A Delicious Recipe Analysis Framework for Exploring Multi-Modal Recipes with Various Attributes

Weiqing Min[1], Shuqiang Jiang[1,2], Shuhui Wang[1], Jitao Sang[3,4], Shuhuan Mei[5,1]
[1] Key Lab of Intelligent Information Processing, Institute of Computing Technology, CAS, Beijing, 100190, China
[2] University of Chinese Academy of Sciences, Beijing, 100049, China
[3] National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing, 100190, China
[4] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China
[5] Shandong University of Science and Technology, Qingdao, 266590, China
{minweiqing,sqjiang}@ict.ac.cn;sangjitao@gmail.com;{shuhui.wang,shuhuan.mei}@vipl.ict.ac.cn

ABSTRACT

Human beings have developed a diverse food culture. Many factors like ingredients, visual appearance, courses (e.g., breakfast and lunch), flavor and geographical regions affect our food perception and choice. In this work, we focus on multi-dimensional food analysis based on these food factors to benefit various applications like summary and recommendation. For that solution, we propose a delicious recipe analysis framework to incorporate various types of continuous and discrete attribute features and multi-modal information from recipes. First, we develop a Multi-Attribute Theme Modeling (MATM) method, which can incorporate arbitrary types of attribute features to jointly model them and the textual content. We then utilize a multi-modal embedding method to build the correlation between the learned textual theme features from MATM and visual features from the deep learning network. By learning attribute-theme relations and multi-modal correlation, we are able to fulfill different applications, including (1) flavor analysis and comparison for better understanding the flavor patterns from different dimensions, such as the region and course, (2) region-oriented multi-dimensional food summary with both multi-modal and multi-attribute information and (3) multi-attribute oriented recipe recommendation. Furthermore, our proposed framework is flexible and enables easy incorporation of arbitrary types of attributes and modalities. Qualitative and quantitative evaluation results have validated the effectiveness of the proposed method and framework on the collected Yummly dataset.

KEYWORDS

Multi-dimensional food analysis, multi-attribute theme modeling, flavor analysis, food summary, recipe recommendation

1 INTRODUCTION

Food is an integral part of our life. Many aspects like ingredients, visual appearance, flavor, meals and geographical regions play important roles in food perception, choice and consumption. For example, the obese subjects demonstrated changes in brain activity elicited by food-related visual cues [34]. Asians use soy sauce widely in their recipes while the ingredients from the Hungarian cuisine usually contain the paprika and lard [2]. Therefore, multi-dimensional food modeling based on these aspects can benefit applications like culinary habit exploration [9, 35], food perception and health [3, 31]. In this work, we focus on modeling these food factors into a unified framework, and then exploit it for various multimedia applications.

The proliferation of food-shared websites (e.g., Instagram, Yummly and Yelp) has provided rich data for food-oriented research, such as food recognition [22, 33], recipe retrieval [8] and culinary practice understanding [2, 35]. For example, Rich *et al.* [33] conducted the large scale content analysis of food images from Instagram. Sajadmanesh *et al.* [35] presented a study of recipes with the ingredients and flavor information from Yummly to explore the culinary habits. However, little work has investigated the problem of taking various types of attributes and multi-modal information into a unified framework, which are critical to enable food-related analysis and understanding.

The food attributes are diverse. For example, each recipe from Yummly is associated with different attributes, such as the cuisine, course and flavor information. Furthermore, the distributions of many attribute features are different. For instance, the cuisine and course attributes are discrete or categorical while the value of each flavor attribute is continuous. Therefore, a multi-dimensional food analysis framework should support arbitrary types of attributes. In addition, besides the text-based food research (e.g., the ingredients) [2, 35], some work like [4, 43] have found the importance of visual information in food related analysis. In many cases, the visual information is able to reduce the high cognitive load for easy understanding [43]. For example, Camacho *et al.* [4] proposed a framework to enhance each review by recommending relevant food images in Yelp. Therefore, different modalities should also be supported for a multi-dimensional food analysis framework.

In this paper, we propose a delicious recipe analysis framework, which is capable of incorporating various types of continuous and discrete attribute features and multi-modal information to meet these requirements. Particularly, we take the recipes from Yummly in our study. As shown in Fig.1, given the multi-modal recipe input with three types of attribute features, including continuous flavor attribute features, discrete cuisine and course attribute features, we first develop a Multiple Attribute Theme Modeling (MATM) method to model the correlation between the ingredients and these food attributes. We then utilize the Multi-Modal Embedding (MME)

MM'17, October 23–27, 2017, Mountain View, CA, USA.

DOI: https://doi.org/10.1145/3123266.3123272

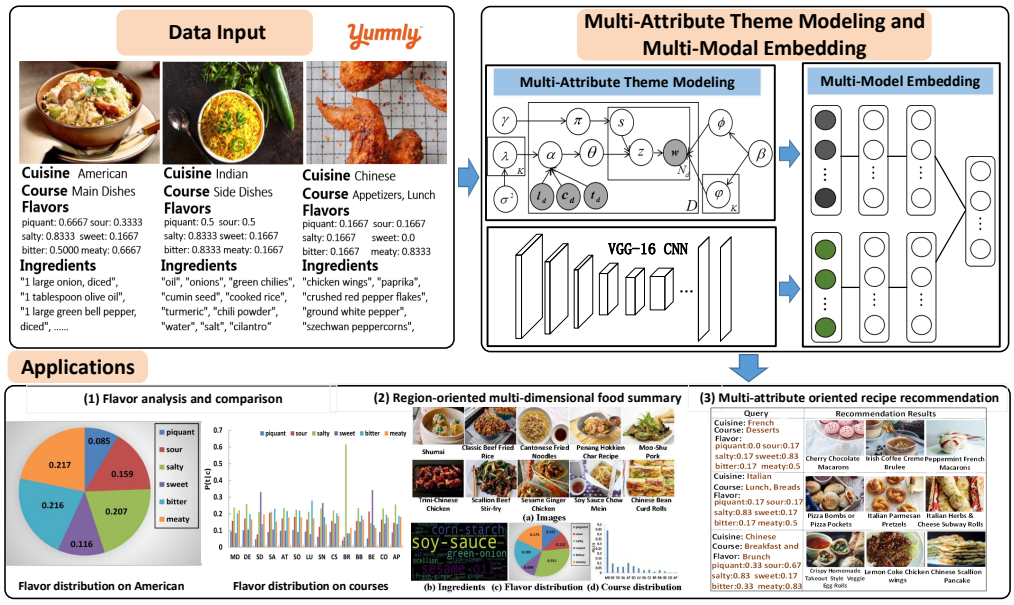


Figure 1: The proposed delicious recipe analysis framework

method to create joint representation based on learned ingredient theme features from MATM and visual features from the deep learning network. After mining attribute-theme patterns and correlating different modalities, we can conduct different tasks: (1) flavor analysis and comparison for better understanding the flavor patterns from different dimensions, such as the region and course, (2) region-oriented multi-dimensional food summary with both multi-modal and multi-attribute information and (3) multi-attribute oriented recipe recommendation for different attribute queries.

