
Discriminating Practical Recipes Based on Content Characteristics in Popular Social Recipes

Yohei Seki

Faculty of Library, Information and Media
Science, University of Tsukuba, 1-2 Kasuga, Tsukuba, Ibaraki,
305-8550
yohei@slis.tsukuba.ac.jp

Kouta Ono

School of Informatics, University of Tsukuba, 1-2 Kasuga,
Tsukuba, Ibaraki, 305-8550

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UbiComp '14, September 13 - 17 2014, Seattle, WA, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3047-3/14/09...\$15.00.

<http://dx.doi.org/10.1145/2638728.2641326>

Abstract

Recipe websites sometimes contain vast collections of recipes, making it time-consuming for users to identify recipes that might suit them. In this study, we aim to support users in their recipe selection by discriminating "practical recipes" that are easy to understand, written concisely with sufficient description, and offer detailed tips and pointers. We performed a content analysis of popular recipes found on Cookpad, focusing on ten types of dishes, and decided to use seven content characteristics, such as "description of the heat level" and "description of the cooking time," as features to discriminate practical recipes. We have implemented a discriminator based on an SVM classifier that uses these features. The results of a discrimination experiment show that the mean value of the accuracy of the ten types of dishes is 0.813. This represents a significant difference from a baseline discriminator.

Author Keywords

Recipe recommendation; SVM; Content analysis

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Natural language.

Introduction

Background and Significance of This Study

Today, there are several recipe websites where users post recipes, such as Cookpad¹ and Rakuten Recipe². On Cookpad, which was launched in 1998 and is now the largest recipe website in Japan, more than 1.6 million recipes posted by users have been published. The Rakuten Recipe website, which was launched in 2010 and is growing rapidly, contains 730,000 recipes. Browsing those recipes one by one is a very time-consuming task.

In considering recipes, we may select one that proves to be inadequate because, for example, either the heat level or the cooking time is not specified, or there is no information on how to measure ingredients. Cooking with that recipe might then pose problems. We believe it would be useful to identify recipes that users can easily understand, to support the user in both recipe selection and the cooking process.

In this study, we aim to support users in their recipe selection and preparation by discriminating “practical recipes” that are easy for the user to understand, are written concisely with sufficient description, and offer detailed tips and pointers.

What is a Practical Recipe?

If we select a recipe from a cooking recipe website without having a detailed explanation of the selected recipe’s contents, we cannot always expect problem-

free cooking. Therefore, if a user finds that it is easy to cook by using a particular recipe, the recipe can be considered “practical.” In addition, if tips and cooking pointers are included in the recipe, enabling us to prepare a tasty dish with a little creativity, we can say that its practicality is enhanced. In this study, we define a practical recipe as one that is:

1. easy to understand,
2. concise,
3. has sufficient description, and
4. offers detailed tips and pointers.

In the article “How to Write a Recipe That is Easy to Understand” (Recipe Blog Magazine [2]), the author states that enriching the contents of a recipe, so as “not to conceal the heat level and the heating time” and “to write the quantity or a rough estimate of the quantity in order to measure ingredients,” will greatly influence the recipe’s understandability. In this study, we discriminate recipes that are easy to understand by taking into account the following four factors:

1. a description of the heat level,
2. a description of the cooking time,
3. a description of how to measure ingredients, and
4. use of onomatopoeia³ (i.e., cooking sounds, such as boiling water or sizzling meat).

Buykx et al. [1] indicated that the clarity and understandability of a recipe varies depending on the size of a cooking step and the number of cooking tasks per cooking step. In this study, we discriminate concise recipes by focusing on the size of a cooking step. Note

¹ The largest recipe website in Japan is “Cookpad.” The number of recipes exceeds 1.6 million. <http://cookpad.com>

² Rakuten Recipe: we can find a recipe, cook, and contribute. Everyone can join. <http://recipe.rakuten.co.jp>

³ Onomatopoeia can make a recipe easy to understand, because it expresses a cooking method and the state of the ingredients.

that a recipe may be too concise to understand. Therefore, we discriminate recipes by considering the number of cooking steps they contain. We consider a recipe as practical if it is sufficiently explanatory and uses several concise cooking steps.

