6-26-2018

Alternative Ingredient Recommendation: A Co-occurrence and Ingredient Category Importance Based Approach

Kuan-Hung Liu
Department of Information Management, National Taiwan University, lkh9543@gmail.com

Hung-Chih Chen
Department of Information Management, National Taiwan University, hairy1016@gmail.com

Kuan-Ting Lai
Department of Information Management, National Taiwan University, r06725007@ntu.edu.tw

Yi-Ying Wu
Department of Information Management, National Taiwan University, r06725013@ntu.edu.tw

Chih-Ping Wei
Department of Information Management, National Taiwan University, cpwei@ntu.edu.tw

Alternative Ingredient Recommendation: A Co-occurrence and Ingredient Category Importance Based Approach

Completed Research Paper

Kuan-Hung Liu
Department of Information Management,
National Taiwan University, Taiwan
lkh9543@gmail.com

Hung-Chih Chen
Department of Information Management,
National Taiwan University, Taiwan
hairy1016@gmail.com

Kuan-Ting Lai
Department of Information Management,
National Taiwan University, Taiwan
r06725007@ntu.edu.tw

Yi-Ying Wu
Department of Information Management,
National Taiwan University, Taiwan
r06725013@ntu.edu.tw

Chih-Ping Wei
Department of Information Management, National Taiwan University, Taiwan
cpwei@ntu.edu.tw

Abstract

As many people will refer to a recipe when cooking, there are several recipe-sharing websites that include lots of recipes and make recipes easier to access than before. However, there is often the case that we could not get all the ingredients listed on the recipe. Prior research on alternative ingredient substitution has built a recommendation system considering the suitability of a recommended ingredient with the remained ingredients. In this paper, in addition to suitability, we also take the diversity of the ingredient categories and the novelty of new combination of ingredients into account. Besides, we combine suitability with novelty as an index, to see whether our method could help find out a new combination of ingredients that is possibly to be a new dish. Our evaluation results show that our proposed method attains a comparable or even better performance on each perspective.

Keywords: Alternative Ingredient Recommendation, Recipe, Cooking, Co-occurrence Frequency, Category Importance

Introduction

Cooking is an activity which requires much experience, knowledge and skills. It is hard for a beginner to make delicious dishes without any assistance. And for those who want to cook a specific dish but have no idea what to do, a recipe is usually a good solution. As the Internet becomes more widespread, instead of reading cookbooks, more and more people turn to recipe-sharing websites where people can share and find information easily (Tsuguya et al. 2011). For example, iCook (https://icook.tw/), one of the most popular recipe sharing websites in Taiwan, has more than 5 million users with about 63 million click through rate (CTR) monthly.

While reading a recipe could be a good start for cooking an unfamiliar dish, there is often the case that we could not follow exactly what the recipe says, especially regarding the ingredients. For example, we might dislike a specific ingredient listed in a recipe or just do not have that in the refrigerator. For a professional cook, there is a need for them to discover new combinations of ingredients that go well together, and further create new dishes, which sometimes can be done by just replacing one ingredient in an existing recipe. In those situations, we will want to find an alternative ingredient that is suitable with the remained ingredients in that recipe for substitution considering taste or flavor. Besides, the diversity of categories of alternative ingredients is also an important factor to be considered. For example, when we want to replace carrot in pumpkin soup, we might desire something other than vegetable such as seafood or meat. In addition, the novelty of the new combination of ingredients will also be taken into account when trying to make a creative dish that is not existing before. However, it is hard to find an ideal one due to the complex relationships between ingredients.

While most of prior research has been working on recipe recommendation, the problems mentioned above still remain unsolved. Since the result of their recommendation is a whole recipe instead of a single ingredient. In a previous work (Shino et al. 2016), they conduct the ingredient replacement based on co-occurrence frequency. However, the idea of only recommending the ingredient in the same category as the replaced ingredient will restrict the diversity of the result and ignore potentially suitable ingredients in different categories. For example, meat would be considered as a suitable replacement of eggs in a salad recipe.

In this paper, we propose a method that utilizes ingredient co-occurrence frequency along with category importance calculated from recipe data for each replacement. Our algorithm can recommend suitable ingredients without loss of diversity. Furthermore, with the diversity of recommended ingredients, we assume that a novel combination of ingredients can be produced, which is potentially to be an attractive, new dish. Therefore, we also take the novelty of the combination into consideration, and combine with suitability to be an index, to see if our idea can be verified.

