Real World Voice Assistant System for Cooking

Takahiko Ito, Shintaro Inuzuka, Yoshiaki Yamada, Jun Harashima
Cookpad Inc.
Yebisu Garden Place Tower 12F, 4-20-3 Ebisu, Shibuya-ku, Tokyo, 150-6012, Japan
{takahi-i, shintaro-inuzuka, yoshiaki-yamada, jun-harashima}@cookpad.com

Abstract

This study presents a voice assistant system to support cooking by utilizing smart speakers in Japanese. This system not only speaks the procedures written in recipes point by point but also answers the common questions from users for the specified recipes. The system applies machine comprehension techniques to millions of recipes for answering the common questions in cooking such as "人参はどうしたらよいですか（How should I cook carrots?）". Furthermore, several machine-learning techniques are applied to generate better responses to users.

1 Introduction

Smart speakers, such as Google Home and Amazon Echo, have drawn attention due to their crucial applications in recent times. Many voice assistant applications that support users have been released in the smart speaker platforms. There are different technologies associated with smart speakers; one of the most important tasks of smart speakers is machine comprehension. In this regard, many studies have been conducted, including (Pranav et al., 2016) and (Yagcioglu et al., 2018).

This study presents a voice assistant system, which supports cooking, in Japanese. This system not only speaks the procedures written in specified recipes point by point but also applies machine comprehension techniques to answer common questions in cooking on particular recipes. The system supports millions of recipes stored on a recipe sharing service. To instantly handle questions from users, we build a *knowledge base* to support the generation of the answer sentences.

2 Voice Assistant System for Cooking

This section describes the basic usage of the voice assistant system. Figure 1 depicts the interaction between the voice assistant system for cooking and the users.

The voice assistant system speaks the procedures (steps) point by point in a recipe specified by the users. In addition, when users ask questions regarding the recipe at any time, the system answers the question in an arbitrary time. To reduce the time of response, we build a knowledge base by which the voice assistant system seeks to generate answers to the questions from users. The knowldege base contains the resources used by the voice assistant to speak each of the procedures and answer the questions from users.

3 Knowledge Base to Support Cooking

The knowledge base contains structured resources, which are dependent on each recipe. The resources contain the title, procedures, ingredients and answers to common questions on recipes. The following sections show the features of the developed kdnowledge for cooking.

3.1 Machine Cmprehension

To make the voice assistant answer the questions, we build a set of resources required to answer the common questions on the target recipes.

In the sequel, we shall describe the processes involved in extracting the resources to answer the questions.

3.1.1 Description of the Related Procedures on Particular Ingredients

When we cook after studying a recipe, we sometimes do not remember the details of the procedures or the quantity of ingredients in the recipe. In this case, users ask a question such as "how should I process potatoes?" Then, our voice assistant replies stating that "the recipe said that you should boil the potatoes and then mush them."
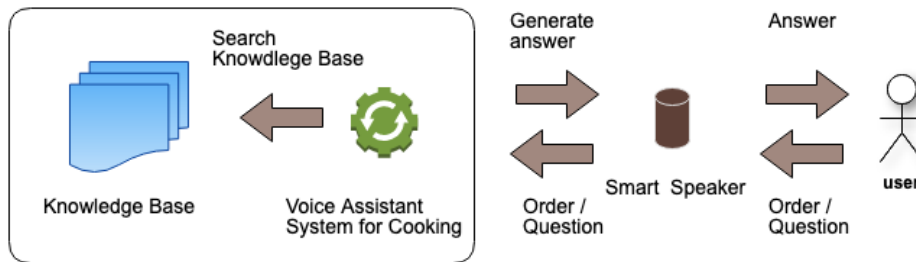
Figure 1: Interaction with a voice assistant system.

The system automatically extracts the procedures of processing each ingredient in the target recipe with simple pattern matching ingredient names and procedures.

3.1.2 Extract other expressions

In addition to the extraction of the related procedures on each ingredient, the system extracts other pieces of knowledge in cooking.

Quantitative expressions Recipes contain quantitative expressions such as temperature or the time range of processing ingredients. For example, when the procedure of a recipe says "heat potatoes for 10 min using a microwave," the system automatically extracts the time and temperature with respect to the cooking instruments.

Seasonings Our voice assistant system extracts the required quantity of the specified ingredients. In addition, it automatically detects whether each ingredient requires seasoning.

3.2 Detect Fake Procedures

In consideration of the recipe data in sharing services, users publish recipes in the form of a series of procedures. Some "procedures" are not actually part of the cooking process but fake. Fake procedures include the advertisements of the recipes themselves or comments. These fake procedures cause problems when they are given by the voice assistant systems.

Therefore, we constructed a LSTM based discriminator that distinguishes fake procedures or those that do not follow the procedures (Inuzuka et al., 2018).

3.3 Normalize the Ingredient Names

Recipes have a list of a blocks containing ingredient names with their quantities. When an ingredient name inserted from the voice assistant system does not match the name in the target recipe, the voice assistant system is not able to return the answer.

Therefore, we apply two types of normalization methods to ingredient names. The first applied method is based on a dictionary. The dictionary contains the ingredient names with different variations and the canonical form. The second applied method is a character-based encoder-decoder model described in (Harashima and Yamada, 2018).

4 Summary and Future Work

This study presented the voice assistant system deployed in smart speakers. We build knowledge base to answer the common questions from users. Practical machine learning techniques are applied to generate better responses to questions from users.

The machine comprehension for recipe is currently implemented with simple matching. In the future we will apply modern deep neural network methods such as BERT.

References

Jun Harashima and Yoshiaki Yamada. 2018. Two-step Validation in Character-based Ingredient Normalization. In *CEA*.

Shintaro Inuzuka, Takahiko Ito, and Jun Harashima. 2018. Step or Not: Discriminator for The Real Instructions in User-generated Recipes. In *W-NUT*.

Rajpurkar Pranav, Zhang Jian, Lopyrev Konstantin, and Liang Percy. 2016. SQuAD: 100,000+ questions for machine comprehension of text. In *EMNLP*.

Semih Yagcioglu, Aykut Erdem, Erkut Erdem, and Nazli Ikizler-Cinbis. 2018. RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes. In *EMNLP*.