The Recipe Blog Magazine article [2] also discussed the subject of tips, noting that “cooking tips and important points should be outside of the recipe itself.” We used tips included in a box outside of the cooking steps as a feature for discrimination, thereby identifying recipes containing tips and pointers that are described in detail.

In this study, therefore, we discriminate practical recipes by focusing on their contents, aiming to help users to save time in recipe selection. In addition, this process enables exclusion of recipes that might be hard to cook because of insufficient detail. Furthermore, if tips and cooking pointers are included in the recipe, we can become creative, which will increase our satisfaction with our meal repertoire.

Related Studies

Buykx et al. [1] showed that a user’s impression of whether a recipe is clear and easy to understand depends on the contents of the recipe and how it is written. In their study, they performed an experiment in which subjects cooked dishes using three versions of a recipe, with varying content and writing style, and compared the results. After the cooking, the subjects evaluated two factors on a five-point scale: “whether the recipe was clear” and “whether it was easy to understand the recipe.” They also commented on each of the recipe versions. Based on the results of the experiment, for the first factor, they found that a recipe was clear if the information required for cooking, such

as ingredient quantities, was included in the recipe directions. For the second factor, they found that a recipe written with shorter, individualized cooking steps was easier to understand than a long recipe written as one step with many cooking tasks. They analyzed the comments, and found that they verified these findings. For example, “it is troublesome to go back to the page that shows the ingredients if the ingredient quantity is not written in the recipe directions” and “it is too complex to understand the recipe if multiple cooking tasks are written in one cooking step.” The comments also revealed new complaints about recipe descriptions. For example, “it is hard to understand unless a detailed explanation of the cooking action is provided.” As expected, these findings show that the contents of a recipe and how it is written greatly influence a user’s ability to use the recipe. In our study, we also attempt to identify recipes that are easy for users to understand by focusing on the number of cooking steps and the contents. In addition, regarding the contents of a recipe, we pay attention to the fact that descriptions of the level of heat, the cooking time, and tips may affect the understandability of a recipe. We have implemented a discriminator of practical recipes, which is based on these features and has potential for expansion.

For recipes in Cookpad, Yamakata et al. [6] proposed a system that provides a typical cooking procedure and describes the characteristics of a recipe. A typical cooking procedure is estimated from a recipe set obtained through a search of recipes, and a characteristic of a recipe is extracted from a comparison of recipes, enabling the user to grasp the required information at first glance. In a study by Tachibana et al. [5], a naming concept is extracted from a recipe, based on a typical element of the dish,

which is used as a characteristic of the recipe. By using this method, a user can grasp the naming concept that is the characteristic of each recipe at first glance.

Those studies attempted to support users in choosing a recipe by extracting and providing characteristics of the recipes. In our study, we share the purpose of supporting users in choosing a recipe. However, even if the characteristics by which a user selects a recipe are provided, its usability may be poor if the contents of the recipe are difficult for the user to understand. In this study, we support users in recipe selection by preselecting and showing easy-to-understand recipes, which will improve the usability of the recipes.

Analysis of Characteristics of Popular Recipes

In this section, we analyze the characteristics of popular recipes. We use those characteristics that are features of popular recipes to discriminate practical recipes. A recipe is considered popular if it appears in the top 20 of a popularity ranking on Cookpad. This ranking is calculated from statistical data such as the number of cooking reports or the number of times it has been accessed. We believe that these measures reflect the recipe's understandability, which is why users want to cook the dish. Therefore, because they share some of the characteristics of practical recipes, we decided to use the characteristics of popular recipes defined by the analysis as features to discriminate practical recipes.

Purpose

The purpose of the analysis is:

1. to clarify the characteristics of popular recipes, and

2. to demonstrate that it is valid to use the characteristics of popular recipes as features to discriminate practical recipes.

Data

For the dishes shown below, we used recipes in the top 20 of a popularity ranking, together with recipes ranked in the bottom 20.

