



# A Neural Network System for Transformation of Regional Cuisine Style

Masahiro Kazama<sup>1</sup>, Minami Sugimoto<sup>2</sup>, Chizuru Hosokawa<sup>2</sup>, Keisuke Matsushima<sup>3</sup>,  
Lav R. Varshney<sup>4</sup> and Yoshiki Ishikawa<sup>1\*</sup>

<sup>1</sup> Habitech Inc., Tokyo, Japan, <sup>2</sup> Department of Public Health, Cancer Scan Inc., Tokyo, Japan, <sup>3</sup> Keisuke Matsushima, Nice, France, <sup>4</sup> Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States

We propose a novel system which can transform a recipe into any selected regional style (e.g., Japanese, Mediterranean, or Italian). This system has two characteristics. First the system can identify the degree of regional cuisine style mixture of any selected recipe and visualize such regional cuisine style mixtures using barycentric Newton diagrams. Second, the system can suggest ingredient substitutions through an extended word2vec model, such that a recipe becomes more authentic for any selected regional cuisine 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, regional cuisine style, newton diagram, neural network, word2vec

## OPEN ACCESS

### Edited by:

Tom Crick,  
Swansea University, United Kingdom

### Reviewed by:

Gokarna Sharma,  
Kent State University, United States  
Hyejin Youn,  
Northwestern University, United States  
Jonathan Gillard,  
Cardiff University, United Kingdom

### \*Correspondence:

Yoshiki Ishikawa  
ishikun@gmail.com

### Specialty section:

This article was submitted to  
Big Data Networks,  
a section of the journal  
Frontiers in ICT

**Received:** 31 January 2017

**Accepted:** 12 June 2018

**Published:** 19 July 2018

### Citation:

Kazama M, Sugimoto M,  
Hosokawa C, Matsushima K,  
Varshney LR and Ishikawa Y (2018) A  
Neural Network System for  
Transformation of Regional Cuisine  
Style. *Front. ICT* 5:14.  
doi: 10.3389/fict.2018.00014

## 1. INTRODUCTION

With growing diversity in personal food preference and regional cuisine style, personalized information systems that can transform a recipe into any selected regional cuisine 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 regional cuisine style mixture of any selected recipe; and 2) developing an algorithm that shifts a recipe into any selected regional cuisine style.

As to the former challenge, with growing globalization and economic development, it is becoming difficult to identify a recipe's regional cuisine style with specific traditional styles since regional cuisine patterns have been changing and converging in many countries throughout Asia, Europe, and elsewhere (Khouri et al., 2014). Regarding the latter challenge, to the best of our knowledge, little attention has been paid to developing algorithms which transform a recipe's regional cuisine style into any selected regional cuisine pattern, cf. (Pinel and Varshney, 2014; Pinel et al., 2014). Previous studies have focused on developing an algorithm which suggests replaceable ingredients based on cooking action (Shidochi et al., 2009), degree of similarity among ingredient (Nozawa et al., 2014), ingredient network (Teng et al., 2012), degree of typicality of ingredient (Yokoi et al., 2015), and flavor (foodpairing.com).

The aim of this study is to propose a novel data-driven system for transformation of regional cuisine style. This system has two characteristics. First, we propose a new method for identifying a recipe's regional cuisine style mixture by calculating the contribution of each ingredient to certain regional cuisine patterns, such as Mediterranean, French, or Japanese, by drawing on ingredient prevalence data from large recipe repositories. Also the system visualizes a recipe's regional cuisine style mixture in two-dimensional space under barycentric coordinates using what we call a

*Newton diagram*. Second, the system transforms a recipe's regional cuisine 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 TRANSFORMATION SYSTEM

