Sukiyaki in French style: A novel system for transformation of dietary patterns

Masahiro Kazama,[1] Minami Sugimoto,[2] Chizuru Hosokawa,[2] Keisuke Matsushima,[3] Lav R. Varshney,[4] and Yoshiki Ishikawa[1]*

[1]Habitech Inc., Tokyo, Japan
[2]Department of Public Health, Cancer Scan Inc., Tokyo, Japan
[3]Keisuke Matsushima, Nice, France
[4]University of Illinois at Urbana-Champaign, Illinois, USA

Abstract

We propose a novel system which can transform a recipe into any selected regional style (e.g., Japanese, Mediterranean, or Italian). This system has three characteristics. First the system can identify the degree of dietary style mixture of any selected recipe. Second, the system can visualize such dietary style mixtures using barycentric Newton diagrams. Third, the system can suggest ingredient substitutions through an extended word2vec model, such that a recipe becomes more authentic for any selected dietary style. Drawing on a large number of recipes from Yummly, an example shows how the proposed system can transform a traditional Japanese recipe, Sukiyaki, into French style.

Keywords: food, big data, dietary style, newton diagram, neural network, word2vec

1 Introduction

Modern dietary styles have been more inclined toward unhealthy eating patterns [1]. As such, there is an emerging consensus that public health efforts should be directed to developing an innovative technology to change dietary style to become more healthy. While there is compelling research in nutrition sciences on what makes a recipe healthy [2], this does not necessarily mean that such a recipe is matched to one's unique food preferences. For example, the Japanese dietary style as measured in the 1970s is reported to have been healthy [3]. However, recipes in such a Japanese style may not be readily acceptable for those that prefer Southern American dietary style.

Furthermore, not limited to such a public health perspective, with growing diversity in personal food preference and dietary style, personalized information systems that can transform a recipe into any selected dietary style that a user might prefer would help food companies and professional chefs create new recipes.

To achieve this goal, there are two significant challenges: 1) identifying the degree of dietary style mixture of any selected recipe; and 2) developing an algorithm that shifts a recipe into any selected dietary style.

As to the former challenge, with growing globalization and economic development, it is becoming difficult to identify a recipe's dietary style with specific traditional styles since dietary patterns have been changing and converging in many countries throughout Asia, Europe, and elsewhere [4]. Regarding the latter challenge, to the best of our knowledge, little attention has been paid to developing algorithms which transform a recipe's dietary style into any selected dietary pattern, cf. [5, 6].

The aim of this study is to propose a novel data-driven system for transformation of dietary style. This system has three characteristics. First, we propose a new method for identifying a recipe's dietary style
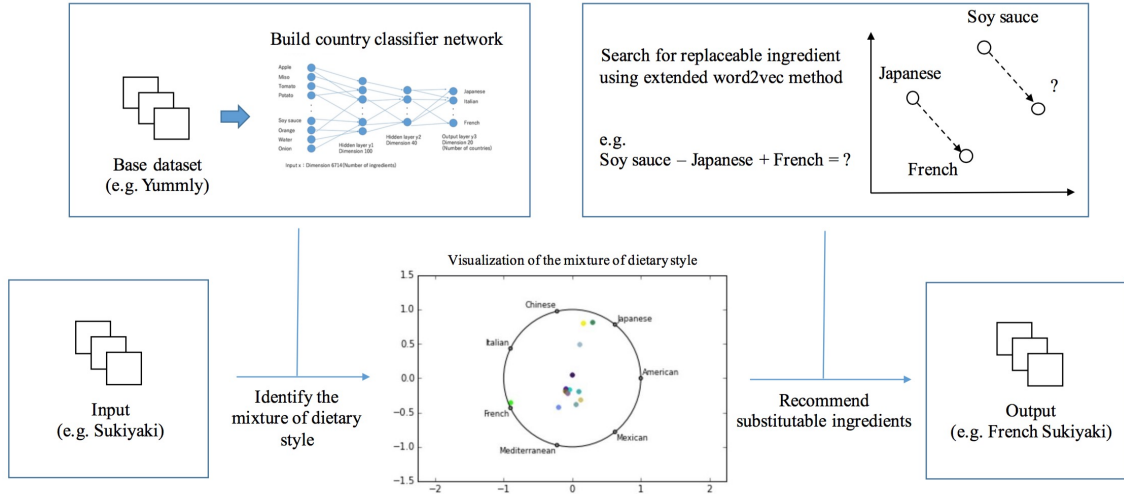
---

