A Proposal of Method of Classifying Words with Diseases for Development of Hospitality in Health Care

Hiroki Kozu, Yukio Maruyama, Tusyoshi Yuyama and Tomoya Hasegawa

Abstract— In recent years, the hospitality in health care is very important service for patients and their families who have some anxious factors with diseases. However, it is difficult that medical staff such as medical doctors and nurses grasp some potential anxious factors of patients and their families in the medical examination and treatment. The purpose of this study is to propose the methods of classifying potential anxious factors for patients and their families not grasped by medical staffs. In this paper, several methods of classifying potential anxious factors for patients and their families are examined using writing contents by Japanese users with diseases on web site of "Yahoo! Answers". At first, the words are extracted from writing contents of each disease by morpheme analysis. Next, the classification of the extracted words is obtained by (1) counting number of occurrences of each word, (2) counting appearance ratio of each word, (3) using correlation analysis, (4) counting number of questions containing each word, and (5) counting number of combinations of each word in the questions. Moreover, obtained correlation between word and each disease is judged according to knowledge of medical staff. As a result, the method of classification of the words by "number of combinations of each word in the questions" shows a tendency to agree with knowledge of medical staff comparing the words classified by other methods.

Index Terms— Hospitality in health care, Potential anxious factors, Morpheme analysis

I. INTRODUCTION

In recent years, a health care service is now changing from "curing" to "looking after" for development of QOL. (Quality of Life) [1]. This "looking after" means developing hospitality such as cooperated medical system within medical team, priority medical system for patients and so on. Therefore, the hospitality in health care is very important service for patients and their families who have some anxious factors with diseases. In previous studies, a questionnaire survey was conducted to grasp some anxious factors of

patients and their families [2] - [4]. The role of nurses and clinical psychologists was reported to improve some anxious factors of patients [5], [6]. In addition, the background factors affecting QOL of patients are measured by a multiple regression model [7]. It is possible to grasp the superficial anxiety of patients and their families, but it is considered difficult to grasp the potential anxiety. This study focuses on the potential anxious factors with diseases of patients and their families. Potential anxiety can be searched by "knowledge search service" on web site. "Knowledge search service" corresponds to "Yahoo! Answers", "OKWAVE" and "Sooda!".

The purpose of this study is to propose the methods of classifying potential anxious factors for patients and their families which not grasped by medical staff. In this paper, several methods of classifying potential anxious factors for patients and their families are examined using writing contents by Japanese users with diseases on web site of "Yahoo! Answers". Moreover, the classified potential anxious factors for patients and their families are judged according to knowledge of medical staff.

II. OUTLINES OF KNOWLEDGE SEARCH SERVICE

A. Overview of "Yahoo! Answers"

The knowledge search service on web site such as "Yahoo! Answers" is a collection of many potential anxious factors. This service is constructed by various questions for anxiety and answers of many users. 13,000,000 questions are accumulated in "Yahoo! Answers" and the number of questions is larger compared to other knowledge search services (as of January 2015) [8].

B. Used data of "Yahoo! Answers"

This study is focused on the questions in "Yahoo! Answers" for classifying potential anxious factors of disease. The answers of "Yahoo! Answers" are not used, only the questions are used. In this study, the questions for three kinds of disease ("High Blood Pressure", "Kidney Disease" and "Diabetes Mellitus") are extracted as datasets from "Yahoo! Answers". The data collection period of each disease is as follows. In case of "High Blood Pressure", the questions were collected from August 19, 2016 to October 18, 2016. In case of "Kidney Disease", the questions will be collected from November 20, 2014 to November 19, 2017. Also, In the case of "Diabetes Mellitus", the questions will be collected from March 20, 2017 to November 19, 2017.

III. PROPOSAL METHOD OF CLASSIFYING WORDS WITH DISEASES

In this session, the procedure of the proposed method is shown in Figure 1. As shown in Figure 1, this proposal method consists of four steps. Each step is described.

A. Dataset of disease (Step1)

In step1, the questions for three kinds of disease ("High Blood Pressure", "Kidney Disease" and "Diabetes Mellitus") are collected as a dataset from "Yahoo! Answers".

