This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

You Are What You Eat: Exploring Rich Recipe Information for Cross-Region Food Analysis

Weiqing Min, Bing-Kun Bao, Shuhuan Mei, Yaohui Zhu, Yong Rui, *Fellow, IEEE*, Shuqiang Jiang, *Senior, IEEE* .

***Abstract*—Cuisine is a style of cooking and usually associated with a specific geographic region. Recipes from different cuisines shared on the web are an indicator of culinary cultures in different countries. Therefore, analysis of these recipes can lead to deep understanding of food from the cultural perspective. In this paper, we perform the first cross-region recipe analysis by jointly using the recipe ingredients, food images and attributes such as the cuisine and course (e.g., main dish and dessert). For that solution, we propose a culinary culture analysis framework to discover the topics of ingredient bases and visualize them to enable various applications. We firstly propose a probabilistic topic model to discover cuisine-course specific topics. The manifold ranking method is then utilized to incorporate deep visual features to retrieve food images for topic visualization. At last, we applied the topic modeling and visualization method for three applications: (1) multi-modal cuisine summarization with both recipe ingredients and images, (2) cuisine-course pattern analysis including topic-specific cuisine distribution and cuisine-specific course distribution of topics, and (3) cuisine recommendation for both cuisine-oriented and ingredient-oriented queries. Through these three applications, we can analyze the culinary cultures at both macro and micro levels. We conduct the experiment on a recipe database Yummly-66K with 66,615 recipes from 10 cuisines in Yummly. Qualitative and quantitative evaluation results have validated the effectiveness of topic modeling and visualization, and demonstrated the advantage of the framework in utilizing rich recipe information to analyze and interpret the culinary cultures from different regions.**

Manuscript received XX; revised November XX. This work was supported in part by the National Natural Science Foundation of China (61532018, 61602437, 61672497, 61572503 and 61620106003), in part by the Beijing Municipal Commission of Science and Technology (D161100001816001), in part by Beijing Natural Science Foundation (4174106 and 4152053), 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).

W. Min, Y. Zhu are with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China email:{weiqing.min, yaohui.zhu }@vipl.ict.ac.cn. S. Jiang is with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China, and also with University of Chinese Academy of Sciences, Beijing, 100049, China email: sqjiang@ict.ac.cn. B. Bao is with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with University of Chinese Academy of Sciences, Beijing, 100049, China email:bkbao@nlpr.ia.ac.cn. S. Mei is with Shandong University of Science and Technology, Shandong, 266590, China, and also with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China shuhuan.mei@vipl.ict.ac.cn. Y. Rui is with the Lenovo Group Ltd, Lenovo Corporate Research, Beijing, 100085, China, and also with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China email:yongrui@lenovo.com

I. Introduction

Understanding the cultural diversity has become imperative in almost every aspect of life. Cuisine has always been a significant aspect of cultures. The recipes from different cuisines are strong signals of the culinary habits of individuals from various parts of the world. For example, East Asian cuisines are dominated by some ingredients such as "soy sauce " and "sesame oil" from the recipes [2]. Some other ingredients uniquely represent a certain cuisine (e.g., mozzarella to the Italian cuisine) [51]. Therefore, the analysis of recipes can facilitate deep understanding of food from not only the health and marketing perspectives, but also the cultural one. Such analysis can further enable various applications, such as food preference learning [62], [27], cuisine classification [56] and health diet analysis [39]. Since the culinary habits have such importance for the culture, we address the topic of investigating and analyzing the culinary cultures of different countries through the recipes.

Existing work mainly focuses on the analysis of recipes based on the textual descriptions (e.g., ingredients) [51] and check-in information from social websites [53]. For example, Sajadmanesh *et al.* [51] used the ingredients and attribute information, such as the taste and cuisine information to understand cuisines and culinary habits around the world. Silva *et al.* [53] analyzed the check-ins from Foursquare to identify cultural boundaries and similarities across populations at different scales. However, little work has investigated how to jointly utilize rich modality and attribute information to enable the analysis and comparison of culinary cultures comprehensively.

While existing studies rely on text-oriented recipe analysis, we believe rich modality and attribute information are critical to comprehensively analyze the diversity of the culinary cultures. (1) Relying exclusively on text-based descriptions can impose high cognitive load in analyzing the culinary habits. Some works such as [37], [14], [62] have found the importance of visual information in eating habits-related applications. For example, Zepeda *et al.* [37] pointed out that photographic food diaries were more effective than written ones at prompting patients to truly understand their eating habits. Yang *et al.* [62] found that the analysis of image features can provide a valuable signal for food preference learning. (2) Different attributes (e.g., the course and cuisine information) reflect respective aspects of the recipes and jointly contribute to comprehensive recipe analysis. Through attribute-based analysis, we can reveal the culinary cultures from different perspectives. However, it is non-trivial to make

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

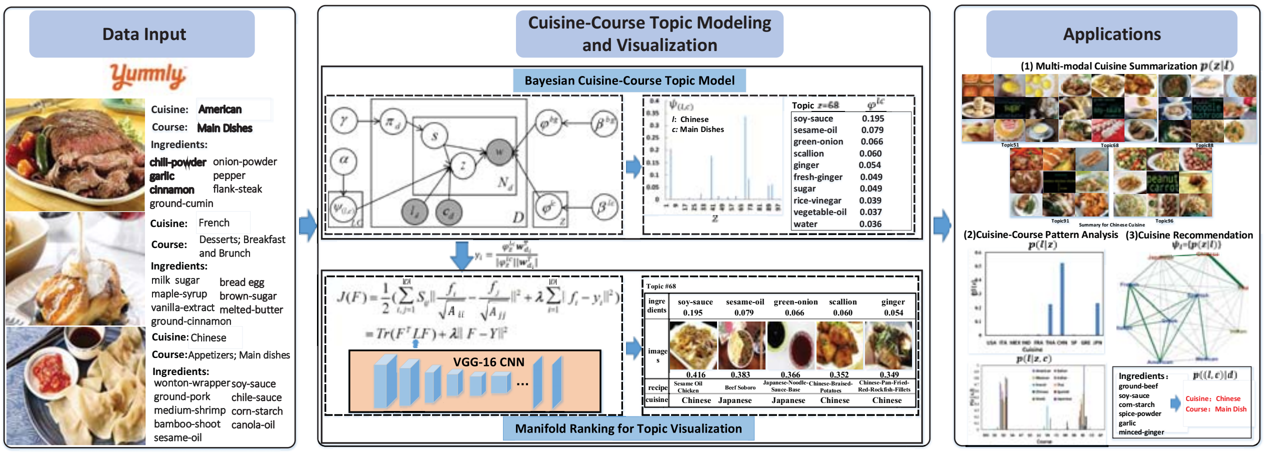


Fig. 1: The proposed culinary culture analysis framework.

the correlation between the content and various attributes. Because this requires us to design a more flexible model, which is able to incorporate arbitrary number of observed attribute information, yet inference remains relatively simple.

Taking all the above-mentioned factors into consideration, we propose a culinary culture analysis framework (Fig. 1), which is capable of discovering topics conditioned on different attributes and visualizing them to enable various applications. In particular, we take Yummly[1], one of the most popular recipe-sharing websites in our study. As shown in the left part of Fig. 1, each recipe includes the ingredients, food images and various attributes (e.g., the cuisine and course). Given the input of rich recipe information, we firstly propose a Bayesian Cuisine-Course Topic Model (BC^{2}TM), which incorporates different attributes to discover the cuisine-course specific topics. Since topic models have been proved successfully in discovering meaningful and interpretable topics or patterns in the text domain, such discovered topics are very suitable for our cross-region food analysis and comparison. For further analysis and better visualization, based on the learned model parameters, namely topic-ingredient distribution, we then utilize the manifold ranking method to integrate both deep visual features and topic-ingredients to retrieve relevant food images for topic visualization. Finally, we exploit the topic modeling and visualization for three applications: (1) Multi-modal cuisine summarization, which summarizes cuisines with both recipe ingredients and food images. Through cuisine summarization, we can analyze and interpret the culinary cultures from the macro level. (2) Cuisine-course pattern analysis including topic-specific cuisine distribution and cuisine-specific course distribution of topics. Such comparative analysis can provide us with more details from the micro level. (3) Cuisine recommendation. It supports flexible queries including both cuisine-oriented and ingredient-oriented queries. We conduct the experiment on one dataset, including about 66K recipes with the ingredients, food images, course and cuisine attributes from 10 cuisines from Yummly. Experiment results demonstrate the advantage of the framework in utilizing rich recipe information to discover and compare different eating habits from different regions.

[1]http://www.yummly.com/

The contributions of the proposed approach can be summarized as follows:

- To our knowledge, this is the first time to jointly utilize rich recipe information including multi-modal information and various attributes into a unified framework to enable comprehensive analysis and comparison of culinary cultures.
- We propose a culinary cultures analysis framework, which is capable of discovering topics conditioned on different attributes and visualizing them for recipe-oriented applications.
- We present a wide variety of applications, including 1) multi-modal cuisine summarization, 2) cuisine-course pattern analysis, and 3) cuisine recommendation.
- We conduct the comprehensive evaluation on a real-world recipe dataset Yummly-66K and the experimental results have validated the effectiveness of our proposed method and framework.

The rest of the paper is organized as follows. Section II reviews the related work. Section III presents the core components of the proposed culinary culture analysis, including the proposed Bayesian Cuisine-Course Topic Model (BC^{2}TM) and manifold ranking based topic visualization. Section IV introduces three derived applications, including multi-modal cuisine summarization, cuisine-course pattern analysis and cuisine recommendation. Experimental results are reported in Section V. Finally, we conclude the paper and give the future work in Section VI.

II. Related Work

Our work is closely related to four research fields: (1) topic modeling, (2) attribute based classification, recounting and summarization, (3) recipe analysis and recommendation, and (4) food image recognition.

A. Topic Modeling

The Probabilistic Topic Models (PTMs) aim to explore a set of topics from a large collection of documents, where a topic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

is a distribution over terms and a document is a distribution over topics. They have been successfully applied to various tasks such as information retrieval and recommendation [26], [42]. Because many documents are generally associated with various metadata, such as the location and the date, some extensions to the basic topic model Latent Dirichlet Allocation (LDA) [7] have been proposed to take these metadata into account [41], [47], [3], [40]. For example, Bauer *et al.* [3] incorporated the location and time information into account to learn different topic space. Qian *et al.* [47] added the collection information into the topic model. We also incorporate the metadata information into the topic model. The difference from them is that we incorporate the cuisine and course information into our topic model to enable the analysis of the recipe-oriented culinary culture. Furthermore, a topic visualization method is utilized to visualize discovered topics and facilitates the analysis and interpretation of discovered topics.