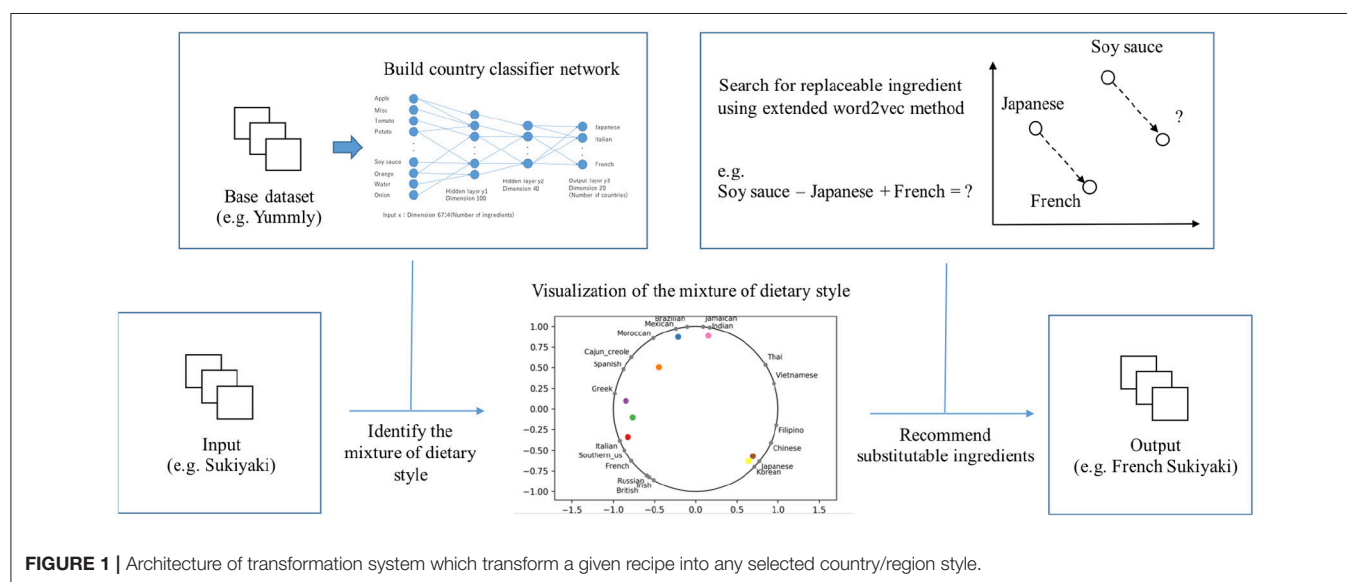
**Figure 1** shows the overall architecture of the transformation system, which consists of two steps: 1) identification and visualization of a recipe's regional cuisine style mixture; and 2) algorithm which transforms a given recipe into any selected regional/country style. Details of the steps are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

### 2.1. Step 1: Identification and Visualization of a Recipe's Regional Cuisine Style Mixture

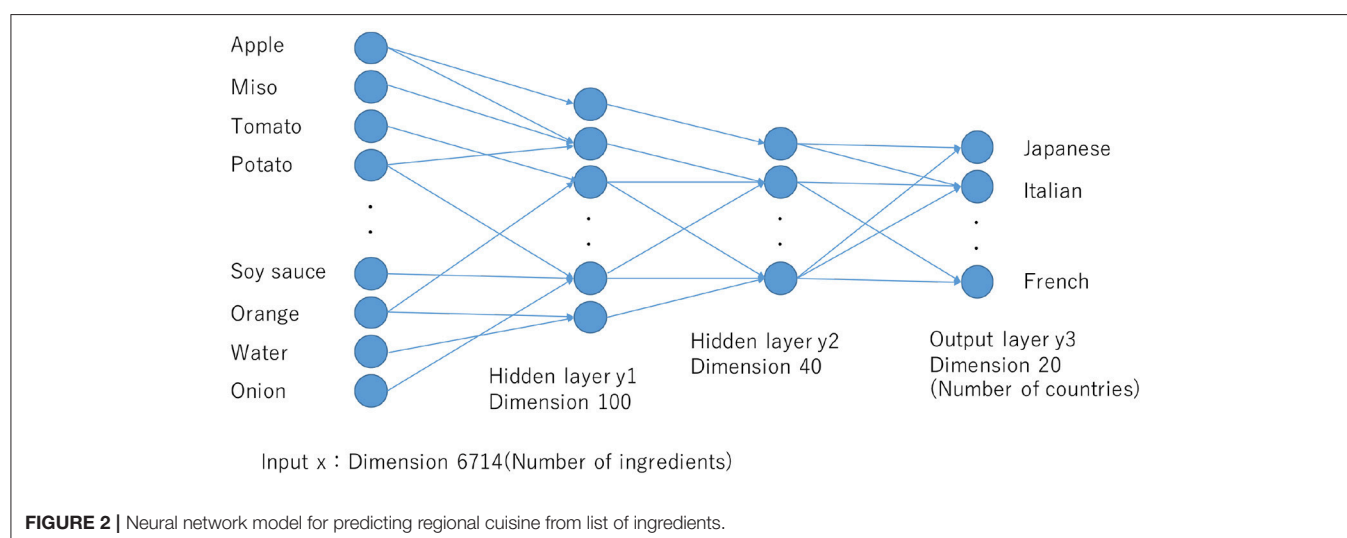
Using a neural network method as detailed below, we identify a recipe's regional cuisine style. The neural network model was constructed as shown in **Figure 2**. The number of layers and dimension of each layer are also shown in **Figure 2**.

