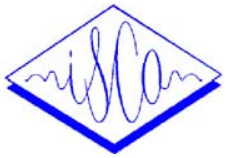
ISCA Archive
http://www.isca-speech.org/archive



INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

Speech Intention Understanding based on Decision Tree Learning

Yuki Irie[†], *Shigeki Matsubara*[††,†††], *Nobuo Kawaguchi*[††,†††],
Yukiko Yamaguchi[††], *Yasuyoshi Inagaki*[††††]

[†] Graduate School of Information Science, Nagoya University, Japan,
[††] Information Technology Center, Nagoya University, Japan,
[†††] Center for Integrated Acoustic Information Research, Nagoya University, Japan,
[††††] Faculty of Information Science and Technology, Aich Prefectural University, Japan,
irie@el.itc.nagoya-u.ac.jp

10.21437/Interspeech.2004-240

Abstract

This paper proposes a method of speech intention understanding based on a spoken dialogue corpus to which the intention tags are given. The intention tag expresses the task-dependent intention of the speaker, and therefore, the proper understanding enables a spoken dialogue system to take appropriate actions. We have tagged about 35000 utterances in the CIAIR in-car speech database. In our method, several decision trees for intention understanding are constructed. By constructing decision trees and using them at the same time, the strong amount of characteristic features related to intentions can be retrieved, and it can also be robustly coped with the diversity of the utterances. An experiment on inference of utterance intentions has shown 73.1% accuracy.

1. Introduction

In order to interact with a user naturally and smoothly, it is necessary for a spoken dialogue system to understand the intention of the user exactly. As a method of speech intention understanding, example-based approaches have been considered so far [1, 3, 6]. In general, these approaches involve comparing a spoken utterance with examples in a correctly-tagged corpus. The intention of the utterance is regarded as the intention tag of the most similar example in the corpus. However, it is difficult to infer the intention of the utterance to which any example in the corpus is not similar.

This paper proposes a method of speech intention understanding based on a spoken dialogue corpus. The method constructs several decision trees for intention understanding. By constructing several decision trees, the strong amount of characteristic features related to intentions can be retrieved, and it can also be robustly coped with the diversity of the utterances. So far, we have designed an organization of the tags which is called Layered Intention Tag(LIT). These tags show more detailed utterance intention rather than the illocutionary act level, and have built the corpus[2, 3]. LIT is divided into several layers considering the relevance between an intention and various phenomena relevant to an utterance, such as a style, a keyword, a sentence structure. This method constructs several decision trees by using this corpus and infers the intention by combining them.

In order to evaluate the effectiveness of our method, an experiment on inference of the utterance intentions was conducted using the driver utterances about restaurant search recorded on a large-scale in-car spoken dialogue corpus of CIAIR[4, 5]. As a result, the effectiveness of the method was confirmed.

2. Dialogue corpus with intention tags

2.1. Design of an organization of intention tags

In order to build a dialogue corpus with intention tags, we used the transcript of a large-scale in-car spoken dialogue corpus of CIAIR[4, 5]. That is, the intention tags were assigned manually into dialogue data on the corpus. We have designed an organization of intention tags by dividing them into 4 layers considering the relevance between intentions and various phenomena relevant to utterances. The intention tag expresses the task-dependent intention of the speaker, and therefore, the proper understanding enables the system to take appropriate actions in the dialogue. Therefore, the level of the intention ranges from the information about a discourse act to the detailed information, such as a "store name" and a "price". Consequently, the tag was divided into layers by considering elements of the utterance strongly connected with an intention. Figure1 shows a part of the organization of intention tags. Discourse act layer denotes the role of the speech unit in the dialogue. Action layer denotes the action of the utterance unit. Object layer denotes the object of the action such as "Shop", "Parking", etc. Argument layer denotes the other miscellaneous information about the utterance unit. Most of the argument layer tags can be decided directly from the specific keywords in the sentence. As Figure1 shows, the lower layered intention tag depends on the upper-layered one.

2.2. Building a dialogue corpus with LIT

For each utterance unit about restaurant search on the corpus, we provided the intention tag[2, 3]. At this time, we have tagged for over 35,000 utterance units. When we built the transcribed corpus, an utterance is divided into utterance units by a pause of 200 msecs or more. So, one intention tag is given to one utterance unit in principle. Table1 shows an example of a dialogue corpus with intention tags.