Our study provides several contributions to the literature on alternative ingredient substitution in a recipe. First, we present an approach that can find appropriate categories automatically when recommending alternative ingredients. Second, we consider not only suitability but diversity into our method for more effective recommendation. Third, we combine novelty of the combination with suitability of an alternative ingredient as an index for evaluation, which we think is vital as a good recommendation system, providing a new perspective to this area of research.

Related Work

Previous recipe related research mainly focuses on three aspects, menu recommendation (Kuo et al. 2012), recipe recommendation (Mino et al. 2009; Teng et al. 2012; Ueda et al. 2011; van Pinxteren et al. 2011) and ingredient replacement (Shidochi et al. 2009; Shino et al. 2016; Yamanishi et al. 2015). For menu recommendation, the goal is to suggest a set of recipes. Kuo et al. (2012) construct a recipe graph to capture the co-occurrence relation between recipes. A menu is generated from the graph and well-accompanied with user's query ingredients.

In terms of recipe recommendation, the goal is to find recipes that are similar to a given one. Since recipes are built upon ingredients, how to generate the ingredient similarity with different kinds of features is often discussed. Co-occurrence relation between ingredients is one of the feature that has often been utilized to illustrate the similarity of ingredients. For example, Teng et al. (2012) construct two types of networks using PMI (i.e. pointwise mutual information) to capture the complementary and substituting relation of ingredients, respectively. Their method discloses the possibility of using PMI and network information in recipe recommendation. Another branch of research incorporates user preference for recipe recommendation (Forbes et al. 2011; Ueda et al. 2011). Nutrition of the ingredients in recipe is also an important factor that may affect one's recipe preference. Mino et al. (2009) proposed a method to recommend recipes under the restriction of carbohydrate, fat, protein. In summary, many features mentioned above depicting ingredient similarity are also useful for ingredient replacement research.

In comparison to recipe recommendation, ingredient replacement has attracted relatively less attention in the field. In order to recommend replaceable ingredients, some research try to find different vector representation for ingredients. Shidochi et al. (2009) proposed a method that finds replaceable materials by first extracting the cooking actions that correspond to each material, then measuring the similarity of the extracted cooking actions in the same recipe group. Their method incurs several limitations. First of all, ingredient vectors generated by cooking actions is considered less direct comparing with those generated with ingredient co-occurrence frequency. Second, similarity measured within the same recipe group will not be able to find replaceable materials across cooking context. Moreover, their algorithm does not take the replacement of seasonings into consideration.

Seasonings are often treated as stop words since their frequency of occurrence is much higher than other types of ingredients. Yamanishi et al. (2015) proposed a method that recommends alternative ingredients except seasonings based on the co-occurrence relation. In our observation, frequency-based algorithm tends to recommend seasonings no matter what ingredient to be replaced because they appear in almost every recipe. In most cases, it is inappropriate to recommend seasonings when replacing meat or other ingredients that are not seasonings. However, if we remove seasonings from data, algorithms will not be able to recommend alternatives for seasonings.

Shino et al. (2016) constructed a system that suggests alternative ingredients based on both co-occurrence frequency and ingredient category information. In their research, they found that ingredients belonging to the same category are much suitable to replace each other in most cases. However, restricting the category of alternative ingredients to be the same as the replaced ingredient is making a strong assumption. And such assumption will hurt the recall of recommendation. For example, vegetables and meat are both considered suitable when replacing egg in a salad recipe, but vegetables and meat are not in the same category as egg. Therefore, we believe that we can take advantage of category information to give better suggestions by incorporating ingredient co-occurrence frequency and category importance calculated from recipes for a substitution.

Our Proposed Method

Our proposed method consists of the following three steps.

Step 1: Find relationship between alternative ingredient and remained ingredients.

Step 2: Calculate category importance for substitution.

Step 3: Combine Step 1 with Step 2 and get final score.

Where Step 1 focuses on co-occurrence between alternative ingredient and remained ingredients to find suitable ingredients. Step 2 focuses on finding suitable alternative category.