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



**FIGURE 1** | Architecture of transformation system which transform a given recipe into any selected country/region style.



**FIGURE 2** | Neural network model for predicting regional cuisine from list of ingredients.

(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. ADAM (Kingma and Ba, 2014) was used as an optimization technique. The number of epochs in training was 200. These network structure and parameters were chosen after preliminary experiments so that the neural network could perform the country classification task as efficiently as possible.

In this study, we used a labeled corpus of Yummly recipes to train this neural network. Yummly dataset has 39,774 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% for testing how precisely it can classify. The final neural network achieved a classification accuracy of 79% on the test set. **Figure 3** shows the confusion matrix of the neural network classification. **Table 2** shows the examples of ingredient classification results. Common ingredients, onions for example, that appear in many regional recipes are assigned to all countries with low probability. On the other hands some ingredients that appear only in specific country are assigned to the country with high probability. For example mirin that is a seasoning commonly used in Japan is classified into Japan with high probability.

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 4**. The basic idea of the visualization, drawing on Isaac Newton's visualization of the color spectrum (Newton, 1704), 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. The countries on the Newton diagram are placed by spectral graph drawing (Koren, 2003), so that similar countries are placed nearby on the circle. The calculation is as follows. First we define the adjacency matrix  $W$  as the similarity between two countries. The similarity between country  $i$  and  $j$  is calculated by cosine similarity of country  $i$  vector and  $j$  vector. These vector are defined in next section.  $W_{ij} = \text{sim}(\text{vec}_i, \text{vec}_j)$ . The degree matrix  $D$  is a diagonal matrix where  $D_{ii} = \sum_j W_{ij}$ . Next we calculate the eigendecomposition of  $D^{-1}W$ . The second and third smallest eigenvalues and corresponded eigenvectors are used for placing the countries. Eigenvectors are normalized so as to place the countries on the circle.

## 2.2. Step 2: Transformation Algorithm for Transforming Regional Cuisine 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

**TABLE 1 |** Statistics of Yummly dataset and some recipe examples.

Country	Recipes	Ingredients
Italian	7,838	2,929
Mexican	6,438	2,684
Southern US	4,320	2,462
Indian	3,003	1,664
Chinese	2,673	1,792
French	2,646	2,102
Cajun Creole	1,546	1,576
Thai	1,539	1,376
Japanese	1,423	1,439
Greek	1,175	1,198
Spanish	989	1,263
Korean	830	898
Vietnamese	825	1,108
Moroccan	821	974
British	804	1,166
Filipino	755	947
Irish	667	999
Jamaican	526	877
Russian	489	872
Brazilian	467	853
ALL	39,774	6,714

RecipeID	Country	Ingredients
34466	British	greek yogurt, lemon curd, confectioners sugar, raspberries
44500	Indian	chili, mayonaise, chopped onion, cider vinegar, fresh mint, cilantro leaves
38233	Thai	sugar, chicken thighs, cooking oil, fish sauce, garlic, black pepper

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.  $P(C = c | r - x_i + x_j)$  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 (Mikolov et al., 2013). As the name implies, it is a method to vectorize words, and similar words are represented by similar vectors. To train word2vec, skip-gram model is used. In the skip-gram model, the objective is to learn word vector representations that can predict the nearby words. The objective function is

$$\sum_{d \in D} \sum_{w_i \in d} \sum_{-n \leq j \leq n, j \neq 0} \log P(w_{i+j} | w_i) \quad (1)$$

where  $D$  is the set of documents,  $d$  is a document,  $w_i$  is a word, and  $n$  is the window size. This model predicts the  $n$  words before

