Evaluation of Japanese Dialogue Processing Method Based on Similarity Measure Using $tf \cdot AoI$

Yasutomo Kimura[1], Kenji Araki[1], and Koji Tochinai[2]

[1] Graduate School of Engineering, Hokkaido University,
Kita 13 Nishi 8, Kita-ku, Sapporo-shi, 060-8628 Japan
{kimu,araki}@media.eng.hokudai.ac.jp
[2] Graduate School of Business Administration, Hokkai-Gakuen University,
Asahi-machi 4-1-40, Toyohira-ku, Sapporo-shi, 062-8625 Japan
tochinai@econ.hokkai-s-u.ac.jp

Abstract. In this paper, we propose a Japanese dialogue processing method based on a similarity measure using $tf \cdot AoI(termfrequency \times AmountofInformation)$. Keywords are specially used in a spoken dialogue system because a user utterance includes an erroneous recognition, filler and a noise. However, when a system uses keywords for robustness, it is difficult to realize detailed differences. Therefore, our method calculates similarity between two sentences without deleting any word from an input sentence, and we use a weight which multiplies term frequency and amount of information($tf \cdot AoI$). We use 173 open data sets which are collected from 12,095 sentences in SLDB. The experimental result using our method has a correct response rate of 67.1%. We confirmed that correct response rate of our method was 11.6 points higher than that of the matching rate measure between an input sentence and a comparison sentence. Furthermore that of our method was 7.0 points higher than that of $tf \cdot idf$.

1 Introduction

Recently, Information Extraction, Information Retrieval and Summarization attract attention in NLP. In these researches, sentences or words are classified into information types by a similarity measure. Similarity measurés are used not only for classification problems but also for comparison of documents. Therefore it is applicable also to a dialogue processing system. From such a background, similarity measures are recognized to be indispensable technology in the applicable field of NLP.

A similarity measure is used as a criterion for comparing either words or sentences. When we calculate similarity between two sentences, the same sentences which consist of perfect matching become the highest similarity. However, two sentences which have high matching rate are not necessarily similar. Each domain should select an expression of a similarity measure. In Information Retrieval, some similarity measure expressions have been proposed such as Boolean model, Vector Space model and so on. Although a Boolean model expresses a search question by the logic formula, it almost becomes the same as a matching comparison. In vector space model, a similarity measure is calculated using the Euclidean distance, a cosine, a Dice coefficient and so on. A similarity measure has been multiplied term frequency and another weight just like idf in order to make the characteristic of an input sentence reflect.

By the way a dialogue processing system has used keywords since 1960s[1]. Especially keywords have been used in a spoken dialogue processing[2][3][4] because a spoken dialogue includes a speech recognition error, an interjection, and noise. However, the sentence that does not include any keyword often has an important meaning.

In information retrieval, $tf \cdot idf$ is used widely[5]. $tf \cdot idf$ means multiplying term frequency and inverse document frequency. The value of $tf \cdot idf$ becomes higher when the term does not exist in other document very much. However, it does not give suitable weight when there is only a little difference of idf.

In this paper, we propose Euclidean distance based on $tf \cdot AoI$ in order to measure similarity of two sentences which do not have many words. AoI which is short for Amount of Information shows as follows:

$$AoI = -log_2 P(x) = -log_2 \frac{f(x)}{N} \quad \cdots (1)$$

N means number of running words, and $f(x)$ means frequency which the word x exists. In our method, high frequency interjection and unpredictable noise can become small weight because noise and interjection tend to repeat. Therefore we give a weight of $tf \cdot AoI$. In a vector space model, a setup of the feature amount has big influence on results. Most of vector models delete stop words. However stop words are sometimes necessarily. Our method calculates weights for all words of input sentence. In this paper, we describe how to calculate a weight, and try to increase correct response number by changing parameter. Furthermore we describe how to apply to dialogue processing.

First we explain $tf \cdot idf$ in Chapter 2, and Chapter 3 describes this technique of our method. In Chapter 4 and Chapter 5, we describe the result of the evaluation experiment by the dialogue processing based on our method. Finally, we describe the effectiveness of our method and a future subject.

