User's Food Preference Extraction for Personalized Cooking Recipe Recommendation

Mayumi Ueda*, Mari Takahata**, and Shinsuke Nakajima***

Kyoto University*
Yoshida Nihonmatsu-cho, Sakyo-ku, Kyoto, Kyoto 606–8501, Japan
Kyoto Sangyo University**
Motoyama, Kamigamo, Kita-Ku, Kyoto-City 603–8555, Japan
`mayumi@mm.media.kyoto-u.ac.jp`,* `g0846741@cc.kyoto-su.ac.jp`,**
`nakajima@cse.kyoto-su.ac.jp`***

Abstract. There are many websites and researches that involve cooking recipe recommendation. However, these websites present cooking recipes on the basis of entry date, access frequency, or the recipe's user ratings. They do not reflect the user's personal preferences. We have proposed a personalized recipe recommendation method that is based on the user's food preferences. For extracting the user's food preferences, we use his/her recipe browsing and cooking history. In this paper, we present a method for extracting the user's preferences. In the experimental results, extracting the user's favorite ingredients were detected with a 60 to 83% of precision. And extracting the unfavorite ingredients were detected with 14.7% of precision, and 58% of recall. Furthermore, the F-measure value for extraction of favorite ingredients was 60.8% when we focused on the top 20 ingredients.

Key words: user's food preferences, preference extraction, recipe recommendation, cooking and browsing history

1   Introduction

As a result of the lifestyle-related disease epidemic, dietary life is now attracting attention. Good eating habits are important for maintaining a healthy life. However, menu planning requires one to take various factors into consideration, such as the nutritional value, food in stock, food preferences, and cost. Thus, people need to expand a lot of effort toward planning their daily menu. Against this background, a number of cooking websites comprising various food recipes have recently been launched, such as Cookpad[1] and Yahoo! Recipe[2]. Many people refer to these websites when planning their menu. Cookpad contains 900,000 recipes and has 10,000,000 monthly users[3]. This data reflects the high demand for recipe-providing services. However, these websites do not reflect user's preferences and conditions, although these two factors need to be considered if the goal is to provide high-satisfactory recipes.

Furthermore, several researches on cooking recipe recommendation for menu planning support have been conducted in the past. Mino et al. propose a method

that takes the user's schedule into consideration[4]. This paper defines the evaluation value of either the intake or consumption calories that are assigned to each event in the user's schedules. Karikome et al. propose a system that helps users plan nutritionally balanced menus and visualize their dietary habits[5]. Their system calculates the nutritional value of each dish, and records this information in the form of a dietary log. Next, the system recommends recipes foster sound nutrition. Freyne et al. show the results of their investigation in which they compare three recommendation strategies: content-based, collaborative, and hybrid[6].

In these circumstances, we have proposed a recipe recommendation method based on the user's food preferences[7]. Our method breaks recipes down into their ingredients, and scores them on the basis of the frequency of use and specificity of the ingredients. Furthermore, our proposed system does not recommend dishes that are similar to the food the users have eaten over the past few days on the grounds that people do not want to eat similar dishes iteratively. Moreover, our system does not require any particular action on the user's past to reflect his/her food preferences: it estimates the user's food preferences automatically through his/hers recipe browsing and cooking history. In this paper, we present a method for extracting the user's preferences.

This paper is structured as follows. Section 2 describes the method of scoring recipes and extracting user's preferences. Section 3 shows experimental results, using precision and recall. Section 4 shows concludes the paper.

2 Scoring Recipes and Extracting User's Preferences

In the recent years, concern over various health issues, such as lifestyle-related diseases and diets, has been growing. It has also been noted that picky eating is one of the main reasons causing these health issues. However, people do not want to eat food that they dislike even if it perfectly addresses their nutritional needs. They hope to derive essential nutrition solely from their favorite foods. We conducted questionnaire to the 20 men and women in their 20s to 40s to survey the key considerations for menu planning. According to the results of the questionnaire, people consider the following elements (1) food preferences, (2) nutritional balance and calories, (3) ingredients they have in stock or they can procure easily, (4) easily to cook, and (5) mood. Therefore, in this paper, we focus on (1) food preferences and try to extract user's food preferences.