Step 1: Find Relationship Between Alternative Ingredient and Remained Ingredients

In general, high co-occurrence frequency between two ingredients means many people think these two ingredients are suitable to be put in a recipe. Therefore, co-occurrence frequency is an important index for choosing alternative ingredient. To calculate co-occurrence score, we use pointwise mutual information (PMI) because PMI will be high if two ingredients usually co-occur with only each other. This paper assumes that ingredients with high PMI scores between remained ingredients will be suitable to be recommended. For PMI between two ingredients, we calculate PMI score by the following equation:

$$\mathrm{PMI}(a, R_{i,n}) = \begin{cases} 0 & (CoOc(a, R_{i,n}) < 2) \\ (\ln \dfrac{P(a, R_{i,n})}{P(a)P(R_{i,n})} / -\ln P(a, R_{i,n}) + 1)/2 & (CoOc(a, R_{i,n}) \geq 2) \end{cases} \tag{1}$$

where

$$rep_i = Replaced\ ingredient\ in\ recipe\ i\ ,$$

$$R_i = Recipe\ i\ without\ rep_i,$$

$$R_{i,n} = n\text{-}th\ remained\ ingredient\ in\ R_i\ ,$$
$$(a\ \in\ All\ ingredient | a\ \neq ex_i, a\ \notin R_i)\ ,$$
$$CoOc(a, R_{i,n}) = \frac{\#\ of\ recipes\ contains\ a, R_{i,n}}{\#\ of\ recipes},$$
$$P(a, R_{i,n}) = \frac{CoOc(a, R_{i,n})}{\#\ of\ recipes}$$
$$P(a) = \frac{\#\ of\ recipes\ contains\ a}{\#\ of\ recipes}$$

We set $PMI = 0\ if\ CoOc(a, R_{i,n}) < 2$ to handle a condition that $(a, R_{i,n})$ occurs only in one recipe. We assume this situation is a special case, so we do not take it into consideration. To prevent having outliers, we employ normalized PMI and adjust it to range [0,1] when calculating mean of PMI (Bouma 2009). Next, we calculate PMI score between a and R_i using the following equation:

$$PMI\_Score(a, R_i) = \frac{\sum_n PMI(a, R_{i,n})}{\#\ of\ ingredients\ in\ R_i} \tag{2}$$

Step 2: Calculate Category Importance for Substitution

Although PMI may find a suitable ingredient for the remained ingredients in a recipe, it does not take into account the combination of categories. For example, it may recommend another cheese when we want to replace chicken from Caesar salad but we may expect another meat or seafood. To solve this problem, we develop this step to find the right category. First of all, we build category frequency vectors for every recipe. Let $C = \{c_1, c_2, \ldots, c_{n_c}\}$ be a set of n_c categories. For each recipe RD_j in the recipe dataset, we then have a category frequency vector:

$$Cat\_Fq\_Vec_{RD_j} = (Freq(c_1, RD_j), Freq(c_2, RD_j)\ , \ldots, Freq(c_{n_c}, RD_j)) \tag{3}$$

where

$$n_c = \#\ of\ categories,$$
$$Freq(c_k, RD_j) = \#\ of\ ingredients\ in\ recipe\ RD_j\ which\ is\ in\ category\ c_k.$$

With all category frequency vectors, we can find some similar vectors (Neighbors) of $Cat\_Fq\_Vec_{R_i}$. By observing difference between $Cat\_Fq\_Vec_{R_i}$ and its Neighbors, we will know in which categories R_i misses some ingredients. Then we can give these categories higher scores. To find neighbors of $Cat\_Fq\_Vec_{R_i}$, we calculate Euclidean distance between each $Cat\_Fq\_Vec_{RD_j}$ and $Cat\_Fq\_Vec_{R_i}$, then pick neighbors whose distance are less than $\frac{NUM(R_i)}{3}$.

$$Neighbors = \{Cat\_Fq\_Vec_{RD_j} | D(Cat\_Fq\_Vec_{RD_j}, Cat\_Fq\_Vec_{R_i}) < \frac{NUM(R_i)}{3}\} \tag{4}$$

where

$$NUM(R_i) = \#\ of\ ingredients\ in\ R_i,$$
$$D(Vec_a, Vec_b) = Euclidean\ distance\ between\ Vec_a and\ Vec_b.$$