2 $tf \cdot idf$

In information retrieval, $tf \cdot idf (termfrequency \times inversedocumentfrequency)$ is used for calculating the weight of each word. Table.1 shows how to calculate $tf \cdot idf$. Each line represents one document, each row represents an indexing word. d_1 line includes each term frequency within the document d_1. The t_1 column shows the term frequency in each document. idf means the following formula

$$idf(t) = log\frac{N}{df(t)} + 1 \quad \cdots (2)$$

Table 1. Example of tf, df, idf.

	t_1	t_2	t_3	t_4
d_1	1	2	0	2
d_2	0	3	2	0
d_3	2	1	1	0
df	2	3	2	1
	$log_2 \frac{D}{df} + 1$			
idf	2.58	1.00	1.58	2.58

Table 2. Example of $tf \cdot idf$.

	t_1	t_2	t_3	t_4
d_1	2.58	1.00	0.00	5.16
d_2	0.00	3.00	3.16	0.00
d_3	5.16	1.00	1.58	0.00
df	2	3	2	1
idf	2.58	1.00	1.58	2.58

At this point, "+1" means a smoothing for taking account of $log\frac{N}{df(t)}$=0. Each term weight is calculated by $tf \times idf$. Table 2 shows the result of each weight of a term. For example, if indexing words could be t_1 and t_2, each document would then be as follows:

d_1 $2.58 + 1.00 = 3.58$
d_2 $0.00 + 3.00 = 3.00$
d_3 $5.16 + 1.00 = 6.16$

Next paragraph, we explain our method with the difference from $tf \cdot idf$.

3 Similarity of Euclidean Distance Using $tf \cdot AoI$

3.1 Basic Idea

Our method calculates similarity between two sentences. Although the comparison using matching rate measure does not deal with the difference between an important word and an unimportant word, our method calculates the weight of each word. We deal with spoken dialogue examples which include interjections and length expression. Since these expressions usually depend on a person, it is difficult to make stop word list. In this paper, we try to resolve this problem by using calculating a word weight. The word weight is multiplied term frequency and amount of information. Furthermore spoken dialogue examples often include inverted sentences. We try to resolve this problem by using a vector space model which has robustness.

By the way, a spoken dialogue system needs some basic research such as a parser and semantic analysis. However, it is difficult to clarify a problem of an erroneous response when a dialogue system uses many techniques. In this paper, we clarify the problem of our method by simplifying a problem. We prepare a data which consists of pairs between a question and an answer. And when our system finds the highest similarity pair between an input sentence and a question of the data, the answer of the selected pair is replied.

3.2 Similarity

We use Euclidean distance for our similarity measure. When there are $(\alpha, \beta) and (\gamma, \delta)$ in the vector space model, Euclidean distance becomes $\sqrt{(\alpha-\gamma)^2+(\beta-\delta)^2}$. Each element of the feature is calculated by multiplication between term frequency and amount of information. We call this weight $tf \cdot AoI$. Table 3 shows examples of calculating the similarity. First, a criteria sentence is divided into by the morphological tool[6]. We call an input sentence the criteria sentence. Here, our system does not use a domestic knowledge which depends on specific language because we consider that our system applies to spoken dialogue system and other languages. When "number of difference words" of the criteria sentence is "n", our system calculates by n dimensions. Sentence A's $tf(term frequency)$ is the word frequency of each word in the document. When different words are $(t_1, t_2, \cdots, t_n)$, features of tf become $tf(tf_1, tf_2, \cdots, tf_n)$ As the features are only the appearance words, tf of the criteria sentence occurs more than once. According to each weight of features, $tf \cdot AoI$ becomes$(tf_1 \cdot AoI(t_1), tf_2 \cdot AoI(t_2), \cdots, tf_n \cdot AoI(t_n))$. We assume that a sentence A is a comparison sentence for comparing with the criteria sentence. The tf of the comparison sentence A becomes $(tf_{a1}, tf_{a2}, \cdots, tf_{an})$. Euclidean Distance between the criteria sentence and the comparison sentence is calculated by $tf \cdot AoI$. The formula is as follows:

$$D = \sqrt{\sum_{i=1}^{n} (tf_i \cdot AoI(t_i) - tf_{ai} \cdot AoI(t_{ai}))^2} \quad \cdots (3)$$

$$Similarity = \frac{1}{D+1} \quad \cdots (4)$$

When there are not the difference between a criteria sentence and a comparison sentence, Euclidean distance becomes " 0 " . Therefore the highest similarity is " 1 " .