B. Attribute based Classification, Recounting and Summarization

As an effective mid-level representation, attributes have been widely used in many applications, such as zero-shot classification [60], [66], [34], recounting [64], [10], [36], [18] and summarization [32], [49], [50]. Zero-shot recognition is used to predict class labels for target domain instances based on the side information (e.g. attributes) of unseen classes in the source domain. For example, Zhang *et al.* [66] proposed a novel general probabilistic method for zero-shot recognition by learning joint latent similarity embeddings for both source and target domains. Xian *et al.* [60] presented a novel latent variable based model for learning a nonlinear compatibility function for zero-shot classification. Besides zero-shot classification, the attributes are useful for video recounting, which is used to explain why this video contains the desired event [64]. For example, Chang *et al.* [10] utilized the semantic evidence based on the concepts and attributes to simultaneously conduct event detection and recounting.

These above-mentioned two tasks focus on visual attribute learning from images or videos. In contrast, attribute based summarization is mainly to use rich metadata as attributes from the social media for visual summarization. For example, Kennedy *et al.* [32] used the location, tags and the visual features of images to generate diverse and representative landmark images. Similar to [32], Rudinac *et al.* [49] also utilized these rich metadata information to generate visual summarization for certain geographic area. Later, they [50] further incorporated the user preferences, such as the popularity and visual aesthetic appeal for visual summarization of image collections. We also use the attributes (cuisine and course) for cuisine summarization in our work, but with three differences. (1) Motivation. We aim to summarize cuisines for cross-region food analysis and comparison from the cultural aspect in the food domain. In contrast, these works mainly focus on visual summarization for travel or image organization. (2) Methodology. Considering the importance of ingredients for distinguishing different cuisines [2], [51], we first use the topic model to extract the patterns of the ingredients, and then select relevant food images to facilitate further analysis. However, these existing works focus on visual summarization. The text information and other attribute information are used to help visual summarization as the auxiliary information. (3) Results display. We show the summarized results in multi-modality by the topics while these works place the emphasis on visual summarized results and show these results with only images.

C. Recipe Analysis and Recommendation

There have been many efforts on the processing of recipes for the analysis of recipes. For example, Ahn *et al.* [2] constructed a data-driven flavor network relating ingredients together to capture the flavor compounds shared by culinary ingredients. Jain *et al.* [28] adopted this flavor network to further analyze culinary practices of specific cultures in the Indian cuisine. In addition, some works [44], [8] applied LDA to the recipes to find a set of ingredient bases. For example, Kusmierczyk *et al.* [33] used a topic model to mine and model online food content by combining text topics with related nutrient facts. Recently, Sajadmanesh *et al.* [51] presented a study of recipes including the ingredients and flavor information from Yummly to understand cuisines and culinary habits from different regions. In addition, Silva *et al.* [53] identified the cultural boundaries by analyzing the check-ins from Foursquare. Mejova *et al.* [39] analysed the metadata associated with images from Instagram to analyze the obesity patterns. Our work is different from them in that we provide a framework to utilize multi-modal (e.g., the ingredients and food images) and rich attribute information for comprehensive understanding of the culinary cultures from different regions.

Besides recipe analysis, some works [17], [35] modeled the hidden factors between users and ingredients for recommendation. For example, Lin *et al.* [35] proposed a content-driven matrix factorization approach to model the latent dimensions of recipes, users, and features. Ge *et al.* [19] leveraged a data set including users' ratings and tags, which signal the food's ingredients that the users like for recipe recommendation. In addition, Zhang *et al.* [65] used the check-ins to recommend the user a list of restaurants for his next dining. Our work is different from them in that we recommend the cuisine and the course information based on the ingredients on hand.

D. Food Image Recognition

A lot of works focus on food image recognition [16], [9], [30], [13]. For example, some works such as [55], [21] employed the deep network for food classification. Some works [6], [4], [61], [24], [23] developed restaurant-specific dish recognition. There are also some works on mobile food recognition [46], [31], [38], [15], [57], mobile food calorie estimation [45] and recording [1]. For example, Tanno *et al.* [57] used the multi-scale network-in-networks to extract the deep features for real time food recognition and implemented multi-threaded mobile applications on both iOS and Android mobile platforms. Recent work [11] proposed deep architectures for simultaneous learning of ingredient recognition and food categorization by exploiting the visual features, ingredients and image categories. In addition, many cross-modal models have been proposed for cross-modal retrieval, such as [12], [?], [29]. For example, Chen *et al.* [12] presented a stacked attention network for learning the commonality between the recipe image and ingredients to enable the cross-modal retrieval between recipes and recipe images. Recently, Salvador *et al.* [52] further proposed a neural network to find a joint embedding of recipes and images for the image-recipe retrieval task. In addition, they released a new large-scale, structured corpus of over 1m cooking recipes and 800k food images. In contrast, Min *et al.* [43] proposed a multitask deep belief network to enable both cross-modal recipe image retrieval given additional attributes and richer image annotations by inferring ingredients and attributes from the food images. Our work is different from [43] in that: (1) Motivation. [43] focuses on multi-modal learning and retrieval while we mainly explore rich recipe information for cross-region food analysis and comparison from the cultural aspect. (2) Methodology. [43] uses Restricted Boltzmann Machine (RBM) based undirected graph model while we adopt the topic model based directed graph model, which is more suitable for our task because of its interpretation. Rich *et al.* [48] performed the first large scale content analysis of food images from Instagram to understand the content of partially-labelled images taken in-the-wild. Different from their work, we focus on utilizing the food images to visualize the discovered ingredient topics for summarizing, analyzing and comparing the culinary habits from different countries.

In addition, there are some food datasets available, such as FoodCam-256 [31], ETHZ Food-101 [9], UPMC Food-101 [59], Geolocation-food [61], VIREO Food-172 [11] and Yummly-28K [43]. FoodCam-256 and ETHZ Food-101 are suitable for the task of food recognition, since there are food images with only the labels. In contrast to ETHZ Food-101, UPMC Food-101 contains additional text information, and can be used for multimodal food recognition. Geolocation-food[2] and VIREO Food-172[3] contain more metadata information, such as the geo-tag information and ingredient information. However, these recipes belong to the Chinese dishes and are not suitable for cross-region food analysis. Our previous work published a Yummly dataset Yummly-28K[4] for recipe retrieval and exploration. Based on this dataset, we further enlarge this dataset to larger scale Yummly-66K for cross-region food analysis.

III. Bayesian Cuisine-Course Topic Modeling and Visualization

In this section, we first propose a Bayesian Cuisine-Course Topic Model (BC^{2}TM) to learn cuisine-course specific topic space and then use the manifold ranking method for topic visualization.

A. Bayesian Cuisine-Course Topic Model (BC^{2}TM)

Intuitively, the ingredients in one recipe can be classified into two categories: 1) common ingredients (e.g., "water" and "flour"), and 2) ingredients related to the topic whose cuisine-course distribution we are interested in analyzing. Correspondingly, the proposed BC^{2}TM is to mine two kinds of topics: 1) Background topic, which represents more common ingredients; (2) Cuisine-Course topics, which are more related to certain cuisine and course. Based on the above concepts, the problem of BC^{2}TM can be defined as follows:

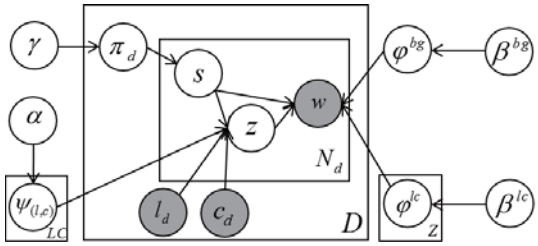


Fig. 2: Bayesian Cuisine-Course Topic Model (BC^{2}TM).

Given the recipe set D, which covers a set of ingredients W, a set of cuisines L and a set of courses C, each recipe d is associated with the cuisine l_d and the course c_d. The goal of BC^{2}TM is to learn (1) two kinds of topic spaces, background topic space φ^{bg} and cuisine-course topic space $\{\varphi_z^{lc}\}_{z\in Z}$, and (2) the distributions over cuisine-course topics $\{\psi_{(l,c)}\}_{(l,c)\in LC}$, where φ^{bg}, φ_z^{lc} and $\psi_{(l,c)}$ denote the background distribution of the corpus, the multinomial distribution over ingredients for cuisine-course topics, and the multinomial distribution over the cuisine-course topics for certain l and c, respectively. z and Z are the cuisine-course topic assignment and the set of cuisine-course topics. LC is the set of cuisine-course pairs and $LC = L \times C$.

The reasons for modeling the background topic are as follows (1) Different countries have different culinary habits, and thus have different ingredients from their recipes. However, there are some ingredients shared by all the countries. Therefore, we choose modeling these shared ingredients using the background topic for more accurate modeling of the recipe data. In the following experiment, we find that there are indeed some ingredients such as salt and water shared from all countries. (2) These ingredients frequently occurred in the recipes happens in many discovered cuisine-course topics with higher probability scores, and thus probably influence the interpretation of the learned cuisine-course topics. Therefore, we use the background topic to represent common ingredients and separate them from the cuisine-course topics to make discovered cuisine-course topics more interpretable.

1) Generative Process of BC^{2}TM: Given the recipe set D, each recipe $d \in D$ is a combination of (i) a background distribution over common ingredients φ^{bg} and (ii) a mixture distribution over cuisine-course topics Z. The switch variable s is sampled from multinomial distribution π_d, which represents the proportion distribution of two kinds of topics in one document. As presented in Fig. 2, the generative process of the BC^{2}TM model is shown in Alg. 1.

