Keyword Extraction for Local Foods from Restaurant Menus of Roadside Stations

Yoko Nishihara
College of Information Science and Engineering
Ritsumeikan University
Shiga, Japan
email: nisihara@fc.ritsumei.ac.jp

Hirofumi Noguchi
College of Information Science and Engineering
Ritsumeikan University
Shiga, Japan

Ryosuke Yamanishi
Faculty of Informatics
Kansai University
Osaka, Japan

***Abstract*—Some foods are only consumed and available in some regions; these foods are called local foods. For example, B.C. roll is one of the local foods in Canada. The local foods are a form of tourism and are expected to attract more tourists. People in the region are so accustomed to the local foods that they may have difficulty in identifying their values, which means they are unsure whether a particular food is local or not. If they success to choose appropriate local foods, they may get more tourists. Therefore, they should know which food should be sold as a local food. Previous studies have proposed statistical methods for extracting keywords for local foods from restaurant menus on the Internet. However, the restaurant menus often exclude local foods. This study applies the previous statistical method to restaurant menus of roadside stations and extracts keywords for their local foods. These roadside stations are government-designated rest areas located along Japanese roads. They sell local foods for promotion, while the restaurants provide menus of the local foods. Thus, if we apply the statistical method to the restaurant menus of roadside stations, we may obtain menus of the local foods. First, we developed a dataset of restaurant menus of roadside stations. Then, we apply the previous method to the dataset to extract keywords for the local foods. We invited participants to our experiment and they evaluated whether the extracted keywords were related to the local foods. The average rate of keywords for the local foods was 21.1%. Furthermore, we discovered that the extracted keywords were not only for foods but also place names, dish names, and their combinations.**

Keywords-local food; roadside station; restaurant menu; food tourism; TF-IDF

I. INTRODUCTION

People enjoy not only sightseeing on their vacation but also food tourism[1][2]. Some foods are only eaten and available in specific regions. These are called local foods. B. C. roll (a kind of sushi containing barbecued salmon and cucumber), Nanaimo bar (butter tarts), and Rocky road (chocolate ice cream with almond and marshmallow) are examples of food only eaten in Canada. Lanzhou Ramen (a kind of Chinese noodle), Tianjin Xiao Long Bao (a type of soup dumpling), and Peking duck (roast duck), which are local Chinese foods, can be found in foreign countries. However, if people want to experience their authentic taste, they would have to get the dishes from their original region. Local foods have been receiving a lot of attention in recent years: the example is the Slow Food movement in Italy [3]. The Slow Food Association in Italy published guidebooks titled "Osterie D'Italia," which list restaurants offering local foods. Tourists expect to enjoy the local food on their vacations. The tourism industry must advertise local foods that can be eaten in their region and consider how to make them appealing to tourists [4].

Although local foods are only eaten and available in specific regions, people may be confused about which foods are local. This is because people in these regions are accustomed to these foods that they are unaware of their origin. If they success to choose appropriate local foods, they may get more tourists. Therefore, they should know which food should be sold as a local food. Several studies have been conducted to survey the local foods in restaurant menus[5][6]. Further studies have been conducted to discover local foods automatically by applying statistical methods to restaurant menus [7]. The previous methods achieved the automatic evaluation of the locality of food.

The previous study [7] applied their method to restaurant menus on the websites of restaurants, such as Yelp. However, the restaurant menus on the website do not necessarily include local foods. We discover more local foods if the previous method is applied to restaurant menus that include more local foods. Thus, this study applies the previous method to restaurant menus in Japanese roadside stations to extract keywords for the local foods. As the roadside stations provide foods produced in those regions, the restaurants may offer menus that include the local foods. We explain how to extract keywords for the local foods and show the experimental results in the next sections.

The contributions of this study are as follows:

1) We developed a dataset of restaurant menus in Japanese roadside stations. The dataset has 8,707 menus from 1,109 roadside stations in 47 prefectures (regions) (Please contact the 1st author if the database is needed.)
2) We discovered that the keywords for the local foods were not only foods but also place names, dish names, and their combinations.

In Section 2, we propose a keyword extraction method. In Section 3, we explain experiments and experimental results. In Section 4, we discuss the experimental results. In Section 5, we conclude the paper.

II. KEYWORD EXTRACTION METHOD

Figure 1 illustrates the outline of the keyword extraction method. The method is based on restaurant menus of roadside stations. Furthermore, the texts of restaurant menus are parsed into words. The words are evaluated based on their locality and assigned evaluation values. The method extracts keywords with a high locality for the local foods. The following sections will introduce how to develop a dataset and extract keywords.