To calculate the importance scores for categories, we think the extra category that R_i lacks is more important than the category that R_i contains. For example, if the remained ingredients in R_i belong to seasoning and vegetable, we may prefer the result that the recommended ingredient belongs to a different category other than the existing two, yet not exclude the possibility of having just these two categories in a recipe. As a result, we take the square root of the ingredient quantity to show the effect:

$$Diff\_Vec_d = \left[\sqrt{neb_{d,c_1}}, \sqrt{neb_{d,c_2}}, \dots, \sqrt{neb_{d,c_{n_c}}}\right] - [\sqrt{r_{i,c_1}}, \sqrt{r_{i,c_2}}, \dots, \sqrt{r_{i,c_{n_c}}}] \tag{5}$$

where

$$neb_{d,c_k} = \ ingredient\ quantity\ in\ category\ c_k\ \ of\ \ Neighbor\ d\ ,$$

$$r_{i,c_k} = ingredient\ quantity\ in\ category\ c_k\ of\ Cat\_Fq\_Vec_{R_i}\ .$$

So that with the same difference between neb_{d,c_k} and r_{i,c_k}, the importance of a category decreases as the r_{i,c_k} increases. For example, $[\sqrt{1}, \sqrt{2}, \sqrt{3}] - [\sqrt{0}, \sqrt{1}, \sqrt{2}] \cong [1, 0.414, 0.318]$. Besides, we only consider the positive value in $Diff\_Vec_d$ by setting the negative values to 0, since our goal is to find a category to "add" into R_i.

With difference of vectors, we calculate an average importance score for each category and normalize them to [0,1] using following equation:

$$Cat\_Impt\_Vec\ = \frac{\sum_{d=1}^{n_d} Diff\_Vec_d}{\#\ of\ Neighbors} \tag{6}$$

$$Cat\_Impt\_Vec\_N = \frac{Cat\_Impt\_Vec}{\sum_{i=1}^{n_c} Cat\_Impt\_Vec_{c_i}} \tag{7}$$

where

$$Cat\_Impt\_Vec = category\ importance\ score\ vector$$

$$Cat\_Impt\_Vec\_N = normalized\ Cat\_Impt\_Vec$$

To prevent some error cases, this paper has two assumptions. First, according to Shino et al. (2016), it is reasonable to recommend ingredients of the same category as rep_i, so we set the score of the category of rep_i to be the same as the maxima in $Cat\_Impt\_Vec\_N$. Second, in most cases, it is inappropriate to recommend seasoning if the category of rep_i is not seasoning. Therefore, we exclude seasoning in our recommendation list if the category of the rep_i is seasoning.

Step 3: Combine Step 1 with Step 2 and Get Final Score

For each ingredient a we calculate its final score (PMI_Cat_Score$_a$) by the following equation:

$$PMI\_Ca\_Socre_a = \alpha PMI\_Score(a, R_i) + (1 - \alpha) Cat\_Impt\_Vec\_N_{Category(a)} \tag{8}$$

where

$$Cat\_Impt\_Vec\_N_{category(a)} = the\ importance\ score\ of\ the\ category\ of\ ingredient\ a$$

The coefficient α can be used to adjust the influences between these two score. If differences between PMI_Score are less than the gaps between $Cat\_Impt\_Vec\_N$, larger α will balance the influences.

Empirical Evaluation

Data Collection

Recipe Dataset

In Taiwan, iCook is the most popular recipe-sharing website where people can upload their cooks. Recipes uploaded to iCook contain cooking actions and used ingredients. In this paper, we collect 138,398 public recipes with their ingredients from iCook. Due to the user generated content, we have to do some data-preprocessing to make raw data more usable.

The first step is to process the freeform text of ingredient data. In raw data, there are 138,266 unique ingredient names, but in fact some ingredients are with the same meaning but presented in different text and some rarely appear in recipes, e.g. “pineapple”, used in 600 recipes is considered same meaning with “fresh pineapple” used in 64 recipes. Therefore, top 1000 frequent ingredients which cover 73.9% ingredients in our recipe data are selected. There is still some duplicate meaning in that, so we manually combine these ingredient names and finally get 720 names as our final ingredient list.