In Alg. 1, $\mathbf{w}_d$ denotes a recipe d with a sequence of N_d ingredients. $\mathrm{Multi}(\cdot)$, $\mathrm{Dir}(\cdot)$, $\mathrm{Bin}(\cdot)$ and $\mathrm{Beta}(\cdot)$ denote the Multinomial distribution, the Dirichlet distribution, the

[2]http://isia.ict.ac.cn/dataset/Geolocation-food/
[3]http://vireo.cs.cityu.edu.hk/VireoFood172/
[4]http://isia.ict.ac.cn/dataset/

Algorithm 1 Generative process of BC^{2}TM

```
draw a distribution over ingredients φ^bg ~ Dir(β^bg) for the background topic
for each cuisine-course topic z ∈ {1, ..., Z} do
    draw a multinomial distribution φ_z^lc ~ Dir(β^lc)
end for
for each cuisine-course pair (l, c) ∈ {1, ..., LC} do
    draw a multinomial distribution ψ_(l,c) ~ Dir(α)
end for
for each recipe d ∈ {1, ..., D} do
    draw π_d ~ Beta(γ)
    for each ingredient w_{d,n} ∈ w_d do
        draw a switch variable s_{d,n} ~ Bin(π_d)
        if s_{d,n} = bg do
            draw an ingredient w_{d,n} ~ Multi(φ^bg)
        else if s_{d,n} = lc do
            draw a topic z_{d,n} ~ Multi(ψ_(l_d,c_d))
            draw an ingredient w_{d,n} ~ Multi(φ^lc_{z_{d,n}})
        end if
    end for
end for
```

Binomial distribution and the Beta distribution, respectively. Similar to [7], we assume the symmetric Dirichlet priors, that is for $Dir(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_Z)$, $\alpha_z = \alpha$, $z \in \{1, ..., Z\}$, similarly for $Dir(\boldsymbol{\beta}^{bg})$ and $Dir(\boldsymbol{\beta}^{lc})$. We also assume the symmetric Beta prior $Beta(\boldsymbol{\gamma})$, $\boldsymbol{\gamma} = (\gamma^{bg}, \gamma^{lc})$, that is $\gamma = \gamma^{bg} = \gamma^{lc}$. $\boldsymbol{\beta}^{bg}$, $\boldsymbol{\beta}^{lc}$, γ, γ^{bg} and γ^{lc} are hyper-parameters.

2) Model Inference and Parameter Estimation: To estimate the parameters of BC^{2}TM, we need to estimate the latent variables conditioned on the observed variables, namely $p(\mathbf{s}, \mathbf{z}|\mathbf{w}, \alpha, \beta, \gamma)$, where $\beta = \{\beta^{bg}, \beta^{lc}\}$. we use the collapsed Gibbs sampling [20] for model inference, where the derivation of model inference is detailed in the appendix. After the Gibbs sampler reaches burn-in, we estimate the following parameters:

$$\begin{aligned}
\hat{\pi}_{d,s} &= \frac{n_{d,s} + \gamma}{n_{d,bg} + n_{d,lc} + 2\gamma} \qquad s \in \{bg, lc\} \\
\hat{\psi}_{(l,c),z} &= \frac{n^{lc}_{(l,c),z} + \alpha}{\sum_{z'} n^{lc}_{(l,c),z'} + Z\alpha} \\
\hat{\varphi}^{bg}_{w} &= \frac{n^{bg}_{w} + \beta^{bg}}{\sum_{w'} n^{bg}_{w'} + W\beta^{bg}} \\
\hat{\varphi}^{lc}_{z,w} &= \frac{n^{lc}_{z,w} + \beta^{lc}}{\sum_{w'} n^{lc}_{z,w'} + W\beta^{lc}}
\end{aligned} \tag{1}$$

where $n_{d,s}$ is the number of times of the words in the document d assigned to the background topics $s = bg$ and cusine-course topics $s = lc$, respectively. ${n^{lc}}_{(l,c),z}$ is the number of times that cuisine-course topic z is assigned to the cuisine-course (l, c). n^{bg}_{w} is the number of times that the word w is assigned to the background distribution. $n^{lc}_{z,w}$ is the number of times that the word w is assigned to the cuisine-course topic $s = lc$.

B. Topic Visualization

After parameter estimation, we obtain a set of cuisine-course topics as the ingredient bases in the recipe set. To visually enrich discovered ingredient bases for better interpreting what can we cook using them, we utilize estimated model parameters and visual information to retrieve relevant food images.

We resort to the manifold ranking [22] to retrieve relevant images. For manifold ranking, given a set of data, $\chi = \{\mathbf{v}_i\}_{i=1}^{|D|}$, where $|D|$ is the number of recipes, we build a graph G on the data. S denotes the adjacency matrix, where S_{ij} is the weight of the edge between node i and j. Let $F : \chi \rightarrow \mathcal{R}$ be a ranking function, which assigns each document $\mathbf{v}_i$ a ranking score f_i, $F = [f_1, ..., f_{|D|}]^\top$. In addition, we need to define an initial vector $Y = [y_1, ..., y_{|D|}]^\top$.

The cost function is defined as

$$J(F) = \frac{1}{2}(\sum_{i,j=1}^{|D|} S_{ij} \| \frac{f_i}{\sqrt{A_{ii}}} - \frac{f_j}{\sqrt{A_{jj}}} \|^2 + \lambda \sum_{i=1}^{|D|} \|f_i - y_i\|^2) \tag{2}$$

$$= \mathrm{Tr}(F^\top U F) + \lambda \|F - Y\|^2 \tag{3}$$

where $\lambda > 0$ is the regularization parameter. The diagonal elements of S, $S_{ii} = 0$. A is a diagonal matrix with $A_{ii} = \sum_{j=1}^{M} S_{ij}$. U is the symmetrical normalization of S, $U = A^{-1/2} S A^{-1/2}$. $\mathrm{Tr}(\cdot)$ is the trace of the matrix.

We finally solved the following optimization problem:

$$\min_{F} \{\mathrm{Tr}(F^\top U F) + \lambda \|F - Y\|^2\} \tag{4}$$

We iterate the following equation to minimize the cost function to obtain the final ranking score F^* [22].

$$F(t+1) = \rho U F(t) + (1 - \rho) Y \tag{5}$$

where $\rho = \frac{1}{1+\lambda}$.

Based on the set of deep features from the VGG deep network [54], we construct the visual correlation graph S, where each recipe image is one node i, with the representation $\mathbf{v}_i$. To effectively construct the edges e_{ij} between node i and other nodes, following the general idea of [5], we connect node i and its k visual neighbors to form the edges from node i and the weight of an edge is defined as the Gaussian kernel $S_{ij} = \exp(-\frac{d^2(\mathbf{v}_i, \mathbf{v}_j)}{2\sigma^2})$, where σ is a parameter and $d(\mathbf{v}_i, \mathbf{v}_j)$ is a Euclidean distance between $\mathbf{v}_i$ and $\mathbf{v}_j$.

For the initial ranking score, each topic has different proportions over different courses and cuisines. In order to retrieve more relevant food images, we first derive $p(l, c|z)$ as follows:

$$\begin{aligned}
p(l, c|z) = \frac{p(l, c, z)}{p(z)} &= \frac{p(l, z, c)}{\sum_{c' \in C} \sum_{l' \in L} p(z|c', l') p(c', l')} \\
&= \frac{\hat{\psi}_{(l,c),z} p(c, l)}{\sum_{c' \in C} \sum_{l' \in L} \hat{\psi}_{(c',l'),z} p(c', l')}
\end{aligned} \tag{6}$$

where $p(c, l)$ is the joint probability of the cuisine c and l, which is given by the ingredient count in cuisine l annotated by the course c divided by the total ingredient count in the collection.

We then select the set of cuisine-course pairs $o = \{(l, c)\}$

when $p(l,c|z) > \epsilon$. For each pair $(l,c) \in o$, the initialized ranking score is

$$y_{(l,c),i} = \begin{cases} \frac{\hat{\boldsymbol{\varphi}}_z^{lc} \mathbf{w}_{d_i}^\top}{|\hat{\boldsymbol{\varphi}}_z^{lc}||\mathbf{w}_{d_i}|} & (l,c) \in o \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{w}_{d_i}$ is represented as TF-IDF.

We use Eqn.4 to obtain $F^*_{(l,c)}$ and the final score is the average of all the $F^*_{(l,c)}$

$$F^* = \frac{1}{|o|} \sum_{(l,c)\in o} F^*_{(l,c)} \tag{7}$$

IV. Applications

A. Multi-modal cuisine summarization

An direct application is to summarize cuisines with ingredients and food images based on discovered topics and corresponding food images. A user (e.g., a foodie) would like to know more details of certain cuisine, like "What are the most representative recipe in Chinese food?" "Can you tell me a few distinct ingredients and representative food images in French?" To summarize representative aspects of a cuisine, we first derive a probability distribution over topics conditioned on certain cuisine l_q as follows:

$$p(z|l_q) = \frac{p(z,l_q)}{p(l_q)} = \frac{\sum_{c\in C} p(z,l_q,c)}{\sum_{c\in C} p(l_q,c)} = \frac{\sum_{c\in C} \psi_{(l_q,c),z} p(l_q,c)}{\sum_{c\in C} p(l_q,c)} \tag{8}$$

For the topics with higher scores, we then remove topics with higher similarity according to their Jensen-Shannon (JS) divergence $D_{JS}(z_i||z_j)$, where $D_{JS}(z_i||z_j) = \frac{1}{2}(D_{KL}(z_i||z_j) + D_{KL}(z_j||z_i))$ and $D_{KL}(\cdot||\cdot)$ denotes the Kullback-Leibler (KL) divergence. For the remaining topics, we select top-ranked ingredients according to $\hat{\varphi}^{lc}_{z,w}$ as the textual representation and retrieve relevant food images using Eqn. 4 as the visual representation. Through cuisine summarization, we can analyze the culinary culture from different regions at the macro level.

B. Cuisine-Course Pattern Analysis

Based on learned parameters from the BC^{2}TM, we can also compute various kinds of topic patterns. For example, we can derive the cuisine distribution on certain topic and the course distribution based on certain cuisine and certain topic.

For the cuisine distribution of topics, the cuisine distribution on certain topic can be computed according to the Bayes' theorem as

$$\begin{aligned} p(l|z) = \frac{p(l,z)}{p(z)} &= \frac{\sum_{c'\in C} p(l,z,c')}{\sum_{c'\in C}\sum_{l'\in L} p(z|c',l')p(c',l')} \\ &= \frac{\sum_{c'\in C} \hat{\psi}_{(l,c'),z} p(c',l)}{\sum_{c'\in C}\sum_{l'\in L} \hat{\psi}_{(c',l'),z} p(c',l')} \end{aligned} \tag{9}$$

With the corresponding topic and cuisine constraint, we then use Eqn.4 to show relevant food images to further facilitate our analysis.

In addition, given one cuisine, we can calculate this from the Bayes' theorem in a similar way by

$$p(c|z,l) = \frac{\hat{\psi}_{(l,c),z} p(c,l)}{\sum_{c'\in C} \hat{\psi}_{(l,c'),z} p(c',l)} \tag{10}$$

Based on $p(l|z)$, $p(c|z,l)$ and retrieved food images, various cuisine-course patterns can be discovered and analyzed. Through these distributions, we can analyze the culinary culture from different regions at the micro level.