2.1 Preferences for Ingredients

We express the user's food preferences I_k by using in the form of the following Eq.(1).
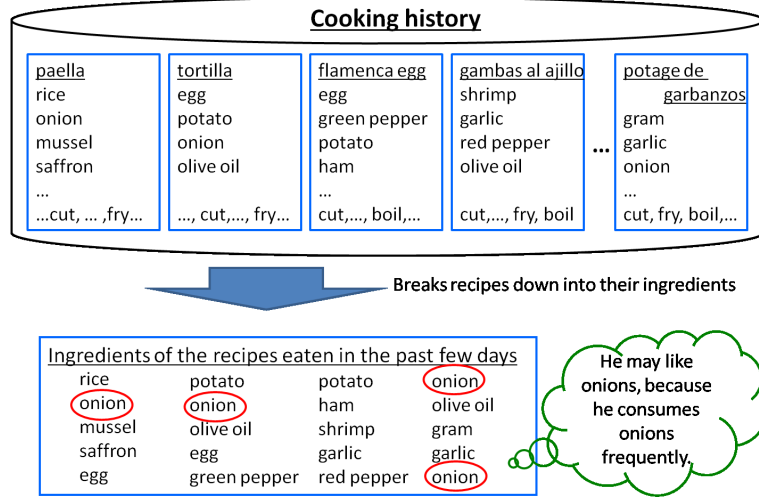
$$I_k = I_k^+ + I_k^-$$ (1)

Fig. 1. Extracting the favorite ingredients using cooking history.

User's favorite ingredients Fig.1 shows the key idea behind estimating user's favorite ingredients by his/her cooking history. Our method considers the ingredient that the user eats repeatedly as his/her favorite ingredients. It breaks recipes down into their ingredient as the outset and calculates *the score of ingredients* I_k^+ by incorporating the frequency of use of the ingredients in the dishes that the target user has eaten(FF_k: Foodstuff Frequency) as well as the specificity of ingredients(IRF_k: Inverted Recipe Frequency) into Eq.(2). This equation is based on the idea of TF-IDF.

$$I_k^+ = FF_k \times IRF_k \tag{2}$$

For estimating the user's favorite ingredients by using *the frequency of use of ingredient k (FF_k)*, we utilize *the simple frequency of use of ingredient k (F_k)* during a definite period D, as shown in Eq.(3).

$$FF_k = \frac{F_k}{D} \tag{3}$$

Then, we calculate —it the specificity of ingredient k (IRF_k) using *the total number of recipe (M)* and *the number of recipes that contain ingredient k (M_k)*, as shown in Eq.(4).

$$IRF_k = log\frac{M}{M_k} \tag{4}$$

User's disliked ingredients We consider that user's food preferences are also influenced by his/her disliked ingredients. We estimate the user's disliked ingredients, by considering the ingredients in the recipes that he/she has never
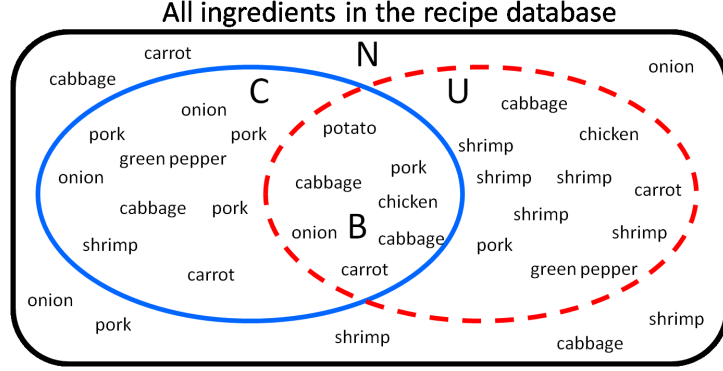
Fig. 2. Extracting the disliked ingredients using browsing and cooking history.

cooked, even if he/she has browsed the recipe details. Fig.2 shows the estimating method for user's disliked ingredients through the user's recipe browsing and cooking history. N corresponds to the set of ingredients in the recipes that the user has not browse. C corresponds to the set of ingredients in the recipes that the user has cooked over the past few days. U corresponds to the set of ingredients in the recipes that the user has not cooked, even if he/she has browse them completely. For example, "shrimp" in Fig.2 corresponds to the user's disliked ingredient. We calculate *the score of disliked ingredient k (I_k^-)* in Eq.(5).

$$I_k^-(x) = \begin{cases} 0 & (0 < \frac{2|U_k|}{|A_k|} \leqq 0.5) \\ (\frac{2|U_k|}{|A_k|} - 1)^x & (0.5 < \frac{2|U_k|}{|A_k|} \leqq 1) \end{cases} \tag{5}$$

$|U_k|$ denotes the presence of ingredient k in U and $|A_k|$ denotes the presence of ingredient k in the recipe database. We should investigate the ratio or frequency of the user's avoidance of the ingredient that he/she dislikes, because he/she will use the ingredients that he/she does not like. x denotes the ratio or frequency of avoiding the ingredients, and we plan to verify x through some preliminary experiments.

