5C-2

Recipe Search Application based on Spoken Dialogue System on Mobile Devices

Xin Xu† Tsuneo Kato†

KDDI R&D Laboratories, Inc. †

1. Introduction

This paper proposes a new recipe search application using a knowledge-based spontaneous dialogue system to assistant users' operation on mobile devices. It aims to help users, especially novices who do not have abundant knowledge about recipes, to easily select satisfactory recipes from among thousands of potential options. The proposed application asks the user a series of questions related to various cooking categories including recipe genres and cooking needs in order to narrow down the potential recipes that meet users' wants. A decision algorithm is proposed to find the best questions to be asked to narrow down the candidates as quickly as possible according to the recipe database. Furthermore, the questions responding to the recipes that users are viewing during the search process are preferentially selected. The experiments were carried out to compare the proposed application with a conventional keyword search application, and a commercial recipe search application.

2. Related Works

Recently, cooking-related application and research becomes more popular [1]. Conventional recipe search systems are mostly based on a domain-specific web search or full-text search with keywords [2]. They cannot provide proper results unless those results include the keyword. Kitamura proposed a prototype to let users interact with multiple spoken-dialogue agents for recipe search [3]. It is easier for the user to input recipe queries and specify their preferences to reduce the number of search results by cooperative interaction among agents. The approach for collecting preference information was not mentioned, and this is generally regarded as difficult.

3. Spoken Dialogue System Cooperative to Find Recipes

Compared to other earlier works, we proposed a recipe search application employing a dialogue system, which has one unique aspect. A novel dialogue strategy offers questions related to the recipe categories as a way to actively refine the search query. It simplifies the search process for novice users, without any user preference information.

Our dialogue management model is designed as a mixture of a user-initiative model and system-initiative model. First, the application prompts an open question to ask the users, “What kind of recipes are you looking for?” Users can for example answer, “A recipe using chicken and cabbage”. The recipe results including those keywords are then refined. At the same time, the application prompts, “Please press here if confused by the number of recipe results” and displays a button. Once the user presses the button, the application starts to ask the user a series of questions related to the recipe categories, such as “Do you use an oven?” The user then simply presses a button or speaks into the microphone to choose yes or no, and the recipe results are consequently narrowed down until the user finds a satisfactory recipe. This design is intended to assist users who have not prepared many keywords but expect an easy search.

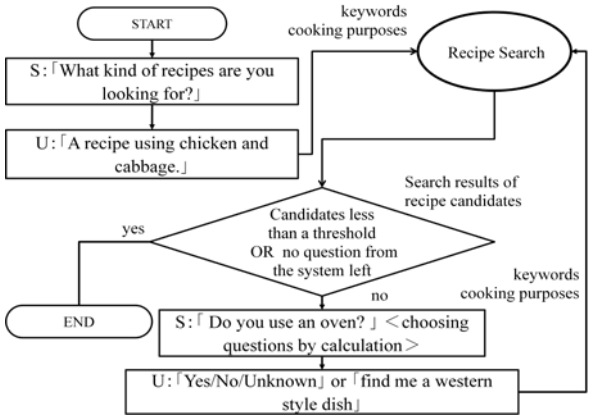

Figure 1 The flowchart of dialogue scenario

Even while being asked questions, users are allowed to view the details of recipes and freely input keywords and sentences to narrow down the number of recipe results.

4. Question Selection for Quick Narrowing Recipes

In order to minimize the number of search refinement steps, the question selection in the dialogue scenario is mainly based on the information gain according to the remaining recipes.

The information gain is calculated by Equation 1 and 2, using the decision algorithm of Iterative Dichotomiser 3 [4]. Entropy *H(S)* represents the average amount of uncertainty in the recipe set *S*. *p(x)* is the proportion of the number of elements in category *x* to the number of elements in *S*. Information gain *IG(A)* is the difference in entropy from before to after the set of recipes *S* is split to create category *t*, which is related to the question *A*. *A* is also regarded as an attribute. In other words, the information gain *IG(A)* means how much uncertainty in *S* was reduced after splitting the original set on the question *A*.

$$H(S) = \sum_{m \in M} p(x)\, log_2\, p(x) \qquad (1)$$

$$IG(A) = H(S) - \sum_{t \in T} p(t) H(t) \qquad (2)$$

Furthermore, as users usually check a number of recipes that they are interested in before they make the final decision. Preferentially selecting the questions responding to the recipes that users are viewing is supposed to be helpful to reflect users’ potential search purpose. Therefore, *P(A)* is induced as a weight of *IG(A)* in Equation 3, which is the proportion of the number of recipes categorized by *A* that have been viewed by the user during the search process to the number of all recipes that have been viewed. *r* is the smoothing parameter. The question related to the category with the highest *PIG(A)* value is regarded as the optimal candidate that keeps balance between the search efficience and users' personalizaiton. After each refinement, *PIG(A)* is calculated for each remaining question. The question with the largest *PIG(A)* is used on this iteration.

$$PIG(A) = (P(A) + r) \times IG(A) \qquad (3)$$

Based on a questionnaire of users, and investigation of recipe websites, 81 recipe categories of cooking purposes are created. They correspond to the cooking situation, cooking utensils, cooking methods, ingredients genres, cooking needs, cooking difficulty, dish origins and dish genres. Besides the normal

Recipe Search Application based on Spoken Dialogue System on Mobiles Devices

†Xin Xu, Tsuneo Kato, KDDI R&D Laboratories, Inc.

categories, the categories adapted to spoken queries, such as outdoor, diet and saving electricity are also included. The learning algorithm named Confidence-weighted linear classification [5] is used for classification and estimation.