C. Cuisine Recommendation

A user generally has some preferences about some cuisines. There are two frequently asked questions :

- Being similar to a given cuisine. "I quite enjoyed the Chinese food. Are there any other cuisines with similar style?"
- Being relevant to a given cooking. "I have some ingredients at hand, what kind of cuisine and course do I can cook?"

For the first question, we firstly define the distance between different cuisines. For certain query cuisine l_q, According to Eqn 8, we obtain the representation for certain cuisine l as

$$\psi_l = \{p(z|l)\}_{z\in Z} \tag{11}$$

The symmetric similarity between two cuisines is measured by the distance between their corresponding multinomial distributions over topics through the following Gaussian kernel

$$\mathrm{CuiSim}(l_i,l_j) = \exp\{-\frac{D_{JS}(\psi_{l_i}||\psi_{l_j})}{2\sigma^2}\} \tag{12}$$

The larger the $\mathrm{CuiSim}(l_i,l_j)$, the larger the similarity.

Given a query cuisine l_q and a candidate destination set L, each cuisine $l \in L$ has a similarity to l_q in the cuisine topic space, defined as $\mathrm{CuiSim}(l_q,l_j)$. The ranking score for recommendation is computed according to Eqn.12

For the second question, we consider the query document d_q. We also need a representation and corresponding similarity metric of d_q, so as to measure the relevance of a $(\text{cuisine}, \text{course})$ pair to given query ingredients. We compute the representation $p(z|d_q)$ as a probability through the Gibbs sampler by maximizing the following likelihood function

$$p(\mathbf{w}_{d_q}|D_{train}) = \prod_{i=1}^{N_{d_q}} [\pi_{d_q,bg}\hat{\varphi}^{bg}_{w_i} + \pi_{d,lc}\sum_{z\in Z} p(z|d_q)\hat{\varphi}^{lc}_{z,w_i}] \tag{13}$$

When a query document d_q is observed, the learned parameters from the training set $\hat{\varphi}^{bg}_{w_i}$ and $\hat{\varphi}^{lc}_{z,w_i}$ are fixed. A Gibbs sampler is run to update the document-specific parameters $p(z|d_q)$,$\pi_{d_q,bg}$ and $\pi_{d_q,lc}$.

We then compute $p((l,c)|d_q)$ as follows:

$$p((l,c)|d_q) = \sum_z p((l,c)|z)p(z|d_q) \tag{14}$$

where $p((l,c)|z)$ is computed through Eqn.6.

Finally, we rank the results according to Eqn.14.

V. Experiment

In this section, we firstly describe the experimental setting including the dataset and implementation details. We then

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

evaluate the performance of the proposed BC^{2}TM qualitatively and quantitatively. Next, we evaluate the performance of topic visualization. Finally, we evaluate the performance of the proposed three applications including multi-modal cuisine summarization, cuisine-course pattern analysis and cuisine recommendation, respectively.

A. Yummly-66K dataset

The experiment is conducted on 10 different cuisines, including *American*, *Italian*, *Mexican*, *Indian*, *French*, *Thai*, *S-panish*, *Chinese*, *Greek*, *Japanese* from different regions. Note that we will use these country names and their corresponding abbreviations[5] interchangeably for convenience. Our dataset consists of two parts. One part is from our previous work [43], including 27,638 items, where each recipe item includes the food image, the recipe name, the ingredients, cuisine and course information. For the other part, we used the URLs in Yummly provided by [51] to crawl corresponding images. The crawled images and other recipe information provided by [51] from these 10 cuisines constitutes the second part of our dataset. We adopt the preprocessing method [43] to preprocess each ingredient line. After processing, the vocabulary of ingredients is 2,416. For the ingredients with more than two words, we represent them by concatenating them using '-'. For example, we use soy-sauce to represent the ingredient "soy sauce". Each recipe item includes the recipe name, preprocessed ingredient line, recipe image, cuisine and course attribute information. There are 14 kinds of course attributes including Main Dishes, Desserts, Side Dishes, Salads, Afternoon Tea, Soups, Lunch, Snacks, Condiments and Sauces, Breads, Breakfast and Brunch, Beverages, Cocktails, Appetizers[6]. Similarly, we use these course name and their abbreviated names (MD, DE, SD, SA, AT, SO, LU, SN, CS, BR, BB, BE, CO and AP) interchangeably for convenience. There are 66,615 recipe items in all, which we denote as the Yummly-66K dataset. The dataset is available[7]. The statistics of the resulting dataset are shown in Table I. Fig. 3 shows some examples.

TABLE I: The statistics of the collected cuisines.

Cuisine	♯items	Cuisine	♯items
American	13,262	Italian	9,401
Greek	4,998	Japanese	4,804
Mexican	7,960	Indian	5,470
French	6,173	Spanish	4,014
Thai	5,282	Chinese	5,251

[5]The corresponding abbreviations are USA, ITA, MEX, IND, FRA, THA, ES, CHN, GRE, JPN.

[6]There are 13 kinds of supported courses in Yummly and consider Lunch and Snacks as one kind of course attribute. However, we found that many recipe items are either only annotated by Lunch or by Snacks. Therefore, similar to [27], we consider Lunch and Snacks as two kinds of attributes. If one recipe item is annotated as Lunch and Snacks, we think they belong to both Lunch course and Snack course.

[7]http://isia.ict.ac.cn/dataset/Yummly-66K.html

Fig. 3: Some examples from different cuisines.

B. Implementation Details

For each image, we represent the visual content by deep learning features. Similar to [25], [43], we first use the food101 dataset [9] to finetine the VGG-16 deep network [54] and then compute the 4,096 dimensional vector representation for each image in Yummly-66K by extracting the network activations in the final fully-connected layer.

As for the hyper-paramerters of the model, without any prior knowledge, we empirically set the fixed values, i.e., $\alpha = 1.0/Z$, Z is the number of topics. $\beta^{bg} = \beta^{lc} = 0.01$, $\gamma = 1.0$. In addition, as shown in Fig. 3, each recipe is associated with one cuisine, but probably with more than one course information. Similar to [63], we consider such recipe as multiple items, where each item is associated with one cuisine and one course.

For recipe image retrieval, the parameter σ of the gaussian kernel function is empirically set as 0.15 while in cuisine recommendation application, the parameter σ of the gaussian kernel function is 0.5. The number of visual neighbors is set to be 100. The tolerance of the iterations for optimization in manifold ranking is $1e-7$. The parameter ρ is empirically set as 0.50. The threshold ϵ for selecting the set of cuisine-course pairs is empirically set as 0.001. For our topic model, we run the training procedure with 3,000 iterations.

C. Evaluation of BC^{2}TM

1) Quantitative Evaluation: To demonstrate the effectiveness of the proposed BC^{2}TM model, we compare it with the following two baselines:

- BC^{2}TM_NC. This baseline is similar to our method, but does not consider the course information.
- BC^{2}TM_NB. This baseline is similar to our method, but does not consider the background distribution.

We resort to the *perplexity* [7] as the evaluation metric, which is a standard measure for estimating how well one generative model fits the data. The lower the perplexity score is, the better the performance. In BC^{2}TM, the perplexity of a test set of ingredient descriptor $\mathbf{w}_d$ is defined as follows:

$$perplexity(D_{test}) = \exp(-\frac{\sum_{d\in D_{test}} \log p(\mathbf{w}_d|D_{train})}{\sum_{d\in D_{test}} N_d}) \tag{15}$$

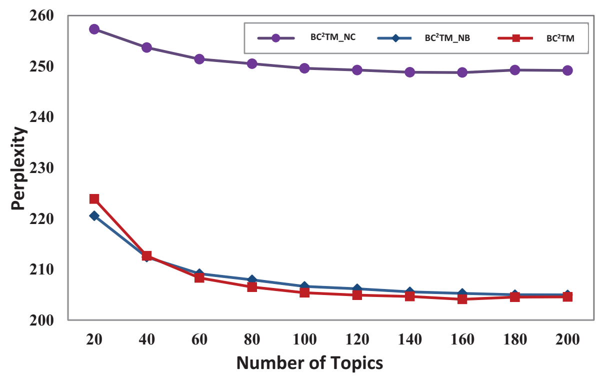


Fig. 4: The perplexity score over different topic number for different models.

where

$$
\begin{aligned}
p(\mathbf{w}_d|D_{train}) &= p(\mathbf{w}_d|\boldsymbol{\pi}, \Psi^{lc}) \\
&= \prod_{i=1}^{N_d}[\pi_{d,bg}\hat{\varphi}^{bg}_{w_i} + \pi_{d,lc}\sum_{z\in Z}\psi_{(l_d,c_d),z}\hat{\varphi}^{lc}_{z,w_i}]
\end{aligned} \quad (16)
$$

where D_{train} is the training set and D_{test} is the test set. N_d is the count of ingredients in the document d.

When the words of a test document d is observed, the parameters $\hat{\varphi}^{bg}_{w_i}$ and $\hat{\varphi}^{lc}_{z,w_i}$ learned from the training set are fixed. A Gibbs sampler is run on the observed words to update the document-specific parameters $\psi_{(l_d,c_d)}$, $\pi_{d,bg}$ and $\pi_{d,lc}$. These updated parameters are used in the computation of perplexity. We divide our dataset into two parts: the training set and the test set. For each cuisine, we randomly select 80% of the recipe items as the training data and the remaining 20% of the recipe items as the test data. We set the number of topics $Z \in \{20, 40, 60, 80, 100, 120, 140, 160, 180, 200\}$.

The perplexity scores over different topic number for different models are shown in Fig. 4. We can see that (1) The perplexity score of both BC^{2}TM and BC^{2}TM_NB is much lower than BC^{2}TM_NC. That means these two models are better than BC^{2}TM_NC. The probable reason is that the added course information provides a better prior for the content of the ingredient document. This attribute information can help improve the generalization performance of the model.(2) the BC^{2}TM model has lower perplexity scores than BC^{2}TM_NB when $Z > 40$. This indicates that the model is using the background words "route" to better learn predictive models for recipe items. (3) It is clear that the perplexity decreases much slower when $Z >= 100$. Since larger topic number requires more computation cost, we choose the desired topic number Z to be 100 for our model in the following experiments. It takes about 24 minutes with 3000 iterations and 100 topics to converge when training the topic model using the Gibbs sampling on our dataset.

