance consisted of many sentences, but only one sentence was usually related to the current utterance. Although many features were derived from the previous utterance, the most of them were irrelevant. Such noisy features might cause the classification error. To overcome this problem, the coherence between the current and previous utterance should be considered. In other words, it is required to introduce a method to choose only the sentence relevant to the current utterance from long previous utterance.

Many errors were also found when both the current and previous utterance were too short. We guessed that the classification errors were caused by the lack of the features. Due to the feature selection, even the word n-gram feature was sometimes not extracted from short utterance. One of the solutions is to apply feature selection only for bi-gram and trigram while remaining all uni-gram features, in order to prevent from extracting no n-gram feature.

We also found that several false negatives were caused by the feature F_{end} . Some of the expressions at the end of the sentence indicate the speaker's sympathy, but not always. Let us suppose such an expression appeared in non-sympathetic utterance and the lengths of both current and previous utterance were short. In such cases, since only a few features were extracted, the end of the sentence feature strongly worked to classify the utterance as the sympathetic. The way to incorporate the end expression into the classifier should be refined.

5 Conclusion

This paper proposed a method to identify the sympathetic utterance in the free conversation. The main contribution of the paper is to propose novel features for sympathy identification. Results of the experiments indicate that (1) the proposed features are effective, especially when the pairs of these features are considered as the additional features, (2) among the proposed features, turn taking and repetition of the content words show strong correlation with the sympathetic utterance, and (3) the filtering of negative samples is important to improve the F-measure.

F-measure of the proposed method was still low in the extremely imbalanced positive and negative sample data. We proposed the filtering method to remove the redundant negative samples, but it was worse than the random filtering. However, since the results on the balanced data were promising, we believe that the filtering of negative samples is a right way to improve the performance. In future, we will continue to explore a better way of negative sample filtering.

References

- Alexander V. Libin and Elena V. Libin 2004. Person-Robot Interactions From the Robopsychologists' Point of View: The Robotic Psychology and Robotherapy Approach. Proceedings of the IEEE 92, pp. 1789– 1803.
- Ryuichiro Higashinaka, Kohji Dohsaka and Hideki Isozaki. 2008. Effects of self-disclosure and empathy in human-computer dialogue. Spoken Language Technology Workshop, pp. 109–112.
- Anderson, C. and Keltner, D. 2002. The role of empathy in the formation and maintenance of social bonds. Behavioral and Brain Sciences 25 (1), pp. 21–22.
- Bo Xiao, Dogan Can, Panayiotis G. Georgiou, David Atkins and Shrikanth S. Narayanan. 2012. Analyzing the Language of Therapist Empathy in Motivational Interview based Psychotherapy. Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference.
- Andreas Stolcke. 2002. SRILM An Extensible Language Modeling Toolkit. Proceedings of International Conference on Spoken Language Processing 2, pp. 901–904.
- Yasuhiro Minami, Ryuichiro Higashinaka, Kohji Dohsaka, Toyomi Meguro, Akira Mori and Eisaku Maeda. 2012. POMDP Dialogue Control Based on Predictive Action Probability Obtained from Dialogue Act Trigram Sequence (in Japanese). The Transactions of the Institute of Electronics, Information and Communication Engineers A, Vol. 95, No. 1, pp. 2–15.
- Takahiro Sekino, Masashi Inoue. 2010. Tagging Extended Conversation Tag to Utterance (in Japanese).Tohoku-Section Convention of Information Processing Society of Japan, 10-6-B3-2.
- D. Jurafsky, E. Shriberg, D. Biasca. 1997. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual Draft 13, University of Colorado, Institute of Cognitive Science, Tech. Rep, pp. 97-102.
- Toyomi Meguro, Ryuichiro Higashinaka, Hiroaki Sugiyama, Yasuhiro Minami 2013. Dialogue act tagging for microblog utterances using semantic category patterns (in Japanese). IPSJ SIG Technical Report, Vol. 2013-SLP-98, No. 1, pp. 1–6.
- Huifang Zhang. 2009. The Semantic Type and Expression Function of YONE in Natural Dialogue (in Japanese). TSUKUBA WORKING PAPERS IN LIN-GUISTICS, No.2, pp. 17–32.
- Masako Itoh and Ryota Nagata. 2009. Rhetorical Function of a Sentence-Final Particle for Constructing Interaction in Discourse (in Japanese). Cognitive studies: bulletin of the Japanese Cognitive Science Society, Vol. 14, No. 3, pp. 282–291.

- Chang, Chih-Chung and Lin, Chih-Jen. C.-C. Chang and C.-J. Lin. 2001. LIBSVM: A library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
- Daelemans, W., Zavrel, J., Van der Sloot, K., and Van den Bosch, A. 2010. TiMBL: Tilburg Memory Based Learner, version 6.3, Reference Guide. Technical Report ILK 03-10, Tilburg University, ILK.
- National Institute for Japanese Language and Linguistics. 2004. Bunruigoihyo. Dainippon tosho.