3. Speech intention understanding

3.1. Outline of a method

LIT, which is an object of inference in this research, is divided into 4 layers, which are strongly dependent on each other. For example, for the utterance including the predicate which is strongly connected with an intention, the predicate would be an

Table 1: Example of a dialogue corpus with LIT

Utterance	speaker	Discourse act	Action	Object	Argument
I'm looking for a Chinese restaurant.	driver	Request	Search	Shop	
There are "Daikokuten" near here.	operator	Statement	Exhibit	SearchResult	ShopName
Does it serve a Chinese noodle?	driver	Request	Exhibit	ShopInfo	Menu
Yes, it does.	operator	Statement	Exhibit	ShopInfo	Menu
Please guide me there.	driver	Request	Guide	Shop	
Ok. I'll navigate to "Daikokuten".	operator	Express	Guide	Shop	

Table 2: Attributes and attribute values

attribute	attribute value	
speaker	driver, operator	
the previous intention tag	"Statement, Exhibit, ShopInfo" etc. 48types	
the previous speaker	driver, operator	
morpheme	appear in the utterance or not	
the tag already obtained	Discourse act	5types
	Action	7types
	Object	12types
	Argument	36types

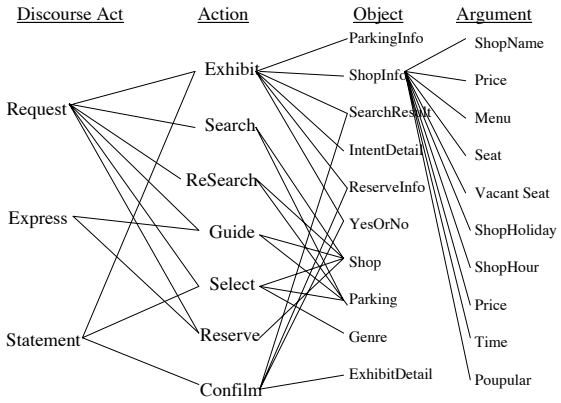


Figure 1: A part of the organization of intention tags

important key of speech intention understanding. In our study, in order to understand speech intention flexibly, decision trees are constructed for every layers of LIT. By considering the elements strongly connected with a certain layer of LIT and ordering decision trees for every utterance, this method can infer the layer of LIT strongly related in utterance first.

3.2. Decision tree learning

The decision tree inferring the tag is constructed for each layer. The software See5[1] was used for decision tree learning . As a set of an attribute, the present speaker, the previous intention tag (Discourse act layer tag, Action layer tag and Object layer tag), the previous speaker and morphemes appeared in a training data are used, because they were effective attributes in a preliminary experiment on speech intention understanding[3]. Furthermore,

[1] See5: http://www.rulequest.com/see5-info.html.

Table 3: Classes and its number of types

class	the number of classes
Discourse act layer tag	5
Action layer tag	7
Object layer tag	12
Argument layer tag	36

the intention tag of the layer that has already obtained is also used as an attribute. Since restrictions exist between layers of an intention tag as Section 2.1 described, the tag of a certain layer would be dependent on the tag of other layers. The tags of each layer were given as a class. Table2 shows attributes and attribute values, and Table3 shows classes and its number of types. Moreover, an example of the data corresponding to an utterance " Does the shop serve a Chinese noodle? " is shown in Figure2.

32 types of decision trees are constructed by the data set obtained from the dialogue corpus.

- No tag of any layer is inferred.The 1st group
  As an attribute, the present speaker, the previous intention tag, the previous speaker and morphemes appeared in a training data are used. 4 types of decision trees are constructed. The decision tree for Discourse act layer inference, the one for Action layer inference, the one for Object layer inference and the one for Argument layer inference.
- The tag of one layer is inferred.The 2nd group
  As an attribute, the present speaker, the previous intention tag, the previous speaker, morphemes appeared in a training data and the tag already obtained are used. Since 3 patterns of the inferring layer exist each layer, a total of 12 types of decision trees are constructed.
- The tags of two layers are inferred.The 3rd group
  As an attribute, the present speaker, the previous inten-

Dialogue Corpus with intention tags

utterance	intention tag
D : Does the shop serve a Chinese noodles	Request＋Exhibit＋ShopInfo＋Menu

Data file

speaker	The previous intention tag	The previous speaker	morpheme: this	morpheme: the	morpheme: shop	morpheme: ・・・	The tag already obtained	class
D	Express＋Exhibit＋SearchResult	O	no	yes	yes	・・・	1-Request	2-Exhibit
D	Express＋Exhibit＋SearchResult	O	no	yes	yes	・・・	1-Request	3-ShopInfo
D	Express＋Exhibit＋SearchResult	O	no	yes	yes	・・・	1-Request	4-Menu