2) Qualitative Evaluation: We further demonstrate the effectiveness of BC^{2}TM by providing some example topics. These examples are shown in Table II, where "#j" denotes the j-th topic index. Each topic is represented by 10 top-ranked ingredients, sorted by their topic-ingredient distributions $\hat{\varphi}^{lc}$. As shown in Table II, we can see that these topics have a reasonable physical interpretation as the ingredient bases. For example, some topics correspond to a particular dish type. Topic #2 is a whole lot of spices. Topic #12 represents the ingredients of a cake. Topic #36 probably denotes the Italy-style ingredients. Topic #43 is a typical kind of drink. Topic #68 with the soy-sauce and sesame-oil gives it an oriental taste because Asians such as Chinese use these plant derivatives widely in their food and other countries do not [2]. Topic #88 is the ingredient set for the noodles. Topic #89 is a kind of fruit drink. Topic #93 is the Spanish-style ingredients. In addition, we show the background distribution learned by our model in Fig. 5. We can see that these ingredients such as salt, egg and water are very common and shared by most recipes from different cuisines.

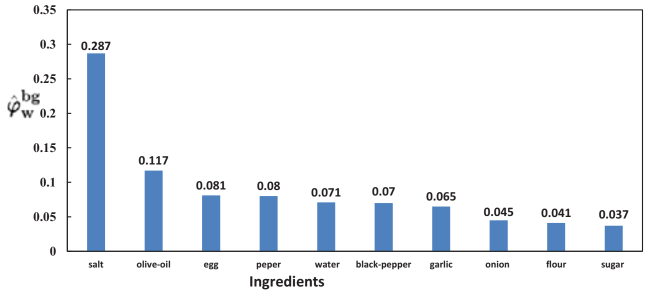


Fig. 5: The background distribution (10 top-ranked ingredients) learned by BC^{2}TM.

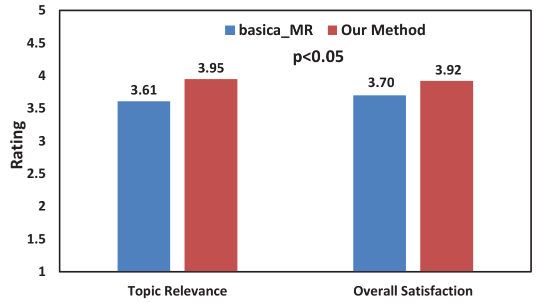


Fig. 6: A subjective evaluation of topic visualization generated by our method and the baseline.

D. Evaluation of Topic Visualization

Since it is usually subjective to evaluate the extent to which a recipe image with ingredients is relevant to one topic, we resorted to an user study to evaluate the relevance of retrieved food images. As a comparison, we implemented a baseline method, basic manifold ranking (basic_MR), which does not consider the proportion over different courses and cuisines $p(l, c|z)$. Thirty participants (twenty graduate students and ten IT workers) were asked to evaluate the relevance to the topic using 1 to 5 ratings, from two aspects namely (1) topic relevance (i.e., to what extent the retrieved food images with associated ingredients are describing this topic), and (2) overall satisfaction. Whether the proposed method can retrieve food images not only relevant to the topic but also informative and comprehensive.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

TABLE II: Some examples of discovered topics.

Topic ♯2	Topic ♯12	Topic ♯36	Topic ♯68	Topic ♯72
ground-cumin 0.106	butter 0.074	ricotta-cheese 0.066	soy-sauce 0.195	milk 0.204
chili-powder 0.074	flour 0.062	mozzarella-cheese 0.059	sesame-oil 0.079	honey 0.062
fresh-cilantro 0.067	ground-cinnamon 0.056	spaghetti 0.059	green-onion 0.066	cinnamon 0.053
vegetable-oil 0.046	vanilla-extract 0.042	ground-beef 0.046	scallion 0.060	vanilla-extract 0.043
sour-cream 0.043	salted-butter 0.039	italian-sausage 0.043	ginger 0.054	lemon 0.033
dried-oregano 0.041	milk 0.037	lasagna-noodle 0.042	fresh-ginger 0.049	ground-cinnamon 0.031
corn-tortilla 0.041	sugar 0.037	tomato-sauce 0.037	sugar 0.049	whipped-cream 0.026
diced-tomato 0.037	chopped-pecan 0.033	parmesan 0.033	rice-vinegar 0.039	brown-sugar 0.025
bell-pepper 0.030	sour-cream 0.029	pasta-sauce 0.024	vegetable-oil 0.037	orange 0.023
canola-oil 0.023	vegetable-oil 0.029	garlic-clove 0.023	water 0.036	coconut-milk 0.018
Topic ♯73	**Topic ♯88**	**Topic ♯89**	**Topic ♯91**	**Topic ♯93**
egg 0.107	noodle 0.059	mango 0.093	boneless-skinless-chicken 0.236	spanish-paprika 0.120
milk 0.099	mushroom 0.046	coconut-milk 0.085	chicken 0.102	pimento 0.058
bread 0.096	cabbage 0.035	milk 0.080	chicken-breast 0.089	smoked-paprika 0.048
cinnamon 0.072	sprout 0.034	ice-cube 0.0691	bell-pepper 0.040	green-olive 0.035
butter 0.063	ground-pork 0.030	water 0.064	onion 0.035	spanish-onion 0.035
vanilla-extract 0.050	star-anise 0.030	tea 0.0377	chicken-thigh 0.035	olive 0.034
maple-syrup 0.039	baby-bok-choy 0.029	coconut 0.034	boneless-chicken-breast 0.029	manchego-cheese 0.032
brown-sugar 0.035	bok-choy 0.028	thai-basil 0.023	shrimp 0.028	spanish-smoked-paprika 0.030
ground-cinnamon 0.035	chicken 0.026	sticky-rice 0.022	ground-pepper 0.027	ground-cumin 0.029
heavy-cream 0.022	firm-tofu 0.025	banana 0.021	beef 0.018	canola-oil 0.027

We randomly select 20 topics as the query topics. For each topic, we show 15 top-ranked food images for each method. Each recipe image is associated with the ingredients, cuisine, course and recipe names. We averaged all the participators evaluations for the query topics as the ratings of all the methods. The results are shown in Fig. 6. As depicted in Fig. 6, our method shows advantages in both aspects due to the constraint of the proportion over different courses and cuisines $p(l, c|z)$, and the differences are significant for both two aspects (two sided sign test: $p < 0.05$).

We further verify the effectiveness of retrieval method and find what kind of cuisine we can do using these discovered topics. These examples are shown in Fig. 7(b). We use 5 top-ranked cuisine images from the dataset for each topic. Meanwhile, we label discovered images with their associated cuisines and recipe names. As shown in Fig. 7(b), we can see some interesting results. For example, as mentioned above, Topic #68 with the soy-sauce is an oriental taste and It is suitable for the Asian cuisine, such as Chinese and Japanese. The ingredients of Topic #91 belong to the meat. From these visualized results, we can easily interpret discovered topics. The ingredients of Topic #93 is suitable for Spanish cuisine, such as "Pork Loin Roast with Paprika" and "Slow Cooker Garlicky Shrimp".

In addition, for cuisine or course specific topics, our method can retrieve more relevant food images, such as Topic#88. This topic is more relevant to Chinese cuisine. Therefore, the top ranked food images from our method all belong to the Chinese cuisine. In contrast, the retrieval results from basic_MR contain images from other cuisines. In addition, the Topic#93 is more relevant to the Spanish cuisine and Main (or Lunch) course. Therefore, our method can retrieve food images about meat, which are more reasonable than basic_MR.

E. Multi-modal cuisine summarization

Based on the proposed framework, we can summarize cuisines using discovered topics and retrieved food images. For each cuisine, we finally empirically select five relevant topics. For each topic, we show 5 top-ranked ingredients and 8 top-ranked food images. Two example cuisine summaries are illustrated in Fig. 8 for Chinese and Italian, respectively. For each topic, the center position is the ingredients, generated by wordcloud[8]. From such summaries, users could easily understand the culinary habits and local characteristics of one region, from a representative and comprehensive overview with both ingredient and visual information. For example, in Chinese cuisine, their ingredient base often includes the plant derivatives such as soy-saurce and sesame-oil. They often use them when they eat dumplings. In order to compare different cuisines in details, we next analyze cuisine-course patterns under different cuisines and course.

F. Cuisine-Course Pattern Analysis

To evaluate the cuisine distribution of certain topic, we show some interesting examples of three topics: Topic#61, Topic#68 and Topic#89 in Fig. 9. Fig. 9(a) shows that there is the highest probability in Italian. As mentioned before, Topic#61 is a typical of Italian-style ingredient base. Therefore, the result is reasonable. Similarly, the ingredient base Topic#68 with the soy-sauce is a dish with the oriental taste because Asians use soy-sauce widely in their food and other ethnic groups do not. From Fig. 9(b), we can see that the top-3 cuisines with the higher probability is indeed Asian cuisines, Chinese, Japanese and Thai. Since the soy sauce originated in China[9], there is no doubt that there is the highest probability in Chinese cuisine. Japan also likes the dish with the soy-sauce. This is because they think umami is as important as sweetness or saltiness and soy sauce is the prime example of umami taste[10]. The soy-sauce is also as one of ten basic sauces in Thai. In order to facilitate further analysis, we retrieve relevant food images with the recipe name using $p(w|z, l)$ and Eqn.4. As shown in Fig. 10, for each cuisine, we show top five food images. We can see that the home cooking of Chinese cuisine such as the chicken and potatoes use the seasonings of Topic#68 including the soy-sauce, sesame oil, grinder, vinegar and so

[8]https://github.com/amueller/word_cloud
[9]https://en.wikipedia.org/wiki/Soy_sauce
[10]http://www.japan-talk.com/jt/new/shoyu

