International Journal of Engineering & Technology, 7 (4.4) (2018) 5-6

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Culinary Recipe Recommendation based on Text Analytics

Jiheon Hong1, Heejung Lee2*

[1]*Graduate School of Technology Innovation and Management*
[2] *Department of Interdisciplinary Industrial Studies*
[1,2]*Hanyang University*
[1,2]*Republic of Korea*

Abstract

Many researchers and practitioners have studied the recipe recommendation, and that problem is not only to find the tasty dishes based on the individual's preference, but also to generate new ones. In the digital age, understanding and utilizing text data is one of the most important part in the knowledge discovery. In this paper, we proposed how to use text analysis in the recipe recommendation problem and provided the insights to design new recipes.

Keywords:Text mining, Recommender system, Food recipe

1. Introduction

People are constantly generating unstructured and the most representative form of unstructured data is the text-based natural language. Text is very useful in the following aspects. First, text is the most natural way to symbolize our knowledge. Second, text is the most common type of information we encounter. Third, text is the most expressive form of information. So understanding and utilizing text is one of the most important part in the knowledge discovery. Considering the vast amount and scope of text in the future, text analysis techniques can find useful hidden knowledge and have applications in a variety of context. In this paper, we studied how to use text analysis in cooking problem, which is one of the most important parts of every life. When people usually cook, most ideas about what to cook will come from their experience or knowledge. If they want to cook a new dish, they can browse the web. Furthermore, if they have some ingredients in their kitchen, there can be a novel way to recommend the recipe that can be made with these available ingredients. The main goal of this research is to devise recipe recommendation method considering the text in the food recipes, which are increasingly shared over the digital world.

2. Recipe Text Processing

2.1 Recipe Feature Tagging

The first step in processing the recipe is to identify the ingredients, cooking methods, and related information from the text in the recipe. Recipes consist of a variety of ingredients and methods for preparing a given dish. For example, "*달군팬에돼지고기 100g 을넣고볶는다(put 100g of pork into a hot pan and fry)*" To identify ingredients, methods, and related information from the recipe sentence, we parsed the recipe sentences and tagged the necessary information using the standard chunking method, BIO encoding[2], which tags each parsed word as being begin-of-entity(B-X) or continuation-of-entity(I-X) or others(O). In this study, we proposed the tagging rules as shown in table 1 and the above example sentence will be tagged like: *[달구]$_{B\text{-}C}$[ㄴ]$_{I\text{-}C}$ [팬]$_{B\text{-}T}$ [에]$_{I\text{-}T}$ [돼지고기]$_{B\text{-}E}$ [100]$_{B\text{-}Q}$ [g]$_{B\text{-}U}$ [을]$_{I\text{-}U}$ [넣]$_{B\text{-}A}$ [고]$_{I\text{-}A}$ [볶]$_{B\text{-}A}$ [는다]$_{I\text{-}A}$ [.]$_{O}$*

Table 1. BIO encoding rule

Meaning	*Activity*	*Ingredients*	*Quantity*	*Unit*	*Tool*	*Comment*
Encoding	*B-A, I-A*	*B-E, I-E*	*B-Q, I-Q*	*B-U, I-U*	*B-T, I-T*	*B-C, I-C*

With the BIO tagging we can extract the structured data from recipe using conditional random fields[1], which models the conditional distribution $p(\boldsymbol{y}|\boldsymbol{x})$ over input sequence $\boldsymbol{x}$(parsed words) and label sequence $\boldsymbol{y}$(BIO encoding) that maximizes $p(\boldsymbol{y}|\boldsymbol{x})$. CRF introduces the feature function, which is defined as $f_k(y_i, y_{i-1}, \boldsymbol{x}_i)$ and its corresponding weight λ_k, based on what information might be useful to determine the relationship between input sequence **x** and output label sequence **y**. The conditional distribution $p(\boldsymbol{y}|\boldsymbol{x})$ is defined as equation (1).

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x})} \prod_{i=1}^{t} exp\left\{\left(\sum_{k=1}^{K} \lambda_k f_k(y_i, y_{i-1}, \boldsymbol{x}_i)\right)\right\} \quad (1)$$

where, $Z(\mathbf{x})$ is normalization function.

In general, there are two fundamental types of relations between words: syntagmatic and paradigmatic associations. If the word pairs, *pour-sauce* or *melt-butter*, co-occur more frequently, there is a syntagmatic relation between two words. If the two words, *eat-drink* or *coffee-milk*, can be substituted for one another in a sentence, the relation between two words is paradigmatic. We found two associations from the parsed words considering the context of *ingredients*, *activities*, and so on. Satisfying both syntagmatic and paradigmatic associations, the novel recipes can be generated by mutating the ingredients of existing recipes. For example, the relation between two words, *돼지고기(pork)*and *쇠고기(beef)*, is paradigmatic and there are syntagmatic relations in theword pairs, *돼지고기(pork)-넣고볶는다(put and fry)* and *쇠고기(beef)-넣고볶는다(put and fry)*. With this associations, we can generating the new sentence by

replacing 돼지고기(*pork*) with 쇠고기(*beef*): "*달군팬에**돼지고기** 100g 을넣고볶는다(put 100g of **pork** into a hot pan and fry)→달군팬에**쇠고기** 100g 을넣고볶는다(put 100g of **beef**into a hot pan and fry)*"