B. Morpheme analysis and Extract words such as nouns (Step2 and Step3)

In this session, procedure of step2 and 3 are described. The words are extracted from writing contents of each disease by morpheme analysis [9].

Result of extract words such as nouns by morpheme analysis is shown in Table 1. As shown in Table 1, the words such as nouns are extracted from writing contents of three kinds of major disease by morpheme analysis. For "High Blood Pressure", 4,530 words such as 'Symptom', 'Blood Pressure', 'Hospital' and so on are extracted from 1,000 question contents. For "Kidney Disease", 4,248 words such as 'Disease', 'Inspection', 'Hospital' and so on are extracted from 631 question contents. And for "Diabetes Mellitus", 4,291 words such as 'Meal', 'Blood glucose', 'Inspection' and so on are extracted from 1,000 question contents. In this study, the words are extracted by using KH Coder 2.00f (third-party software).
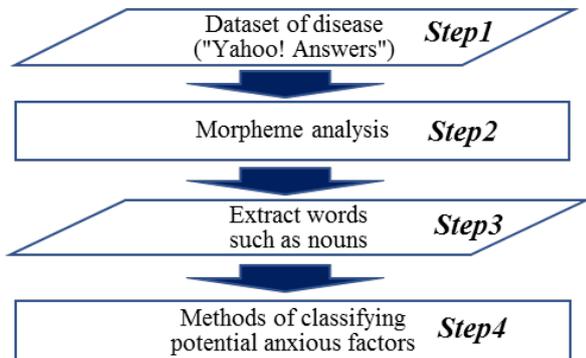

Figure 1: Analytical procedure

Table 1: Result of morpheme analysis

High Blood Pressure	Kidney Disease	Diabetes Mellitus
Symptom	Disease	Meal
Blood Pressure	Inspection	Blood glucose
Hospital	Hospital	Inspection
Inspection	Kidney	Hospital
Diagnosis	Insurance	Pregnancy

C. Methods of classifying potential anxious factors (Step4)

In step4, the classification of an extracted word is obtained by number of occurrences of each word (Method1), appearance ratio of each word (Method2), correlation analysis (Method3), number of questions containing each word (Method4) and number of combinations of each word in the questions (Method5).

The procedure of Method1 is shown in Figure 2. As shown in Figure 2, the number of occurrences of each word extracted by using morpheme analysis is calculated. For example, when 'Blood pressure' is used 3 times, the number of occurrences of 'Blood pressure' is 3.

In Method2, the appearance ratio of each word extracted by using morpheme analysis is defined in the following equation.

$$\text{(Appearance ratio)} = \frac{\text{(Number of counts per word in each question)}}{\text{(Total number of words in each question)}} \quad (1)$$

In Method3, the correlation coefficient among each word is calculated by using appearance ratio of Method2.

The procedure of Method4 is shown in Figure 3. As shown in Figure 3, the Number of questions containing of each word extracted by using morpheme analysis is calculated. For example, when there are three questions that included 'Inspection', the number of questions is 4.

The procedure of Method5 is shown in Figure 4. As shown in Figure 4, the number of combinations of each word in questions extracted by using morpheme analysis is calculated. For example, when 'hospital' and 'Inspection' appear both in question No. 3 and No. 4, the number of combinations of them is 2.
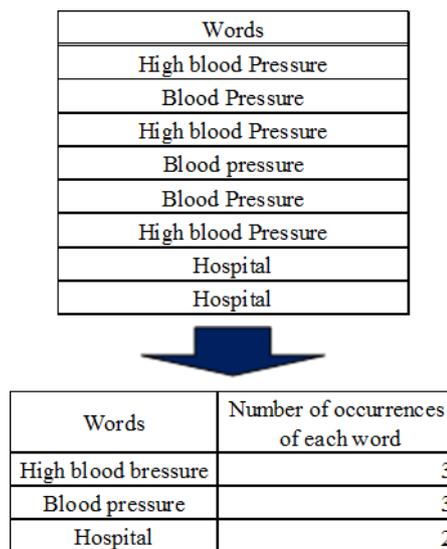
Figure 2: Number of occurrences of each word