*ishikun@gmail.com

Figure 1: Architecture of recommendation system which transform a given recipe into any selected country/region style.

mixture by calculating the contribution of each ingredient to certain dietary patterns, such as Mediterranean, French, or Japanese, by drawing on ingredient prevalence data from large recipe repositories. Second, the system visualizes a recipe's dietary style mixture in two-dimensional space under barycentric coordinates using what we call a *Newton diagram*. Third, the system transforms a recipe's dietary pattern into any selected regional style by recommending replaceable ingredients in existing recipes.

As an example of this proposed system, we transform a traditional Japanese recipe, Sukiyaki, into French style.

2 Architecture of recommendation system

Figure 1 shows the overall architecture of recommendation system, which consists of two steps: 1) identification and visualization of a recipe's dietary style mixture; and 2) algorithm which transforms a given recipe into any selected regional/country style. Details of the steps are described as follows.

2.1 Step 1: Identification and visualization of a recipe's dietary style mixture

Using a neural network method as detailed below, we identify a recipe's dietary style. The neural network model was constructed as shown in Figure 2.

When we enter a recipe, this model classifies which country or regional cuisine the recipe belongs to. The input is a vector with the dimension of the total number of ingredients included in the dataset, and only the indices of ingredients contained in the input recipe are 1, otherwise they are 0.

There are two hidden layers. Therefore, this model can consider a combination of ingredients to predict the country probability. Dropout is also used for the hidden layer, randomly (20%) setting the value of the node to 0. So a robust network is constructed. The final layer's dimension is the number of countries, here 20 countries. In the final layer, we convert it to a probability value using the softmax function, which represents the probability that the recipe belongs to that country.

In this study, we used a labeled corpus of Yummly recipes to train this neural network. Yummly dataset has 39774 recipes from the 20 countries as shown in Table 1. Each recipe has the ingredients and country information. Firstly, we randomly divided the data set into 80% for training the neural network and 20%
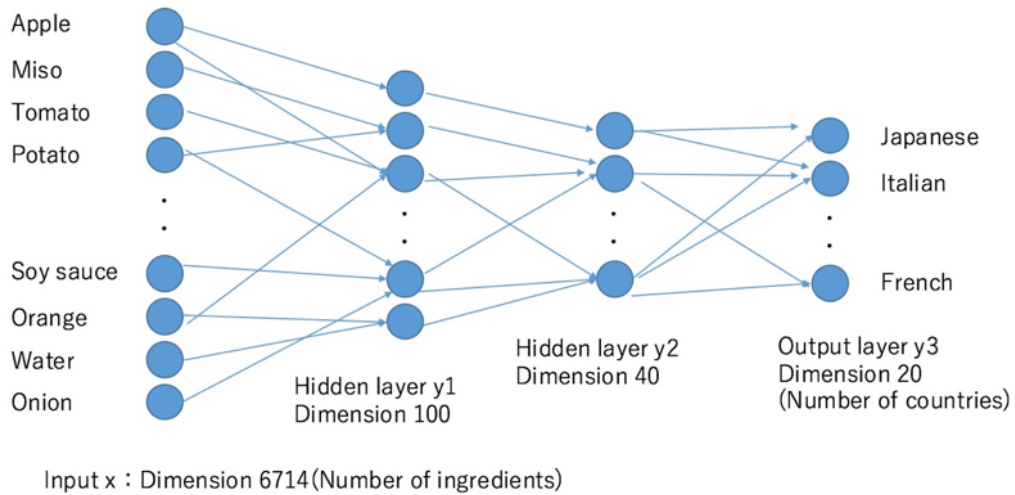
Figure 2: Neural network model for predicting regional cuisine from list of ingredients.