		Predict																			
		Irish	Mexican	Chinese	Filipino	Vietnamese	Moroccan	Brazilian	Japanese	British	Greek	Indian	Jamaican	French	Spanish	Russian	Cajun Creole	Thai	Southern US	Korean	Italian
		Irish	Mexican	Chinese	Filipino	Vietnamese	Moroccan	Brazilian	Japanese	British	Greek	Indian	Jamaican	French	Spanish	Russian	Cajun Creole	Thai	Southern US	Korean	Italian
True	Irish	0.53	0.02	0.00	0.01	0.00	0.01	0.00	0.00	0.09	0.01	0.01	0.01	0.14	0.00	0.01	0.00	0.00	0.11	0.00	0.07
	Mexican	0.00	0.93	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.02	0.00	0.02
	Chinese	0.00	0.01	0.83	0.02	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.01	0.02	0.01
	Filipino	0.01	0.01	0.11	0.67	0.04	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.04	0.03	0.00	0.04
	Vietnamese	0.00	0.01	0.08	0.03	0.61	0.00	0.01	0.04	0.01	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.17	0.00	0.00	0.01
	Moroccan	0.00	0.02	0.00	0.00	0.00	0.81	0.00	0.01	0.00	0.02	0.05	0.00	0.02	0.02	0.01	0.00	0.00	0.01	0.00	0.04
	Brazilian	0.01	0.12	0.00	0.01	0.01	0.00	0.64	0.00	0.02	0.00	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.08	0.00	0.02
	Japanese	0.00	0.00	0.10	0.01	0.02	0.00	0.00	0.72	0.00	0.00	0.08	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.03	0.01
	British	0.06	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.47	0.00	0.01	0.00	0.13	0.01	0.03	0.01	0.00	0.19	0.00	0.06
	Greek	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.69	0.02	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.21
	Indian	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.89	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00
	Jamaican	0.02	0.08	0.00	0.03	0.00	0.00	0.02	0.01	0.00	0.00	0.01	0.68	0.01	0.00	0.01	0.01	0.02	0.05	0.00	0.03
	French	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.68	0.03	0.01	0.01	0.00	0.05	0.00	0.15
	Spanish	0.00	0.09	0.00	0.01	0.00	0.02	0.01	0.00	0.01	0.01	0.00	0.01	0.12	0.48	0.02	0.01	0.01	0.04	0.00	0.19
	Russian	0.04	0.02	0.00	0.01	0.00	0.02	0.01	0.01	0.07	0.02	0.03	0.00	0.10	0.02	0.49	0.00	0.00	0.06	0.00	0.08
	Cajun Creole	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.03	0.01	0.01	0.70	0.00	0.16	0.00	0.05
	Thai	0.00	0.01	0.07	0.02	0.10	0.00	0.00	0.02	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.74	0.01	0.00	0.00
	Southern US	0.01	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.04	0.01	0.01	0.04	0.00	0.76	0.00	0.06
	Korean	0.00	0.01	0.10	0.01	0.01	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.01	0.78	0.01
	Italian	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.05	0.01	0.00	0.00	0.00	0.01	0.00	0.89

FIGURE 3 | Confusion matrix of neural network classification.

TABLE 2 | Example of ingredient classification by the neural network.

Ingredient	Top1	Top2	Top3
Onions	French	Italian	Mexican
	0.130	0.126	0.126
Soy sauce	Japanese	Chinese	Filipino
	0.246	0.233	0.122
Mirin	Japanese	French	Korean
	0.890	0.040	0.020

Three top countries are listed with the probability that the ingredient are classified into.

and after the input word, as described in left side of **Figure 5**. The objective function is to maximize the likelihood of the prediction of the surrounding word  $w_{i+j}$  given the center word  $w_i$ . The probability is

$$P(w_j|w_i) = \frac{\exp(v_{w_i}^T v'_{w_j})}{\sum_{w \in W} \exp(v_{w_i}^T v'_w)} \quad (2)$$

where  $v_w \in \mathbb{R}^K$  is an input vector of word  $w$ ,  $v'_w \in \mathbb{R}^K$  is an output vector of word  $w$ ,  $K$  is the dimension of the vector, and  $W$  is the set of all words. To optimize this objective function, hierarchical softmax or negative sampling method (Mikolov et al., 2013) are used. After that we get the vectors of words and

we can calculate analogies by using the vectors. For example, the analogy of “King – Man + Women = ?” yields “Queen” by using word2vec.