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

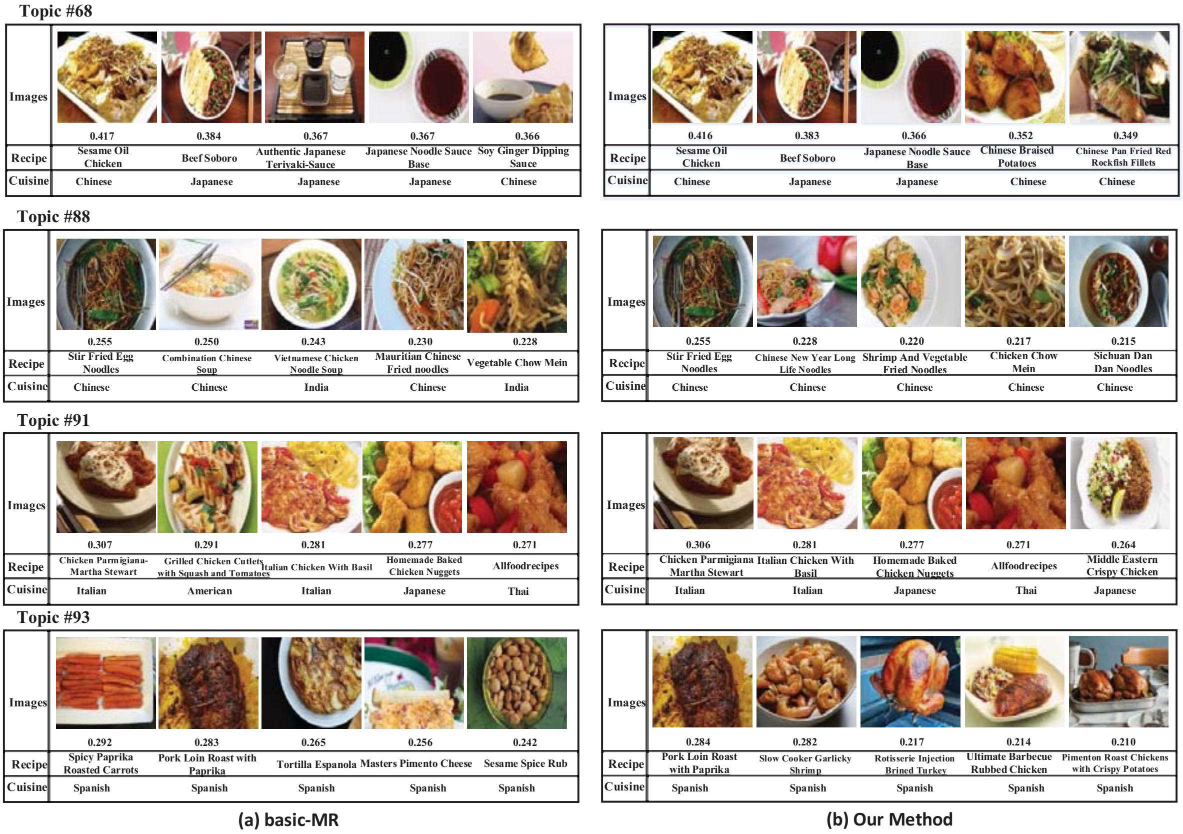


Fig. 7: Some examples of topic visualization, where top five images are retrieved by (a) basic_MR and (b) our method.

Fig. 10: Visualized results of Topic #68 for top 3 ranked cuisines.

on. We can find that when Chinese eat dumplings, they often use the soy-sauce, vineries and other ingredients as the sauce base. In addition, from Fig. 10, we can find that Asians use the ingredients from Topic#68 when eating noodles, such as Japanese and Thai.

In addition, we can also plot the course distribution from different cuisines given one topic according to Eqn.10 to analyze meal-cuisine relationships in detail. Fig. 11 shows two examples of topics: Topic#72 and Topic#91. Because Topic#72 includes milk, lemon and other drink ingredients, it is clear that such ingredient base is fit for the Side Dish and Beverages, which is consistent with Fig. 11(a). We can also find French use these ingredients for the Snacks while Beverages for Chinese cuisine. Topic #91 is about chicken, it is obvious that such ingredients mainly belong to the Main Dishes and Lunch, which is also consistent with the discovered results in Fig. 11(b).

G. Cuisine Recommendation

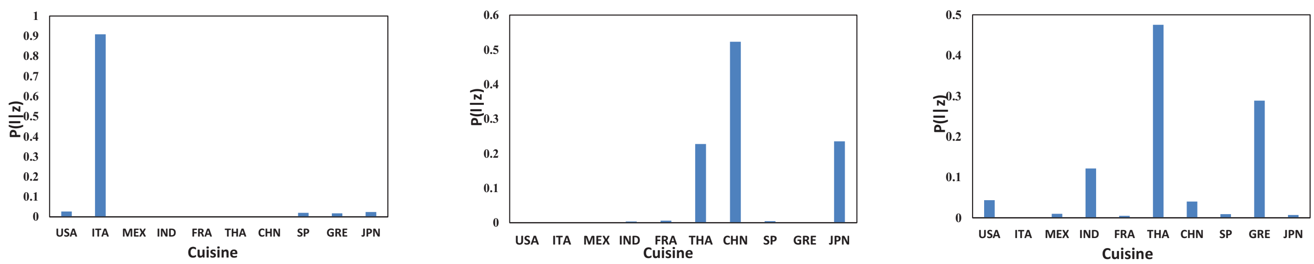
Since the effectiveness of similarity-oriented cuisine recommendation highly relies on the pair-wise similarity metric of different cuisines, we directly evaluate this metric's capability of discovering cuisines from a given set. we computed pairwise cuisine similarities of all the 10 cuisines in our database using Eqn.12 to form a cuisine similarity graph and use the Grephi software[11] to visualize this graph. Nodes with the same color are those that belong to the same group, clustered by the K-means algorithm and the thickness of the edges indicates the pairwise similarity, as shown in Fig. 12.

We find the following interesting results: (1) the geographical distance between countries plays an important role in determining the similarity of the eating habits. For example, some Asian countries, such as Chinese, Japanese and Thai are more similar than other countries. Indeed, historically there were more trade and cultural exchanges between spatially-neighboring countries, so they would share more similar eating

[11] https://gephi.org/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

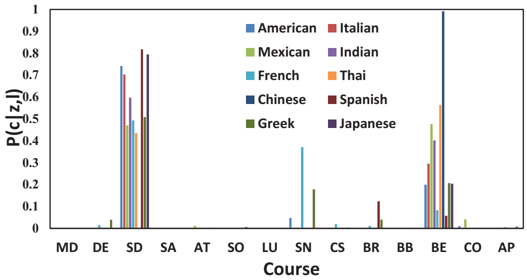
Fig. 8: An illustration of the generated cuisine summaries for (a) Chinese cuisine and (b) Italian cuisine, each composed of 5 topics including 5 top-ranked ingredients (the center position of each topic) and 8 top-ranked images for each topic. In the summaries, the larger an ingredient is, the larger the probability is.



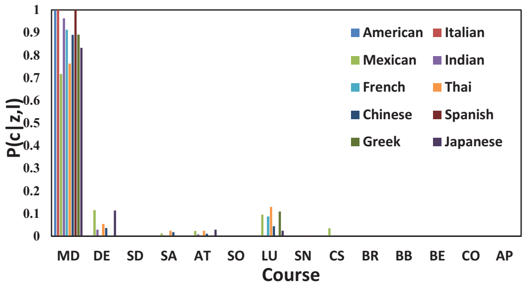
(a) Topic#61 (b) Topic#68 (c) Topic#89

Fig. 9: Cuisine distribution conditioned on Topic #61, Topic #68 and Topic #89.

habits in their food culture. (2) Due to the similarity of cultures in Europe and America, it can bee seen that ingredients from these regions are categorized as one cluster. (3) There is an outlier detected in our result, Indian. This is because the eating habits of Indian is quite different from other countries of the Asian countries. Based on the analysis, these different categories of cuisines are roughly differentiated by our similarity metric. his verified the effectiveness of our BC^{2}TM model. In addition, we found that the work [51] also reported their cuisine similarity based on two kinds of ingredient-based similarity, namely TF-IDF and Jensen-Shannon divergence. We compared our cuisine-similarity graph with theirs and found that our cuisine similarity graph is basically consistent with their results. However, our proposed framework can also



(a) Topic#72



(b) Topic#91

Fig. 11: Course distribution conditioned on Topic #72, Topic #91 for different cuisines.

Fig. 12: The cuisine similarity graph generated by our model. Each node indicates one cuisine. Nodes with the same color are those that belong to the same group and the thickness of the edges indicates the pairwise similarity.

be applied to multi-modal cuisine summarization and the ingredient topic discovery.

For ingredient-relevant cuisine and course recommendation, our task is to recommend both relevant cuisine and course information given the ingredients. We split the dataset into the training set and test set, as the same split in quantitative evaluation of BC^{2}TM. Each recipe item probably has more than one label. For example, the last example of Fig. 3 consist of two kinds of labels (Chinese,Appetizers) and (Chinese,Main Dishes). Therefore, we consider this task as the annotation task. Similar to [58], we use the top-N F1-measure, denoted

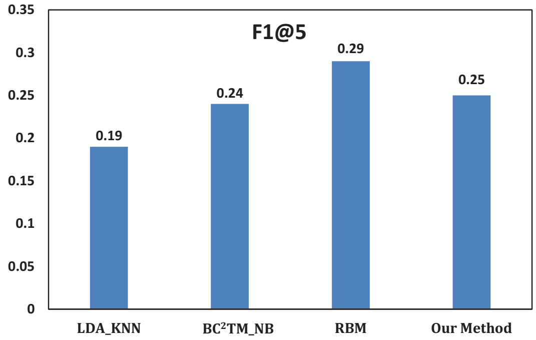


Fig. 13: Top-5 F1-measure for cuisine and course recommendation.

as F1@N, where we set $N = 5$ as the evaluation metrics.

$$F1@N = 2 * \frac{precision@N * recall@N}{precision@N + recall@N} \quad (17)$$

We consider the following methods for comparison.

- LDA-Based K-Nearest Neighbors (LDA_KNN). LDA_KNN retrieves all recipe items to find the nearest neighbors by computing the cosine similarity between the query's text topic vector representation and the topic vector from the training set.
- BC^{2}TM_NB. This baseline is similar to our method, but does not consider the background distribution.
- RBM: Restricted Boltzmann Machine [43]. Compared with directed topic model, this work used another kind of graph model, namely undirected RBM based graph model. In this work, we consider the ingredients to recommend cuisines and courses and use a standard binary RBM with 100 hidden layers, which is the same as the number of topics in our topic model. Based on the learned latent features h, we use two softmax classifiers to predict the cuisine score $p(l|h)$ and the course score $p(c|h)$, respectively.We finally obtain the final score $p(l, c|h) = p(l|h)p(c|h)$.

The evaluation results are shown in Fig. 13. We can see that our method outperforms LDA_KNN and BC^{2}TM_NB, mainly because of its good latent representations on cuisine-course information. Also on this dataset, the performance of RBM is the best in all the methods. However, such methods are not suitable for our previous two applications based on discovered interpreted topics.

