Development of a machine learnable discourse tagging tool

Masahiro Araki†, Yukihiko Kimura†, Takuya Nishimoto†, and **Yasuhisa Niimi†**

†Department of Electronics and Information Science
Kyoto Institute of Technology
Matsugasaki Sakyo-ku Kyoto 606-8585, Japan
{araki,kimu,nishi,niimi}@dj.kit.ac.jp

Abstract

We have developed a discourse level tagging tool for spoken dialogue corpus using machine learning methods. As discourse level information, we focused on dialogue act, relevance and discourse segment. In dialogue act tagging, we have implemented a transformation-based learning procedure and resulted in 70% accuracy in open test. In relevance and discourse segment tagging, we have implemented a decision-tree based learning procedure and resulted in about 75% and 72% accuracy respectively.

1 Introduction

In dialogue research communities, the need of dialogue corpora with various level of annotation is recognized. However, creating annotated dialogue corpora needs considerable cost in recording, transcribing, annotating, and checking the consistency and reliability of the annotated data.

Considering such situation, we focused on annotation step and developed discourse level tagging tool for spoken dialogue corpus using machine learning methods. In this paper, we explain the detail of tagging scheme and describe machine learning algorithm suitable for each level of tagging.

2 Multiple level tagging scheme for Japanese dialogue

It is widely recognized that making annotated spoken dialogue corpora is labor-intensive effort. To this end, the Discourse Research Initiative (DRI) was set up in March of 1996 by US, European, and Japanese researchers to develop standard discourse annotation schemes (Carletta et al., 1997; Core et al., 1998). In line with the effort of this initiative, Japanese Discourse Research Initiative has started and created annotation scheme for various level of information of dialogue, that is JDTAG (Japanese Dialogue TAG) (JDRI, 2000). Our aim is to develop tagging tools in line with the JDTAG.

In the following of this section, we explain the part of tagging scheme which are relevant to our tools.

2.1 Dialogue act

In JDTAG, slash unit is defined following Meteer and Taylor (Meteer and Taylor, 1995). Dialogue act tagging scheme is a set of rules to identify the function of each slash unit from the viewpoint of speech act theory (Searle, 1969) and discourse analysis (Coulthhard, 1992; Stenstrom, 1994). These dialogue act tag reflect a local structure of the dialogue.

To improve the agreement score among the annotators, we assume basic structure of dialogue shown in Figure 1.

Typical exchange pattern is shown in Figure 2.

In this scheme, the tags (Figure 3) need to be an element of exchange structure except for those of dialogue management.

2.2 Relevance

Dialogue act tag can be regarded as a function of utterance. Therefore, we can see the sequence of dialogue act tag as the flat structure of the dialogue. It is insufficient to express

- Task-orientedDialogue → (Opening) ProblemSolving (Closing)

- ProblemSolving → Exchange$^+$

- Exchange → Initiate (Response)/Initiate* (Response)* (FollowUp) (FollowUp)

'*':repeat more than 0 time！$+$':repeat more than 1 time, ( ): the element can be omitted.

Figure 1: Exchange structure

```
-------------------------------------------
(I) 0041 A: chikatetsu no ekimei ha?
  (What's the name of the subway station?)
(R) 0042 B: chikatetsy no teramachi
            eki ni narimasu.
  (The subway station is Teramachi.)
(F) 0043 A: hai.
  (I see.)
-------------------------------------------
I: Initiate, R: Response, F:Follow-up
```

Figure 2: Typical exchange pattern

- Dialogue management

  Open, Close

- Initiate

  Request, Suggest, Persuade, Propose, Confirm, Yes-no question, Wh-question, Promise, Demand, Inform, Other assert, Other initiate.

- Response

  Positive, Negative, Answer, Other response.

- Follow up

  Understand

- Response with Initiate

  The element of this category is represented as *Response Type / Initiate Type.*

Figure 3: Tag set of dialogue act

tree-like structure, such as embedded subdialogue. In order to represent such higher level information, we use a relevance tag.

There are two types of relevance between slash units. The one is the relevance of the inside of exchange. The other is the relevance of between neighboring exchanges. We call the former one as relevance type 1, and the latter one as relevance type 2.

Relevance type 1 represents the relation of initiate utterance and its response utterance by showing the ID number of the initiate utterance at the response utterance. By using this tag, the initiate-response pair which strides over embedded subdialogue can be grasped.