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

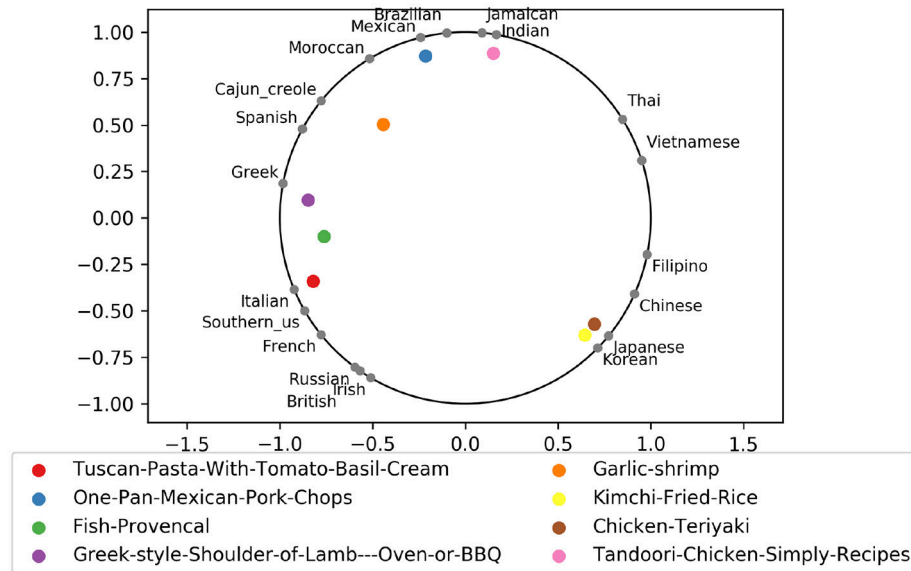
$$\sum_{r \in R} \sum_{w_i \in r} \sum_{j \neq i} \log P(w_j|w_i) \quad (3)$$

where  $R$  is a set of recipes,  $r$  is a recipe, and  $w_i$  is the  $i$ th ingredient in recipe  $r$ . The architecture is described in middle of **Figure 5**. The objective function is to maximize the likelihood of the prediction of the ingredient  $w_j$  in the same recipe given the ingredient  $w_i$ . The probability is defined below.

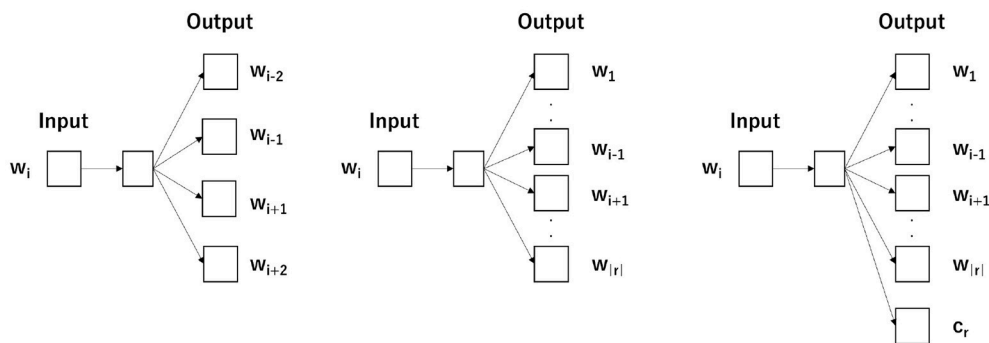
$$P(w_j|w_i) = \frac{\exp(v_{w_i}^T v'_{w_j})}{\sum_{w \in W} \exp(v_{w_i}^T v'_w)} \quad (4)$$

where  $w$  is an ingredient,  $v_w \in \mathbb{R}^K$  is an input vector of ingredient,  $v'_w \in \mathbb{R}^K$  is an output vector of ingredient,  $K$  is the dimension of the vector, and  $W$  is the set of all ingredients.





**FIGURE 4 |** Newton diagram: visualization of probability that the recipe belongs to the several regional cuisine style. Countries are placed by spectral drawing.



**FIGURE 5 |** The word2vec (skip-gram) architecture. The left panel is the traditional word2vec with window size  $n = 2$ . The middle panel is the word2vec for recipe data. The right panel is the word2vec for recipe data with country information.

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.

