

# Step or Not: Discriminator for The Real Instructions in User-generated Recipes

Shintaro Inuzuka, Takahiko Ito and Jun Harashima

Cookpad Inc.

Yebisu Garden Place Tower 12F, 4-20-3 Ebisu, Shibuya-ku, Tokyo, 150-6012, Japan  
{shintaro-inuzuka, takahi-i, jun-harashima}@cookpad.com

## Abstract

In a recipe sharing service, users publish recipe instructions in the form of a series of steps. However, some of the “steps” are not actually part of the cooking process. Specifically, advertisements of recipes themselves (e.g., “introduced on TV”) and comments (e.g., “Thanks for many messages”) may often be included in the step section of the recipe, like the recipe author’s communication tool. However, such *fake* steps can cause problems when using recipe search indexing or when being spoken by devices such as smart speakers.

As presented in this talk, we have constructed a discriminator that distinguishes between such a fake step and the step actually used for cooking. This project includes, but is not limited to, the creation of annotation data by classifying and analyzing recipe steps and the construction of identification models. Our models use only text information to identify the step. In our test, machine learning models achieved higher accuracy than rule-based methods that use manually chosen clue words.