Table 1: Description of Yummly dataset.

Country	Recipes	Ingredients
Italian	7838	2929
Mexican	6438	2684
Southern US	4320	2462
Indian	3003	1664
Chinese	2673	1792
French	2646	2102
Cajun Creole	1546	1576
Thai	1539	1376
Japanese	1423	1439
Greek	1175	1198

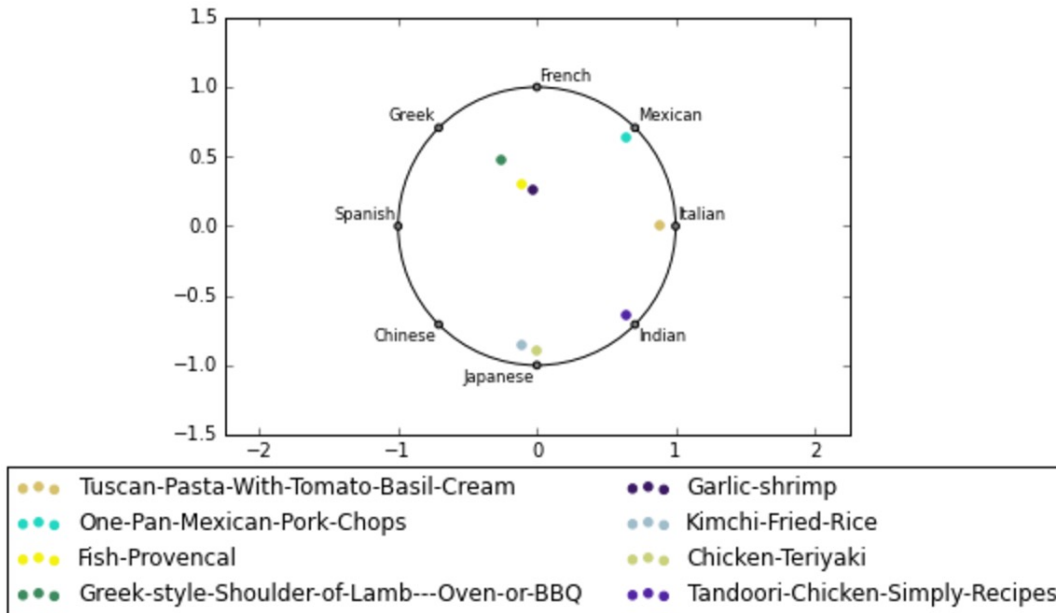| Country	Recipes	Ingredients
Spanish	989	1263
Korean	830	898
Vietnamese	825	1108
Moroccan	821	974
British	804	1166
Filipino	755	947
Irish	667	999
Jamaican	526	877
Russian	489	872
Brazilian	467	853
ALL	39774	6714

Figure 3: Newton diagram: visualization of probability that the recipe belongs to the several regional dietary styles. We pick 8 recipes and 8 countries as an example.

for testing how precisely it can classify. The final neural network achieved a classification accuracy of 79% on the test set.

By using the probability values that emerge from the activation function in the neural network, rather than just the final classification, we can draw a barycentric Newton diagram, as shown in Figure 3. The basic idea of the visualization, drawing on Isaac Newton's visualization of the color spectrum [7], is to express a mixture in terms of its constituents as represented in barycentric coordinates. This visualization allows an intuitive interpretation of which country a recipe belongs to. If the probability of Japanese is high, the recipe is mapped near the Japanese.

2.2   Step 2: Recommendation algorithm for transforming dietary style

If you want to change a given recipe into a recipe having high probability of a specific country by just changing one ingredient, which ingredient should be alternatively used?

When we change the one ingredient x_i in the recipe to ingredient x_j, the probability value of country likelihood can be calculated by using the above neural network model. If we want to change the recipe to have high probability of a specific country c, we can find ingredient x_j that maximizes the following probability.

$$P(C = c | r - x_i + x_j) \tag{1}$$

where r is the recipe. However, with this method, regardless of the ingredient x_i, only specific ingredients having a high probability of country c are always selected. In this system, we want to select ingredients that are similar to ingredient x_i and have a high probability of country c. Therefore, we propose a method of extending word2vec as a method of finding ingredients resembling ingredient x_i.