Figure 2: Example of the data corresponding to utterance "Does the shop serve a Chinese noodle?"（in the case of the 2nd group）

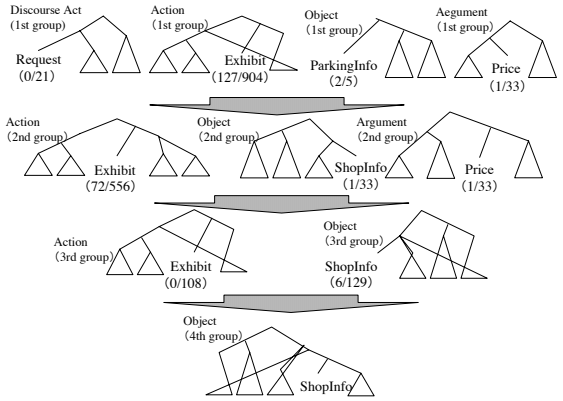


Figure 3: Flow of the intention inference processing

Table 4: Inference accuracy（training data：2972utterances；closed test）

tag	Discourse act	Action	Object	Argument
accuracy（%）	92.2	76.8	80.6	85.8

tion tag, the previous speaker, morphemes appeared in a training data and the tag already obtained are used. Since 3 patterns of the inferring layer exist each layer, a total of 12 types of decision trees are constructed.

- The tags of three layers are inferred.（The 4th group） As an attribute, the present speaker, the previous intention tag, the previous speaker, morphemes appeared in a training data and the tag already obtained are used. Since 1 pattern of the inferring layer exist each layer, a total of 4 types of decision trees are constructed.

3.3. Inference algorithm

The tag of each layer is inferred using several decision trees obtained in 3.2. The order of the inferring layer determines by decision tree ordering rule. This ordering rule was defined as follows.

Decision tree ordering rule

1. A decision tree with the lowest rate of a re-classification error (a rate of an error in training data) of the leaf reached in inference is chosen.
2. When the rate of a re-classification error is the same, the decision tree which had much training dates reaching the leaf is chosen.
3. When a decision tree cannot be chosen above rules, it is chosen in the order as follows. : Discourse Act, Argument, Object, and Action tag. This based on the accuracy (see Table4) when the tags are inferred (closed test) using See5 in advance.

According to the following algorithm, an utterance intention infers using this ordering rule.

Inference algorithm

(1) The candidate of the tag of each layer is outputted using 4 types of decision trees of the 1st group.

(2) One decision tree is chosen according to the ordering rule, and let the tag that it outputted be an inferred result.

(3) The tag obtained by (2) is added to attribute values, and the tag of undecided layers is inferred using the decision tree of the 2nd group.

(4) The tag obtained by (3) is further added to an attribute values, and the tag of undecided layers is inferred using the decision tree of the 3rd group.

(5) The tag obtained by (4) is further added to attribute values, and the tag of undecided layers is inferred using the decision tree of the 4th group.

3.4. Example of speech intention understanding

Figure3 shows an example of speech intention understanding of "How much is it?" .

First, the candidate of the tag of each layer is outputted using 4 types of decision trees of the 1st group. In Figure3, "Request" is outputted as a candidate of a Discourse act tag, "Exhibit" is outputted as one of an Action tag, "ParkingInfo" is outputted as one of an Object tag and "Price" is outputted as one of an Argument tag. "Request"of a Discourse act tag regard as an inferred result, because its rate of a re-classification error is the lowest (decision tree ordering rule 1). Next, "Request" tag is added to attribute values, and the tag of undecided layers is inferred. The candidates of the tag of undecided layers are outputted using the decision tree of the 2nd group. Then, "Exhibit" is outputted as a candidate of an Action tag, "ShopInfo" is outputted as one of an Object tag and "Price" is outputted as one of an Argument tag. "Price" tag regarded as an inferred result (decision tree ordering rule 3). Similarly, an Argument tag is added to attribute values, and the tag of undecided layers is inferred using the decision tree of the 3rd group. Then, "Exhibit" regarded as an inferred result. Finally, Object tag is added to attribute values and "ShopInfo" is inferred using the decision tree of the 4th group. As a result, the system can arrive at a correct utterance intention "Request+Exhibit+ShopInfo+Price" .