The second step is to filter non-suitable recipes out. There are two main types of food in our recipe data: sweet and savoury food. For sweet food, some ingredients are difficult to be replaced, e.g. “flour” and “egg” used in “strawberry cake” are coagulation characteristic. So, this kind of recipes are not considered in this paper. Then, we filter the recipes which contain ingredients not in our ingredient list out. Finally, there are 54,728 recipes left.

Ingredient Category

“Taiwan Food Nutrition Database” is a database which contains ingredients with their category and nutrition information. It categorizes food into 18 groups and we further add 2 categories as “Spices” and “Non-ingredient”.[1] In Chinese food, shallot, ginger, garlic are usually used to flavor food rather than eat them. Also, some cookware may appear in our ingredient data, so we need “Non-ingredient” category to handle them.

Baseline Model

Shino et al. (2016) proposed a method that applies Naïve Bayes classifier to calculate the co-occurrence frequency, and recommends ingredients belonging to the same category of rep_i. We implement their algorithm as our baseline model and compare its effectiveness with that of our proposed method.

Evaluation

To evaluate the effectiveness of our proposed method, we define “suitability” and “novelty” as follows: whether the recommended ingredient is suitable for remained ingredients, and whether the combination of the recommended ingredient and the remained ingredients is unexpected. Ingredient replaced experiments are conducted with twelve recipes. Table 1 shows our twelve experiment recipes. The twelve recipes are the most common recipes in our recipe data and replaced ingredients are selected by maximizing the diversity of ingredient categories.

Table 1. Experiment Recipes

Recipe Name	Recipe Ingredients	Replaced Ingredient
Pumpkin Soup	pumpkin, white mushroom, onion, stock, fresh cream, salt	pumpkin
Three Cup Chicken	chicken legs, garlic, soy sauce, basil, sugar, ginger, cooking rice wine, sesame oil	chicken legs
Three Cup Eryngii	eryngii, basil, crystal sugar, chili, cooking rice wine, ginger, sesame oil, oil, soy sauce	eryngii
Kung Pao Chicken	chicken breast, pricklyash peel, garlic, egg, cooking rice wine, shallot, sugar, corn flour, soy sauce, peanut, dry chili	chicken breast
Chinese Pickled Cucumber	vinegar, cucumber, garlic, sugar, sesame oil	vinegar

[1] The original 18 groups are: Cereals, Grains, Nuts, Fruits, Vegetables, Laminaria, Mushrooms, Beans, Meat, Seafood, Egg, Milk, Oil, Sugar, Drink, Seasoning, Snack, and Processed Food.

Ants Climbing a Tree	green bean noodle, ground pork, sesame oil, ginger, chili bean sauce, soy sauce, water, shallot, sugar	green bean noodle
Potato Salad	cucumber, mayonnaise, egg, salt, potato, ham	cucumber
Potato Stew	potato, mirin, Japanese stock, say sauce, onion, sake, carrot, garlic, shallot, pork	potato
Shrimp Pot with Bean Threads	shrimp, garlic, soy sauce, green bean noodle, thick soy, chili, water, cooking rice wine, shallot	shrimp
Scrambled Eggs and Tomatoes	beef tomato, ketchup, ginger, soy sauce, egg, water, salt	beef tomato
Mapo Tofu	tofu, potato starch, vinegar, pricklyash peel, garlic, shallot, ginger, ground pork, chili bean sauce, soy sauce	tofu
Japanese Dressing Salad	lettuce, carrot, salt, apple, egg, white sesame, Japanese style sauce	egg

With the purpose to compare the influence of α in our proposed method (referred to as the PMI_Category algorithm), we make $\alpha = 0.5$ and $\alpha = 0.8$ and recommend 5 ingredients for the twelve recipes respectively. Along with 5 recommended ingredients of baseline model, we randomly sorted all recommended ingredients to get finalized recommended ingredient list for each recipe. Table 2 show recommendations of "Pumpkin Soup" for three methods.