Next, we extend word2vec to be able to incorporate information of the country. When we 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 maximize objective function (5).

$$\sum_{r \in R} \sum_{w_i \in r} \left( \log P(w_i | c_r) + \log P(c_r | w_i) + \sum_{j \neq i} \log P(w_j | w_i) \right) \quad (5)$$

where  $R$  is a set of recipes,  $r$  is a recipe,  $w_i$  is the  $i$ th ingredient in recipe  $r$ , and  $c_r$  is the country recipe  $r$  belongs to. The architecture is described in right of **Figure 5**. The objective function is to maximize the likelihood of the prediction of the ingredient  $w_j$  in the same recipe given the ingredient  $w_i$  along with the prediction of the ingredients  $w_i$  given the country  $c_r$  and the prediction of the country  $c_r$  given the ingredient  $w_i$ . The probability is defined below.

$$P(b|a) = \frac{\exp(v_a^T v'_b)}{\sum_{c \in W} \exp(v_a^T v'_c)} \quad (6)$$

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 the dimension of vector, and  $W$  is the set of all ingredients and all countries.

We can use hierarchical softmax or negative sampling (Mikolov et al., 2013) to maximize objective function (5) and find the vectors of ingredients and countries in the same vector space.

**Table 3** shows the ingredients around each country in the vector space, and which could be considered as most authentic for that regional cuisine (Ahn et al., 2011). Also, **Figure 6** shows the ingredients and countries in 2D map by using t-SNE method (van der Maaten and Hinton, 2008).

### 3. EXPERIMENT

Our substitution strategy is as follows. First we use extended word2vec and train it by Yummy dataset. After that all

ingredients and countries are vectorized into 100 dimensional vector space. Second we find substitution by analogy calculation. For example, to find french substitution of Mirin, we calculate “Mirin – Japanese + French” in the vector space and get the vector as result. After that we find similar ingredients around the vector by calculating the cosine similarity.

As an example of our proposed system, we transformed a traditional Japanese “Sukiyaki” into French style. **Table 4** 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 cabbage. **Figure 7** shows the Sukiyaki in French style

**TABLE 3 |** Authentic ingredients for each country.

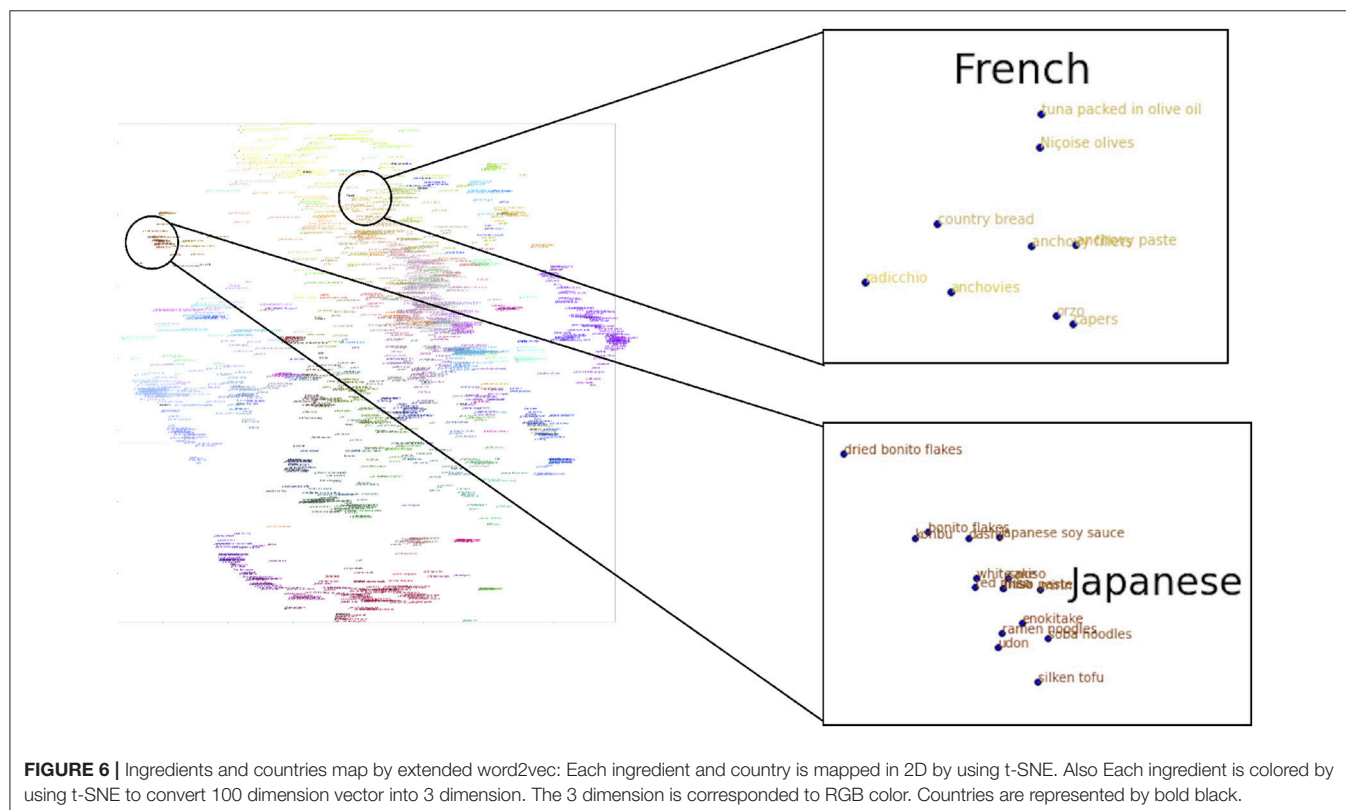
	French	Japanese	Italian	Mexican
Top1	Cognac	Mirin	Grated parmesan cheese	Corn tortillas
Top2	Calvados	Dashi	pecorino romano cheese	Salsa
Top3	Thyme springs	Nori	prosciutto	Tortilla chips
Top4	Gruyere cheese	Wasabi paste	marinara sauce	Guacamole
Top5	Nicoise olives	Bonito flakes	Sweet italian sausage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