3. Recipe Recommendation Method

The main goal of this research is to suggest the recipe recommender, which recommends recipes that can be made using the ingredients when we have the available ingredients list. For example, mackerel, potato, cucumber, and onion are available, which recipe would you recommend? We can probably think that mackerel and potato go together. That is because we have a set of beliefs based on our experience or knowledge. But we cannot expect to put together watermelon and potato. This recipe disagrees with our beliefs about food. To solve this problem, we defined *modified-PMI* (*MPMI*) as equation (2).

$$MPMI = \frac{1}{n}\sum_{i\neq j} PMI(E_i, E_j), \quad where PMI = \log\left(\frac{P(X,Y)}{P(X)P(Y)}\right), \qquad (2)$$

P(X, Y) = (*number of recipes containing X and Y*) / (*total number of recipes*),

P(X) = (*number of recipes containing X*) / (*total number of recipes*),

P(Y) = (*number of recipes containing Y*) / (*total number of recipes*).

The PMI gives the probability that two ingredients occur together against the probability that they occur separately. That is, the co-occurrence of ingredients provides information about which ingredients go well together or not[3].For each user problem, we recommend the foods that have the high MPMI (>0) value and contain as many ingredients as possible. However, what if the recommended recipe does not include the some ingredients we want? Or, is there any way to find a substitute among supplies at hand when one is lacking a certain ingredient? For that case, we can consider the recipe feature association rule, modify the ingredients and generate the new recipe containing all the ingredients we want. At this point, this recipe feature association rule can provide a surprising result, but cannot guarantee it goes well.

4. Experiments

This section shows our experiments on recipe recommendation. We collected 1,105 recipes from one Korean website (www.naver.com) and parsed all sentences for each recipe, and tagged information based on CRF algorithm. Assume that we have two ingredients lists: (*beef, shrimp, egg, garlic*) and (*pork, tofu, pepper, jujube*). Given the ingredients lists, we can calculate *MPMI* values as shown in table 2.

Table 2:. MPMI examples

MPMI(beef, shrimp, egg)	0.31	MPMI(pork, tofu, pepper)	0.80
MPMI(beef, shrimp, garlic)	-0.11	MPMI(pork, tofu, jujube)	-0.22
MPMI(beef, egg, garlic)	0.27	MPMI(pork, pepper, jujube)	-0.05
MPMI(shrimp, egg, garlic)	0.12	MPMI(tofu, pepper, jujube)	-0.77
MPMI(beef, shrimp, egg, garlic)	0.15	MPMI(pork, tofu, pepper, jujube)	-0.06

Given the first list (*beef, shrimp, egg, garlic*), we searched the recipe containing the all the ingredients because *MPMI*(*beef, shrimp, egg, garlic*) is positive, and found the one dish, i.e. *seasoned beef shrimp garlic chives*. While given the second list (*pork, tofu, pepper, jujube*), we searched the recipe containing only three ingredients, *pork, tofu, pepper*, and found the four dishes as shown in table 3. Now we might add or replace *jujube* with garlic in the text of original recipe, considering recipe feature association rule: "*original*: 애호박,양파,마늘을넣고끓이다(*put pumpkin, onion, garlic and boil*) → *add:* 애호박,양파,마늘,대추를넣고끓이다(*put pumpkin, onion, garlic, jujube and boil*) or *replace*: 애호박,양파,대추를넣고끓이다(*put pumpkin, onion, jujube and boil*)"

Table 3:. Recommeded dishes

No	Ingredients	dishes	Missing Ingredients
1	beef, shrimp, egg, garlic	seasoned beef shrimp garlic chives (쇠고기대하실부추무침)	-
2	pork, tofu, pepper, jujube	Tofu stew(두부찌개), Kimch stew(김치찌개), Kimch pork stew(돼지고기고추장찌개), Mapo tofu(마파두부)	jujube

5. Conclusion

We performed the text analytics on the food recipes, and proposed the recipe recommendation method considering our preference and ingredients at hand. More generally this study provides how text data driven information can provide the insights in food technology.

Acknowledgment

This research was supported by the National Research Foundation of Korea funded by the Korea government (NRF-2017R1D1A1B03033387)

References

[1] J. Lafferty, A. McCallum, and F. Pereira, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, Proceedings of International Conference on Machine Learning, 2001.
[2] F. Shan and F. Pereira, Shallow Parsing with Conditional Random Fields, Proceedings of NAACL, 2013.
[3] C. Y. Teng, Y. R. Lin, and L. A. Adamic, Recipe recommendation using ingredient networks, Proceedings of the 4th ACM Web Science Conference, 2012.