Table 2. Recommendations of "Pumpkin Soup" for different models

	1	2	3	4	5
Baseline	Carrot	tomato	broccoli	lettuce	sweet pepper
PMI_0.5	tomato paste	tomato	celery	broccoli	carrot
PMI_0.8	tomato paste	tomato	celery	ground beef	carrot

In the experiment, three subjects were invited to rate the seven-point scale questionnaires with two indices: suitability and novelty. The subjects all have over thirty-year experience of cooking and at least cook four times a week. During the experiment, recipe names, replaced ingredients, remained ingredients were displayed and each subject independently rated the suitability and novelty of ingredients in our recommended ingredient list from 1 to 7 points. More points in suitability means that the ingredient is more suitable for the replacement. Similarly, more points in novelty means that the ingredient more infrequently appears with remained ingredients. Table 3 shows the rating of "Pumpkin Soup" by one participant. After collecting raw rating data, we examined the interrater reliability (Koo et al. 2016) with interclass correlation ICC(2, 1) for absolute agreement, having significance level of 0.05. The ICC value of suitability and novelty are 0.3883 and 0.4429 respectively, which shows our subjects has a certain degree of agreement.

Table 3. Rating of "Pumpkin Soup" by One Participant

	Carrot	Tomato	Broccoli	Lettuce	Sweet Pepper	Tomato Paste	Celery	Ground Beef
Suitability	6	7	7	4	6	6	6	6
Novelty	3	2	2	5	4	3	2	4

Having the ratings ready, we obtain each recommended ingredient score by averaging rating of three subjects. To compare different models, the average suitability and novelty scores of all recommended ingredients for each method are calculated. Because one of our purpose is to find new combination of ingredients, we expect not only the new recommended ingredients suitability but also novelty score high. So, we further calculate harmonic mean of suitability and novelty scores. To know the ability of recommending different ingredient categories, category diversity is also calculated by averaging numbers of distinct categories across the twelve recipes.

Table 4 shows the average score of four indices mentioned above. The suitability scores of the PMI_Category algorithm with $\alpha = 0.5$ and $\alpha = 0.8$ are 3.8% and 4.2% less than the baseline model respectively. It means that the recommended ingredients by our proposed PMI_Category algorithm are slightly less suitable than by the baseline model. The novelty scores of the PMI_Category algorithm with $\alpha = 0.5$ and $\alpha = 0.8$ are 11.3% and 12.1% higher than the baseline model respectively. it means that the PMI_Category algorithm has better ability to recommend new combinations of ingredients. As for the result of harmonic mean, the scores of the PMI_Category algorithm with $\alpha = 0.5$ and $\alpha = 0.8$ are 4.4% and 5.7% higher than the baseline model respectively. This comprehensive index indicates that the PMI_Category algorithm has better ability to consider both suitability and novelty. The scores of category diversity indicate the PMI_Category algorithm with $\alpha = 0.5$ and $\alpha = 0.8$ can recommend 1.416 and 2.175 different categories of ingredients on average. It shows that the higher α value can improve category diversity.

Table 4. Average Score of Three Models

	Suitability	Novelty	Harmonic Mean	Category Diversity
Baseline	5.233	3.478	3.848	1
PMI_0.5	5.033	3.872	4.017	1.416
PMI_0.8	5.011	3.9	4.075	2.175

In order to realize whether our models outperform the baseline model in novelty and harmonic mean, we conducted independent sample T-test on average score of different models. Table shows the p-values between different models. The result indicates that the suitability of our models and that of baseline model are comparable (p-values are greater than 0.1), and the novelty and harmonic mean of PMI_Category with $\alpha = 0.8$ are significantly greater than those of the baseline model (p-values are less than 0.1).

Table 5. P-Values of T-test Result

	Suitability	Novelty	Harmonic Mean
Baseline / PMI_0.5	0.2904	0.1305	0.2511
Baseline / PMI_0.8	0.2312	0.0943	0.0947

In terms of recommendation system, the order of recommended ingredients matters. The high-score ingredients ranked at front position deserve more points, so the normalized discounted cumulative gain (Järvelin, K et al. 2002) scores of different methods are calculated:

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

Because in our experiment we only recommend five ingredients for each recipe, we set $p = 5$ which means we calculate the nDCG score at rank position 5. DCG is discounted cumulative gain computed as:

$$DCG@5 = \sum_{i=1}^{5} \frac{score_i}{log_2(i+1)}$$

IDCG is ideal discounted cumulative gain which sorts all recommended ingredients of a recipe by its scores reversely and calculates its top 5 DCG scores.