The contributions of this paper can be summarized as follows:

- We propose a delicious recipe analysis framework, which utilizes various types of attributes and multi-modal information to enable multi-dimensional food analysis and applications.
- We propose a multiple attribute theme modeling method, which can incorporate arbitrary attributes to model the correlation between the content and attributes.
- We present a wide variety of applications, including 1) flavor analysis and comparison, 2) region-oriented multi-dimensional food summary, and 3) multi-attribute oriented recipe recommendation.

2 RELATED WORK

Related work includes food recognition [7, 23, 26, 33, 41, 42], recipe retrieval and recommendation [8, 30], and culinary culture understanding [2, 20, 35, 38]. (1) **Food recognition**. Bossard *et al.* [7] used the random forest method to mine discriminative parts of food images for recognition. Some work [23, 41] utilized the convolutional neural network to extract deep visual features for recognition. Xu *et al.*[42] further introduced the geo-location information to improve the performance of dish recognition. In contrast, Rich *et al.*[33] performed the large scale content analysis of food images from Instagram taken in-the-wild. (2) **Recipe retrieval and recommendation**. Some work [12, 24] employed the traditional matrix factorization methods for recipe recommendation. Recently, Chen *et al.* [8] exploited visual features, ingredients and categories for recipe retrieval. Min *et al.* [30] further utilized both categorical cuisine and course attributes for retrieval. Different from them, we incorporate various types of discrete and continuous attribute features into a unified framework. Furthermore, we apply the proposed framework into different applications, such as multi-dimensional summary and multi-attribute based recommendation. (3) **Culinary culture understanding**. Sajadmanesh *et al.* [35] exploited the ingredients and flavor information from Yummly to understand worldwide culinary culture. Ahn *et al.* [2] constructed a flavor network to understand the culinary practice. Simas *et al.* [38] utilized this flavor network to analyse the food-bridging hypothesis behind the traditional cuisine. Howell *et al.* [20] relied on the food's name and nutritional content to predict the food's taste. Different from [20], we analyzed and compared different flavor patterns based on the ingredients and various food attributes. Furthermore, we proposed a unified framework to enable other food analysis and applications including summary and recommendation.

In addition, our work is also related to the probabilistic topic model [5, 17], such as Latent Dirichlet Allocation (LDA) [6] and Restricted Boltzmann Machine (RBM)[11], which have been applied to many tasks such as classification and recommendation [17, 29, 40]. In addition, some work incorporated additional feature information like the location and time into LDA for user interest modeling [36], landmark analysis [28] and social event analysis [32]. Srivastava *et al.* [39] proposed RBM based deep network for multi-modal data modeling.In this work, we extend the topic model to incorporate different types of continuous and discrete recipe attribute features,ingredients and recipe images for multi-dimensional food analysis and applications.

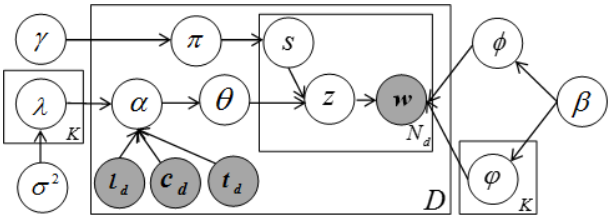


Figure 2: The multi-attribute theme model

3 DELICIOUS RECIPE ANALYSIS

3.1 Multi-Attribute Theme Modeling

We develop a theme modeling method to build the correlation between the ingredients and different types of attributes. Given the recipe set D, each recipe $d \in D$ is a tuple $[\mathbf{w}_d, \mathbf{l}_d, \mathbf{c}_d, \mathbf{t}_d]$. $\mathbf{w}_d$ is the ingredient vector. $\mathbf{l}_d$ is a L-dimensional vector containing features that encode cuisine values. $\mathbf{l}_d$ includes 1 in the positions for each cuisine from the recipe d, and 0 otherwise. Similarly for C-dimensional vector $\mathbf{c}_d$ and T-dimensional vector $\mathbf{t}_d$. L, C, T are the set of cuisines, courses and flavors, respectively. The goal of Multi-Attribute Theme Modeling (MATM) is to utilize the ingredients $\mathbf{w}_d$ and three types of attributes $\mathbf{l}_d, \mathbf{c}_d, \mathbf{t}_d$ to learn the theme space of the ingredients $\{\boldsymbol{\varphi}_k\}_{k \in K}$. K is the set of themes. In addition, we introduce the background distribution $\boldsymbol{\phi}$. As presented in Fig. 2, the generative process of MATM is:

(1) Draw a background distribution over ingredients $\boldsymbol{\phi} \sim Dir(\boldsymbol{\beta})$
(2) For each theme $k \in \{1, ..., K\}$,
  (a) draw $\boldsymbol{\lambda}_k \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$.
  (b) draw $\boldsymbol{\varphi}_k \sim Dir(\boldsymbol{\beta})$.
(3) For each recipe $d \in \{1, ..., D\}$
  (a) Draw $\boldsymbol{\pi}_d \sim Beta(\boldsymbol{\gamma})$
  (b) for each theme k, let $\alpha_{dk} = \exp(\mathbf{f}_d^T \boldsymbol{\lambda}_k)$
  (c) Draw $\boldsymbol{\theta}_d \sim Dir(\boldsymbol{\alpha}_d)$
  (d) for each ingredient $w_{d,n} \in \mathbf{w}_d$
    (i) Draw a switch variable $s_{d,n} \sim Binomial(\boldsymbol{\pi}_d)$
    (ii) if $s_{d,n} = 0$, draw an ingredient $w_{d,n} \sim Multi(\boldsymbol{\phi}^{bg})$
    (iii) if $s_{d,n} = 1$, draw a theme $z_{d,n} \sim Multi(\boldsymbol{\theta}_d)$, draw an ingredient $w_{d,n} \sim Multi(\boldsymbol{\varphi}_{z_{d,n}})$

where $Beta(\cdot)$, $Binomial(\cdot)$, $Multi(\cdot)$, $Dir(\cdot)$, $\mathcal{N}(\cdot)$ denote the Beta distribution,binomial distribution, multinomial distribution, Dirichlet distribution and normal distribution, respectively. $\mathbf{I}$ is the identity matrix. $\sigma^2\mathbf{I}$ is a diagonal matrix with the diagonal element σ^2.

Similar to LDA, we assume the symmetric Dirichlet priors with $\boldsymbol{\beta}$, that is $\beta_w = \beta$, $w \in \{1, ..., W\}$, where W is the size of the vocabulary. However, the theme distribution $\boldsymbol{\theta}_d$ of each recipe is no longer drawn from a Dirichlet prior with fixed hyper-parameters α. Instead, the elements of $\boldsymbol{\alpha}_d, \alpha_{dk} = \exp(\mathbf{f}_d^T \boldsymbol{\lambda}_k)$, where $\boldsymbol{\lambda}$ is the feature matrix with $K \times (L + C + T + 1)$. $\mathbf{f}_d$ is the concatenation of different attribute vectors. $\mathbf{f}_d = [\mathbf{l}_d; \mathbf{c}_d; \mathbf{t}_d; 1]$. Note that the representation of $\mathbf{l}_d$ and $\mathbf{c}_d$ are discrete while $\mathbf{t}_d$ is continuous.The score s_{dt} for each flavor t is in $[0, 1]$. Introducing $\exp(\mathbf{f}_d^T \boldsymbol{\lambda}_z)$ is able to incorporate arbitrary types of observed continuous and discrete features [27]. The last 1 is a default feature to account for the mean value of each theme. Therefore, for different combination of attributes, the resulting $\boldsymbol{\alpha}$ values are distinct. The theme distributions extracted from the recipe set are induced by both the ingredient patterns and various attribute features.