2.2 Recipe Scoring

Our method scores cooking recipes in accordance with the estimation results regarding favorite/disliked ingredients, and then provides recipes in decreasing order of the scores. In general, people do not like eating dishes similar to those they have eaten in the past few days. Therefore, our method weights recipes to avoid the repetition of similar dishes. *The score of cooking recipes* are defined as shown in Eq.(6).

$$Score(R) = \sum_{k \in R} I_k - \alpha \sum_{d=1} (w_d \cdot sim(R, R_d)) \tag{6}$$

d denotes the weight for avoiding repeating similar dishes iteratively. $sim\,(R, R_d)$ denotes the similarities between the considered recipe R and the recipe of the dish eaten d days ago R_d. The weight w_d for avoiding similar dishes eaten d days ago is defined as shown in Eq.(7).

$$w_d = 1 - \frac{d-1}{7} \qquad (1 \leqq d \leqq 7) \tag{7}$$

3 Evaluation of the Accuracy of Extracting User's Food Preference

3.1 Experimental Condition

In order to verify the extracting accuracy of the user's food preferences, we conducted simple experiments. We used 100 recipes extracted from Cookpad[1] that is a most popular recipe search website in Japan. We used randomly selected recipes categorized as main dish.

We conducted experiment as follows.

1. We present a list of 10 recipe titles to subjects.
2. He/She chooses recipes which he/she would like to browse completely, such as ingredients, procedures, and so on.
3. He/She chooses one recipe that he/she would like to cook.
4. Repeat this sequence(Step 1 to 3) 10 times.

We gathered each user's browsed recipes and recipes he/she would like to cook through the above procedure. 6 men and women in their 20s to 40s participated in this experiment as subjects.

We calculated the specificity of ingredient k (IRF_k) in the target 100 recipes, using Eq.(4). Table 1 shows the examples of IRF_k. Furthermore, we collected the labeled data as the user's preferences via questionnaire. Responses are coded on a 6-point scale, ranging from "love" to "hate".

3.2 Evaluation of Extracting the Favorite Ingredients

We evaluated the extraction accuracy of the user's favorite ingredients. We extracted individual user's favorite ingredients by calculating I_k^+, using Eq.(2). For evaluating the accuracy, we calculated precision, recall, and F-measure for top N ingredients which sorted by I_k^+. The results are shown in Table 2 and Fig.3. The precision, recall, and F-measure were calculated by the number of extracted user's favorite ingredients in the top N (E) and the number of user's favorite ingredients via questionnaire (Q), as follows.

$$Precision = \frac{E}{N}$$
$$Recall = \frac{E}{Q}$$

Table 1. Examples of the specificity of ingredient k (IRF_k)

ingredient	IRF_k	ingredient	IRF_k	ingredient	IRF_k
pumpkin	1.70	oyster	1.52	white wine	1.10
cucumber	1.70	eggplant	1.40	mayonnaise	1.01
burdock root	1.70	parsley	1.40	tomato	1.00
konjac	1.70	honey	1.40	sesame oil	0.96
snow crab	1.70	beef	1.40	carrot	0.92
green pepper	1.70	shrimp	1.30	butter	0.92
yellowtail	1.70	bacon	1.30	milk	0.92
lettuce	1.70	Japanese radish	1.22	mushroom	0.89
yam	1.70	minced meat	1.22	egg	0.77
soy milk	1.70	potato	1.15	pork	0.74
pickled plum	1.70	miso	1.15	ginger	0.68
chinese chive	1.52	lemon	1.15	garlic	0.57
bean sprout	1.52	tofu	1.15	cibol	0.57
salmon	1.52	cabbage	1.10	chicken	0.54
Colocasia esculenta	1.52	cheese	1.10	onion	0.44

Table 2. Elicitation accuracy of user's food preferences

N	Recall	Precision	F-measure
1	0.045	0.833	0.086
3	0.124	0.778	0.213
5	0.196	0.733	0.309
10	0.410	0.767	0.535
15	0.521	0.656	0.580
20	0.609	0.607	0.608

$$F\text{-}measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$
$$(N = 1, 3, 5, 10, 15, 20) \qquad (8)$$