Table 2: Classifying words according to knowledge of medical staff (Excerpt)

High Blood Pressure		Kidney Disease		Diabetes Mellitus	
Strong relationship	Weak relationship	Strong relationship	Weak relationship	Strong relationship	Weak relationship
Low	Inspection	Protein	Chronic	Blood glucose	Dialysis
Glycosuria	High	Blood	Meal	Life	Family
Pregnancy	Diagnosis	Glycosuria	Bladder	Kidney	Stress
Heart	Hospitalization	Dialysis	Blood vessel	Insulin	Rice
Hospital	Reason	Bleeding	Water	Weight	Snack foods

Question No.	Contents
1	*High blood pressure* , …*inspection* …consultation
2	*Hospital* …**inspection**…consultation…
3	…*inspection* …no abnormality…
4	*High blood pressure* …Diagnosis…
5	*Hospital* …*instection* …
6	Degradation of diabetes...Diagnosis

Words	Number of questions containing each word
Inspection	4
High blood pressure	2
Hospital	2

Figure 3: Number of questions containing each word

Question No.	High blood pressure	Blood pressure	Hospital	Inspection
1	○		○	
2	○	○		○
3	○		○	○
4	○		○	○
5	○			

	High blood pressure	Blood pressure	Hospital	Inspection
High blood pressure	100	1	3	3
Blood pressure	-	80	0	1
Hospital	-	-	50	2
Inspection	-	-	-	301

Figure 4: Number of pairs of each word in questions

IV. EVALUATION OF PROPOSAL METHOD

A. Knowledge of medical staff

Result of selected words by knowledge of medical staff is shown in Table 2. As shown in Table 2, each word extracted using morpheme analysis is classified into 2 levels (Strong relationship and Weak relationship) according to knowledge of medical staff. In the case of "High Blood Pressure", the words 'Low', 'Glycosuria', 'Pregnancy' and so on have a strong relationship with the disease, and 'Inspection', 'High', 'Diagnosis' and so on have a weak relationship with the disease. In the case of "Kidney Disease", the words 'Protein', 'Blood', 'Glycosuria' and so on have a strong relationship with the disease, and 'Chronic', 'Meal', 'Bladder' and so on have a weak relationship with the disease. In the case of "Diabetes Mellitus", the words 'Blood glucose', 'Life', 'Kidney' and so on have a strong relationship with the disease, and 'Dialysis', 'Family', 'Stress' and so on have a weak relationship with the disease. The words strongly related to diseases are defined as being able to imagine diseases and relevance other than the medical staff. And the words weakly related are defined as difficult for people other than medical staff to understand the relationship with diseases. The medical staff who provided the opinions here are specialists in "High Blood Pressure", "Kidney Disease" and "Diabetes Mellitus".

B. Comparison of each word extracted and knowledge of medical staff

Result of comparison result between words extracted in relation to each disease and knowledge of medical staff is shown in Table 3, Table 4 and Table 5. For each disease, the classification words of the proposed method have been compared with the knowledge of medical staff. In any disease, the proposed method (see Method4 below) shows a characteristic tendency as comparing with other proposed methods (see Method1, Method2, Method3 and Method5 below). As shown in Table 3, in the case of "High Blood Pressure", the words 'High' and 'Low' are included in the top rank. As shown in Table 4, cases of "Kidney Disease", 'Kidney' and 'Insufficiency' are not included in the top rank. As shown in Table 5, in the case of "Diabetes Mellitus" the words 'Blood glucose' is not in top rank, but 'Life' is included in top rank.

In the top 50 words classified by the proposed method, the content ratio of words considered to be "Strong relationship" or "Weak relationship" by medical staff is defined as equation (2) and (3) below.