Relevance type 2 represents the meso-structure of the dialogue such as chaining, coupling, elliptical coupling as introduced in (Stenstrom, 1994). Chaining is a pattern of [A:I B:R] [A:I B:R] (speaker A initiates the exchange and speaker B responds it). Coupling is a pattern of [A:I B:R] [B:I A:R]. Elliptical coupling is a pattern of [A:I] [B:I A:R] which omits the response in the first exchange. Relevance type 2 tag is attached to the each initiate response in showing whether such meso-level dialogue structure can be observed (yes) or not (no).

The follow-up utterance has no relevance tag. It is because follow-up necessarily has a relevance to the preceded response utterance.

The example of dialogue act tagging (first element of tag) and relevance tagging (second element) is shown Figure 4.

```
----------------------------------------
[<Yes-no question> <relevance no>]
0027 A: hatsuka no jyuuji kara
        ha aite irun de syou ka
  (Is it available from 10 at 20th?)

[<Yes-no question> <relevance yes>]
0028 B: kousyuu shitsu desu ka?
  (Are you mentioning the seminar room?)

[<Positive> <0028>]
0029 A: hai
  (Yes.)

[<Negative> <0027>]
0030 B: hatsuka ha aite orimasen
  (It is not available in 20th.)

[<Understand>]
0031 A: soudesu ka
  (OK.)
----------------------------------------
```

Figure 4: An example dialogue with the dialogue act and relevance tags

```
---------------------------------------
[2: room for a lecture: ]
38 A: {F e} heya wa dou simashou ka?
(How about meeting room?)


[1: small-sized meeting room: clarification]
39 B: heya wa shou-kaigishitsu wa aite masu ka?
(Can I use the small-sized meeting room?)


40 A: {F to} kayoubi no {F e} 14 ji han kara
      wa {F e} shou-kaigisitsu wa aite imasen
(The small meeting room is not available
 from 14:30 on Tuesday.)


[1:the large-sized meeting room: ]
41 A: dai-kaigishitsu ga tukae masu
(You can use the large meeting room.)


[1: room for a lecture: return]
42 B: {D soreja} dai-kaigishitsu de onegai
      shimasu
(Ok. Please book the large meeting room.)
---------------------------------------
[TBI:topic name:segment relation]
```

Figure 5: An example dialogue with the dialogue segment tags

2.3 Dialogue segment

Dialogue segment of JDTAG indicates boundary of discourse segment introduced in (Grosz and Sidner, 1986). A dialogue segment is identified based on the exchange structure explained above. A dialogue segment tag is first inserted before each initiating utterance. After that, a topic break index, a topic name, and a segment relation are identified.

Topic break index (TBI) takes the value of 1 or 2: the boundary with TBI=2 is less continuous than the one with TBI=1 with regard to the topic. The topic name is labeled by annotators' subjective judgment for the topics of that segment. The segment relation indicates the one between the preceding and the following segments, which is classified as clarification, interruption, and return.

Figure 5 shows an example dialogue with the dialogue segment tags.

3 Dialogue act tagger

Considering the limitation of amount of corpus with dialogue level annotations, a promising dialogue act tagger is based on machine learning method with limited amount of training data rather than statistical method, which needs large amount of training data. Rule-based and example-based learning algorithms are suitable to this purpose. In this section, we compare our implementation of transformation-based rule learning algorithm and example-based tagging algorithm.

3.1 Transformation-based learning

Transformation-based learning is a simple rule-based learning algorithm. Figure 6 illustrates the learning process.
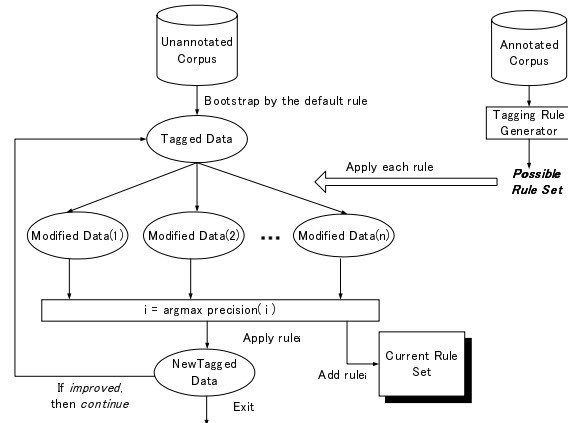

Figure 6: Learning procedure of dialogue act tagging rule by TBL

First, initial tagged data was made from unannotated corpus by using bootstrapping method. In our implementation, we use default rule which assigns the most frequent tag to all the utterance as a bootstrapping method. All the possible rules are constructed from annotated corpus by combining conditional parts and their consequence. All the possible rule are applied to the data and the rule whose transformation results in the greatest improvement is selected. This rule is added to the current rule set and this iteration is continued until no improvement is observed. In the previous research, TBL showed successful performance in many annotation task, e.g. (Brill, 1995), (Samuel et al., 1998).

