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, Hokkaido-Gakuen University, Asahi-machi 4-1-46, Toyohira-ku, Sapporo-shi, 062-8625 Japan  
tochinai@econ.hckkai-s-u.ac.jp

Abstract. In this paper, we propose a Japanese dialogue processing method based on a similarity measure using \( tf \cdot AoI(\text{term frequency} \times \text{Amount of Information}) \). 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 12,995 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 7.6 points higher than that of \( tf \cdot iaf \).

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 measures 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 $i\alpha f$ 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, $tf \cdot i\alpha f$ is used widely[5]. $tf \cdot i\alpha f$ means multiplying term frequency and inverse document frequency. The value of $tf \cdot i\alpha f$ 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 $i\alpha f$.

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 $tf \cdot i\alpha f$ 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 $tf \cdot i\alpha f$

In information retrieval, $tf \cdot i\alpha f (\text{term frequency } \times \text{inversedocument frequency})$ is used for calculating the weight of each word. Table.1 shows how to calculate $tf \cdot i\alpha f$. 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. $i\alpha f$ means the following formula

$$i\alpha f(t) = log \frac{N}{f(t)} + 1 \quad \cdots (2)$$