 ISBN: 978-1-61208-982-9

TABLE I: DATASET EXAMPLE OF RESTAURANT MENUS IN ROADSIDE STATIONS. THE AUTHORS TRANSLATED JAPANESE MENUS INTO ENGLISH.

Menu	Roadside station	Prefecture
Chinese noodle with horse meat	Shichinohe	Aomori
Chinese noodle with boiled strawberry	Hashikami	Aomori
Curry with Murakami (place name) beef	Asahi	Nigata
Rice with tea poured and a sea bream in Kashiwazaki (place name)	Nigata furusato mura	Nigata
Meal of fried pork with Japanese miso	Fuji	Shizuoka
Japanese mustard flavored ice cream	Amagi goe	Shizuoka
Meal of fried pork of Fujisakura (place name)	Mitomi	Yamanashi
Row salmon of Shinshu (place name)	Kotani	Yamanashi
Meal of fried chicken of Choshu (place name) with vinegar and tartar sauce	Abucho	Yamaguchi
Meal of rice bowl with fried shrimp of Hagi (place name) and Japanese noodle	Hagi jyukan	Yamaguchi
Cold Japanese soba of Reiwa (era name of Japan's official calendar)	Adachi	Fukushima
Meal of Aizu (place name) DE Jyaran	Kita no sato	Fukushima
Rice bowl of fried gamecock with vinegar and tartar sauce	Nangoku furari	Kochi
Rice bowl of shiitake mushroom	Birafu	Kochi
Ice cream with fig flavored	Buzen okoshikake	Fukuoka
Ice cream with salt and Yame (place name) tea flavored	Tachibana	Fukuoka
Curry with Japanese pepper	Shimizu	Wakayama
Fried chicken	San Pin Nakatsu	Wakayama
Curry with loquat	Tomiura	Chiba
Meal of preserved fish	Kamogawa ocean park	Chiba

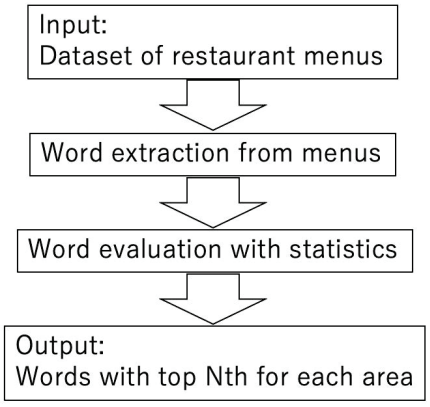

Figure 1. Outline of the proposed method.

TABLE II: STATISTICAL DATA OF DATASET.

Item	Value
Number of prefectures	47
Number of roadside stations	1,109
Number of menus	8,707
Average of menus per a prefecture	184.8
Minimum of menus among prefectures	0
Maximum of menus among prefectures	561
Average of menus per a roadside station	24.9
Minimum of menus among roadside stations	0
Maximum of menus among roadside stations	102

A. Dataset of restaurant menus in roadside stations

We focus on the restaurant menus at roadside stations that may provide menus of local foods. These roadside stations are government-designated rest areas located along Japanese roads [8]. There are shops, rest areas, and restaurants that people can use 24 hours a day. Drivers use the facilities since these stations are along roads. However, these stations are used to promote and make local areas active. They provide services using the local culture, famous places, and specialties; the local food is a specialty. Although ingredients for the local foods can be bought at the market, these items must be cooked before they can be eaten. For example, B. C. roll is indigenous to Canada. Salmon, cucumber, rice, and some sauce can be bought at markets but must be cooked before being eaten as B. C. roll. Since the roadside stations provide the ingredients for the local foods and have restaurants, their restaurants may offer menus that include these ingredients for the local foods. Thus, we use restaurant menus at roadside stations to extract keywords for the local foods.

These roadside stations have Web pages that provide information about their services. We develop a dataset of restaurant menus in roadside stations using the following attributes:

1) Restaurant menu.
2) Roadside station of the restaurant offering the menu.
3) Prefecture, where the roadside station is located.

We choose prefectures in Japan as the regions. The dataset has the three attributes in one line that are separated by commas. TABLE I shows a part of the dataset.

TABLE II shows the statistical information of the dataset. The dataset has 8,707 menus from 1,109 roadside stations in 47 prefectures. The average of menus per prefecture and roadside station is 184.8 and 24.9, respectively.