4 Experiment

4.1 Purpose of Experiment

The purpose of this experiment is to compare our method with other methods, and we clarify effectiveness of our method. First the compared method is the

Table 3. How to calculate a similarity.

criteria sentence	$(t_1$	$, t_2$	$, \cdots\cdots,$	$t_n)$
tf	$(tf_1$	$, tf_2$	$, \cdots\cdots,$	$tf_n)$
AoI	$(-log_2 \frac{tf_1}{N}$	$, -log_2 \frac{tf_2}{N}$	$, \cdots\cdots,$	$-log_2 \frac{tf_n}{N})$
$tf \cdot AoI$	$(tf_1 \cdot AoI(t_1)$	$, tf_2 \cdot AoI(t_2)$	$, \cdots\cdots,$	$tf_n \cdot AoI(t_n))$
Comparison sentence A	$(t_{a1}$	$, t_{a2}$	$, \cdots\cdots,$	$t_{an})$
tf_a	$(tf_a 1$	$, tf_a 2$	$, \cdots\cdots,$	$tf_a n)$
$AoI(t_a)$	$(-log_2 \frac{tf_{a1}}{N}$	$, -log_2 \frac{tf_{a2}}{N}$	$, \cdots\cdots,$	$-log_2 \frac{tf_{an}}{N})$
$tf_a \cdot AoI(t_a)$	$(tf_{a1} \cdot AoI(t_{a1})$	$, tf_{a2} \cdot AoI(t_{a2})$	$, \cdots\cdots,$	$tf_n \cdot AoI(t_{an}))$
Euclidean Distance	$\sqrt{\sum_{i=1}^{n} (tf_i \cdot AoI(t_i) - tf_{ai} \cdot AoI(t_{ai}))^2}$			
Similarity	$\frac{1}{EuclideanDistance+1}$			

Euclidean Distance by adding weight of $tf \cdot idf$. Second the compared method is the matching rate between a criteria sentence and a comparison sentence.

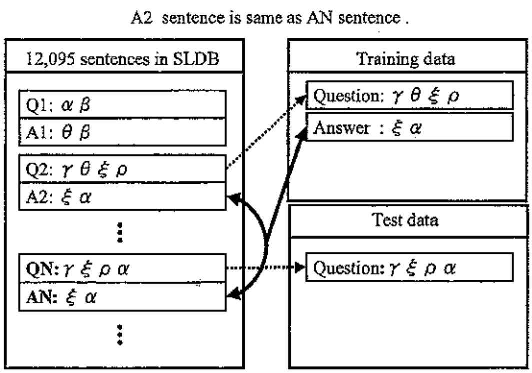

Fig. 1. How to collect data.

How to collect experiment data We would like to collect data which consist of two different questions of the same meaning and one answer for two questions. Therefore, we searched for the training data and the test data from SLDB[7] which includes 12,095 sentences. Figure.1 shows how to get the experiment data. In order to collect the experiment data, we use the following procedure:

Step1 Find the same sentences in SLDB, such as A2 and AN in Figure 1.
Step2 Determine the difference between the sentences communicated prior to the matching sentences such as Step1.
Step3 Having less than four different previous sentences because the meanings between previous sentences tend to become different in many same sentences.
Step4 Collect a set which consists of two different sentences that generated the same response by human observation.

Figure 2 shows examples of the collected data. Here, Question 1 and Answer are the training data, and Question 2 is the test data. These Questions does not necessarily become interrogative sentences in terms of the collected procedure. We could collect 173 sets from 12,095 sentences. However, They are not many set because there are many cases which have different meaning between previous sentences. 173 sets are regarded as the training data and the test data. The number of difference words is 745 and the number of running words is 5,190.

Question 1	先ほどチェックインしました百七号室の鈴木和子です。 (I'm Kazuko Suzuki . I've just checked in your room one o seven .)
Question 2	先ほどチェックインしました一〇七号室の鈴木和子と申しますが。 (My name is Kazuko Suzuki . I've just checked in to room one o seven .)
Answer	はい、鈴木様、どういった御用件でしょうか。 (Yes , Miss Suzuki , how can I help you ?)