Word2vec is a technique proposed in the field of natural language processing [8]. As the name implies, it is a method to vectorize words, and similar words are represented by similar vectors. To train word2vec, sentences with their implicit structure among words are used as data. One sentence is made up of a set of

words, and learned on the assumption that the words appearing in the vicinity are similar. After vectorization, word2vec can calculate analogies. For example, the analogy of "King - Man + Women = ?" can be "Queen" by using word2vec.

In this study, word2vec is applied to the data set of recipes. Word2vec can be applied by considering recipes as sentences and ingredients as words. We do not include a window size parameter, since it is used to encode the ordering of words in sentences where it is important. In recipes, the listing of ingredients is unordered.

Each ingredient is vectorized by word2vec, and the similarity of each ingredient is calculated using cosine similarity. Through vectorization in word2vec, those of the same genre are placed nearby. In other words, by using the word2vec vector, it is possible to select ingredients with similar genres. Figure 4 shows ingredient maps by neural network and word2vec.

Next, we extend word2vec to be able to incorporate information of the country. When we can vectorize the countries, we can calculate the analogy between countries and ingredients. For example, this method can tell us what is the French ingredient that corresponds to Japanese soy sauce by calculating "Soy sauce - Japan + French = ?".

The detail of our method is as follows. We will maximize objective function (2).

$$\sum_{r \in R} \sum_{w_i \in r} \left(\log P(w_i|c_r) + \log P(c_r|w_i) + \sum_{j \neq i} \log P(w_j|w_i) \right) \tag{2}$$

where R is a set of recipes, r is a recipe, w_i is i th ingredient in the recipe r, and c_r is a country the recipe r belongs to. The probability is defined below.

$$P(b|a) = \frac{\exp(v_a^T v_b^{'})}{\sum_{c \in W} \exp(v_a^T v_c^{'})} \tag{3}$$

where a is a ingredient or country, b, c are also, $v_a \in \mathbb{R}^K$ is an input vector of ingredient or country, $v_a^{'} \in \mathbb{R}^K$ is an output vector of ingredient or country, K is a dimension of vector, and W is a set of all ingredients and all countries.

We can use hierarchical softmax or negative sampling method [8] to maximize objective function (2) and find the vectors of ingredients and countries.

Table 2 shows the ingredients around each country in the vector space, and which could be considered as most authentic for that regional cuisine [9]. Also, Figure 4 shows the ingredients and countries in 2D map by using t-SNE method [10].

Table 2: Authentic ingredients for each country. Top 5 ingredients around each country in the vector space.

	French	Japanese	Italian	Mexican
Top1	Cognac	Mirin	Grated parmesan cheese	Corn tortillas
Top2	Calvados	Dashi	pecorino romano cheese	Salsa
Top3	Thyme springs	Nori	prosciutto	Tortilla chips
Top4	Gruyere cheese	Wasabi paste	marinara sauce	Guacamole
Top5	Nicoise olives	Bonito flakes	Sweet italian sausage	Poblano peppers