Table 6 shows the nDCG score of different models. The suitability scores of the PMI_Category algorithm with α = 0.5 and α = 0.8 are 4% and 4.7% less than the baseline model respectively. However, the harmonic mean scores of the PMI_Category algorithm with α = 0.5 and α = 0.8 are 4.8% and 4.9% higher than the baseline model respectively. It implies that the recommendation of the baseline model is slightly more suitable than the PMI_Category algorithm, but the recommendation of the PMI_Category algorithm has better ability to consider both suitability and novelty.

Table 6. nDCG Score of Three Models

	Suitability	Harmonic Mean
Baseline	0.898	0.65
PMI_0.5	0.862	0.681
PMI_0.8	0.856	0.682

Discussion

In this paper, we discuss the performance by four indices which are "suitability," "novelty," "harmonic mean of suitability and novelty," and "category diversity." We found their relationship with algorithms below:

The baseline model achieves the highest suitability score and lowest novelty score in Table 4. After the observation of the baseline model's recommendations, we found that the baseline model usually recommends the most common ingredients in the category of rep_i since it multiplies $P(a)$ to the co-occurrence score. Therefore, this algorithm puts it at a lower risk of recommending unsuitable ingredients but sacrifices novelty score. However, this way also limits the category diversity. If they allow their algorithm to recommend other categories, it may only recommend the most commonly used ingredients like garlic and pepper which are so common that they are not worth recommending.

Our proposed PMI_Category algorithm with α = 0.8 attains the highest novelty score and lowest suitability score in Table 4. Compared with the baseline model, PMI puts $P(a)$ in the denominator, so ingredients with high co-occurrence frequency but low individual occurrence frequency will get higher PMI score. Therefore, PMI score tends to recommend less common ingredients than the baseline model does. However, recommending novel combination of ingredients may cause lower suitability because people may consider it is unsuitable when only a few people use. Furthermore, our category score takes the combination of categories into account, which will enhance the suitability. Setting α to 0.8 causes less weight on the category score, which results in higher category diversity at the cost of suitability.

The PMI_Category algorithm with α = 0.8 attains higher category diversity score than α = 0.5. Table 7 shows that the differences between PMI scores is less than the gaps between category scores, so larger α can reduce the influence of category scores which may bring some ingredients with higher PMI scores but a little lower category scores to top. Therefore, the PMI_Category algorithm with α = 0.8 has the highest category diversity score.

Table 7. Top 6 Alternative Ingredients When Replacing Pumpkin in "Pumpkin Soup" Recipe

Ingredient	PMI score	Category Importance	α = 0.5 score	α = 0.8 score
Tomato Paste	0.57577 (1st)	0.252	0.41392 (1st)	0.51103 (1st)
Ground Beef	0.55963 (2nd)	0.195	0.37731 (6th)	0.48670 (4th)
Tomato	0.55119 (3rd)	0.252	0.40163 (2nd)	0.49137 (2nd)

Coriander	0.55091	(4th)	0.252	0.39990	(3rd)	0.4886	(3rd)
Broccoli	0.54207	(5th)	0.252	0.39708	(4th)	0.48407	(5th)
Carrot	0.53166	(6th)	0.252	0.39187	(5th)	0.47572	(6th)

Table 8 suggests that this study proposes a way to recommend ingredients which combines a certain degree of suitability (0.22 points (4.25%) less than the baseline model) with a greater variety (novelty is 0.42 points (12.14%) more than the baseline model).

Table 8. Difference Between the Average Scores of Different Models

	Suitability	Novelty	Harmonic Mean
PMI_0.8 minus Baseline	-0.22 (-4.25%)	0.42 (12.14%)	0.23 (5.89%)
PMI_0.8 minus PMI_0.5	-0.022 (-0.44%)	0.028 (0.7%)	0.06 (1.44%)

Observing Table 7 and Table 9, our proposed PMI_Category algorithm with $\alpha = 0.8$ attains a greater harmonic mean and a comparable suitability. As a result, this empirical result suggests that our PMI_Category algorithm with $\alpha = 0.8$ is more suitable for someone who wants to try new combination of ingredients, because our algorithm is able to recommend diversified and novel ingredients.

Table 9. Difference Between the nDCG Scores of Different Models.