3.1.1 Model Inference and Parameter Estimation. We train this model using a Stochastic Expectation Maximization (SEM) method [10] and use an iterative procedure, which alternates between (i) Stochastic E step and (ii) M step.

(i) **Stochastic E step**: at the m iteration, sample z^m and s^m given current estimate $\boldsymbol{\lambda}^{m-1}$. We use the collapsed Gibbs sampling [15] for model inference.

$$p(z_i^m = k, s_i^m = 1 | \mathbf{z}_{\neg i}^m, \mathbf{s}_{\neg i}^m, \mathbf{w}_{\neg \mathbf{i}}, w_i = w, \boldsymbol{\alpha}_d^{m-1}, \beta, \gamma) \propto$$
$$\frac{n_{d,1,\neg i} + \gamma}{\sum_s n_{d,s,\neg i} + 2\gamma} \frac{n_{d,k,\neg i} + \exp(\mathbf{f}_d^T \boldsymbol{\lambda}_k^{m-1})}{\sum_k (n_{d,k,\neg i} + \exp(\mathbf{f}_d^T \boldsymbol{\lambda}_k^{m-1}))} \frac{(n_{k,w,\neg i} + \beta)}{\sum_{w'} n_{k,w',\neg i} + W\beta}$$
$$p(s_i = 0 | \mathbf{z}_{\neg i}^m, \mathbf{s}_{\neg i}^m, \mathbf{w}_{\neg \mathbf{i}}, w_i = w, \boldsymbol{\alpha}_d^{m-1}, \beta, \gamma) \propto$$
$$\frac{n_{w_i,\neg i}^{bg} + \beta}{\sum_{w'} n_{w',\neg i}^{bg} + W\beta} \frac{n_{d,0,\neg i} + \gamma}{\sum_s n_{d,s,\neg i} + 2\gamma} \tag{1}$$

where $i = (d, n)$ is the current index. The superscript $\neg i$ denotes a counting variable that excludes the i-th ingredient index in the corpus. $n_{d,k,\neg i}$ is the number of times that theme k is assigned to the recipe d. $n_{k,w,\neg i}$ is the number of times that the ingredient w is assigned to the theme k. $n_{w_i,\neg i}^{bg}$ is the number of times that the ingredient w_i is assigned to the background distribution. $n_{d,s_i,\neg i}$ is the number of times of the ingredient in the recipe d assigned to the background distribution $s_i = 0$ and themes $s_i = 1$, respectively.

(ii) **M step**: After stochastic E step, we obtain $n_d^m = \{n_{d,1}^m, ..., n_{d,k}^m, ..., n_{d,K}^m\}$. In this step, we estimate $\boldsymbol{\lambda}^m$ by maximizing the likelihood function $\mathcal{L} = \ln(\prod_{d=1}^{D} p(\mathbf{w}_d, \mathbf{f}_d, \boldsymbol{\lambda}, \boldsymbol{\alpha}_d, \mathbf{z}, \mathbf{s}))$. We compute $\frac{\partial \mathcal{L}}{\partial \lambda_{kj}^m} = 0$,

$$\frac{\partial \mathcal{L}}{\partial \lambda_{kj}^m} = \sum_{d=1}^{D} \exp(\mathbf{f}_d^T \boldsymbol{\lambda}_k^m) \mathbf{f}_{dj} (\Psi(\sum_{k'=1}^{K} \exp(\mathbf{f}_d^T \boldsymbol{\lambda}_{k'}^m)) - \Psi(\sum_{k'=1}^{K} \exp(\mathbf{f}_d^T \boldsymbol{\lambda}_{k'}^m)) + n_d^m) + \Psi(\exp(\mathbf{f}_d^T \boldsymbol{\lambda}_k^m) + n_{d,k}^m) - \Psi(\exp(\mathbf{f}_d^T \boldsymbol{\lambda}_k^m))) - \frac{\lambda_{kj}^m}{\sigma^2} \tag{2}$$

where $\Psi(\cdot)$ is the Digamma function and is defined as the derive of the log gamma function.

Similar to [27], we use the L-BFGS optimizer to compute $\boldsymbol{\lambda}^m$.

After the model inference, we can obtain the theme-attribute feature matrix $\{\hat{\boldsymbol{\lambda}}_k\}_{k=1}^K$ and the following parameters:

$$\hat{\pi}_{d,s} = \frac{n_{d,s} + \gamma}{\sum_{s'=0}^{1} n_{d,s'} + 2\gamma} \quad \hat{\varphi}_{k,w} = \frac{n_{k,w} + \beta}{\sum_{w'=1}^{W} n_{k,w'} + W\beta}$$
$$\hat{\phi}_w^{bg} = \frac{n_w^{bg} + \beta}{\sum_{w'=1}^{W} n_{w'}^{bg} + W\beta} \quad \hat{\theta}_{d,k} = \frac{n_{d,k} + \exp(\mathbf{f}_d^T \hat{\boldsymbol{\lambda}}_k)}{\sum_{k=1}^{K} (n_{d,k} + \exp(\mathbf{f}_d^T \hat{\boldsymbol{\lambda}}_k))} \tag{3}$$

3.2 Multi-Modal Embedding

We obtain the theme distribution $\hat{\boldsymbol{\theta}}_d = (\hat{\theta}_{d,1}, ..., \hat{\theta}_{d,k}, ..., \hat{\theta}_{d,K})$ for each recipe d via MATM. In order to correlate multi-modal content through the learned theme space, we resort to RBM based

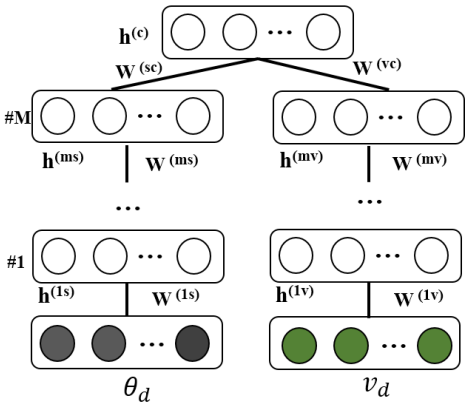


Figure 3: The multi-modal embedding

model for joint learning because of the powerful inference of RBM. Particularly, we utilize multi-modal DBM [39] to learn common space.

As shown in Fig. 3, there are two pathways: text pathway and image pathway, and each pathway contains M layers. For the text pathway, since $\hat{\boldsymbol{\theta}}_d$ is multinomial distribution, the connections between $\hat{\boldsymbol{\theta}}_d$ and $\mathbf{h}^{(1v)}$ are modeled as Multinomial RBM [37]. The conditional probability of the input unit is

$$P(\hat{\theta}_{dk} = 1|\mathbf{h}^{(1s)}) = \frac{\exp(b_k + \sum_j W_{kj}^{(1s)} h_j^{(1s)})}{\sum_{k'} \exp(b_{k'} + \sum_j W_{k'j}^{(1s)} h_j^{(1s)})} \tag{4}$$

where the weight matrix $\mathbf{W}^{(1s)} = \{W_{k,j}^{(1s)}\}$ is associated with the connection between the visible units $\hat{\boldsymbol{\theta}}_d$ and the hidden units $\mathbf{h}^{(1s)} = \{h_j^{(1s)}\}$, as well as bias weights $\mathbf{b} = \{b_k\}$ for the visible units and $\mathbf{c}^{(1s)} = \{c_j^{(1s)}\}$ for the hidden units.