Name of category	Number of recipes	Percentage (%)
Vegetable dishes	23,997	33.9
Meat dishes	10,560	14.9
Rice dishes	7,153	10.1
Fish dishes	6,615	9.3
Eggs and processed soybeans	4,152	5.9
Pasta and gratin	3,965	5.7
Stew and soup	3,371	4.8
Salad	3,127	4.4
Total for the eight categories	62,940	89.0
Total for all the categories in "Today's dishes"	79,690	100.0

Table 1. Recipe categories found in Cookpad (data as of December 10, 2013)

We now explain our dish selection process. First, we selected ten recipe categories, and then picked one dish from each category. A recipe category is a set of recipes classified by type of dish, such as "meat dish" or "vegetable dish." The selection was based on the recipe categories found in Cookpad⁴, where there are three major categories: "Today's dishes," "Sweets," and "Bread." We first selected one dish from each of

⁴ Cookpad recipe category : <http://cookpad.com/category/list>

the major categories “Sweets” and “Bread.” Under “Today’s dishes,” there are many subcategories, such as “Meat dishes.” We chose the eight categories shown in Table 1 from this subcategory list. These were the top eight categories as determined by the number of recipes in each category. We then selected one major dish type (decided by the authors) from each of the eight categories, giving ten dish types from ten categories, as shown in Table 2.

Dish name	Number of recipes
Nikujaga	6,128
Oyakodon	2,469
Carbonara	5,217
Cream Stew	1,623
Shrimp with chili sauce	1,460
French toast	4,355
Simmered kabocha squash	1,541
Mapo tofu	3,499
Potato salad	7,711
Sweet potato cake	3,659

Table 2. The ten types of dishes used to analyze characteristics (data as of December 22, 2013)

Methodology

In the analysis, we focused on the contents of the recipes and how they are written. In particular, we:

1. analyzed the characteristics of each recipe manually, allocating the recipe of the ten types of dishes to them, and
2. compared the recipes in the top 20 and the ones in the bottom 20 as ranked in popularity on Cookpad, and identified their characteristics.

After observing the contents of recipes and how they were written, we counted the number of recipes in each

group of 20 that included each of the characteristics: “the contents are easy to understand,” “each cooking step is short,” “the number of cooking steps is large,” and “tips are written in detail.”

Results

Table 3 lists the four characteristics of popular recipes.

	Top/bottom 20 recipes	The contents are easy to understand	Tips are written in detail	The number of cooking steps is large	Each cooking step is short
Nikujaga	Top 20	15	12	10	12
	Bottom 20	5	4	1	5
Oyakodon	Top 20	12	10	12	13
	Bottom 20	8	2	2	8
Carbonara	Top 20	12	14	8	19
	Bottom 20	3	5	1	9
Cream Stew	Top 20	13	13	9	13
	Bottom 20	6	4	1	16
Shrimp with chili sauce	Top 20	14	12	13	16
	Bottom 20	5	4	6	17
French toast	Top 20	17	13	11	17
	Bottom 20	2	11	1	16
Simmered kabocha squash	Top 20	16	13	10	19
	Bottom 20	6	5	4	11
Mapo tofu	Top 20	14	12	12	17
	Bottom 20	7	4	1	9
Potato salad	Top 20	16	12	11	15
	Bottom 20	6	3	8	17
Sweet potato cake	Top 20	17	12	15	17
	Bottom 20	7	3	4	8

Table 3. Comparison of characteristics between the top 20 recipes and the bottom 20 recipes (count)

We see that the top 20 recipes have these characteristics in more than six of the ten kinds of dishes when compared with the bottom 20 recipes. Note that the term “cooking step” is used to indicate an individual phase of a cooking process found in a Cookpad recipe.

Discussion

In this subsection, we discuss in detail the four characteristics of popular recipes: “the contents are easy to understand,” “tips are written in detail,” “each cooking step is short,” and “the number of cooking steps is large.”

(1) THE CONTENTS ARE EASY TO UNDERSTAND

As shown in Table 3, this characteristic appears far more often in the top 20 recipes than it does in the bottom 20 recipes, for all ten kinds of dishes. This indicates that popular recipes contain contents that are easy to understand. In addition, we analyzed this characteristic in detail, dividing it into the four description factors shown in Table 4.