3   Experiment

As an example of our proposed system, we transformed a traditional Japanese Sukiyaki into French style. Table 3 shows the suggested replaceable ingredients and the probability after replacing. "Sukiyaki" consists of soy sauce, beef sirloin, white sugar, green onions, mirin, shiitake, egg, vegetable oil, konnyaku, and chinese
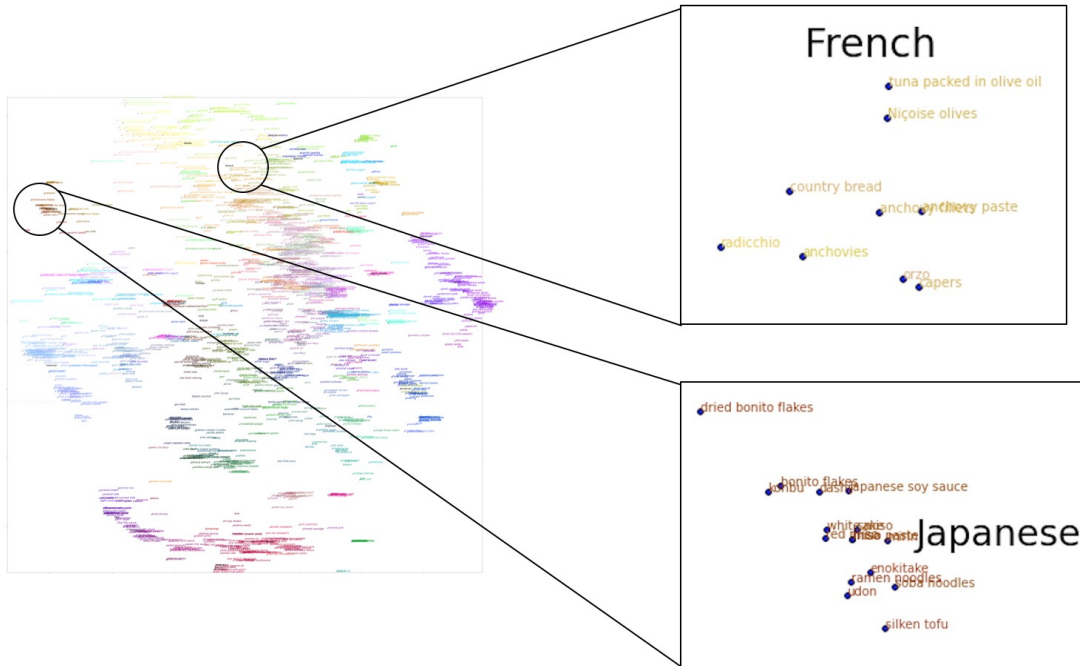
Figure 4: Ingredients and countries map by extended word2vec: Each ingredient and country is mapped in 2D by using t-SNE. Also Each ingredient is colored by using t-SNE to convert 100 dimension vector into 3 dimension. The 3 dimension is corresponded to RGB color. Countries are represented by bold black.

cabbage. Figure 5 shows the Sukiyaki in French style cooked by professional chef KM who is one of the coauthors of this paper.

Figure 5: Sukiyaki in French style. Professional chef KM who is one of the coauthors of this paper cooked the recipe suggested by our system.

4 Discussion

Unhealthy dietary style is one of the most important public health issues to be tackled. Given that past research has identified what makes a recipe healthy, future efforts should pay attention to the development of data-driven systems which can transform such healthy recipes into any given dietary style. Then a scientifically proven healthy dietary style could be acceptable in accordance with a user's unique food preferences. Also, with growing diversity in personal food preference and dietary style, such an algorithm might be of

Table 3: Alternative ingredients suggested by extended word2vec model and country probability of changing food ingredients in order from the top. Professional chef KM who is one of the coauthors of this paper chose one alternative ingredient from top 10 suggested ingredients each.

Original Ingredient	Alternative Ingredient	P(Japanese)	P(French)	# of replacement
-	-	0.974	0.000	0
Mirin	Calvados	0.552	0.009	1
Vegetable oil	Olive oil	0.393	0.031	2
Soy sauce	Bouquet garni	0.011	0.976	3
Green onions	Fresh tarragon	0.000	0.997	4
Egg	Melted butter	0.000	0.999	5

value for food companies or professional chefs to create new recipes.

In this regard, this study adds two important contributions to the literature. First, this is to the best of our knowledge, the first study to identify a recipe's mixture of dietary country/regional styles from the large number of recipes around the world. Previous studies have focused on assessing degree of adherence to a single dietary pattern. For example, Mediterranean Diet Score is one of the most popular diet scores. This method uses 11 main items (e.g., fruit, vegetable, olive oil, and wine) as criteria for assessing the degree of one's Mediterranean style [11]. However, in this era, it is becoming difficult to identify a recipe's dietary style with specific country/regional style. For example, should Fish Provencal, whose recipe name is suggestive of Southern France, be cast as French style? The answer is mixture of different country styles: 32% French; 26% Italian; and 38% Spanish (see Figure 3).