For the image pathway, the input units $\mathbf{v}_d$ are visual features, and therefore the connections between $\mathbf{v}_d$ and $\mathbf{h}^{(1v)}$ are modeled as Gaussian RBM [18]. The conditional probability of the input unit is

$$v_{di} = 1|\mathbf{h}^{(1v)} \sim \mathcal{N}(b_i + \sigma_i \sum_j W_{ij}^{(1v)} h_j^{(1v)}, \sigma_i^2) \tag{5}$$

where $\mathbf{v}_d$ is the visual feature vector of recipe d. σ_i^2 is the variance of the Gaussian distribution. The remaining layers from two pathways are the standard binary RBM [11]. The conditional distributions are written similarly and omitted here for the space limit.

Similar to [39], we train the network with stochastic gradient descent using the greedy layer-wise pre-training strategy. After training, for given recipe-theme vector $\hat{\boldsymbol{\theta}}_d$ as the input, we can use the mean-field method to sample the hidden modalities from $p(\mathbf{v}|\hat{\boldsymbol{\theta}}_d)$ by updating each hidden layer given the states of the adjacent layers. Finally, we obtain the visual representation based on the learned ingredient representation.

4 APPLICATIONS

4.1 Flavor Analysis and Comparison

In this task, we analyze the flavor patterns from different attributes, such as the region and course.

For flavor distributions across regions, we compute $p(t|l)$:

$$p(t|l) = \sum_k p(t|k)p(k|l) \tag{6}$$

We obtain the cuisine representation $p(k|l)$ as

$$p(k|l) = \psi_{l,k} = \frac{\exp(\mathbf{f}_l^T \hat{\boldsymbol{\lambda}}_k)}{\sum_{k'} \exp(\mathbf{f}_l^T \hat{\boldsymbol{\lambda}}_{k'})} \tag{7}$$

where for $\mathbf{f}_l$, the position corresponding to this cuisine feature l is 1 and 0, otherwise. The proportion of theme k over the cuisine l is $\exp(\mathbf{f}_l^T \hat{\boldsymbol{\lambda}}_k)$.

Similarly, we obtain the flavor representation $p(k|t)$ as

$$p(k|t) = v_{t,k} = \frac{\exp(\mathbf{f}_t^T \hat{\boldsymbol{\lambda}}_k)}{\sum_{k'} \exp(\mathbf{f}_t^T \hat{\boldsymbol{\lambda}}_{k'})} \tag{8}$$

We then compute $p(t|k)$

$$p(t|k) = \frac{p(t,k)}{\sum_{t' \in T} p(k|t')p(t')} = \frac{v_{t,k}p(t)}{\sum_{t' \in T} v_{t',k}p(t')} \tag{9}$$

where $p(t) = \frac{\sum_d s_{dt}}{\sum_{t'} \sum_d s_{dt'}}$.

Therefore, the flavor distribution on certain cuisine $p(t|l)$ is

$$p(t|l) = \sum_k \frac{v_{t,k}p(t)}{\sum_{t' \in T} v_{t',k}p(t')} \psi_{l,k} \tag{10}$$

Similarly, the flavor distributions from different courses $p(t|c)$ is computed as

$$p(t|c) = \sum_k \frac{v_{t,k}p(t)}{\sum_{t' \in T} v_{t',k}p(t')} \xi_{c,k} \tag{11}$$

where $\xi_{c,k} = \frac{\exp(\mathbf{f}_c^T \hat{\boldsymbol{\lambda}}_k)}{\sum_{k'} \exp(\mathbf{f}_c^T \hat{\boldsymbol{\lambda}}_{k'})}$.

4.2 Region-Oriented Multi-dimensional Food Summary

In this task, we summarize regional foods based on representative ingredients, representative recipe images, flavor and course patterns.

The representativeness score of ingredient w for region l_q is measured according to the expected number of times ingredient w, which is projected onto the theme space.

$$Rep_{l_q}(w) = n(l_q, w) \frac{\sum_k \hat{\varphi}_{kw} \psi_{l_q,k}}{\sum_k \hat{\varphi}_{kw} \psi_{l_q,k} + \hat{\phi}_w^{bg}}, p(w|l_q) = \frac{Rep_{l_q}(w)}{\sum_{w'} Rep_{l_q}(w')} \tag{12}$$

where $n(l_q, w)$ is the frequency of w in recipes with the cuisine l_q, $\psi_{l_q,k}$ is calculated by Eqn.7.

We can obtain a list of region-representative ingredients by ranking the ingredients according to $p(w|l_q)$.

In order to select representative recipe images from one region, we consider both textual and visual information. We obtain the representation $\boldsymbol{\psi}_{l_q}$ for ingredients using Eqn. 7, and then use MME

to infer the visual representation $\mathbf{v}_{l_q}$ based on $\boldsymbol{\psi}_{l_q}$. The image is sorted by the weighted sum of cosine similarity between l_q and the image I_d:

$$sim(l_q, I_d) = \tau \frac{\boldsymbol{\psi}_{l_q}^T \hat{\boldsymbol{\theta}}_d}{\|\boldsymbol{\psi}_{l_q}\| \|\hat{\boldsymbol{\theta}}_d\|} + (1-\tau) \frac{\mathbf{v}_{l_q}^T \mathbf{v}_d}{\|\mathbf{v}_{l_q}\| \|\mathbf{v}_d\|} \tag{13}$$

where l_d represents the region information of the recipe d. The first term is the semantic similarity and the second term is the visual similarity. τ is the weight parameter.

We compute $p(t|l_q)$ and $p(c|l_q)$ to obtain the flavor distribution and course distribution using Eqn.10-11, respectively.

4.3 Multi-Attribute Oriented Recipe Recommendation

Given the multiple query attributes including the cuisine l, course c and flavor t, this task is to recommend relevant recipes to match these query attributes.

The attribute representation $\boldsymbol{\vartheta}_{\mathbf{f}_q} = \{p(k|\mathbf{f}_q)\}_{k \in K}$ with specified cuisine, course and flavor information is calculated as

$$p(k|\mathbf{f}_q) = \frac{\exp(\mathbf{f}_q^T \hat{\boldsymbol{\lambda}}_k)}{\sum_k \exp(\mathbf{f}_q^T \hat{\boldsymbol{\lambda}}_k)} \tag{14}$$

For each recipe d with the ingredient information, we compute the representation of the recipe $p(k|d)$ through the Gibbs sampler by maximizing

$$\mathcal{L}(\mathbf{w}_d) = \prod_{i=1}^{N_d} [\pi_{d,0} \hat{\phi}_{w_i} + \pi_{d,1} \sum_{k \in K} p(k|d) \hat{\varphi}_{k,w_i}] \tag{15}$$

where N_d is the count of ingredients in the recipe d.