B. Keyword extraction for a local food

Restaurant menus in the dataset are parsed into words. The words are evaluated in their locality and extracted as keywords for the local foods. Keyword extraction is performed by the previous method [7] using a statistical index called Term Frequency and Inverse Document Frequency (TF-IDF). TF-IDF is an index for word importance in a document [9]; it is a product of two indices: TF (term frequency) and IDF (inverse document frequency).

Although a document contains numerous words, word distributions are not uniform. Some words appear frequently while others do not. Since a document consists of messages by the

 ISBN: 978-1-61208-982-9

author, words of high frequency must be more significant for a document. Thus, the word frequency is used to evaluate word importance in a document; this is called a TF value. However, when looking at a set of documents, some words appear in most of the documents while others do not. Therefore, words that appear in particular documents must be more important for the documents. Thus, the inverse document frequency is used to evaluate the word importance for a document; this is called an IDF value. The two values are multiplied to obtain a TF-IDF value using (1).

$$TF - IDF(w, d) = tf(w, d) \times log \frac{N + 1}{df(w, D) + 1}, \quad (1)$$

where w is a word, d is a document that includes word w, N denotes the number of documents, and D is a set of N documents.

The previous study [7] applied TF-IDF to extract keywords for the local foods. It is assumed that the keywords for the local foods frequently appear in the menus of the area. In contrast, the keywords for the local foods do not frequently appear in those from other areas. Based on this assumption, Restaurant Frequency (RF) and Inverse Local Frequency (ILF) were adopted instead of TF and IDF. The product of the two values becomes RF-ILF, which is given by (2).

$$RF - ILF(w, a) = rf(w, a) \times log \frac{M + 1}{lf(w, A) + 1}, \quad (2)$$

where a is a menu that includes the word w, M denotes the number of prefectures, and A is a set of M prefectures.

In this study, we employ words that are nouns, verbs, adjectives, and adverbs. For parsing, we used MeCab and the NeologD as a dictionary. Words with a high locality for each prefecture are extracted as keywords for the local foods.

III. Experiment

We conducted experiments with participants.

A. Experimental settings

Experimental procedures were as follows:

1) An experimenter extracts keywords using the method introduced in Section II.
2) The experimenter asks participants to evaluate whether or not the extracted keywords are for the local foods.
3) The experimenter evaluates the rate of keywords for the local foods.

The experimenter is the second author of this study. We chose prefectures as the regions. In procedure 1), the experimenter extracted 20 keywords for each of the 47 prefectures in Japan, totaling 940 keywords.

In procedure 2), 92 Japanese participated in the experiment. Their ages ranged from 10 to 80, which includes 45 men and 47 women. The experimenter asked the participants to choose prefectures from which they were born and resided for a long time. Then, the participants evaluated the keywords of the chosen prefectures based on their locality. We obtained 100 answers for 29 prefectures because some participants chose a few prefectures. The average answer was 1.08 for each participant. The experimenter evaluated the rate of keywords for the local foods for each prefecture in Japan (in procedure 3)). The experiments were conducted on December, 2021.

TABLE III: EXPERIMENTAL RESULT: RATES OF KEYWORDS FOR LOCAL FOODS. THE TOP AND BOTTOM FIVE PREFECTURES ARE DEPICTED.

Prefecture	Rate
Hokkaido	68%
Nagasaki	64%
Okinawa	39%
Gifu	30%
Aomori	28%
Hiroshima	4%
Ibaragi	5%
Miyazaki	6%
Kanagawa	8%
Yamaguchi	10%
Average	21.1%

TABLE IV: EXAMPLES OF EXTRACTED KEYWORD FOR HOKKAIDO PREFECTURE AND THEIR APPROVAL RATES. PLACE NAME MEANS A KEYWORD TYPE.

Keyword	Rate	Type
Siretoko	100%	Place name
Rausu	66%	Place name
Chinese noodle	100%	Dish name
Ice cream	33%	Dish name
Curry	0%	Dish name
Oyster	100%	Food name
Scallop	66%	Food name
Fried noodles with starchy sauce	33%	Dish name
Hanamaka	100%	Place name
Kelp	100%	Food name
Crab	66%	Food name
Traditional Fried Chicken	100%	Dish name
Dosan	100%	Place name
Lunch	33%	Others (meal style)
Salmon roe	100%	Food name
Mongolian mutton barbecue	100%	Dish name
Sun flower	33%	Others (plant name)
Atsukeshi	0%	Place name
Uryu	66%	Place name
Vegetable	66%	Food name

B. Experimental results

TABLE III shows experimental results, which illustrates the top and bottom five prefectures of the approval rate of keywords for the local food. We obtained answers from 29 of the 47 prefectures. The highest, lowest, and average approval rates were 68%, 4%, and 21.1%, respectively. The approval rate was the rate of approval numbers versus answer numbers.