	Top/bottom 20 recipes	Description of the level of heat	Description of the cooking time	Description of how to measure ingredients	Use of onomatopoeia
Nikuja ga	Top 20	16	16	19	10
	Bottom 20	6	5	15	7
Oyakodon	Top 20	11	14	19	3
	Bottom 20	7	6	12	6
Carbonara	Top 20	6	11	13	6
	Bottom 20	3	6	13	1
Cream Stew	Top 20	15	11	14	5
	Bottom 20	9	9	10	2
Shrimps	Top 20	5	3	16	9
	Bottom 20	3	4	14	1

with chili sauce					
French toast	Top 20	14	10	9	13
	Bottom 20	3	5	6	1
Simmered kabocha squash	Top 20	14	18	15	8
	Bottom 20	7	7	16	1
Mapo tofu	Top 20	10	10	20	6
	Bottom 20	3	9	15	4
Potato salad	Top 20	5	10	15	8
	Bottom 20	0	3	14	3
Sweet potato cake	Top 20	15	18	14	8
	Bottom 20	10	14	12	4

Table 4. Detail of “the contents are easy to understand” (count)

■ Description of the level of heat

Table 4 shows that the number of the top 20 recipes that include a description of the level of heat was greater than that of the bottom 20 recipes, for all ten kinds of dishes. We expect that it will have an effect if we use it as a feature to discriminate practical recipes.

■ Description of the cooking time

As shown in Table 4, the number of recipes ranked in the top 20 that include a description of the cooking time was greater than the number in the bottom 20 recipes, for nine kinds of dishes. Again, we believe that it will have an effect if we use it as a feature to discriminate practical recipes.

■ Description of how to measure ingredients

Table 4 shows that, overall, recipes contain descriptions of how to measure ingredients. Only

recipes for oyakodon had a significant variation between the top 20 and the bottom 20. However, there are some differences, and it may be effective in discriminating practical recipes.

■ Use of onomatopoeia

As Table 4 shows, a small number of recipes had this characteristic. However, because this expresses a cooking method and the state of the ingredients, it can make a recipe easy to understand. Recipes for shrimp with chili sauce, French toast, and simmered kabocha squash indicate a large difference in the number of times this characteristic appears in the top 20 recipes versus the bottom 20 recipes. In French toast case, the state of the ingredients is expressed for cooking understandability in top 20 recipes using onomatopoeia. In shrimp with chili sauce and simmered kabocha squash cases, onomatopoeia is used to modify cooking method in detail. For these types of dishes, we expect that this will be important in making recipes easy to understand.

(2) TIPS ARE WRITTEN IN DETAIL

In recipes found in Cookpad, there is a column entitled “Helpful Hints,” in which the recipe’s contributor offers tips and pointers to accompany the main recipe. We can prepare a tasty dish with a little ingenuity if such detailed, practical tips are provided. We therefore believe that this feature can enhance the practicality of the recipe. In Table 3, we can see that the number of recipes with the feature included in the top 20 was greater than the number in the bottom 20, for all ten kinds of dishes. In fact, many of the recipes in the top 20 had detailed tips or practical tips. Examples of tips include “the trick is to simmer it vigorously at high heat

without a lid” and “everything will go well if you do not beat it too much.” We believe that tips and pointers are characteristics found only in popular recipes, and they are effective features with which to discriminate practical recipes.

(3) THE NUMBER OF COOKING STEPS IS LARGE

As shown in Table 3, recipes in the top 20 usually have a large number of cooking steps. For eight of the ten kinds of dishes, there were significant differences between the numbers in the top 20 and those in the bottom 20. Based on this, we can say that popular recipes tend to have several cooking steps.

(4) EACH COOKING STEP IS SHORT

As shown in Table 3, recipes in the top 20 usually have a concise description of each cooking step. For five of the ten kinds of dishes, there were significant differences between the numbers in the top 20 and those in the bottom 20. We can say that this is also a particular characteristic of popular recipes.