We rank the results according to the similarity of the query attributes and the recipes from the dataset, which is computed as:

$$JsSim(\boldsymbol{\vartheta}_{\mathbf{f}}, \boldsymbol{\theta}_d) = \exp\{-D_{js}(\boldsymbol{\vartheta}_{\mathbf{f}} || \boldsymbol{\theta}_d)\} \tag{16}$$

where $D_{js}(\cdot||\cdot)$ denotes the Jensen-Shannon divergence and $\boldsymbol{\theta}_d = \{p(k|d)\}_{k \in K}$.

5 EXPERIMENT

5.1 Experimental Settings

5.1.1 Dataset. The experimental dataset is crawled from Yummly. Each recipe includes the ingredients, the image and three attributes: cuisine, course and flavor. There are 44,204 recipe items, 10 cuisines, 14 courses[1] and 6 flavors. Table 1 lists the values of different attributes. Note that we use course names and their abbreviated ones interchangeably. The name in the brackets is the abbreviated one of the course. Each cuisine represents one region. The value of each flavor is continuous and they are in [0, 1]. The other two attributes are categorical. We preprocess each ingredient line using the method [30]. For the ingredients with more than two words, we represent them by concatenating them using "-". For example, we use chili-powder to represent the ingredient "chili powder". After preprocessing, the vocabulary of ingredients is 2,407.

[1]Different from predefined 13 courses in Yummly, we consider Lunch and Snacks as two kinds of courses.

Table 1: Values of different attributes

♯ Type	♯Value
Cuisine	American, Italian, Mexican, Indian, French, Thai, Chinese, Spanish, Greek, Japanese
Course	Main Dishes (MD), Desserts (DE), Side Dishes (SD), Salads (SA),Lunch (LU), Snacks (SN), Soups (SO), Afternoon Tea (AT),Condiments and Sauces (CS), Breads (BR), Breakfast and Brunch (BB), Beverages (BE), Cocktails (CO), Appetizers (AP)
Flavor	Piquant, Sour, Salty, Sweet, Bitter, Meaty

5.1.2 Implementation Details. For each image, we use the VGG-16 deep network to extract the 4,096-D features according to [16]. For MATM, $\beta = 0.01$. $\gamma = 1.0$. Similar to [27], the variance σ^2 is set to 0.5 for all attribute features and 10.0 for the default features. For the initialization of SEM, the parameter $\alpha = 1.0/K$. We run 2,000 iterations for training MATM. After an initial burn-in period of 200 iterations, we optimize $\boldsymbol{\lambda}$ every 50 iterations. For MME, the learning rate of multinomial layer and Gaussian layer is 0.10 and 0.001. The learning rate of other layers is 0.01. For Gaussian-RBM, each Gaussian visible unit is empirically set to have unit variance to guarantee the stability of training [39]. In the inference, the times for mean-field inference is 100. In Eqn. 13, τ is empirically set as 0.6.

5.2 Evaluation of MATM

5.2.1 Theme Number K Selection. In theme modeling, the selection of theme number K is important. We resort to the *perplexity* [6] as the metric, which is a standard measure for estimating how well one generative model fits the data. The lower the perplexity is, the better the performance. In MATM, the perplexity of the test set is defined as:

$$perplexity(D_{test}) = \exp(-\frac{\sum_{d \in D_{test}} \log p(\mathbf{w}_d, \mathbf{f}_d | D_{train})}{\sum_{d \in D_{test}} N_d}) \tag{17}$$

$$p(\mathbf{w}_d, \mathbf{f}_d | D_{train}) = \prod_{i=1}^{N_d} [\pi_{d,0} \hat{\phi}_{w_i} + \pi_{d,1} \sum_{k \in K} \frac{n_{d,k} + \exp(\mathbf{x}_d^T \hat{\boldsymbol{\lambda}}_k)}{\sum_{k'=1}^{K} (n_{d,k'} + \exp(\mathbf{x}_d^T \hat{\boldsymbol{\lambda}}_{k'}))} \hat{\varphi}_{k,w_i}] \tag{18}$$

where D_{train} is the training set and D_{test} is the test set.

A Gibbs sampler is run on the test data D_{test} to calculate the counts $n_{d,k}$ and π_d given $\hat{\varphi}_{k,w_i}$ and $\hat{\boldsymbol{\lambda}}$, learned from D_{train}.

For the dataset, we randomly select 90% of the recipe items as the training data and the remaining 10% as the test data. We set $K \in \{40, 60, 80, 100, 120, 140\}$. The perplexity scores over different theme number for different iterations are shown in Fig. 4. We can see that (1) The perplexity decreases slowly and converges to a stable level after 1,000 iterations. (2) The perplexity decreases much slower when $K >= 100$. Therefore, we choose $K = 100$ in the following experiments.

5.2.2 Illustration of Discovered Themes. In addition, we provide some discovered themes in Table 2, where "#j" denotes the j-th theme index. Each theme is represented by 10 top-ranked ingredients, sorted by $\hat{\varphi}$. We observe that some themes denote the classic

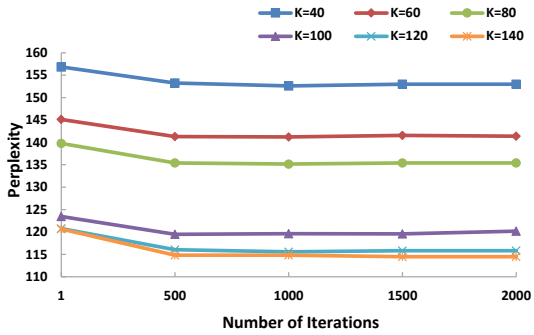


Figure 4: The perplexity score over different theme number for different iterations.

combination of ingredients, such as Theme #11. Theme #12 denotes the ingredients of fruit drinks while Theme #44 denotes the Italy-style ingredients. From these examples, we can see that these themes have a reasonable interpretation as the ingredient base.

Table 2: Some examples of discovered themes.

Theme ♯11	Theme ♯12	Theme ♯44
egg 0.366	honey 0.094	mozzarella-cheese 0.124
milk 0.178	mango 0.073	parmesan-cheese 0.110
salt 0.111	fresh-lime-juice 0.064	ricotta-cheese 0.098
butter 0.060	chopped-fresh-mint 0.041	egg 0.079
flour 0.044	fresh-mint 0.041	lasagna-noodle 0.058
salted-butter 0.032	juice 0.039	marinara-sauce 0.053
melted-butter 0.016	navel-orange 0.036	pasta-sauce 0.044
egg-white 0.015	pineapple 0.035	italian-seasoning 0.043
bread 0.014	orange 0.029	italian-sausage 0.037
honey 0.012	strawberry 0.027	tomato-sauce 0.030

5.3 Evaluation of MME

We investigate how many hidden layers M should we put between the recipe textual ingredients and visual modalities. As shown in Fig. 3, we could have one intervening layer, creating an RBM (theme input|joint hidden layer|image input) as 1-layer. A two-layered MME would have 3 intervening layers (theme input|theme hidden 1|joint hidden|image hidden 1|image input) as 2-layer, and so on. The set of units of different layers for the image-pathway is $\{4096, 1024, 1024, 1024, 1024\}$ and the text pathway is $\{100, 100, 100, 100, 100\}$. The units of the joint layer under the different selection of layers from the image-pathway and text-pathway are all $1024 + 100 = 1124$ for fair comparison.

