RESPONSE EVALUATION OF SPOKEN DIALOGUE PROCESSING METHOD USING INDUCTIVE LEARNING BASED AMOUNT OF INFORMATION

YASUTOMO KIMURA KENJI ARAKI

Graduate School of Engineering, Hokkaido University, Kita 13 Nishi 8, Kita-ku, Sapporo-shi, 060-8628 Japan

KOJI TOCHINAI

Graduate School of Business Administration, Hokkai-Gakuen University, Asahi-machi 4-1-40, Toyohira-ku, Sapporo-shi, 062-8625 Japan

To be effective, a spoken dialogue system must be able to process and accurately respond to conversation as well as adapting to background noise, interjections and unnecessary vocabulary. To do so the system needs to obtain information from the dialogue automatically. Therefore we propose a learning method using AoI(Amount of Information). We define the rule of spoken dialogue based on AoI. In this paper, we describe the experiment and results which show a 44% improvement over the Inductive Learning method.

Key words: Dialogues, Spoken languages, Inductive Learning, Amount of Information

1. INTRODUCTION

Recently speech recognition techniques have improved with high rates of recognition. Spoken dialogue systems are now beginning to use web information. For example, JUPITER [Victor Zue and Hetherington2000] provides weather information using a phone. JUPITER consists of speech recognition, language understanding and dialogue modeling. TOOT, which is another spoken dialogue system, can respond with information concerning train timetables. These systems use a frequently updated data of limited domain. However if the systems are unable to obtain the necessary information from the web, they will not be able to respond appropriately.

A system that is required to carry out a dialogue with an individual and without limited domain is unable to obtain personal information about the individual on the web. The system needs to obtain information directly from the dialogue itself. Therefore the system needs to have a learning process in order to adapt to a topic.

We have proposed a spoken dialogue method of Inductive Learning using Genetic Algorithm (GA-IL)[Y. Kimura2001]. The most important aspect of a learning system is its acquisition capability which can acquire the rules of spoken dialogue from a dialogue example. The main capability of learning is to distinguish common parts and different parts. However GA-IL cannot efficiently acquire the rules if the spoken dialogue contains few matching rules. Therefore we propose a learning method using "Amount of Information" (hereafter abbreviated as AoI). In this paper, We define the common parts and the different parts based on AoI and describe the experiment and results.

2. COMMON AND DIFFERENT PARTS

We describe the definition of the common and different parts based on GA-IL. The system seeks the common and different parts from the rules. The common parts are the matching parts, and the others parts are the different parts. The pairs of the different parts have a characteristic relationship. Theses pairs are "**response rules**". The rules, which replace

© 2003 Pacific Association for Computational Linguistics

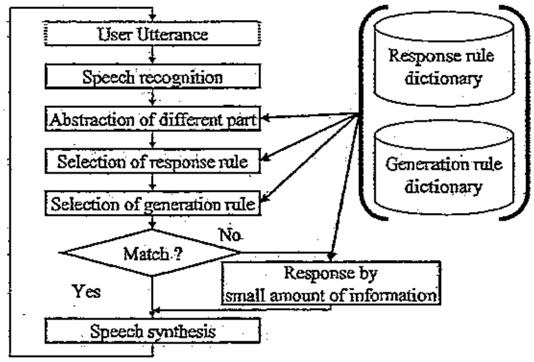

FIGURE 1. Flow of process.

the different part with a variable, are "**generation rules**" and act as templates for making a sentence. The GA-IL system acquires the relationship, being the different part pair, which is obviously important. The pair of the different part is characterized by its relationship. However it is ineffective in acquiring these relationships.

Therefore we define the common and different parts based on AoI. The common part consists of a small AoI because the words of common parts frequently exist in the dialogue example. The different part includes a large AoI because of characteristic parts. AoI, $I(x) = -logP(x) \cdots (1)$ $P(x) = \frac{F(x)}{N}$, $F(x)$ is the frequency of word x, N is sum of word number. According to the formula (1), low frequency indicates a large AoI.

3. PROCESS

3.1. Learning process

The system acquires the response rule and the generation rule using AoI. The response rule is the pair of different parts between user A utterance and user B utterance. A response rule consists of both a left hand side(LHS), which includes the previous different part and the current different part, and a right hand side(RHS) which is the different part for the response. The system makes a sentence by combining the RHS of the response rule and the generation rule. A generation rule consists of common parts and a variable which replaces the different parts. The generation rule then exchanges the variable for the RHS of the selected response rule.