Question 1	かしこまりました。そう致します。鈴木様、明朝のモーニングコールは差し上げましょうか。(Yes , of course , we'll certainly do so . And uh will you be requiring a wake-up call , Miss Suzuki ?)
Question 2	はい、えーではどなたかお見えになったら、お電話差し上げます。はい、えー明朝のモーニングコールはいかがなさいますか。 (Yes , I'll call you if someone should come to visit . Shall I give you a wake-up call tomorrow morning ?)
Answer	ええ、七時にお願いします。(Yes , at seven o'clock , please .)

Fig. 2. Examples of the collected data.

4.2 Experiment Procedure

Figure 3 shows how to reply using our similarity measure which calculates Euclidean Distance using $tf \cdot AoI$. Here, the training data consists of 173 pairs between a question and an answer. In the foregoing paragraph, we collected such a open data. When our system calculates the Euclidean Distance between a question of the test data and an input sentence of training data, the system replies the answer of high similarity which is the nearest distance in the training data. If the system could find a suitable answer for each question, we evaluate the correct answer. If not, it becomes an erroneous answer. The comparison experiment is conducted by the three following method.

- Similarity of Euclidean Distance using $tf \cdot AoI$
- Similarity of Euclidean Distance using $tf \cdot idf$
- Matching rate between an input sentence and a comparison sentence.

Similarity of Euclidean Distance using $tf \cdot AoI$ is our method. Here, an input sentence is treated as a criteria sentence. Similarity of Euclidean Distance using $tf \cdot idf$ is different in terms of the weight. Matching rate measure is selected by high matching rate.

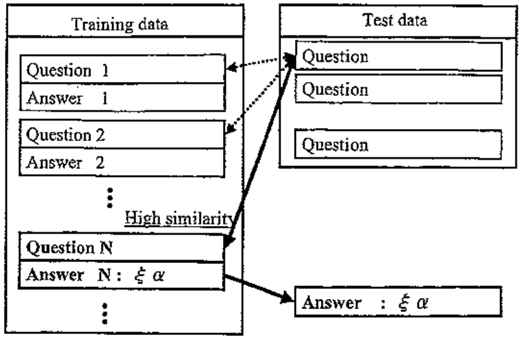

Fig. 3. Dialogue processing.

4.3 Experiment Result

Table 4 shows the experiment results. Here, a correct answer is defined as the correct response which has the only best value since the system has to choose the best answer when there are two or more responses which have the best same value. Our method performed the correct selection of 112 answers for 173 questions The correct response rate of our method had 64.7%, and we confirmed 9.2 points higher than that of the matching rate measure which has 55.5%. Furthermore the correct response rate of our method was 4.6 points higher than that of the similarity which uses Euclidean Distance using $tf \cdot idf$. In the matching rate measure, there are 3 correct responses which our method cannot select correctly. In the similarity which uses Euclidean Distance using $tf \cdot idf$, there are 3 correct responses which our method cannot select correctly. We consider that our system's ability includes other method's merit because there are not almost the correct responses which our system could not select in the correct responses of other systems.

Table 4. Experiment result

	Our method	$tf \cdot idf$	Matching rate measure
Correct response number	112/173	104/173	96/173
Correct response rates	64.7%	60.1%	55.5%
Correct response number which excludes our correct responses	-	3	3

<table>
<tr><td>Input</td><td colspan="23">はい、三人部屋でしたらございますが、あしかしえーお子様の年齢はおいくつですか。(Well Yes we do have rooms ah for three people. However , um how old is your child?)</td></tr>
<tr><td>Sentence A</td><td colspan="23">え お子様 は お 幾つ で いらっしゃい ます か 。(And how old is your child ?)</td></tr>
<tr><td>Sentence B</td><td colspan="23">はい 、そうします と、うん「のぞみ九号」博多行きというの が 有ります ね。こちらの 電車です と、東京を九時五十六分に出まして、京都には 十二時十一分に着きます 。お 値段の 方は 乗車券と特急券を併せて、お 一人 一万三千九百二十円です ね。(I see , then , we have a Nozomi number nine that's bound for Hakata. This one leaves Tokyo at nine fifty-six and arrives at Kyoto at twelve eleven. The price will be thirteen thousand nine hundred and twenty yen per person , which includes both the boarding ticket and the limited express ticket.)</td></tr>
<tr><td>Words</td><td>はい</td><td>、</td><td>三</td><td>人</td><td>部屋</td><td>でし</td><td>たら</td><td>ござい</td><td>ます</td><td>が</td><td>あ</td><td>しかし</td><td>えー</td><td>お子様</td><td>の</td><td>年齢</td><td>は</td><td>お</td><td>いくつ</td><td>です</td><td>か</td><td>。</td></tr>
<tr><td>f(x)</td><td>87</td><td>262</td><td>42</td><td>13</td><td>2</td><td>4</td><td>9</td><td>101</td><td>221</td><td>71</td><td>9</td><td>0</td><td>14</td><td>1</td><td>140</td><td>0</td><td>83</td><td>62</td><td>0</td><td>156</td><td>118</td><td>530</td></tr>
<tr><td>tf_input</td><td>1</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>tf_A</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>tf_B</td><td>1</td><td>5</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>2</td><td>2</td><td>0</td><td>2</td><td>0</td><td>3</td></tr>
</table>