Table 6: Comparison of each word extracted and knowledge of medical staff

Methods	High Blood Pressure		Kidney Disease		Diabetes Mellitus	
	Strong relationship	Weak relationship	Strong relationship	Weak relationship	Strong relationship	Weak relationship
1	0.636	*0.609*	*0.778*	0.500	0.733	*0.800*
2	0.636	*0.609*	*0.778*	0.500	0.733	*0.800*
3	0.636	0.478	*0.778*	0.400	*0.800*	*0.800*
4	*0.682*	*0.609*	0.556	*0.600*	*0.800*	*0.800*
5	*0.682*	0.522	*0.778*	0.400	*0.800*	0.600

Table 3: Comparison result between words extracted in relation to "High Blood Pressure" and knowledge of medical staff (Excerpt)

Method1	Method2	Method3	Method4	Method5
High blood pressure	High blood pressure	High blood pressure	*Hight*	*Blood pressure*
Blood pressure	*Blood pressure*	*Blood pressure*	*Low*	High blood pressure
Hospital	*Hospital*	*Hospital*	*Blood*	*Hospital*
Inspection	*Inspection*	Request	*Inspection*	Request
Request	Request	*Inspection*	Confinement	*Inspection*
Symptom	*Symptom*	*Diagnosis*	*Pregnancy*	*Diagnosis*
Diagnosis	*Diagnosis*	Self	*Work*	*Factor*
Self	Self	*Symptom*	Mentalis	*Glycosuria*
Hospitalization	*Hospitalization*	*Factor*	*Blood pressure*	Self
Glycosuria	*Glycosuria*	Disease	Child	Question

: Strong relationship   : Weak relationship

Table 5: Comparison result between words extracted in relation to "Diabetes Mellitus" and knowledge of medical staff (Excerpt)

Method1	Method2	Method3	Method4	Method5
Glycosuria	Glycosuria	Glycosuria	Seeking diagnosis	Inspection
Blood glucose	*Blood glucose*	Request	*Life*	Glycosuria
Inspection	Inspection	Hospital	*Kidney*	*Blood glucose*
Hospital	Hospital	*Blood glucose*	Grandmother	Hospital
Request	Request	Inspection	Parents	Request
Self	Self	*Meal*	Work	Symptom
Meal	*Meal*	Self	Condition	*Pregnancy*
Family	*Family*	*Family*	Inter net	*Calory*
Relationship	Relationship	Medical doctor	Feeling	Talks
Stress	*Stress*	*After meals*	Prescription	Relationship

: Strong relationship   : Weak relationship

Table 4: Comparison result between words extracted in relation to "Kidney Disease" and knowledge of medical staff (Excerpt)

Method1	Method2	Method3	Method4	Method5
Patient	Patient	Patient	Custom	Request
Inspection	Inspection	Request	Life	Patient
Hospital	Hospital	*Insufficiency*	function	*Insufficiency*
Kidney	*Kidney*	Inspection	Lowering	Inspection
Insufficiency	*Insufficiency*	Hospital	Cancer	Hospital
Insurance	Insurance	*Kidney*	Life-long	*Kidney*
Hospitalization	Hospitalization	Function	Medical	function
Life	Life	*Dialysis*	*Insufficiency*	Lowering
Meal	*Meal*	Answer	*Chronic*	Patient
Self	Self	Lowering	*Albuminoid*	Life

: Strong relationship   : Weak relationship

(Content ratio of "Strongly related words") =

$$\frac{\text{(Number of “Strongly related words” in the top 50 words)}}{\text{(Number of "Strongly related words")}} \quad (2)$$

(Content ratio of "Weak related words") =

$$\frac{\text{(Number of “Weak related words” in the top 50 words)}}{\text{(Number of "Weak related words")}} \quad (3)$$

Result of comparison of each word extracted and knowledge of medical staff is shown in Table 6. As shown in Table 6, the case of "High Blood Pressure", it is found that the highest content ratio of the strong relationship words is the proposed Method4 and Method5, and the highest content ratio of the weak relationship words is proposed to be Method1, Method2 and Method4. And in the case of "Kidney Disease", the most highest content ratio of the strong relationship words is the proposed Method1, Method2, Method3 and Method5, and that of the weak relationship words, the proposed method4 is the most highest. Last in the

case of "Diabetes Mellitus", it is found that the most highest content ratio of the strong relationship words is proposed to be Method3, Method4 and Method5, and that of the weak relationship words, the proposed Method1, Method2, Method3, and Method4 is the most highest. Therefore, the proposed Method4 is shown to be consistent with the knowledge of medical staff.