Top 5 ingredients around each country in the vector space.

**TABLE 4 |** Alternative ingredients suggested by extended word2vec model and country probability of changing food ingredients in order from the top.

Original Ingredient	Alternative Ingredient	P(Japanese)	P(French)	No. 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

Professional chef KM who is one of the authors of this paper chose one alternative ingredient from top 10 suggested ingredients each.





**FIGURE 7 |** Suki-yaki in French style. Professional chef KM who is one of the authors of this paper cooked the recipe suggested by our system.

cooked by professional chef KM who is one of the authors of this paper. He assesses the new recipe as valid and novel to him in terms of Suki-yaki in French. Here our task is in generating a new dish, for which by definition there is no ground truth for comparison. Rating by experts is the standard approach for assessing novel generative artifacts, e.g., in studies of creativity (Jordanous, 2012), but going forward it is important to develop other approaches for assessment.

## 4. DISCUSSION

With growing diversity in personal food preference and regional cuisine style, the development of data-driven systems which can transform recipes into any given regional cuisine style might be of 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 regional cuisine style from the large number of recipes around the world. Previous studies have focused on assessing degree of adherence to a single regional cuisine 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 (Panagiotakos et al., 2006). However, in this era, it is becoming difficult to identify a recipe's regional cuisine style with specific country/regional style. For example, should Fish Provençal, 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 4**).

Furthermore, our identification algorithm can be used to assess the degree of personal regional cuisine 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 regional cuisine style of one's daily eating pattern. As such, a future research direction would be developing algorithms that can transform personal regional cuisine patterns to a healthier style by providing a series of recipes that are in accordance with one's unique food preferences.

Our transformation 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 algorithms for transforming regional cuisine 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 regional cuisine 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 (Kettenring, 1971), 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 regional cuisine pattern has been revised to adapt to current dietary patterns (Serra-Majem et al., 2004; Kinouchi et al., 2008). Therefore, future research should employ time-trend recipe data to distinctively specify a recipe's mixture of regional cuisine style and its date cf. Varshney et al. (2013). Third, we did not consider the cooking method (e.g., baking, boiling, and deep frying) 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. Fourth, 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 (Ahn et al., 2011; Varshney et al., 2013; Zhu et al., 2013; Jain et al., 2015; Tallab and Alrazgan, 2016). 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 regional cuisine style. This system has two characteristics: 1) the system can identify a degree of regional cuisine style mixture of any selected recipe and visualize such regional cuisine style mixture using a barycentric Newton diagram; 2) the system can suggest ingredient substitution through extended word2vec model, such that a recipe becomes more authentic for any selected regional cuisine style. Future research directions were also discussed.

## AUTHOR CONTRIBUTIONS

MK, LV, 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

Varshney's work was supported in part by the IBM-Illinois Center for Cognitive Computing Systems Research (C3SR), a research collaboration as part of the IBM AI Horizons Network.

## ACKNOWLEDGMENTS

This study used data from Yummly. We would like to express our deepest gratitude to everyone who participated in this services. We thank Kush Varshney for suggesting the spectral graph drawing approach to placing countries on the circle.