MME is used to generate the visual modality given specified theme features, we therefore select the cross-modal retrieval task for evaluation. Since there is only one groundtruth match between the recipe ingredients and the recipe image, we use Top R % for evaluation [21], which is the relative number of images correctly retrieved in the first R% of the ranked list. Specifically, we set $R \in \{20, 40, 60, 80\}$. Based on the split in the evaluation of MATM, we further select 10% from the training data as the validation set. Based on the learned representation, the cosine similarity is calculated to obtain the ranked results. Table. 3 shows the results of these models. Comparing the performance of MME with different layers, we can see that the 3-layer MME leads to the best performance. Therefore, we select $M = 3$ for the following experiment.

Table 3: The performance for different layers with Top R%.

Method	R=20	R=40	R=60	R=80
1-layer	0.199	0.399	0.598	0.798
2-layer	0.202	0.399	0.600	0.800
3-layer	**0.207**	**0.411**	**0.615**	**0.810**
4-layer	0.201	0.401	0.600	0.800
5-layer	0.203	0.404	0.603	0.799

5.4 Evaluation of Applications

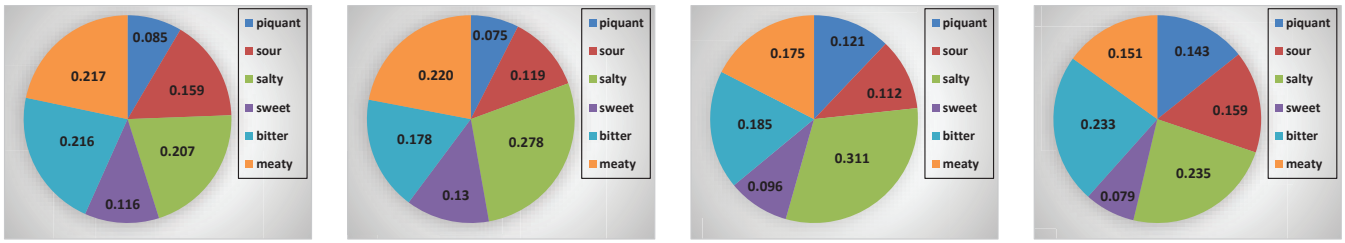
5.4.1 Flavor Analysis and Comparison. The results of flavor distributions in different regions are demonstrated in Fig. 5. For the space limit, we show the flavor distributions of four countries including American, French, Chinese and Mexican. We can analyze and compare different flavor patterns in one country or among different countries from these results. For example, American is the country with the high consumption of meat foods, with the probability of meaty flavor 0.217, when compared to foods of other flavors from this country. As we all known, American is one country, where meat is one of most-eat food. Besides the meaty food, French also likes sweet foods, with the probability of 0.130, which is the most proportion compared with other countries. The significant difference in the flavor distributions between these four countries is the high consumption of salty flavor labeled foods among Chinese, with the probability of 0.31. This is because Chinese is one of countries with the highest consumption of salty foods in the world [19].

In addition, we show the flavor differences among different courses in Fig. 6. We can see that some courses such as Beverages (BE), Side Dishes (SD) and Snacks (SN) are observed to have comparatively higher proportions of sweet foods compared to other courses, perhaps reflecting the prevalence of sugary items, like fruit juices. Meanwhile, salty and meaty foods dominate some courses, such as Main Dishes (MD) and lunch (LU), when more meat might be consumed.

The results can be insightful for future exploration to enable different applications like exploring culinary habits [2] and providing better recommendation service.

5.4.2 Region-Oriented Multi-dimensional Food Summary. As it is subjective to evaluate the summarized results, we resort to the user study and consider the following baselines for comparison.

- MATM based Textual Summary (MATM-TS). MATM-TS uses Eqn.12 derived from MATM to select representative ingredients for summary.
- MATM based Multi-modal Summary (MATM-MS). MATM-MS uses both representative ingredients and images for summary. This baseline selected the recipe images based on the the similarity of ingredients.



(a) American (b) French (c) Chinese (d) Mexican

Figure 5: Flavor distributions in different regions

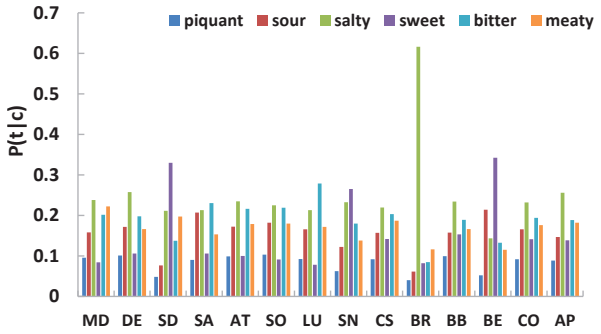


Figure 6: Flavor distributions on different courses.

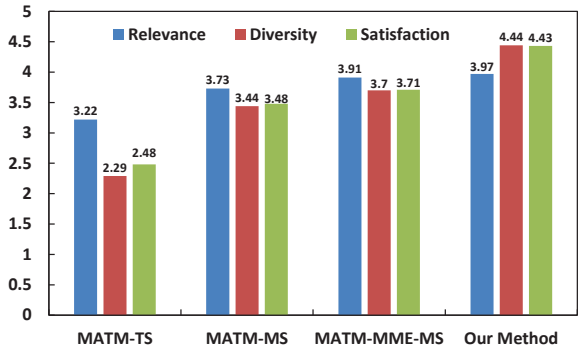


Figure 7: Subjective evaluation for food summary.

- MATM and MME based Multi-modal Summary (MATM-MME-MS). This baseline also uses both the representative ingredients and images for summary. However, it considers both textual and visual information using Eqn.13 to select representative images.

Compared with MATM-MME-MS, our framework additionaly introduced the flavor and course distribution. We asked 15 graduate students to assess the summarized results using 1 to 5 ratings from three aspects including (1) Relevance (i.e., to what extent the summary is depicting the regional food), (2) Diversity (i.e., depicting the regional food from multiple aspects), and (3) Satisfaction. How satisfied are you with the summarized results? They can consult the internet (e.g., Wikipedia) and other cooking books to help make judgement.

We show 50 top-ranked ingredients in a tag-cloud way, generated by wordcloud[2] and 10 top-ranked food images. We averaged all the scores as the final rating. The results are shown in Fig.7. We can see (1) After incorporating the visual information into the selection of images, MATM-MME-MS has higher scores in all aspects than MATM-MS. We further observed that there are some recipe images, which are not the dishes corresponding to the ingredients, but cooking books or other irrelevant information. MATM-MME-MS can exclude these images and improve the ranked results; and (2) Note that for the relevance, the difference of MATM-MME-MS and our method is small. Because introducing the flavor and course distribution does not affect the relevance score, while in other two aspects, our method exhibits advantages due to the introduced flavor and course distributions from our method. In addition, two example summaries are illustrated in Fig. 8 for Chinese and Mexican, respectively. The recipe name is showed below each image. From such multi-dimensional regional food summary, users could understand the local culinary characteristics of a region efficiently and comprehensively with the textual information, visual information and different attribute patterns. For example, we can see that Mexican likes the food with sour-and-hot flavor, since their ingredients include chili-powder, sour-cream and so on. Correspondingly, the total probability of piquant and sour is the highest compared with other flavor propositions. From the representative recipe images, we can further easily understand their culinary habits.