V. CONCLUSION

The purpose of this research is to propose a classification method of potential anxious factors of patients and the families which medical staff do not grasp. In this paper, several classification methods of concerned potential anxious factors have been analyzed using sentences contents of Japanese users with diseases on web site of "Yahoo! Answers". The classification of potential anxious factors of patients and their families is judged based on the knowledge of the medical staff. In this study, there are following two points by examination results.

1. The proposed Method4 tends to have more words matching with the knowledge "Strong relationship" and "Weak relationship" of the medical staff.

2. The words classified by the proposed methods and the knowledge of medical staff tend to be different depending on the disease.

The definition of "Strong relationship" is a term indicating the relevance to diseases even if it is not medical staff. Also, the definition of "Weak relationship" is a term difficult to understand in relation to diseases other than medical staff. In this research, it shows that words with "Weak relationship" are important to understand potential anxious factors. Therefore, the proposed Method4 can be the considered as a good classification.

On the other hand, words such as 'Insurance' or 'Life-long' in "Kidney disease" are not included in knowledge of medical staff, but it is considered important to grasp potential anxious factors of patients and their families indirectly. Therefore, it is possible to explain that the classified words contain important terms which medical workers do not grasp. Future tasks are as follows.

1. Consider words which are not included in the knowledge of medical staff.

2. Classify potential anxious factors by using component analysis, factor analysis.

3. Consider the effects on the differences with diseases.

REFERENCES

[1] T. Mono, Medical Marketing (Japan edition). Tokyo, Japan hyoron-Sha, 2011.

[2] M. Tanaka, K. Ooi, K. Yanagisawa, T. Yoshioka, T. Gomi, T. Miyashita1 and S. Takamizawa, "Benefits of Pharmacists' Assessment of Depression and Anxiety in Outpatients Undergoing Chemotherapy," J. Pharm. Health Care Sci., vol. 34, no.12, 2008, pp. 1086-1090

[3] A. Takei, T. Itou, T. Kanou, J, Onozeki, M, Maeda, S. Tutumi, T. Asao, H. Kuwano and K, Kanda, "Analysis of Anxiety in Cancer Patient Received Outpatient Chemotherapy", Kitakanto Med J, vol.55, 2005, pp.133-139

[4] M. Hisamatsu and S. Niwa, "Support Factors of Coping with Anxiety in Families of Patients with Terminal Cancer", J. Jpn. Acad. Nurs. Sci., vol.31, no.1, 2011, pp.58–67

[5] N. Yasuda, H. Ito, H, Mori, M, Ushiro, M. Miura, Y, Odaka, Y. Ishimaru, K. Hashimoto and M. Miyazaki, "Nursing elderly hemodialysis patients following surgery under general anesthesia", Journal of Japanese Society for Dialysis Therapy, vol.24, no.8, 1991, pp.1167-1170

[6] N. Nagai, T, Nomura, T, Morimoto, and Y. Sasaki, "A Role of Clinical Psychologists in Terminal Care for Cancer Patients," Palliat Care Res, vol.11, no.3, 2016, pp.534-37

[7] H. Sano, K. Asao, M. Matsushima, T. Agata, M. Kusaka, T. Sasaki, Y. Tanishima, I. Yamamoto, H. Shimizu and N. Tajima, "Measurement of Quality of Life in Patient with Type 2 Diabetes (2): The Influence of Patients' Profiles and Diabetic Complications on Quality of Life", J. Japan Diab. Soc. vol.44, no. 1, 2001, pp.57-62,

[8] Core Marketing Blog. Keyword analysis from Yahoo answers. Available: https://blog.core-j.co.jp/keyword-analytics-yahoo-chiebukuro

[9] K. Higuchi, Quantitative text analysis for social researchers: a contribution to content analysis. Tokyo, Nakanishi Press, 2014