In our experiment, the selected features in the conditional part of the rule are words (the notation in the rule is *include*), sentence length (*length*) and previous dialogue act tag (*prev*). Although each feature is not enough to use as a clue in determining dialogue act,

the combination of these features works well. We used four types of combinations, that is, *include + include, include + length, include + prev* and *length + prev.*

The result of the learning process is a sequence of rules. For example, in dialogue act tagging, acquired rules in scheduling domain are shown in Figure 7.

```
#1 condition: default,
     new_tag: wh-question
#2 condition: include="yoroshii(good)"
              & include="ka(*1)",
     new_tag: yes-no-question
#3 condition: include="hai(yes)"
              & prev=yes-no-question,
     new_tag: affirmative
#4 condition: include="understand"
              & length < 4,
     new_tag: follow-up
 ...

(*1 "ka" is a functional word
    for interrogative sentence)
```

Figure 7: Acquired dialogue act tagging rules

3.2 Example-based learning

Example-based learning is suitable for classification task. It stores up example of input and corresponding class, calculates the distance between these examples and new input, and classifies it to the nearest class.

In our dialogue act tagging, the example is consists of word sequence (partitioned by slash unit tag) and part of speech information. Corresponding dialogue act tag is attached to all the example.

The distance between example and new input is calculated using the weighted agreement of elements shown in Table 1.

Table 1: Elements for calculating a distance.

element	weight
dialogue act of before two sentence	1
dialogue act of previous sentence	3
postpositional word of end of sentence	3
clue word for dialogue act	3
another word	2

3.3 Experimental results

We have compared above two dialogue act tagging algorithms in two different tasks: a route direction task and a car trouble shooting task. We used 4 dialogues for each task (268 and 184 sentences) as a training data and 2 dialogues as a test data (113 and 63 sentences). The results are shown in Table 2.

Table 2: Comparison of TBL and example-based method.

algorithm	task	closed	open
TBL	route direction	85.1	72.6
	car trouble shooting	90.2	66.7
	average	87.7	69.7
ex-based	route direction	93.8	62.6
	car trouble shooting	89.7	52.4
	average	91.8	57.5

We got similar average score for closed test. Therefore, we regard the tuning level of parameter of each algorithm as a comparable level. In open test in the same task, we got 69.7% in TBL and 57.5 % in example-based method. As a result, we can conclude TBL is more suitable method for dialogue act tagging learning in limited amount of training data.

4 Relevance tagger using decision tree

4.1 Decision tree learning

Decision tree learning algorithm is one of classification rule learning algorithm. The training data is a list of attribute-value pair. The output of this algorithm is a decision tree whose nodes are regarded as set of rules. Each rule tests the value of an attribute and indicates the next node.

A basic algorithm is as follows:

1. create root node.

2. if all the data belong to the same class,

   create a class node and exit.

   otherwise,

   - choose one attribute which has the maximum mutual information and create nodes corresponding values.
   - divide and assign the training data according to the values and create link to the new node.

3. apply this algorithm to all the new nodes recursively

We also used post-pruning rule hired in C4.5 (Quinlan, 1992).

4.2 Relevance tagging algorithm and results

Relevance type 1

Relevance type 1 tag is automatically annotated according to the exchange structure which is identified in dialogue act tagging stage. The accuracy of the relevance type 1 tag is depend on whether a given dialogue or task domain is follow the assumption of exchange structure explained above. In well formed dialogue, the accuracy is above 95%. However, in ill-formed case, it is around 70%.

Relevance type 2

We have used decision tree method in identifying relevance type 2, which identifies whether neighboring exchange structures have a certain kind of relevance. The attributes of training data are as follows.

1. relevance type 2 tag of previous exchange
2. initiative dialogue act tag of previous exchange
3. response dialogue act tag of previous exchange
4. initiative dialogue act tag of current exchange

We used 9 dialogue (hotel reservation, route direction, scheduling, and telephone shopping) as training and test data. The results are shown in Table 3. We got this results after 10 cross validation. In cross domain experiment, we got 84% accuracy in closed test (average 47 nodes) and 75% in open test. Using post-pruning method, we got 82% of accuracy (average 22 nodes; estimated accuracy 76%) in closed test and 77% in open test.

5 Dialogue segment tagger

5.1 TBI tagger

We used decision tree method in identifying the value of topic break index because the target attribute have only two values; 1 (small topic change) or 2 (large topic change). In case of target attribute has small number of values, decision tree method can be estimated to outperform transformation-based learning. The attributes of training data are as follows.