VI. Conclusions

This paper presented a culinary culture analysis framework, which is capable of jointly utilizing the ingredients, food images and various attributes to enable various applications. We proposed a Bayesian Cuisine-Course Topic Model (BC^{2}TM) to incorporate the cuisine and course information to discover cuisine-course topics. A manifold ranking method is utilized to retrieve topic-relevant food images for topic visualization. We applied the topic modeling and visualization method into three applications, including multi-modal cuisine summarization, cuisine-course pattern analysis and cuisine recommendation. Through our proposed framework, we indeed found some interesting results, which reveal the diversity of the culinary cultures. For example, some topics such as the one with the ingredients soy-sauce and sesame-oil exclusively belong to some regions in Asia, such as Chinese, Thai and Japanese. As an outlier, the eating habits of Indian is quite different from other countries.

This work can be extended in the following four directions: 1) Exploring more recipe elements from Yummly for supporting more applications. For example, we can utilize the taste information associated with one recipe in Yummly to model the correlation between the taste and multi-modal information for taste prediction. 2) Enlarging our multi-modal recipe dataset to deeply understand cuisines and culinary cultures around the world. 3) We have analyzed the cuisine similarity presented in the undirected graph. As the future work, it is very interesting to find a way to turn that into a directed graph, which showcases the cuisine to cuisine influence. 4) The fourth direction is on a method dimension for better multi-modal modeling on the recipe data. For example, recent work such as [11][12] has proposed multi-modal deep learning method such as CNNs and the attention network for cross-modal recipe retrieval and recognition, which can be utilized for better modeling of rich recipe information.

Appendix A

For BC^{2}TM, we use the collapsed Gibbs sampling [20] for model inference, that allows sampling of z_i and s_i alternatively for each word token w_i,

(i) sample $\mathbf{s}$ given the current estimate of $\mathbf{z}$

$$p(s_i = bg|w = w_i, \mathbf{s}_{\neg i}, \mathbf{z}_{\neg i}, \mathbf{w}_{\neg i}) \propto \frac{n^{bg}_{w_i,\neg i} + \beta^{bg}}{\sum_{w'} n^{bg}_{w',\neg i} + W\beta^{bg}} \frac{n_{d,s_i,\neg i} + \lambda}{n_{d,bg,\neg i} + n_{d,lc,\neg i} + 2\lambda} \tag{18}$$

$$p(s_i = lc|z_i, w = w_i, \mathbf{s}_{\neg i}, \mathbf{z}_{\neg i}, \mathbf{w}_{\neg i}) \propto \frac{n^{lc}_{(l_d,c_d),z_i,\neg i} + \alpha}{\sum_z n^{lc}_{(l_d,c_d),z,\neg i} + Z\alpha} \frac{n_{d,s_i,\neg i} + \lambda}{n_{d,bg,\neg i} + n_{d,lc,\neg i} + 2\lambda} \tag{19}$$

(ii) sample $\mathbf{z}$ given the current estimate of $\mathbf{s}$

$$p(z_i|s_i = lc, w = w_i, \mathbf{z}_{\neg i}, \mathbf{s}_{\neg i}, \mathbf{w}_{\neg i}) \propto \frac{n^{lc}_{z_i,w_i,\neg i} + \beta^{lc}}{\sum_{w'} n^{lc}_{z_i,w',\neg i} + W\beta^{lc}} \frac{n^{lc}_{(l_d,c_d),z_i,\neg i} + \alpha}{\sum_{z'} n^{lc}_{(l_d,c_d),z',\neg i} + Z\alpha} \tag{20}$$

where $i = (d, n)$ is the current index. the superscript $\neg i$ denotes a counting variable that excludes the i-th word index in the corpus. $n^{lc}_{(l_d,c_d),z_i,\neg i}$ is the number of times that cuisine-course topic z_i is assigned to the document d associated with cuisine-course (l_d, c_d). $n^{bg}_{w_i,\neg i}$ is the number of times that the word w_i are assigned to the background distribution. $n^{lc}_{z_i,w_i,\neg i}$ is the number of times that the word w_i is assigned to the cuisine-course topic z_i. $n_{d,s_i,\neg i}$ is the number of times of the words in the document d assigned to the background distribution $s_i = bg$ and cusine-course topics $s_i = lc$, respectively.

References

[1] Z. Ahmad, M. Bosch, N. Khanna, D. A. Kerr, C. J. Boushey, F. Zhu, and E. J. Delp. A mobile food record for integrated dietary assessment. In *International Workshop on Multimedia Assisted Dietary Management*, pages 53–62, 2016.

[2] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási. Flavor network and the principles of food pairing. *Scientific reports*, 1, 2011.

[3] S. Bauer, A. Noulas, D. O. Seaghdha, S. Clark, and C. Mascolo. Talking places: Modelling and analysing linguistic content in foursquare. In *IEEE International Conference on Social Computing*, pages 348–357, 2012.

[4] O. Beijbom, N. Joshi, D. Morris, S. Saponas, and S. Khullar. Menu-match: restaurant-specific food logging from images. In *Applications of Computer Vision, IEEE Winter Conference on*, pages 844–851, 2015.

[5] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7(1):2399–2434, 2006.

[6] V. Bettadapura, E. Thomaz, A. Parnami, G. D. Abowd, and I. Essa. Leveraging context to support automated food recognition in restaurants. In *Applications of Computer Vision, IEEE Winter Conference on*, pages 580–587, 2015.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[8] C. Boscarino, V. Nedović, N. J. Koenderink, and J. L. Top. Automatic extraction of ingredient's substitutes. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 559–564. ACM, 2014.

[9] L. Bossard, M. Guillaumin, and L. Van Gool. Food-101–mining discriminative components with random forests. In *European Conference on Computer Vision*, pages 446–461. 2014.

[10] X. Chang, Y. Yu, Y. Yang, and A. G. Hauptmann. Searching persuasively: Joint event detection and evidence recounting with limited supervision. In *ACM on Multimedia Conference*, pages 581–590, 2015.

[11] J. Chen and C.-W. Ngo. Deep-based ingredient recognition for cooking recipe retrieval. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 32–41, 2016.

[12] J. Chen, L. Pang, and C. W. Ngo. Cross-modal recipe retrieval: How to cook this dish? In *International Conference on Multimedia Modeling*, pages 588–600, 2017.

[13] S. Christodoulidis, M. Anthimopoulos, and S. Mougiakakou. Food recognition for dietary assessment using deep convolutional neural networks. In *International Conference on Image Analysis and Processing*, pages 458–465, 2015.

[14] F. Cordeiro, E. Bales, E. Cherry, and J. Fogarty. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3207–3216. ACM, 2015.

[15] J. Dehais, M. Anthimopoulos, and S. Mougiakakou. Dish detection and segmentation for dietary assessment on smartphones. In *International Conference on Image Analysis and Processing*, pages 433–440, 2015.

[16] G. M. Farinella, M. Moltisanti, and S. Battiato. Classifying food images represented as bag of textons. In *IEEE International Conference on Image Processing*, pages 5212–5216, 2014.

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

[18] C. Gan, N. Wang, Y. Yang, D. Yeung, and A. G. Hauptmann. Devnet: A deep event network for multimedia event detection and evidence recounting. In *Computer Vision and Pattern Recognition*, pages 2568–2577, 2015.

[19] M. Ge, M. Elahi, Ferna, ndez Tob, I. As, F. Ricci, and D. Massimo. Using tags and latent factors in a food recommender system. In *International Conference on Digital Health*, pages 105–112, 2015.

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

[21] H. Hassannejad, G. Matrella, P. Ciampolini, I. D. Munari, M. Mordonini, and S. Cagnoni. Food image recognition using very deep convolutional networks. In *International Workshop on Multimedia Assisted Dietary Management*, pages 41–49, 2016.

[22] J. He, M. Li, H.-J. Zhang, H. Tong, and C. Zhang. Manifold-ranking based image retrieval. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 9–16. ACM, 2004.

[23] L. Herranz, S. Jiang, and R. Xu. Modeling restaurant context for food recognition. *IEEE Transactions on Multimedia*, 19(2):430 – 440, 2017.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

[24] L. Herranz, R. Xu, and S. Jiang. A probabilistic model for food image recognition in restaurants. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2015.

[25] J. Hessel, N. Savva, and M. J. Wilber. Image representations and new domains in neural image captioning. *Computer Science*, 2015.

[26] M. D. Hoffman, D. M. Blei, and P. R. Cook. Content-based musical similarity computation using the hierarchical dirichlet process. In *International Conference on Music Information Retrieval*, pages 349–354, 2008.

[27] P. D. Howell, L. D. Martin, H. Salehian, C. Lee, K. M. Eastman, and J. Kim. Analyzing taste preferences from crowdsourced food entries. In *International Conference on Digital Health Conference*, pages 131–140, 2016.

[28] A. Jain, N. K. Rakhi, and G. Bagler. Spices form the basis of food pairing in indian cuisine. *Quantitative Biology*, 2015.

[29] X. Jiang, F. Wu, X. Li, Z. Zhao, W. Lu, S. Tang, and Y. Zhuang. Deep compositional cross-modal learning to rank via local-global alignment. pages 69–78, 2015.

[30] H. Kagaya, K. Aizawa, and M. Ogawa. Food detection and recognition using convolutional neural network. In *Proceedings of the ACM International Conference on Multimedia*, pages 1085–1088, 2014.

[31] Y. Kawano and K. Yanai. Foodcam-256: a large-scale real-time mobile food recognitionsystem employing high-dimensional features and compression of classifier weights. In *Proceedings of the ACM International Conference on Multimedia*, pages 761–762, 2014.

[32] L. S. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. In *International Conference on World Wide Web*, pages 297–306, 2008.

[33] T. Kusmierczyk and K. Norvag. Online food recipe title semantics: Combining nutrient facts and topics. In *the 25th ACM International on Conference on Information and Knowledge Management*, pages 2013–2016, 2016.

[34] C. H. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2014.

[35] C.-J. Lin, T.-T. Kuo, and S.-D. Lin. A content-based matrix factorization model for recipe recommendation. In *Advances in Knowledge Discovery and Data Mining*, pages 560–571. 2014.

[36] Y. Lu. Zero-example multimedia event detection and recounting with unsupervised evidence localization. In *ACM on Multimedia Conference*, pages 1464–1468, 2016.

[37] Z. Lydia and D. David. Think before you eat: photographic food diaries as intervention tools to change dietary decision making and attitudes. *International Journal of Consumer Studies*, 32(6):692–698, 2008.

[38] T. Maruyama, Y. Kawano, and K. Yanai. Real-time mobile recipe recommendation system using food ingredient recognition. In *Proceedings of the 2nd ACM international workshop on Interactive multimedia on mobile and portable devices*, pages 27–34, 2012.