3.2. Dialogue processing

Figure 1 shows the outline of dialogue processing. The system recognizes the user's utterance using a speech recognition tool. The system extracts the different parts from the recognition, and selects a response rule. The selected response rule is combined with the generation rule. The combined generation rule is the response to the input. If the system could not make a response, the system uses either the relationship of the common parts or the smallest AoI. The User is given the response by a speech synthesis tool.

The response with a small AoI is used for continuing a conversation. In human conversation, even if a speaker could not understand his/her partner's utterance, the speaker would reply giving a small AoI. Using a small AoI to respond to a large AoI means the details of the topic or information are unknown. If the system could not find the appropriate different part to respond to a large AoI, the system replies using a small AoI.

Though the response of a small AoI provides a vague answer, the system tries to keep the consistency of conversation using common parts. We consider that the response using the relationship of common parts is better than the response using a small AoI.

4. SPOKEN DIALOGUE EXPERIMENT

4.1. Experiment procedure

The purpose is to evaluate the system with real spoken dialogue(open data). The learning data is based upon conversations with free topics between two people. The participants are three university students. The system acquired the rules by listening to the conversation between a pair of speakers. We call the three speakers, A,B and C respectively. After the system has acquired the rules, each speaker individually communicates with the system. Each response is divided into 4 patterns, being correct, semi correct, vague and erroneous.

- Correct : a satisfactory response.
- Semi correct :comprehensible, but not a satisfactory response.
- Vague :some meanings, but able to understand.
- Erroneous : not understood.

4.2. Experiment result and consideration

Table 1 shows the results being the response patterns as determined by each speaker. Excluding erroneous responses the results show a comprehension between 51% and 65%. In open data of spoken dialogue, these results show a high success response rates because the spoken dialogue includes erroneous recognitions and the user's erroneous utterance.

TABLE 1. Result of response patterns.

Speaker	Correct	Semi Correct	Vague	Erroneous
A	6.12%	31.36%	27.55%	34.69%
B	17.02%	22.34%	11.70%	48.94%
C	17.89%	32,36%	3.16%	46.32%

5. COMPARISON EXPERIMENT

5.1. Experiment procedure

We compared our method and that of Inductive Learning method. The contents of the experiment are the acquired number and the accuracy of the response. The experiment data is made up of 400 sentences in the SLDB.

5.2. Experiment and consideration

Table 2 shows the number of acquired rules by each method. With the Inductive Learning method, the system acquired 29 different parts and 18 generation rules which consist of a common part and a variable. Though it is possible to generate 522(= 29 × 18) sentences, the total are 452 sentences once the same sentences in the learning data are excluded. In our method, our system acquired 364 different parts and 367 generation rules, making it possible to generate 133,588(=364 × 367)sentences. However the system does not generate sentences in the case of "0" co-occurrence. Therefore the total number of sentences generated by this method is 3,574.

We evaluate these generated sentences between our method and the Inductive Learning method. Each sentence is evaluated either as being readable or unreadable by a subject. To evaluate and compare systems, 100 sentences were randomly selected from each method. Table 2 shows the number of both readable and unreadable sentences. The success rate of our proposed method was 62% as opposed to 43% using Inductive Learning. In comparison, our system obtained a 44% improvement over the Inductive Learning system. The improvement confirms that our system can efficiently acquire rules and generate correct sentences.

TABLE 2. Comparison experiment.

The number of acquired rules in each method		
	Inductive Learning	Our method
Acquired different part	29	364
Acquired "common part and variable"	18	367
Evaluation of generated sentences		
Generation sentence	452	3,592
Correct generation sentences	43%	62%
Erroneous generation sentences	57%	38%

6. CONCLUSION

Our purpose is to create a spoken dialogue system which is able to cope with frequent changes in conversational topics as occurs in daily life. Therefore we proposed a learning method using AoI(Amount of Information). In this paper, we extended the definition of common and different parts by AoI, co-occurrence and mutual information.

In comparison, our system obtained a 44% improvement over the Inductive Learning system. We confirmed that our system could efficiently acquire rules and generate sentences.

Future work, we will improve division points and the correct response rate.

REFERENCES

[Victor Zue and Hetherington2000] James Glass Joseph Polifroni Christine Pao Timothy J. Hazen Victor Zue, Stephanie Seneff and Lee Hetherington. 2000. Jupiter: A telephone-based conversational interface for weather information. *IEEE Transactions on Speech and Audio Processing, Vol.8 , No.1.*

[Y. Kimura2001] Y. Momouchi K. Tochinai Y. Kimura, K. Araki. 2001. Spoken dialogude processing method using inductive learning with genetic algorithm. *The transactions of the Institute of Electronics, Information and Communication Engineers D-II, Vol.J84-D-II, No.9,* pages 2079–2091, Sep.