Fig. 4. Different example of a response selection.

4.4 Consideration

Figure 4 shows the difference between the matching rate measure and our similarity. The example of input sentence means "Well yes we do have room ah for three persons. However , um how old is your child?". This sentence becomes the criteria sentence, and includes filler such as 「あ しかし えー」.[3] Our method selected Sentence A which means "And how old is your child?", and this sentence is correct answer. The matching rate measure selected $SentenceB$ which means "I see , then , we have a Nozomi number nine that's bound for Hakata. This one leaves Tokyo at nine fifty-six and arrives at Kyoto at twelve eleven. The price will be thirteen thousand nine hundred and twenty yen per person , which includes both the boarding ticket and the limited express ticket." Here, a denominator of the matching rate is words number of the criteria sentence. Notice that the maximum word number of the denominator depends on the criteria sentence. For example, although 22 words of $SentenceB$ are matching for the criteria's words of Figure 4, the matching words become 12 words. Therefore the matching rate of $SentenceA$ is $\frac{6}{23} = 26.1\%$, and that of $Sent_B$ is $\frac{12}{23} = 52.2\%$. Here, we explain about calculation of our method. The criteria ' s frequency of " はい (Yes) " is "1", and $SentenceA$ ' s frequency of it is "0". AoI of " はい (Yes) " is $-log_2 \frac{87}{5190}$ because the number of running words is 5,190 and frequency of " はい (Yes) " is 87. The calculation expression is as follows:

$$\sqrt{((1-0) \cdot -log_2 \frac{87}{5190})^2 + ((2-0) \cdot -log_2 \frac{262}{5190})^2 + \cdots}$$

[3] The underline means filler.

The calculation expression for $Sent_B$ is as follows:
$\sqrt{((1-1)\cdot -log_2\frac{87}{5190})^2 + ((1-5)\cdot -log_2\frac{262}{5190})^2 + \cdot\cdot}$
In our method, the similarity for $Sent_A$ was 0.0271(35.9), the similarity for $Sent_B$ was 0.0240(40.6). Here, our method could select correctly using "$tf \cdot AoI$". Although there were many same values in the calculation of the matching rate measure, our system could calculate difference for each sentence.

There were 47 questions which three systems could not select correctly. We collected 173 sets according to processing of 4.1. Therefore it is an example as follows Figure 5.

Here, Question 1 and Response became the training data, Question 2 became the test data. The similarity between Question 1 and Question 2 was 0.0777(11.87). However there were the sentences which have very similarity for Question 2. In other words, their sentences were almost the same as Question 2. For example, there was " はい。えーじゃ、よろしくお願いします。". This sentence is almost same as Question2. Most of erroneous results included this reason. In this experiment, we collected the experiment data fairly. Therefore we consider this erroneous reason shows fairness of the experiment data.

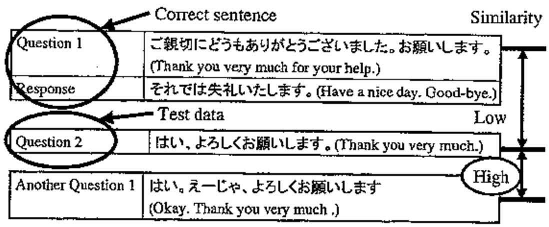

Fig. 5. The reason of erroneous responses.

The following paragraph explains the parameter setting in order to improve our method.