[39] Y. Mejova, H. Haddadi, A. Noulas, and I. Weber. # foodporn: Obesity patterns in culinary interactions. In *Proceedings of the 5th International Conference on Digital Health 2015*, pages 51–58. ACM, 2015.

[40] D. Mimno and A. Mccallum. Topic models conditioned on arbitrary features with dirichlet-multinomial regression. *University of Massachusetts - Amherst*, 2008:411–418, 2012.

[41] W. Min, B. K. Bao, and C. Xu. Multimodal spatio-temporal theme modeling for landmark analysis. *IEEE Multimedia*, 21(3):20–29, 2014.

[42] W. Min, B. K. Bao, C. Xu, and M. S. Hossain. Cross-platform multi-modal topic modeling for personalized inter-platform recommendation. *IEEE Transactions on Multimedia*, 17(10):1787–1801, 2015.

[43] W. Min, S. Jiang, J. Sang, H. Wang, X. Liu, and L. Herranz. Being a super cook: Joint food attributes and multi-modal content modeling for recipe retrieval and exploration. *IEEE Transactions on Multimedia*, 19(5):1100 – 1113, 2017.

[44] V. Nedovic. Learning recipe ingredient space using generative probabilistic models. In *Proceedings of Cooking with Computers Workshop (CwC)*, volume 1, pages 13–18, 2013.

[45] K. Okamoto and K. Yanai. An automatic calorie estimation system of food images on a smartphone. In *International Workshop on Multimedia Assisted Dietary Management*, pages 63–70, 2016.

[46] L. Oliveira, V. Costa, G. Neves, T. Oliveira, E. Jorge, and M. Lizarraga. A mobile, lightweight, poll-based food identification system. *Pattern Recognition*, 47(5):1941–1952, 2014.

[47] S. Qian, T. Zhang, and C. Xu. Multi-modal multi-view topic-opinion mining for social event analysis. In *ACM on Multimedia Conference*, pages 2–11, 2016.

[48] J. Rich, H. Haddadi, and T. M. Hospedales. Towards bottom-up analysis of social food. In *Proceedings of the 6th International Conference on Digital Health Conference*, pages 111–120. ACM, 2016.

[49] S. Rudinac, A. Hanjalic, and M. Larson. Generating visual summaries of geographic areas using community-contributed images. *IEEE Transactions on Multimedia*, 15(4):921–932, 2013.

[50] S. Rudinac, M. Larson, and A. Hanjalic. Learning crowdsourced user preferences for visual summarization of image collections. *IEEE Transactions on Multimedia*, 15(6):1231–1243, 2013.

[51] S. Sajadmanesh, S. Jafarzadeh, S. A. Ossia, H. R. Rabiee, H. Haddadi, Y. Mejova, M. Musolesi, E. De Cristofaro, and G. Stringhini. Kissing cuisines: Exploring worldwide culinary habits on the web. *arXiv preprint arXiv:1610.08469*, 2016.

[52] A. Salvador, N. Hynes, Y. Aytar, J. Marin, F. Ofli, I. Weber, and A. Torralba. Learning cross-modal embeddings for cooking recipes and food images. In *Computer Vision and Pattern Recognition*, pages 3020–3028, 2017.

[53] T. H. Silva. You are what you eat (and drink): Identifying cultural boundaries by analyzing food drink habits in foursquare. *Eprint Arxiv*, 2014.

[54] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[55] A. Singla, L. Yuan, and T. Ebrahimi. Food/non-food image classification and food categorization using pre-trained googlenet model. In *International Workshop on Multimedia Assisted Dietary Management*, pages 3–11, 2016.

[56] H. Su, T.-W. Lin, C.-T. Li, M.-K. Shan, and J. Chang. Automatic recipe cuisine classification by ingredients. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 565–570, 2014.

[57] R. Tanno, K. Okamoto, and K. Yanai. Deepfoodcam: A dcnn-based real-time mobile food recognition system. In *International Workshop on Multimedia Assisted Dietary Management*, pages 89–89, 2016.

[58] C. Wang, D. Blei, and F. F. Li. Simultaneous image classification and annotation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1903–1910, 2009.

[59] X. Wang, D. Kumar, N. Thome, M. Cord, and F. Precioso. Recipe recognition with large multimodal food dataset. In *Multimedia & Expo Workshops, IEEE International Conference on*, pages 1–6, 2015.

[60] Y. Xian, Z. Akata, G. Sharma, Q. Nguyen, M. Hein, and B. Schiele. Latent embeddings for zero-shot classification. In *Computer Vision and Pattern Recognition*, pages 69–77, 2016.

[61] R. Xu, L. Herranz, S. Jiang, S. Wang, X. Song, and R. Jain. Geolocalized modeling for dish recognition. *Multimedia, IEEE Transactions on*, 17(8):1187–1199, 2015.

[62] L. Yang, Y. Cui, F. Zhang, J. P. Pollak, S. Belongie, and D. Estrin. Plateclick: Bootstrapping food preferences through an adaptive visual interface. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 183–192, 2015.

[63] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen. Lcars: a location-content-aware recommender system. In *ACM International Conference on Knowledge Discovery and Data Mining*, pages 221–229, 2013.

[64] Q. Yu, J. Liu, H. Cheng, A. Divakaran, and H. Sawhney. Multimedia event recounting with concept based representation. In *ACM International Conference on Multimedia*, pages 1073–1076, 2012.

[65] F. Zhang, N. J. Yuan, K. Zheng, D. Lian, X. Xie, and Y. Rui. Exploiting dining preference for restaurant recommendation. In *International Conference on World Wide Web*, pages 725–735, 2016.

[66] Z. Zhang and V. Saligrama. Zero-shot learning via joint latent similarity embedding. In *Computer Vision and Pattern Recognition*, pages 6034–6042, 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2017.2759499, IEEE Transactions on Multimedia

Weiqing Min received the B.E. degree from Shandong Normal University, Jinan, China, in 2008 and M.E. degree from Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, in 2015, respectively. He is currently a research associate at the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences. His current research interests include multimedia search and mining, location based multimedia analysis and applications. He has authored and co-authored more than 10 peer-referenced papers in multimedia related journals and conferences, including IEEE Transactions on Multimedia, IEEE Multimedia Magazine, ACM Multimedia, ACM TOMM, etc. He is the reviewer of some international journals including IEEE Transactions on Cybernetics, IEEE Multimedia Magazine, Multimedia Tools and Application, Multimedia System Journal, Neurocomputing, etc. He is the recipient of 2016 ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMM) Nicolas D. Georganas Best Paper Award and the 2017 IEEE Multimedia Magazine Best Paper Award.

Bing-Kun Bao received the Ph.D. degree in control theory and control application from the University of Science and Technology of China, Hefei, China, in 2009. She is currently an Associate Researcher with the Institute of Automation, Chinese Academy of Sciences, Beijing, China. Her current research interests include cross-media cross-modal image search, social event detection, image classification and annotation, and sparse/low rank representation.She was the recipient of the 2016 ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMM) Nicolas D. Georganas Best Paper Award, IEEE Multimedia 2017 Best Paper Award, and the Best Paper Award from ICIMCS09.

Shuhuan Mei received the B.E. degree from Shandong University of Science and Technology, Qingdao, China, in 2015 and is pursuing the M.E. degree in Shandong University of Science and Technology, Qingdao, China. His current research interests include multimedia retrieval and applications.

Yaohui Zhu received the M.E. degree in computer science from Shenyang Aerospace University, Shenyang, China, in 2016, and is currently a PhD candidate at the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. his current research interests include computer vision, multimedia content analysis and machine learning.

Yong Rui (SM'04-F'10) received the B.S. degree from Southeast University, the M.S. degree from Tsinghua University, and the Ph.D. degree from the University of Illinois at Urbana-Champaign. He is currently the Chief Technology Officer and Senior Vice President of Lenovo Group. He is responsible for overseeing Lenovos corporate technical strategy, research and development directions, and Lenovo Research organization, which covers intelligent devices, big data analytics, artificial intelligence, cloud computing, 5G and smart lifestyle-related technologies. He has authored 2 books, 12 book chapters, and 260 refereed journal and conference papers. With over 19,000 citations, and an h-Index of 59, his publications are among the most referenced. He holds 62 issued U.S. and international patents. He is a recipient of many awards, including the 2016 IEEE Computer Society Technical Achievement Award, the 2016 IEEE Signal Processing Society Best Paper Award and the 2010 Most Cited Paper of the Decade Award from Journal of Visual Communication and Image Representation. He is the Editor-in-Chief of IEEE Multimedia Magazine, an Associate Editor of ACM Trans. on Multimedia Computing, Communication and Applications (TOMM), and a founding Editor of International Journal of Multimedia Information Retrieval. He was an Associate Editor of IEEE Trans. on Multimedia (2004-2008), IEEE Trans. on Circuits and Systems for Video Technologies (2006-2010), ACM/Springer Multimedia Systems Journal (2004-2006), International Journal of Multimedia Tools and Applications (2004-2006), and IEEE Access (2013-2016). He is a Fellow of IEEE, IAPR and SPIE, a Distinguished Member of the ACM.

Shuqiang Jiang (SM'08) is a professor with the Institute of Computing Technology, Chinese Academy of Sciences(CAS), Beijing and a professor in University of CAS. He is also with the Key Laboratory of Intelligent Information Processing, CAS. His research interests include multimedia processing and semantic understanding, pattern recognition and computer vision. He has authored or coauthored more than 100 papers on the related research topics. He was supported by the New-Star program of Science and Technology of Beijing Metropolis in 2008, NSFC Excellent Young Scientists Fund in 2013, Young top-notch talent of Ten Thousand Talent Program in 2014. He won the Lu Jiaxi Young Talent Award from Chinese Academy of Sciences in 2012, and the CCF Award of Science and Technology in 2012. He is the senior member of IEEE and CCF, member of ACM, Associate Editor of IEEE Multimedia, Multimedia Tools and Applications. He is the general secretary of IEEE CASS Beijing Chapter, vice chair of ACM SIGMM China chapter. He is the general chair of ICIMCS 2015, program chair of ICIMCS2010, special session chair of PCM2008, ICIMCS2012, area chair of PCIVT2011, publicity chair of PCM2011, web chair of ISCAS2013, and proceedings chair of MMSP2011. He has also served as a TPC member for more than 20 well-known conferences, including ACM Multimedia, CVPR, ICCV, ICME, ICIP, and PCM.