A Proposed Discrimination Method for Practical Recipes

Overview of the Proposal

To solve the problems found in recipe websites, as discussed in the *Introduction*, this study proposes a discrimination method for practical recipes using machine learning. If it is possible to select a practical recipe that has the four features “it is easy to understand,” “it is concise,” “the contents are sufficiently descriptive,” and “tips and pointers are written in detail,” it would be possible to greatly enhance recipe usability. However, as it is too costly to manually construct training data for practical recipes, we used popular recipes, which could be obtained much more inexpensively, as training data. This approach is

generally referred to as distant supervision [3, 4] (supervised learning by using distant training data), and is frequently used. In addition, we have already confirmed that popular recipes share common characteristics with practical recipes to some extent. Figure 1 illustrates the method that we propose.

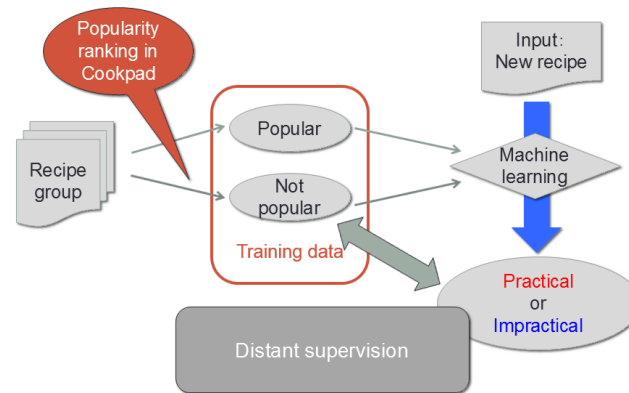


Figure 1. Overview of the proposed method

User support in the proposed method comprises:

1. support in choosing a recipe from vast recipe collections, and
2. support in choosing a recipe that is easy to cook.

These types of support are not prominent on conventional recipe websites. In particular, we have found that ease of cooking is barely considered in the popularity rankings. We expect that providing these two types of support will enable users not only to select a recipe easily, but also to prepare the dish easily without stopping.

Features Used to Discriminate Recipes

In the previous section, we analyzed the characteristics of popular recipes, and, based on the results, we discussed features that can be used to discriminate practical recipes. We chose seven features, and discriminated practical recipes based on those features using machine learning. We now explain how we treated each characteristic as a feature.

(1) THE CONTENTS ARE EASY TO UNDERSTAND

For this characteristic, the frequency with which the following item appeared was used as a feature. The parentheses below specify the relevant keyword.

- Description of the level of heat (heat)
- Description of the cooking time (minutes/seconds)
- Description of how to measure ingredients (measuring/measure)
- Use of onomatopoeia (rattling sound, etc.)

(2) TIPS ARE WRITTEN IN DETAIL

For this characteristic, the frequency with which tips appeared was used as a feature. We used 20 keywords and phrases, such as "... is a trick/... is the point, etc." The authors select these clues with the manual analysis of popular recipes.

(3) THE NUMBER OF COOKING STEPS IS LARGE

For this characteristic, we counted the number of cooking steps in each recipe, and used that number as a feature.

(4) EACH COOKING STEP IS SHORT

For this characteristic, we calculated the mean number of letters in a cooking step per recipe (the total number of letters in a recipe) divided by the number of cooking steps in the recipe and used that number as a feature.

Experiment: Discriminating Practical Recipes

Experiment Methodology

The purpose of the experiment was to examine the extent to which we can discriminate practical recipes. For the ten types of dishes, we used the top 500 and the bottom 500 recipes in the Cookpad popularity ranking for each type of dish, giving 10,000 pieces of data for the ten types of dishes (as published on November 18, 2013). For each type of dish, we performed tenfold cross-validation with 900 supervision data and 100 test data. We evaluated the test data manually, discriminating them as follows.

First, we chose 50 recipes that were discriminated as practical and 50 recipes that were discriminated as impractical for each type of dish, giving 1,000 recipes for the ten types of dishes. In selecting them, we sorted the result of the discrimination in decreasing order of margin width, selecting 50 recipes from the top and 50 recipes from the bottom. In this way, we chose 50 recipes of high practicality and 50 recipes of low practicality. Groups of 50 were selected because we considered that presenting 50 recipes to users in an actual search for practical recipes would be reasonable. We then evaluated manually whether those recipes were truly practical or impractical. As evaluation criteria, we used the four definitions of a practical recipe defined above in “*What is a practical recipe?*” We evaluated a recipe as practical if the evaluators deemed it so, even if the recipe did not meet the criteria. We then calculated the accuracy of the discrimination result and that of the manual evaluation. As a baseline for comparison, we calculated the results when using only the number of letters in one cooking step as the discriminating feature.