IV. Discussion

We discuss a prefecture with the highest rate of keywords for the local foods. TABLE IV shows the extracted keywords and approval rates by the participants. We obtained answers from three participants. More than half of the participants approved 14 of the 20 keywords. We extracted the top 20 keywords with a high locality. However, some keywords were judged as not the local foods (for example, curry and ice cream).

 ISBN: 978-1-61208-982-9

TABLE V: TYPES OF KEYWORDS WITH THE HIGHEST LOCALITY FOR 47 PREFECTURES.

Type	Rate (total 47 prefectures)
Place name	40%
Dish name	26%
Food name	17%
Combination of place and food names	11%
Person name	4%
Adjective	2%

The extracted keywords are for food and non-food items. We analyzed different types of keywords. The third column of TABLE IV shows the types of keywords. We observed that the keywords were not only for foods but also for place names, dish names, meal styles, and plant names. Furthermore, we assumed that other prefectures had similar trends in keyword types. The same analysis of keywords with the highest locality in each prefecture was conducted. TABLE V shows the analysis result. We observed that the different types of keywords were place names (40%), dish names (26%), food names (17%), combinations of place and food names (11%), and others (person names and adjectives were 4% and 2%, respectively). The results indicated that the keywords for the local foods were not necessarily foods. People can feel the locality from keywords for place and dish names.

V. CONCLUSIONS

This study applied a basic statistical method for extracting keywords for local foods from restaurant menus of roadside stations. First, we developed a dataset of restaurant menus of roadside stations. The dataset included 8,707 menus from 1,109 roadside stations in 47 Japanese prefectures. Then, we applied the previous method to the dataset to extract keywords for the local foods. We asked participants to evaluate whether or not the extracted keywords were for the local foods. The average of keywords for the local foods was 21.1%. We found that the results might be significant because keywords for local foods could be extracted but the extraction rates were not high. We would like to improve the proposed method to obtain keywords for local foods with a higher accuracy. Furthermore, we observed that the extracted keywords were not only for foods but also place names, dish names, and their combinations. We haven't compare the results with those of previous study [7] because the evaluation method was not the same. We will evaluate the consistency of the proposed method by comparing the results of [7].

ACKNOWLEDGEMENT

This paper was partly supported by Ritsumeikan Global Innovation Research Organization. We show our great appreciation.

REFERENCES

[1] T. Sengel, A. Karagoz, G. Cetin, F. I. Dincer, S. M. Ertugral, and M. Balik, "Tourists' Approach to Local Food," Procedia - Social and Behavioral Sciences, vol. 195, pp. 429-437, 2015.

[2] F. J. Jimenez Beltran, F. Santa-Cruz, and T. Lopez-Guzman, "Gastronomy as a Factor of Tourism Development: Case Study of the City of Cordoba in Spain, " Journal of Gastronomy and Tourism, vol.2. pp.1-14, 2016

[3] J. Chrzan, "Slow food: What, why, and to where?" Food, Culture & Society, vol. 7, no. 2, pp. 117-132, 2004.

[4] E. Latham, "Food tourism in oceania: Telling the stories," Tourism Culture & Communication, vol. 21, no. 3, pp. 221-234, 2021.

[5] H. R. Yurtseven and O.Kaya, "Local food in local menus: The case of gokceada," Tourismos: An International Multidisciplinary Journal of Tourism, vol. 6, no. 2, pp. 263-275, 2011.

[6] M. Nummedal and C. M. Hall, "Local food in tourism: An investigation of the new zealand south island's bed and breakfast sector's use and perception of local food," Tourism Review International, vol. 9, no. 4, pp. 365-378, 2006.

[7] Y. Tanaka, R. Yamanishi, and Y. Nishihara, "A study on regional characteristics of food by statistical analysis of menus on restaurant search websites (in japanese)," In ARG the 13th Web Intelligenceand Interaction Domestic Conference, no. 13, pp. 61-64, 2018.

[8] Roadside station, https://en.wikipedia.org/wiki/Roadside station.

[9] G. Salton and M. J. McGill, "Introduction to modern information retrieval," McGrow-Hill Book Company, 1983.

 ISBN: 978-1-61208-982-9