Furthermore, our identification algorithm can be used to assess the degree of personal dietary style mixture, using the user's daily eating pattern as inputs. For example, when one enters the recipes that one has eaten in the past week into the algorithm, the probability values of each country would be returned, which shows the mixture of dietary styles of one's daily eating pattern. As such, a future research direction would be developing algorithms that can transform personal dietary patterns to a healthier style by providing a series of recipes that are in accordance with one's unique food preferences.

Second, this study proposes an algorithm that can transform a given recipe into any selected regional pattern. Previous studies have focused on developing recommendation algorithm which suggests replaceable ingredients Based on cooking action [12], degree of similarity among ingredient [13], ingredient network [14], degree of typicality of ingredient [15], and flavor (foodpairing.com). Our recommendation algorithm can be improved by adding multiple datasets from around the world. Needless to say, lack of a comprehensive data sets makes it difficult to develop recommendation algorithms for transforming dietary style. For example, Yummly, one of the largest recipe sites in the world, is less likely to contain recipes from non-Western regions. Furthermore, data on traditional dietary patterns is usually described in its native language. As such, developing a way to integrate multiple data sets in multiple languages is required for future research.

One of the methods to address this issue might be as follows: 1) generating the vector representation for each ingredient by using each data set independently; 2) translating only a small set of common ingredients among each data set, such as potato, tomato, and onion; 3) with a use of common ingredients, mapping each vector representation into one common vector space using a canonical correlation analysis [16], for example.

Several fundamental limitations of the present study warrant mention. First of all, our identification and transformation algorithms depend on the quantity and quality of recipes included in the data. As such, future research using our proposed system should employ quality big recipe data. Second, the evolution of regional cuisines prevents us from developing precise algorithm. For example, the definition of Mediterranean dietary pattern has been revised to adapt to current dietary patterns [17, 18]. Therefore, future research should employ time-trend recipe data to distinctively specify a recipe's mixture of dietary style and its date cf. [19]. Third, we did not consider the cooking method (e.g., baking, boiling, and deep flying) as a characteristic of country/regional style. Each country/region has different ways of cooking ingredients and this is one of the important factors characterizing the food culture of each country/region. Forth, the combination of

ingredients was not considered as the way to represent country/regional style. For example, previous studies have shown that Western recipes and East Asian recipes are opposite in flavor compounds included in the ingredient pair [20, 19, 21, 22, 9]. For example, Western cuisines tend to use ingredient pairs sharing many flavor compounds, while East Asian cuisines tend to avoid compound sharing ingredients. It is suggested that combination of flavor compounds was also elemental factor to characterize the food in each country/region. As such, if we analyze the recipes data using flavor compounds, we might get different results.

In conclusion, we proposed a novel system which can transform a given recipe into any selected dietary style. This system has three characteristics: 1) the system can identify a degree of dietary style mixture of any selected recipe; 2) the system can visualize such dietary style mixture using a barycentric Newton diagram; and the system can suggest ingredient substitution through extended word2vec model, such that a recipe becomes more authentic for any selected dietary style. Future research directions were also discussed.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Author Contributions

MK, LRV, and YI had the idea for the study and drafted the manuscript. MK performed the data collection and analysis. MS, CH, and KM participated in the interpretation of the results and discussions for manuscript writing and finalization. All authors read and approved the final manuscript.

Funding

Acknowledgments

References

[1] F. Imamura, R. Micha, S. Khatibzadeh, S. Fahimi, P. Shi, J. Powles, and D. Mozaffarian, "Dietary quality among men and women in 187 countries in 1990 and 2010: a systematic assessment," *The Lancet Global Health*, vol. 3, no. 3, pp. e132–e142, Mar. 2015.

[2] S. Khatibzadeh, R. Micha, A. Afshin, M. Rao, M. Y. Yakoob, and D. Mozaffarian, "Major dietary risk factors for chronic diseases: A systematic review of the current evidence for causal effects and effect sizes," *Circulation*, vol. 125, no. Su. 10, p. AP060, Mar. 2012.

[3] T. Tsuduki, "Influence of Japanese food on senility and health maintenance," *Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan*, vol. 135, no. 1, pp. 57–65, 2014.