## REFERENCES

- Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P., and Barabási, A.-L. (2011). Flavor network and the principles of food pairing. *Sci. Reports* 1:196. doi: 10.1038/srep00196
- Jain, A., Rakhi N. K., and Bagler, G. (2015). Analysis of food pairing in regional cuisines of India. *PLoS ONE* 10:e0139539.
- Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cogn Comput* 4, 246–279. doi: 10.1007/s12559-012-9156-1
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika* 58, 433–451. doi: 10.1093/biomet/58.3.433
- Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., et al. (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4001–4006. doi: 10.1073/pnas.1313490111
- Kingma, D., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv[preprint] arXiv:1412.6980*.
- Kinouchi, O., Diez-Garcia, R. W., Holanda, A. J., Zambianchi, P., and Roque, A. C. (2008). The non-equilibrium nature of culinary evolution. *N. J. Phys.* 10:073020. doi: 10.1088/1367-2630/10/7/073020
- Koren, Y. (2003). "On spectral graph drawing," in *International Computing and Combinatorics Conference* (Berlin; Heidelberg: Springer), 496–508.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems* 26, eds C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Cambridge, MA: MIT Press), 3111–3119.
- Newton, I. (1704). *Opticks: Or, a Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*. London: Sam. Smith and Benj. Walford.
- Nozawa, K., Nakaoka, Y., Yamamoto, S., and Satoh, T. (2014). Finding method of replaceable ingredients using large amounts of cooking recipes. *IEICE Tech. Report* 114, 41–46. Available online at: <https://ci.nii.ac.jp/naid/110009950250/>
- Panagiotakos, D. B., Pitsavos, C., and Stefanadis, C. (2006). Dietary patterns: a Mediterranean diet score and its relation to clinical and biological markers of cardiovascular disease risk. *Nutr. Metab. Cardiovasc. Dis.* 16, 559–568. doi: 10.1016/j.numecd.2005.08.006
- Pinel, F., Shao, N., and Varshney, L. R. (2014). Using physicochemical correlates of perceptual flavor similarity to enhance, balance and substitute flavors. US Patent App. 14/458,315.
- Pinel, F., and Varshney, L. R. (2014). Substitution of work products. US Patent App. 14/269,353.
- Serra-Majem, L., Trichopoulou, A., de la Cruz, J. N., Cervera, P., Garcí Alvarez, A., La Vecchia, C., et al. (2004). Does the definition of the Mediterranean diet need to be updated? *Public Health Nutr.* 7, 927–929. doi: 10.1079/PHN2004564
- Shidochi, Y., Takahashi, T., Ide, I., and Murase, H. (2009). "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* (Beijing), 9–14.
- Tallab, S. T., and Alrazgan, M. S. (2016). Exploring the food pairing hypothesis in Arab cuisine: A study in computational gastronomy. *Proc. Comput. Sci.* 82, 135–137. doi: 10.1016/j.procs.2016.04.020
- Teng, C.-Y., Lin, Y.-R., and Adamic, L. A. (2012). "Recipe recommendation using ingredient networks," in *Proceedings of the 3rd Annual ACM Web Science Conference (WebSci'12)* (Evanston, IL), 298–307.
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. Available online at: [https://www.bibsonomy.org/bibtex/28b9aebb404ad4a4c6a436ea413550b30/lopuz\\_kdd](https://www.bibsonomy.org/bibtex/28b9aebb404ad4a4c6a436ea413550b30/lopuz_kdd)
- Varshney, K. R., Varshney, L. R., Wang, J., and Myers, D. (2013). "Flavor pairing in Medieval European cuisine: A study in cooking with dirty data," in *International Joint Conference on Artificial Intelligence Workshop* (Beijing), 3–12.
- Yokoi, S., Doman, K., Hirayama, T., Ide, I., Deguchi, D., and Murase, H. (2015). "Typicality analysis of the combination of ingredients in a cooking recipe for assisting the arrangement of ingredients," in *2015 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)* (Turin).
- Zhu, Y. X., Huang, J., Zhang, Z. K., Zhang, Q. M., Zhou, T., and Ahn, Y. Y. (2013). Geography and similarity of regional cuisines in China. *PLoS ONE* 8:e79161. doi: 10.1371/journal.pone.0079161

**Conflict of Interest Statement:** MK and YI were employed by company Habitech Inc. MS and CH were employed by company Cancer Scan Inc. KM was employed by company Keisuke Matsushima.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kazama, Sugimoto, Hosokawa, Matsushima, Varshney and Ishikawa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.