5 Parameter Evaluation Experiment

5.1 Purpose of Experiment

The calculation result using our method changes by setting up a criteria sentence and adding a parameter. In this experiment, we try to improve correct response rate by changing them. First, There are two Euclidean distances by the criteria sentence. Two distances are as follows:

D_1 A criteria sentence is an input sentence.
D_2 A criteria sentence is each sentence in the training data.

In the case of D_1, feature number becomes the number of difference words in a input sentence, On the other hand, feature number of D_2 change by each sentence in the training data. For example, we assume that Ex.1 is "I am a boy" and Ex.2 is "I am Sam". Feature number becomes four when Ex.1 is set to the criteria sentence. It becomes three when Ex.2 is set to the criteria sentence. Thus, a similarity is difference between D_1 and D_2. Therefore we reflect two distances as follows:

$$D_1 + \alpha \times D_2$$

In this experiment, we find the best value of the coefficient α.

Table 5. Experiment results.

	D_1	D_2	D_1 + D_2
Correct response number	112/173	28/173	83/173
Correct response rates	64.7%	16.2%	48.0%
Correct response number which exlude our correct responses	-	7	8

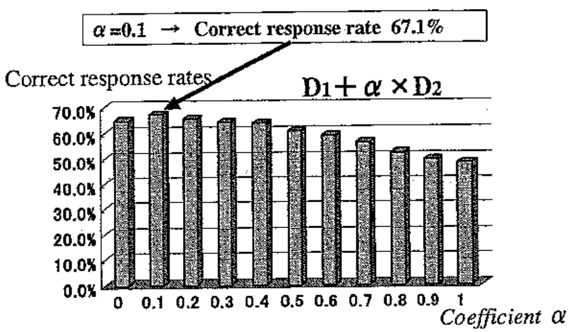

Fig. 6. Evaluation experiment of α .

5.2 Experiment Results and Consideration

Table 5 shows the comparison result by changing the criteria sentence. We added the result of D_1+1×D_2 in Table 5. The correct response rate of D_1 was the highest in three calculation method. That of D_1 was the worst of three, and correct response number became 28 responses. However, there were 7 correct

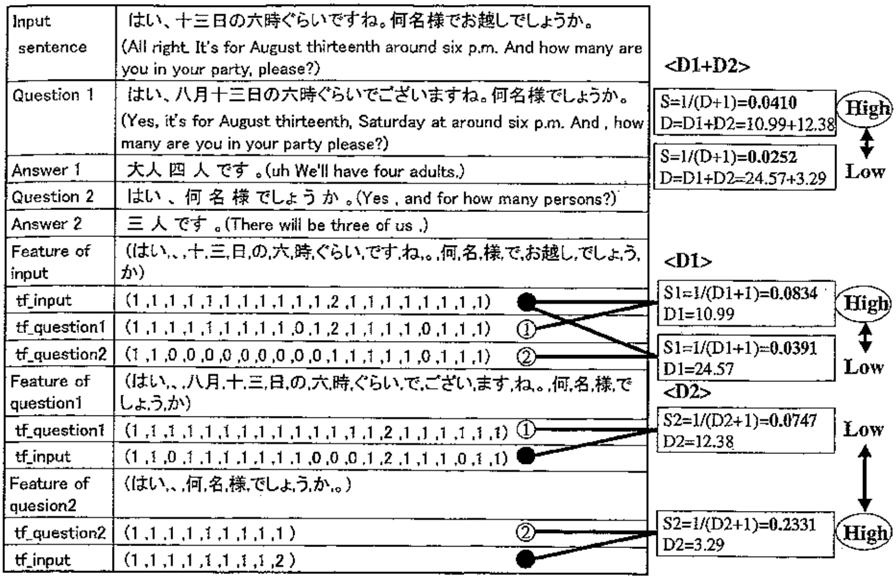

Fig. 7. Example of a calculation.

responses which does not include correct responses of D_1. We can confirm there were the different selection between D_1 and D_2.