[4] C. K. Khoury, A. D. Bjorkman, H. Dempewolf, J. Ramirez-Villegas, L. Guarino, A. Jarvis, L. H. Rieseberg, and P. C. Struik, "Increasing homogeneity in global food supplies and the implications for food security," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 11, pp. 4001–4006, Mar. 2014.

[5] F. Pinel, N. Shao, and L. R. Varshney, "Using physicochemical correlates of perceptual flavor similarity to enhance, balance and substitute flavors," Aug. 13 2014, US Patent App. 14/458,315.

[6] F. Pinel and L. R. Varshney, "Substitution of work products," May 5 2014, US Patent App. 14/269,353.

[7] I. Newton, *Opticks: Or, a Treatise of the Reflexions, Refractions, Inflexions and Colours of Light.* London: Sam. Smith and Benj. Walford, 1704.

[8] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Cambridge, MA: MIT Press, 2013, pp. 3111–3119.

[9] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási, "Flavor network and the principles of food pairing," *Scientific Rerports*, vol. 1, p. 196, Dec. 2011.

[10] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, Nov. 2008.

[11] D. B. Panagiotakos, C. Pitsavos, and C. Stefanadis, "Dietary patterns: a Mediterranean diet score and its relation to clinical and biological markers of cardiovascular disease risk," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 16, no. 8, pp. 559–568, Dec. 2006.

[12] Y. Shidochi, T. Takahashi, I. Ide, and H. Murase, "Finding replaceable materials in cooking recipe texts considering characteristic cooking actions," in *Proceedings of the ACM Multimedia 2009 Workshop on Multimedia for Cooking and Eating Activities*, Oct. 2009, pp. 9–14.

[13] K. Nozawa, Y. Nakaoka, S. Yamamoto, and T. Satoh, "Finding method of replaceable ingredients using large amounts of cooking recipes," *IEICE Technical Report*, vol. 114, no. 204, pp. 41–46, Sep. 2014.

[14] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic, "Recipe recommendation using ingredient networks," in *Proceedings of the 3rd Annual ACM Web Science Conference (WebSci'12)*, Jun. 2012, pp. 298–307.

[15] S. Yokoi, K. Doman, T. Hirayama, I. Ide, D. Deguchi, and H. Murase, "Typicality analysis of the combination of ingredients in a cooking recipe for assisting the arrangement of ingredients," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Jun. 2015.

[16] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, Dec. 1971.

[17] L. Serra-Majem, A. Trichopoulou, J. N. de la Cruz, P. Cervera, A. G. Álvarez, C. La Vecchia, A. Lemtouni, and D. Trichopoulos, "Does the definition of the Mediterranean diet need to be updated?" *Public Health Nutrition*, vol. 7, no. 7, pp. 927–929, Oct. 2004.

[18] O. Kinouchi, R. W. Diez-Garcia, A. J. Holanda, P. Zambianchi, and A. C. Roque, "The non-equilibrium nature of culinary evolution," *New Journal of Physics*, vol. 10, no. 7, p. 073020, 2008.

[19] K. R. Varshney, L. R. Varshney, J. Wang, and D. Myers, "Flavor pairing in Medieval European cuisine: A study in cooking with dirty data," *Int. Joint Conf. Artif. Intell. Workshops*, pp. 3–12, Aug. 2013.

[20] Y.-X. Zhu, J. Huang, Z.-K. Zhang, Q.-M. Zhang, T. Zhou, and Y.-Y. Ahn, "Geography and similarity of regional cuisines in China," *PloS ONE*, vol. 8, no. 11, p. e79161, Nov. 2013.

[21] A. Jain, Rakhi N. K., and G. Bagler, "Analysis of food pairing in regional cuisines of India," *PLoS ONE*, vol. 10, no. 10, p. e0139539.

[22] S. T. Tallab and M. S. Alrazgan, "Exploring the food pairing hypothesis in Arab cuisine: A study in computational gastronomy," *Procedia Computer Science*, vol. 82, pp. 135–137, 2016.