5.4.3 Multi-Attribute Oriented Recipe Recommendation. We divided the test set into two parts: one part contains the course, cuisine and flavor information, and the other one contains the recipe id with the ingredients. Our goal is to use the flavor, course and cuisine attributes as the query to retrieve recipes. We choose S@H [44] as the metric, which is the success rate of finding the target within the top-H recommendation results and $H \in \{1, 3, 5\}$. As a comparison, we consider LDA [6] as the baseline. This baseline is trained without incorporating the attribute information. For the attribute representation, it is the average of the theme representation of trained recipes annotated by the query attributes. As shown in Fig. 9, we can see that our method outperforms the baseline due to the introduction of the attributes. Because some recipes have the same attributes, we further choose MAP@5 for evaluation. The MAP@5

[2]https://github.com/amueller/word_cloud

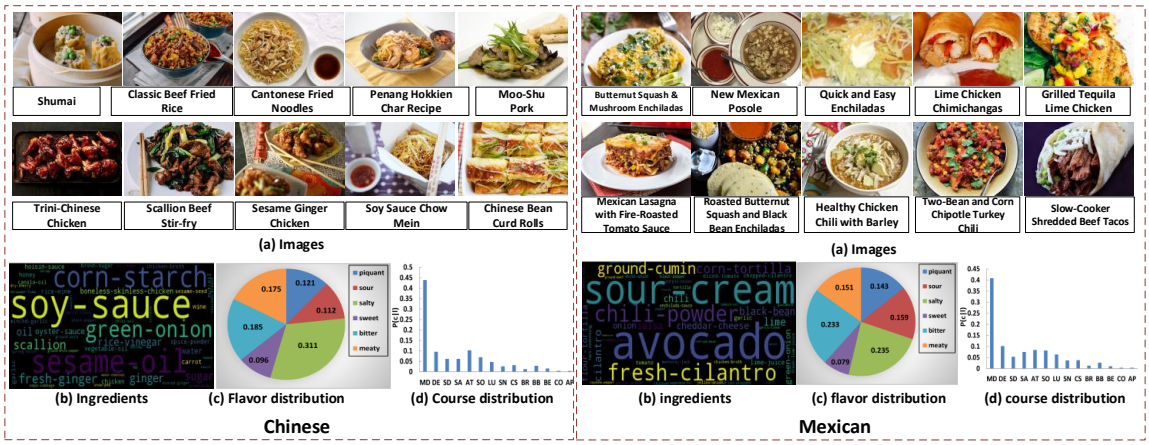


Figure 8: Multi-dimensional summary for Chinese and Mexican food.

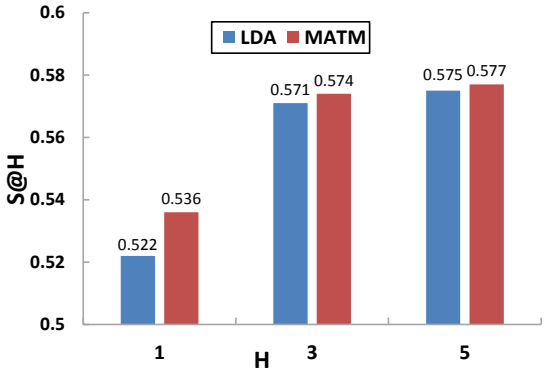


Figure 9: The evaluation of recipe recommendation.

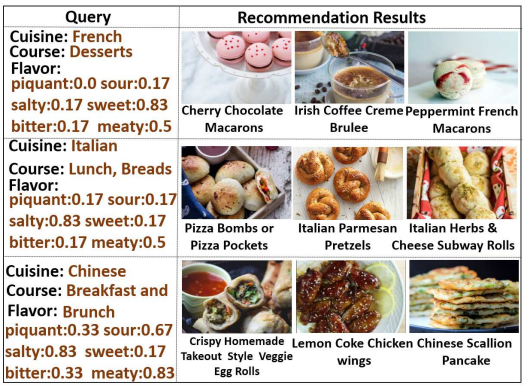


Figure 10: Some recommendation examples from MATM.

of LDA and MATM are 0.571 and 0.578, respectively. Our method again outperforms the baseline. Fig. 10 shows some examples of recommendation results from MATM, where the recipe image and name associated with each recipe id are present.

6 CONCLUSIONS

This work has presented a recipe analysis framework to incorporate multi-modal information, various types of continuous and discrete attribute features for multi-dimensional food analysis. A multi-attribute theme modeling method is proposed to jointly model the ingredients and arbitrary types of recipe attributes, such as continuous flavor attribute, discrete cuisine and course attributes. The derived attribute-theme representation and multi-modal correlation has demonstrated its effectiveness via our proposed three applications in flavor analysis, food summary and recipe recommendation. We hope that this framework could further the agenda of food-relative study and meanwhile form a contribution to other fields, such as computational gastronomy and food science [1].

There are four directions needing further investigation. The first one is to enlarge our data set for more comprehensive and deeper multi-dimensional food analysis. The second one is how to select the recipe images with higher quality for commerical applications. For example, Alex M. [25] has attempted to understand the photo qualities from Yelp for cover photo sorting application. As the third direction, we have demonstrated the potential of our framework via three applications. Based on the proposed framework, more real applications could be designed, such as food tourism [14] and so on. The fourth one is on a product dimension for building better personalized food recommendation system. For example, Ge *et al.* [13] has designed a recipe recommender system, that can suit both the user's preference and their health. Such system should offer healthier options but still satisfy the individual preferences.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (61532018, 61322212, 61602437 and 61672497), in part by the Beijing Municipal Commission of Science and Technology (D161100001816001), in part by Beijing Natural Science Foundation (4174106), in part by the Lenovo Outstanding Young Scientists Program, in part by National Program for Special Support of Eminent Professionals and National Program for Support of Top-notch Young Professionals, and in part by China Postdoctoral Science Foundation (2016M590135, 2017T100110).

REFERENCES

[1] Yong Yeol Ahn and Sebastian Ahnert. 2013. The Flavor Network. *Leonardo* 46, 3 (2013), 272–273.

[2] Yong-Yeol Ahn, Sebastian E Ahnert, James P Bagrow, and Albert-László Barabási. 2011. Flavor network and the principles of food pairing. *Scientific reports* 1 (2011).

[3] Kiyoharu Aizawa and Makoto Ogawa. 2015. FoodLog: Multimedia Tool for Healthcare Applications. *Multimedia IEEE* 22, 2 (2015), 4–8.

[4] Roberto Camacho Barranco, Laura M. Rodriguez, Rebecca Urbina, and M. Shahriar Hossain. 2016. Is a Picture Worth Ten Thousand Words in a Review Dataset? *arXiv:1606.07496* (2016).

[5] David M Blei. 2012. Probabilistic topic models. *Communications of The ACM* 55, 4 (2012), 77–84.

[6] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.

[7] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101–mining discriminative components with random forests. In *European Conference on Computer Vision*. 446–461.

[8] Jingjing Chen and Chong-Wah Ngo. 2016. Deep-based ingredient recognition for cooking recipe retrieval. In *Proceedings of the 2016 ACM on Multimedia Conference*. 32–41.

[9] Munmun De Choudhury, Sanket Sharma, and Emre Kiciman. 2016. Characterizing Dietary Choices, Nutrition, and Language in Food Deserts via Social Media. In *ACM Conference on Computer-Supported Cooperative Work and Social Computing*. 1157–1170.