4. Experiment

4.1. Outline of experiment

In our experiment, 1218 dialogues on a corpus with intention tags were used. Chasen [2] was used for the morphological analysis. Nouns and proper nouns, which are characteristic of a dialogue task, and filler, which is peculiar to spoken language, added to chasen's dictionary.

We have divided 3143 driver's utterances into two groups. One is training data of 2972 utterances, the other is test data of 171 utterances. The decision trees were constructed by the decision tree learning software See5. In our experiment, the following two patterns of decision trees were made, and their results were compared.

1. Inference of all layers at once
   An intention tag considered like "Statement + Exhibit + ShopInfo + Menu".
2. Inference using several decision trees (our method)

4.2. Experimental result

The experimental result is shown in Table 5. 67.8% as the precision was obtained by the inference based on the above method 1, and 73.1% was done in the method 2. This result shows our method using several decision trees to be more effective.

4.3. Discussion

4.3.1. Comparison with example-based approach

We have so far proposed a method of speech intention understanding based on dialogue examples[3]. This method uses a spoken dialogue corpus with LIT to regard the intention of each input utterance as that of the utterance to which it is the most similar in the corpus. The degree of similarity is calculated according to the degree of correspondence in morphemes between utterances, and an intention sequence. An experiment using the same data with the experiment described above has shown 70.8% accuracy.

When the example which is similar to input utterance exists, this method has a high possibility that an intention can be inferred correctly, because of regarding the intention of each input utterance as that of the utterance to which it is the most similar in the corpus. On the other hand, a possibility of mistaking becomes high, when the diversity of an example is small compared with the diversity of input utterance. Since our method learned decision trees with a corpus and was creating the model, the better result was obtained in the amount of data of this experiment.

4.3.2. Comparison with annotators' concordance rate

On the other hand, we have investigated more than one annotators' concordance rate of LIT[3]. This experiments were based on the following two groups of annotators.

1. Four annotators, who are not expert, provided the tags for 528 utterances. Each dialogue was tagged by two annotators.
2. Two designers of LIT provided the tags for 296 utterances.

70.5% as a concordance rate of the group 1 was obtained, and 85.3% as a concordance rate of the group 2 was done. Compared with Table 5, the precision of a proposal method is less than the concordance rate between designers. But it is better than the concordance rate between annotators who are not expert.

[2] Chasen, http://chasen.aist-nara.ac.jp/.

Table 5: Experimental result

	Inference of all layer at once	Inference using several decision trees
precision	67.8 (116/171)	73.1 (125/171)

5. Concluding remarks

This paper proposes a method of speech intention understanding based on a spoken dialogue corpus. This method infers intention by constructing several decision trees and combining them. By constructing several decision trees, the strong amount of characteristic features related to intentions can be retrieved, and it can also be robustly coped with the diversity of the utterances. The experimental result, which uses the decision tree obtained from 2972 driver's utterances of a corpus with intention tags, has shown the effectiveness of our method.

6. References

[1] T.Daniel Midgley: Discourse chunking: a tool in dialogue act tagging, *ACL-03 Companion Volume to the Proc. of the Conf.*, pp.58-63, 2003.

[2] Y. Irie, N. Kawaguchi, S. Matsubara, I. Kishida, Y. Yamaguchi, K. Takeda, F. Itakura and Y. Inagaki : An Advanced Japanese Speech Corpus for In-car Spoken Dialogue Research, *Proc. of Oriental COCOSDA-2003*, pp.209-216, 2003.

[3] Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi and Y. Inagaki : Speech Intention Understanding based on Spoken Dialogue Corpus, SIG-SLUD-A301-03, pp.7-12, 2003. (In Japanese)

[4] N. Kawaguchi, S. Matsubara, I. Kishida, Y. Irie, Y. Yamaguchi, K. Takeda and F. Itakura: Construction and Analysis of the Multi-Layered In-Car Spoken Dialogue Corpus, *Proc. of Workshop on DSP in Mobile and Vehicular Systems*, 2003.

[5] I. Kishida, Y. Irie, Y. Yamaguchi, S. Matsubara, N. Kawaguchi and Y. Inagaki: An Advanced Japanese Speech Corpus for In-car Spoken Dialogue Research, *Proc. of 8th European Conf. on Speech Communication and Technology*, pp.1581-1584, 2003.

[6] S.Matsubara, S.Kimura, N.Kawaguchi, Y.Yamaguchi and Y.Inagaki: Example-based Speech Intention Understanding and Its Application to In-Car Spoken Dialogue System, *Proc. of 17th Int. Conf. on Computational Linguistics*, Vol.1, pp.633-639, 2002.