MIAIS: A Multimedia Recipe Dataset with Ingredient Annotation at Each Instructional Step

Yixin Zhang\(^1\), Yoko Yamakata\(^2\) and Keishi Tajima\(^3\)
\(^1\)Kyoto University and \(^2\)The University of Tokyo, Japan

**RESEARCH BACKGROUND**

More and more datasets on recipes are constructed.

Food-101 dataset, Recipe1M+, UECFood256:
- mainly focus on completed dish images after cooking rather than images in recipe instructions.

Cookpad dataset:
- 1,642,450 images of completed dishes
- 3,105,594 instructional images

**RESEARCH PROBLEM AND PURPOSE**

Most of the datasets:
- No sufficient multimedia information (text and image) on the cooking instructions of the recipe.

However, the instructional information is one of the most important and unique characteristics of recipe data.

**MIAIS Dataset**
- **Goal**: construct a recipe dataset with sufficient multimedia data and the annotations to them for every cooking step (based on Cookpad dataset)
- **Applications**: cooking flow graph generation, recipe text generation, and cooking action recognition

**ANNOTATION AND COOKING FLOW GRAPH**

**Challenge and Common Solutions**

- **A challenge** in understanding the cooking process from recipes: present the complicated operations of cooking activities in a succinct workflow.
- **A common solution** is to generate a flow graph to help them understand the cooking process.
- Focus on the flow of ingredients in recipes.
- However, the ingredient information in text tends to be insufficient or omitted when written by users.

**Quick and moist Nikujaga (meat and potatoes)**

**Ingredients**
- potatoes
- water
- olive oil
- beef
- soy sauce
- Shibazaki
- mirin
- honey

**Cooking Instructions**
1. Wash the potatoes with a bowl.
2. Roughly cut all the potatoes into quarters. After washing with water, add 4 tablespoons of water and cover over with a wrap.
3. Microwave for 4 minutes. Meanwhile, heat a frying pan and sauté the beef in olive oil.
4. After boiling, the potatoes are placed in a bowl of water and peeled them by hand.
5. Sprinkle 4 tablespoons of sugar on the beef. When blended, add a teaspoon of salt. Season the beef well.
6. Add Shibazaki to add 4 tablespoons of mirin and soy sauce.
7. When blended, add all the potatoes and shake the whole to let the potatoes soak in the flavor.
8. Taste and if it seems too thin, add soy sauce and honey to taste. Ready to serve.

**Annotation based on the appearance in text**
- Potatoes is omitted.

**Our new annotation method**
- The flow graph of this recipe can be generated based on the above information.
- The order in which the actual cooking is performed is not sequential, but with several branches.

**Flow Graph Generation**
- We track the paths of ingredients in order to generate the cooking flow graph.
- If a specific ingredient:
  - is contained in a previous step;
  - can also be detected in one of the subsequent steps;
  - we can consider that there is continuity between the two steps.
- The path is created for each ingredient and the final flow graph is a union set of all the paths.
- E.g., "Potato": 1-2-3-4-7-8; "Beef": 3-5-6-7-8

**MIAIS DATASET**

***Multimedia recipe data***
12,000 recipes in Japanese
- The text and image data are originated from Cookpad.
- **Recipe selection rule**:
  - Recipes that have at least five cooking steps
  - Recipes that contain a text description and an image for every step
- 109,547 text-image paired instructional steps in total

**CONCLUSION**

- A multimedia recipe dataset with ingredient annotation for every instructional step
- Containing **both text and image data for every cooking step**
- Supplementary information about text-image pair of every step
- Cooking flow graphs
- Directly derived based on the annotation information
- These flow graphs can become the ground truth for the method of recipe flow graph generation