[10] JosÃĺ G. Dias and Michel Wedel. 2004. An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. *Statistics and Computing* 14, 4 (2004), 323–332.

[11] Asja Fischer and Christian Igel. 2014. Training restricted Boltzmann machines: An introduction. *Pattern Recognition* 47, 1 (2014), 25–39.

[12] Peter Forbes and Mu Zhu. 2011. Content-boosted matrix factorization for recommender systems: experiments with recipe recommendation. In *Proceedings of the fifth ACM conference on Recommender systems*. 261–264.

[13] Mouzhi Ge, Francesco Ricci, and David Massimo. 2015. Health-aware Food Recommender System. In *Proceedings of the Conference on Recommender Systems*. 333–334.

[14] Andrea Giampiccoli and Janet Hayward Kalis. 2012. Tourism, Food, and Culture: Community-Based Tourism, Local Food, and Community Development in Mpondoland. *Culture and Agriculture* 34, 2 (2012), 101–123.

[15] T. L. Griffiths and M. Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America* 101, Suppl 1 (2004), 5228–5235.

[16] Jack Hessel, Nicolas Savva, and Michael J. Wilber. 2015. Image Representations and New Domains in Neural Image Captioning. *Computer Science* (2015).

[17] Geoffrey E Hinton and Ruslan Salakhutdinov. 2009. Replicated softmax: an undirected topic model. In *Advances in neural information processing systems*. 1607–1614.

[18] Geoffrey E Hinton and Ruslan R Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science* 313, 5786 (2006), 504–507.

[19] D. B. Hipgrave, S. Chang, X. Li, and Y. Wu. 2016. Salt and Sodium Intake in China. *Jama* 315, 7 (2016), 703.

[20] Patrick D Howell, Layla D Martin, Hesamoddin Salehian, Chul Lee, Kyler M Eastman, and Joohyun Kim. 2016. Analyzing Taste Preferences From Crowdsourced Food Entries. In *International Conference on Digital Health Conference*. 131–140.

[21] Yangqing Jia, Mathieu Salzmann, and Trevor Darrell. 2011. Learning cross-modality similarity for multinomial data. In *Computer Vision, IEEE International Conference on*. 2407–2414.

[22] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. 2014. Food Detection and Recognition Using Convolutional Neural Network. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 1085–1088.

[23] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. 2014. Food detection and recognition using convolutional neural network. In *Proceedings of the ACM International Conference on Multimedia*. 1085–1088.

[24] Chia-Jen Lin, Tsung-Ting Kuo, and Shou-De Lin. 2014. A content-based matrix factorization model for recipe recommendation. In *Advances in Knowledge Discovery and Data Mining*. 560–571.

[25] Alex M. 2016. Finding Beautiful Yelp Photos Using Deep Learning. *https://engineeringblog.yelp.com/2016/11/finding-beautiful-yelp-photos-using-deep-learning.html* (2016).

[26] Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P Murphy. 2015. Im2Calories: towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision*. 1233–1241.

[27] David Mimno and Andrew Mccallum. 2012. Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression. *University of Massachusetts - Amherst* 2008 (2012), 411–418.

[28] Weiqing Min, Bing Kun Bao, and Changsheng Xu. 2014. Multimodal Spatio-Temporal Theme Modeling for Landmark Analysis. *IEEE Multimedia* 21, 3 (2014), 20–29.

[29] Weiqing Min, Bing Kun Bao, Changsheng Xu, and M. Shamim Hossain. 2015. Cross-Platform Multi-Modal Topic Modeling for Personalized Inter-Platform Recommendation. *IEEE Transactions on Multimedia* 17, 10 (2015), 1787–1801.

[30] Weiqing Min, Shuqiang Jiang, Jitao Sang, Huayang Wang, Xinda Liu, and Luis Herranz. 2017. Being a Super Cook: Joint Food Attributes and Multi-Modal Content Modeling for Recipe Retrieval and Exploration. *IEEE Transactions on Multimedia* 19, 5 (2017), 1100 – 1113.

[31] Ferda Ofli, Yusuf Aytar, Ingmar Weber, Raggi Al Hammouri, and Antonio Torralba. 2017. Is Saki delicious? The Food Perception Gap on Instagram and Its Relation to Health. *arXiv preprint arXiv:1702.06318* (2017).

[32] Shengsheng Qian, Tianzhu Zhang, and Changsheng Xu. 2016. Multi-modal Multi-view Topic-opinion Mining for Social Event Analysis. In *ACM on Multimedia Conference*. 2–11.

[33] Jaclyn Rich, Hamed Haddadi, and Timothy M Hospedales. 2016. Towards Bottom-Up Analysis of Social Food. In *Proceedings of the 6th International Conference on Digital Health Conference*. ACM, 111–120.

[34] M Rosenbaum, M Sy, K Pavlovich, R. L. Leibel, and J Hirsch. 2008. Leptin reverses weight loss-induced changes in regional neural activity responses to visual food stimuli. *Journal of Clinical Investigation* 118, 7 (2008), 2583–2591.

[35] Sina Sajadmanesh, Sina Jafarzadeh, Seyed Ali Ossia, Hamid R Rabiee, Hamed Haddadi, Yelena Mejova, Mirco Musolesi, Emiliano De Cristofaro, and Gianluca Stringhini. 2016. Kissing Cuisines: Exploring Worldwide Culinary Habits on the Web. *arXiv preprint arXiv:1610.08469* (2016).

[36] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. 2007. Restricted Boltzmann machines for collaborative filtering. In *Machine Learning, Proceedings of the Twenty-Fourth International Conference*. 791–798.

[37] R. Salakhutdinov, J. B. Tenenbaum, and A. Torralba. 2013. Learning with Hierarchical-Deep Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 8 (2013), 1958.

[38] Tiago Simas, Michal Ficek, Albert DiazGuilera, Pere Obrador, and Pablo R. Rodriguez. 2017. Food-bridging: a new network construction to unveil the principles of cooking. *arXiv preprint arXiv:1610.08469* (2017).

[39] Nitish Srivastava and Ruslan Salakhutdinov. 2014. Multimodal learning with deep Boltzmann machines. *The Journal of Machine Learning Research* 15, 1 (2014), 2949–2980.

[40] Chong Wang, D. Blei, and Fei Fei Li. 2009. Simultaneous image classification and annotation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1903–1910.

[41] Hui Wu, Michele Merler, Rosario Uceda-Sosa, and John R Smith. 2016. Learning to Make Better Mistakes: Semantics-aware Visual Food Recognition. In *ACM on Multimedia Conference*. 172–176.

[42] Ruihan Xu, Luis Herranz, Shuqiang Jiang, Shuang Wang, Xinhang Song, and Ramesh Jain. 2015. Geolocalized Modeling for Dish Recognition. *Multimedia, IEEE Transactions on* 17, 8 (2015), 1187–1199.

[43] Longqi Yang, Yin Cui, Fan Zhang, John P Pollak, Serge Belongie, and Deborah Estrin. 2015. PlateClick: Bootstrapping Food Preferences Through an Adaptive Visual Interface. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 183–192.

[44] Wanlei Zhao, Yu Gang Jiang, and Chong Wah Ngo. 2006. Keyframe Retrieval by Keypoints: Can Point-to-Point Matching Help? 4071 (2006), 72–81.