	Suitability	Harmonic Mean
PMI_0.8 minus Baseline	-0.041 (-4.58%)	0.032 (4.84%)
PMI_0.8 minus PMI_0.5	-0.006 (-0.64%)	0.001 (0.14%)

Conclusion and Future Research Directions

Our study provides two contributions to the literature on recipe ingredient substitution. First, we propose an alternative ingredient recommendation system which has following features: 1) incorporating category importance to effectively suggest ingredients, 2) Recommending ingredients that have higher harmonic mean on average, and 3) Recommended ingredients are richer in category diversity. Besides those features, since we added alpha as a parameter, users are able to try out different values for alpha to come up with preferable ingredients. Second, our experiment results show that, for a query recipe, our model successfully recommends ingredients that are less frequently used in the recipe and at the same time rich in category diversity by slightly trading off suitability.

In this study, we only focus on recommending one alternative ingredient for a focal recipe. To make our system more flexible, we can extend our proposed algorithm by recommending more than one alternative ingredients in a recipe in the future. Besides, we think that by adjusting our model to produce personalized recommendation, we can further improve our system performance. User preferences to recipes and ingredients can be employed to adjust ingredient frequency information. On the other hand, we also plan to discuss the possibility of utilizing our proposed method in other application contexts. System incorporating frequency and category information is also possible to solve the problem of alternatives recommendation to an element within in some other topics such as recipe set or music playlist.

Acknowledgements

This work was supported in part by the Ministry of Science and Technology of the Republic of China (Taiwan) under the grants 104-2410-H-002-143-MY3 and 106-2410-H-002-068-MY3.

References

Bouma, G. 2009. "Normalized (pointwise) mutual information in collocation extraction." *Proceedings of the Biennial GSCL Conference*, Potsdam, Germany, pp. 31-40

Forbes, P. and Zhu, M. 2011. "Content-boosted matrix factorization for recommender systems: experiments with recipe recommendation." *Proceedings of the fifth ACM conference on Recommender systems*, pp. 261-264.

Järvelin, K. and Kekäläinen, J. 2002. "Cumulated gain-based evaluation of IR techniques." *ACM Transactions on Information Systems* (TOIS), 20(4), pp. 422-446.

Koo, Terry K., and Mae Y. Li. 2016 "A guideline of selecting and reporting intraclass correlation coefficients for reliability research." *Journal of Chiropractic Medicine 15.2*, pp. 155-163.

Kuo, F. F., Li, C. T., Shan, M. K. and Lee, S. Y. 201). "Intelligent menu planning: Recommending set of recipes by ingredients." *Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities*, pp. 1-6.

Mino, Y. and Kobayashi, I. 2009. "Recipe recommendation for a diet considering a user's schedule and the balance of nourishment." *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*. Vol. 3, pp. 383-387.

Shidochi, Y., Takahashi, T., Ide, I. and Murase, H. 2009. "Finding replaceable materials in cooking recipe texts considering characteristic cooking actions." *Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities*, pp.9-14.

Shino, N., Yamanishi, R., and Fukumoto, J. 2016. "Recommendation System for Alternative-Ingredients Based on Co-occurrence Relation on Recipe Database and the Ingredient Category." *Advanced Applied Informatics (IIAI-AAI), 2016 5th IIAI International Congress on.* IEEE, pp. 173-178.

Tsuguya Ueta, Masashi Iwakami, and Takayuki Ito, "A recipe recommendation system based on automatic nutrition information extraction." *Lecture Notes in Computer Science,* 7091, pp.79-90, Springer, 2011.

Teng, C. Y., Lin, Y. R. and Adamic, L. A. 2012. "Recipe recommendation using ingredient networks." *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 298-307.

Ueda, M., Takahata, M. and Nakajima, S. 2011. "User's food preference extraction for personalized cooking recipe recommendation." *Workshop of ISWC*, pp. 98-105.

van Pinxteren, Y., Geleijnse, G., and Kamsteeg, P. 2011. "Deriving a recipe similarity measure for recommending healthful meals." *Proceedings of the 16th international conference on Intelligent user interfaces*, pp. 105-114

Yamanishi, R., Shino, N., Nishihara, Y., Fukumoto, J. and Kaizaki, A. 2015. "Alternative-ingredient Recommendation Based on Co-occurrence Relation on Recipe Database." *Procedia Computer Science* 60, pp. 986-993.