Here, we explain three methods in Figure 7. We assume that an input sentence is "はい、十三日の六時ぐらいですね。何名様でお越しでしょうか". When the input sentence is the criteria sentence, feature becomes (はい,、, 十, 三, 日, の, 六, 時, ぐらい, です, ね,。, 何, 名, 様, で, お越し, でしょ, う, か). Therefore, the vector is expressed like (1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,2 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1). The vector of Question 1 is (1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,0 ,1 ,2 ,1 ,1 ,1 ,1 ,0 ,1 ,1 ,1) because the Question 1 is "はい、八月十三日の六時ぐらいでございますね。何名様でしょうか。" However, when Question 1 is the criteria sentence, features become (はい,、, 八月, 十, 三, 日, の, 六, 時, ぐらい, で, ござい, ます, ね,。, 何, 名, 様, でしょ, う, か). The vector of Question 1 is expressed like (1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,2 ,1 ,1 ,1 ,1 ,1 ,1). The vector of the input sentence is expressed like (1 ,1 ,0 ,1 ,1 ,1 ,1 ,1 ,1 ,1 ,0 ,0 ,0 ,1 ,2 ,1 ,1 ,1 ,0 ,1 ,1). In the case of D_1, Question1 was higher than Question 2 because the similarity between Question1 and the input sentence became 0.0834(10.99)[4], the similarity between Question2 and the input sentence became 0.0391(24.57). In the case of D_2, Question2 was higher than Question 1 because the similarity between Question1 and the input sentence became 0.0747(12.38), the similarity between Question2 and the input sentence became 0.2331(3.29). In the case of " $D_1 + D_2$ " , Question1 was higher than

[4] A value in the parenthesis is the Euclidean distance.

Question 2 because the similarity between Question1 and the input sentence became 0.0410(23.37), the similarity between Question2 and the input sentence became 0.0347(27.86). Although the selection of D_1 is same as that of D_1+D_2 in this example, we confirmed the correct response rates fall by the result of D_2. The results of D_2 tend to become high similarity when Question 2 is shorter than Question 1. From this result, it is possible to improve correct response rate by changing the coefficient.

We find the best coefficient value of $D_1+\alpha\times D_2$. We changed the coefficient from 0 to 1. Figure 6 shows the experiment result. The best correct response rate was 67.1% when α was 0.1. We confirmed that correct response rate of our method was 11.6 points higher than that of the matching rate between an input sentence and a comparison sentence.

6 Conclusion

We proposed Euclidean Distance using $tf \cdot AoI$ for similarity of dialogue processing. We compared with "Similarity of Euclidean Distance using $tf \cdot iaf$" and "Matching rate". As a result, the correct response rate of our method had 64.7%, and we confirmed 9.2 points higher than that of the matching rate measure which has 55.5%. Furthermore we improved our method. The best correct response rate was 67.1% when α was 0.1. We confirmed that correct response rate of our method was 11.6 points higher than that of the matching rate between an input sentence and a comparison sentence.

References

1. J. Weizenbaum, "ELIZA - A Computer Program for the Study of Natural Language Comunication Between Man And Machine," Communications of the Association for Computing Machinery, vol.9, no.1, pp.36-45, 1966.
2. Takebayashi, Y., H. Tsuboi, H. Kanazawa, Y. Sadamoto, H. Hashimoto, and H. Shinichi. 1993. A real-time speech dialog system using spontaneous speech understanding. IEICE Trans. Inf. and Syst. E-76D(1)(Jan).
3. Victor Zue, Stephanie Seneff, James Glass, Joseph Polifroni, Christine Pao, Timothy J. Hazen and Lee Hetherington, "JUPITER: A telephone-based conversational interface for weather information," IEEE Transactions on Speech and Audio Processing, Vol. 8 , No. 1, January 2000.
4. Diane Litman, Shimei Pan and Marilyn Walker, "Evaluating Response Strategies in a Web-Based Spoken Dialogue Agent." In Proceedings of ACL/COLING 98 , 1998.
5. A. Aizawa, "The Feature Quantity: An Information Theoretic Perspective of Tfidf-like Measures," proc. of ACM SIGIR 2000, pp.104–111, 2000.
6. Japanese Morphological Analysis System ChaSen version 2.2.1" ,Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka, Masayuki Asahara, Dec, 2000.
7. T. Morimoto, N. Uratani, T. Takezawa, O. Furuse, Y. Sobashima, H. Iida, A. Nakamura, Y. Sagisaka, N. Higuchi and Y. Yamazaki, "A speech and language database for speech translation research," Proc. ICSLP'94, pp.1791–1794, 1994.