As shown in Table 2, when we focused on the only top one ingredient ($N = 1$), our method extracted with precision of 83.3%. However, at the same time, the recall was only 4.5%. When we focused on the top 20 ingredients ($N = 20$), the recall was increased to 61%. Since the average number of individual user's favorite ingredients was 19.2, the recall was not enough value when we focused on a few ingredients such as $N = 1, 3, 5$. It was found from the result that even if our system focused on the top 20 ingredients ($N = 20$), the precision did not reduce very much. Furthermore, the highest value of F-measure, at this experiment, was 60.8% when we focused on the top 20 ingredients (Table 2). Therefore, the results show that our system should focus on the top 20 ingredients sorted by I_k^+ for recipe recommendation.
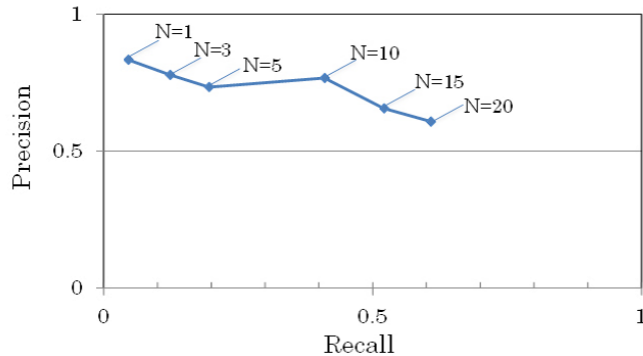
Fig. 3. The precision and recall for extracting the favorite ingredients.

3.3 Evaluation of Extracting the Disliked Ingredients

We evaluated the accuracy of extracting user's disliked ingredients. We extracted individual user's disliked ingredients by calculating I_k^-, using Eq.(5). The precision of the extracting disliked ingredients was 14.7%, and the recall was 58.3%. In this experiment, we considered that the ingredients which were contained only in U, were the user's disliked ingredients. In other words, in this experiment, our method estimated the user's disliked ingredients which were in the recipes that he/she has never cooked, even if he/she has browsed the complete recipe.

$$disliked\ ingredients = \{U \cap \overline{C}\} \tag{9}$$

In this experiment, we could not extract satisfactory accuracy for disliked ingredients, because of the lack of the number of experiments. Since the number of experiments was not enough, our method determined that the ingredient which happened to include in U was his/her disliked ingredient. We consider that our method can improve the accuracy of extracting favorite ingredients, by increasing the number of experiments.

4 Conclusion

In this paper, we presented a method for extracting the user's food preferences for recipe recommendation. Our method estimates a user's preferences from his/her past actions, such as through their recipe browsing and menu planning history. For extracting the preferences, our method breaks recipes down into their ingredients and scores the recipes using the frequency and specificity of ingredients. Since our method can estimate the preferences through their browsing and cooking history, the user convey his/her preferences to the system without having to carry out any particular operation. Furthermore, the user can convey the changes in his/her preferences to the system on a daily basis.

In order to verify the extracting accuracy of the user's food preferences, we conducted simple experiments. In the experimental results, extracting the user's favorite ingredients were detected with a 60 to 83% of precision. And the F-measure was 60.8% when we focused on the top 20 ingredients. Since the average number of user's favorite ingredients was 19.2, our system should focus on the top 20 ingredients sorted by I_k^+ to score recipes for recommendation. And extracting the disliked ingredient were detected with 14.7% of precision, and 58% of recall. In this time, we could not extract satisfactory accuracy. However, we consider that our method can improve the accuracy of extraction, by increasing the number of experiments.

As future work, we plan to consider the ingredient ontology for estimating favorite/disliked ingredients. Furthermore, we plan to investigate the various weights. For example, we plan to verify x in Eq.(5), the ratio or frequency of avoiding the ingredients, through experiments. Moreover, we want to investigate the length of time for which the system should avoid recommending similar dishes for d in Eq.(6),(7), and the degree of similarity that the system should consider while refraining from recommending similar dishes for α in Eq.(6).

Acknowledgments

References

1. Cookpad. http://cookpad.com/, (Accessed 12 August 2011)
2. Yahoo! Recipe. http://recipe.gourmet.yahoo.co.jp/, (Accessed 12 August 2011)
3. Choose dishes using websites of cooking. (in Japanese), http://otona.-yomiuri.co.jp/news/news110118_03.htm, (Accessed 12 August 2011)
4. Yoko Mino and Ichiro Kobayashi. Recipe Recommendation for a Diet Considering a User's Schedule and the Balance of Nourishment. Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems 2009, pp.383-387. 2009.
5. Shihono Karikome and Atsushi Fujii. A System for Supporting Dietary Habits: Planning Menus and Visualizing Nutritional Intake Balance. Proceedings of the 4th International Conference on Ubiquitous Information Management and Communication, pp.386-391. 2009.
6. Jill Freye and Shlomo Berkovsky. Intelligent Food Planning: Personalized Recipe Recommendation. In IUI(2010) 321-324.
7. Mayumi Ueda, Mari Takahata and Shinsuke Nakajima. Recipe Recommendation Method Based on User's Food Preferences. Proceedings of the IADIS International Conference on e-Society 2011, pp.591-594. 2011.