1. relevance type 2 tag of previous exchange
2. relevance type 2 tag of current exchange
3. topic break index tag of previous exchange
4. dialogue act tag of previous slash unit
5. dialogue act tag of current slash unit

We used same data set with the experiment of dialogue act tagging. We got 87% accuracy in closed test (average 61 nodes) and 80% in open test. Using post-pruning method, we got 82% of accuracy (average 12 nodes; estimated accuracy 76%) in closed test and 78% in open test (see Table 4).

5.2 Topic name tagger

In JDTAG topic name tagging scheme, annotators can assign a topic name subjectively to the given dialogue segment. Certainly it is an appropriate method for this scheme to use for the dialogue of any task domain. But, even in the almost same pattern of exchange, different annotators might annotate different topic names. It prevent the data from a practical usage, e.g. extracting exchange pattern in asking a route to certain place.

We prepare a candidate topic name list and assign to dialogue segment as a topic name. Because candidate topic name is around 10 to 30 according to the task domain, we use transformation-based learning method for acquiring a topic name tagging rule set.

The selected features in the conditional part of the rule are words of current segment (up to 2), dialogue act tag of the first slash unit of the segment, topic name tag of previous segment.

As a result, in the above data set, the candidate rules are 5588. And we got 98% accuracy in the closed test and 56% in open test.

5.3 Segment relation tagger

The number of segment relation types are 4 (clarification, interruption, return, and none). Therefore, we used decision tree for acquiring rules for identifying segment relation types.

In making decision tree, we did not use post-pruning because a great many of segment relation tag is none (about 85%). Post-pruning makes a tree too general (only one top node which identifies none or else).

Table 3: Results of relevance type 2 tagging

	not pruned		pruned		
	# of nodes	accuracy	# of nodes	accuracy	estimated error rate
Training	47.3	83.7%	22.0	81.9%	23.7%
Test	47.3	75.4%	22.0	77.2%	23.7%

Table 4: Results of topic break index tagging

	not pruned		pruned		
	# of nodes	accuracy	# of nodes	accuracy	estimated error rate
Training trouble shooting	53.0	90.0%	10.2	82.4%	22.6%
Test trouble shooting	53.0	85.2%	10.2	78.2%	22.6%
Training route direction	69.0	84.1%	14.5	82.4%	25.5%
Test route direction	69.0	73.8%	14.5	77.5%	25.5%

As a result, also in the same data set, we got 92% accuracy in the closed test.

6 Conclusion

We have developed a discourse level tagging tool for spoken dialogue corpus using machine learning methods. We use transformation-based learning method in case of many target values, and decision tree method otherwise.

Our future work is to develop an environment in which annotators can easily browse and post-edit the output of the tool.

Acknowledgement

References

E. Brill. 1995. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21(4):543–566.

J. Carletta, N. Dahlback, N. Reithinger, and M. A. Walker. 1997. Standards for dialogue coding in natural language processing. Dagstuhl-Seminar-Report:167 (ftp://ftp.cs.uni-sb.de/pub/dagstuhl/ reporte/97/9706.ps.gz).

M. Core, M. Ishizaki, J. Moore, C. Nakatani, N. Reithinger, D. Traum, and S. Tutiya. 1998. *The Report of the Third Workshop of the Discourse Research Initiative*. Chiba Corpus Project. Technical Report 3, Chiba University.

M. Coulthhard, editor. 1992. *Advances in Spoken Discourse Analysis*. Routledge.

B. J. Grosz and C. L. Sidner. 1986. Attention, intention and the structure of discourse. *Computational Linguistics*, 12:175–204.

The Japanese Discourse Research Initiative JDRI. 2000. Japanese dialogue corpus of multi-level annotation. In *Proc. of the 1st SIGDIAL Workshop on discourse and dialogue*.

M. Meteer and A. Taylor. 1995. Dysfluency annotation stylebook for the switchboard corpus. Linguistic Data Consortium (ftp://ftp.cis.upenn.edu/pub/treebank/ swbd/doc/DFL-book.ps.gz).

J. R. Quinlan. 1992. *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

K. Samuel, S. Carberry, and K. Vijay-Shanker. 1998. Dialogue act tagging with transformation-based learning. In *Proc. of COLING-ACL 98*, pages 1150–1156.

J. R. Searle. 1969. *Speech Acts*. Cambridge University Press.

A. B. Stenstrom. 1994. *An Introduction to Spoken Interaction*. Addison-Wesley.