SVM (Support Vector Machine)

An SVM is an identification method that uses supervised learning. In this study, we implemented a classifier for the practicality of a recipe in LibSVM-3.17⁵, which is an SVM library. We used the radial basis function as a kernel, setting its C parameter as 1.0 after varying it from 1.0 to 10.0 in steps of 0.1.

Results of the Experiment

Table 5 shows the accuracy of the results for the proposed discriminator and the manual evaluation.

Dish name	Accuracy	
	Proposed method	Baseline
Nikujaga	0.850	0.590
Oyakodon	0.790	0.620
Carbonara	0.810	0.620
Cream Stew	0.830	0.650
Shrimp with chili sauce	0.810	0.600
French toast	0.830	0.630
Simmered kabocha squash	0.780	0.580
Mapo tofu	0.810	0.600
Potato salad	0.800	0.550
Sweet potato cake	0.820	0.610
Macro Average	0.813**	0.605

Table 5. Evaluation of the discrimination of practical recipes

** : There is a significant difference at the significance level of 1% from the baseline.

⁵ LIBSVM – A Library for Support Vector Machine. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

The macro average value for the ten types of dishes is 0.813 for the proposed method and 0.605 for the baseline. We can observe that the accuracy of the proposed method is superior. We performed a two-sided t-test at a significance level of 1%, concluding that the difference is significant.

Conclusion

For this study, we had the aim of supporting users in choosing recipes and in being able to prepare them smoothly. We therefore proposed and implemented a method for discriminating practical recipes using machine learning. To achieve this, we first analyzed the most popular recipes on Cookpad to identify the characteristics found in popular recipes. The characteristics of popular recipes are very similar to the characteristics of practical recipes, including their understandability and conciseness. We therefore decided to use seven characteristics found in popular recipes, as determined by an analysis, as the features used to discriminate practical recipes. We performed an experiment with the proposed discriminator, which was implemented as an SVM. We compared its accuracy with that of a baseline manual evaluation. The results confirm that the mean value of the accuracy for the ten types of dishes was 0.813 for the proposed method, representing a significant difference from the baseline method.

In future work, we plan to apply ranking strategy to the recipes from different social website with semi-supervised learning approach. We also plan to apply our proposal to find practical social recipe creators by the degree of practical cooking abilities of their recipes.

Acknowledgements

We would like to acknowledge the use of data provided by Cookpad Inc. for this study and to express our gratitude. This study was partially supported by Grant-in-Aid for Scientific Research (C) (24500291), Grant-in-Aid for Scientific Research (B) (25280110), and Grant-in-Aid for Exploratory Research (25540159).

References

- [1] Buykx, L. and Petrie, H. What cooks need from multimedia and textually enhanced recipes. *Proc. of IEEE International Symposium on Multimedia* (2011), 387-392.
- [2] Fusosha (Editing). *Recipe Blog Magazine 2013 Autumn/Winter*, Vol. 1, Fusosha Publishing, Tokyo, Japan, 2013. (In Japanese)
- [3] Mintz, M., Bills, S., Snow, R. and Jurafsky, D. Distant supervision for relation extraction without labeled data. *Proc. of the 47th Annual Meeting of the ACL* (2009), 1003-1011.
- [4] Purver, M. and Battersby, S. Experimenting with distant supervision for emotion classification. *Proc. of the 13th European Chapter of the Association for Computational Linguistics* (2012), 482-491.
- [5] Tachibana, A., Wakamiya, S., Namba, H. and Sumiya, K. Extraction of naming concepts based on modifiers in recipe title. *Proc. of the International MultiConference of Engineers and Computer Scientists* (2014), 507-512.
- [6] Yamakata, Y., Imahori, S., Sugiyama, Y., Mori, S. and Tanaka, K. Feature extraction and summarization of recipes using flow graph. *Proc. of the 5th International Conference on Social Informatics* (2013), 241-254.