5. Experimental Evaluations

To evaluate the usability of the proposed application, it was compared with a conventional keyword-based search application (Application1). Furthermore, a major commercial recipe search application (Application2) was also compared as a reference. The mobile device in the experiment was Google Nexus 7 tablet (OS: Android 4.1). The comparisons of subjective evaluations and operating time were both carried out.

5.1 Experimental Setup

Experiments were conducted with 7 subjects (3 females, 4 males) aged from 20 to 40 years, including engineers and clerks. All of them were experienced smart phone users and novice users of recipe search services.

Application1 is almost the same with the proposed application only except the dialogue interface. It only allows keyword input with a keyword suggest function. Both the proposed application and Application1 stored 54,277 recipes in Japanese and they are implemented on the open-source search engine [6]. The Application2 is a widely-used recipes search application with a commercial database over 1 million recipes. All of the applications support speech input, in which the same Android standard speech input module is used [7].

The experiments were carried out in a quiet room in our lab. The subjects were asked to imagine that they were in their own living room, and trying to search for a satisfactory recipe with the mobile devices for cooking dinner 1 hour later. No time limitation was set. For each application, a list of 2 designated tasks and a free task was given in which subjects were asked to finally decide on one recipe for each of these requirements. For examples, one designed task is searching a recipe of western food with spinach as an ingredient. They were allowed to add or change keywords that were related to the cases given. Before subjects started the experiment, a brief guidance video was shown, and a few minutes preliminary practice was allowed. The order in which each application was used was randomly determined for each subject.

5.2 Results of Experiment

Once subjects finished the operations of an application, they were asked to rate six aspects of their experience by using a 5-point Likert scale, on which 5 points were awarded for the most satisfactory experience, and 1 point for an unsatisfactory experience. The list of aspects is as below:

1. Did you feel easy to operate the application?
2. Did you become clear about the interface after read the guide of the application?
3. Did you feel smooth to execute the next operation when you try to find the recipe?
4. Did you feel easy when you were considering the keywords?
5. Did you feel ok with the number of operations for inputting a keyword?
6. Are you satisfied with the recipe that you decided?

The average scores are shown in Table1. The greater the score is, the more positive the subject evaluates the application. Besides the score, uses' free comments of each aspect are also recorded. On aspect 1, the proposed application achieved worse scores than the others. It is because the operation on spoken dialogue interface is quite new for the subjects, getting used to when and how to answer was not easy. For aspects 2,

Table 1 Results of subjective evaluation

Score	Proposed Application	Application 1	Application 2
Aspect 1	4.29	4.43	4.43
Aspect 2	4.29	4.71	3.57
Aspect 3	4.29	3.86	4.14
Aspect 4	3.29	3.00	3.00
Aspect 5	3.71	3.57	3.29
Aspect 6	4.14	3.86	3.86

Application1 achieved the best evaluation because of its simplest interface. However, the subject most positively evaluates the proposed application on aspect 3. The reason is supposed to be the speech guidance and the simple operation to choose yes or no to refine the search results.

On the other hand, the proposed application received scores that were higher than those of other conventional applications on aspect 4 and 5. It verified that the strategy using dialogue interface to simplify and to refine the search is successful, since it saved subjects' trouble to reconsider keywords. Finally, the proposed application achieved the best score on aspect 6. It is ascribed to the proper question selection that provides useful keywords and information for the subjects to search recipes. In addition, especially for the proposed application, an aspect "Did you feel the question asked by application helpful?" was asked. The average score is 4, which is relatively positive.

The operating times including both real times and sensory times of subjects are also recorded. As is shown in Table 3, the proposed application cost more time than others. However, different from Application2, the sensory time of the proposed application is shorter than the real time. It may reflect that subjects felt less stress when using the proposed application.

Table 2 Average operating time for one search task

Time (m)	Proposed Application	Application 1	Application 2
Real Time	3.80	3.74	3.47
Sensory Time	3.14	2.95	3.55

6. Conclusion

This paper introduces a novel spoken dialogue system for recipe search. The proposed recipe search application actively asks the user questions to help their refine search. Based on the results of evaluation experiments, we can conclude that the proposed application performs better than the conventional applications in terms of their satisfaction with search results and the effort spent for reconsidering search keywords.

References

[1]COOKPAD. http://cookpad.com/
[2]Satoshi Oyama, Takashi Kokubo, Toru Ishida. 1983. "Domain-Specific Web Search with Keyword Spices". IEEE Transactions on Knowledge and Data Engineering, VOL. 16, NO. 1, pages 17-27.
[3]Yasuhiko Kitamura et al. 2001. "Interactive Integration of Information Agents on the Web". Proceedings of the 5th International Workshop on Cooperative Information Agents,pages 1-13. J. R.
[4]Quinlan 1986. "Induction of Decision Trees". Journal Machine Learning Volume 1 Issue 1,pages 81-106
[5]Mark Dredze, Koby Crammer, Fernando Pereira. 2008. "Confidence-weighted linear classification". Proceedings of the 25th international conference on Machine learning,pages 264-271
[6]Michael McCandless, Erik Hatcher, Otis Gospodnetic 2010. "Lucene in Action (2nd ed.)". Manning Publications., pages 475
[7]On Ballinger, Cyril Allauzen, Er Gruenstein, Johan Schalkwyk 2010. "On-Demand Language Model Interpolation for Mobile Speech Input". Proceedings of Interspeech 2010,pages 1812-1815