Cross-Platform Multi-Modal Topic Modeling for Personalized Inter-Platform Recommendation

Weiqing Min, Bing-Kun Bao, Changsheng Xu, *Fellow, IEEE*, and M. Shamim Hossain, *Senior Member, IEEE*

***Abstract*—In this paper, we investigate a novel cross-platform multimedia problem: given two platforms, Flickr and Foursquare, we conduct the recommendation between these two platforms, namely the photo recommendation from Flickr to Foursquare users and the venue recommendation from Foursquare to Flickr users. Such inter-platform recommendations enable users from one single platform to enjoy different recommendation services effectively. To solve the problem, we propose a cross-platform multi-modal topic model ($\mathrm{CM}^3\mathrm{TM}$), which is capable of: 1) differentiating between two kinds of topics, i.e., platform-specific topics only relevant to a certain platform and shared topics characterizing the knowledge shared by different platforms and 2) aligning multiple modalities from different platforms. Specifically, $\mathrm{CM}^3\mathrm{TM}$ can not only split the topic space into the shared topic space and platform-specific topic space and learn them simultaneously, but also enable the alignment among different modalities through the learned topic space. Given the location information, we applied the proposed $\mathrm{CM}^3\mathrm{TM}$ into two inter-platform recommendation applications: 1) personalized venue recommendation from Foursquare to Flickr users and 2) personalized image recommendation from Flickr to Foursquare users. We have conducted experiments on the collected large-scale real-world dataset from Flickr and Foursquare. Qualitative and quantitative evaluation results validate the effectiveness of our method and demonstrate the advantage of connecting different platforms with different modalities for the inter-platform recommendation.**

***Index Terms*— Cross-platform, topic model, recommendation.**

I. Introduction

With the fast development of Web2.0, various social media platforms are gaining more and more popularity with their different types of services. For example, Flickr is an image hosting website for users to share personal photographs, while Foursquare provides a personalized location search service and recommends great places around a user's current location. All these social media platforms host huge and ever-growing data which are of different modalities and for their own services. How to utilize those data without the boundaries of platforms has become a hot research area because of its great potential in various applications, such as social link prediction on Foursquare and Twitter [43] and novelty-seeking trait mining on SinaWeibo and Taobao [42]. In this work, we investigate the problem of cross-platform multi-modal data correlation and analysis. Specifically, we propose a model to correlate the data from two social platforms, i.e., Flickr and Foursquare, and apply this model to the personalized inter-platform recommendation. In these applications, Flickr users will be able to quickly find out nearby venues what they are interested in when they come to new places, thanks to the recommendations from Foursquare. Foursquare users will also enjoy more interesting pictures from Flickr. These recommendations are mutual and simultaneous.

Existing cross-platform based work [29], [15], [43] mainly designed multimedia applications on a single platform using the information from auxiliary platforms. For example, Zhong *et al.* [43] used the overlapping users as the bridge to conduct friend recommendation in Foursquare by employing users' information and relationship information in Twitter. However, it is not easy to obtain or identify the same user account from different platforms in many cases. Roy *et al.* [29] utilized the event information to correlate Twitter and YouTube for socialized video search and recommendation in Youtube by exploiting the textual information in Twitter. However, little work has investigated how to design a model to use general topic information to correlate different platforms with different modalities for enabling inter-platform applications.

Each platform hosts rich content information shared by users, such as images with tags in Flickr and the venues with tags in Foursquare. Generally, the knowledge mined from the content information on different platforms can be divided into two parts: shared knowledge from all platforms and platform-specific knowledge. We use a toy example to elaborate this. Fig. 1 shows some exemplary images with tags from Flickr and venues with tags from Foursquare. We can see that both "music concert" (shown in red) and "animal" (shown in purple) are the shared knowledge by both platforms. The images with tags about the sunset in Flickr (shown in green) represent Flickr-specific knowledge while the checked bookstores in Foursquare (shown in blue) correspond to Foursquare-specific

Manuscript received January 21, 2015; revised May 13, 2015 and July 03, 2015; accepted July 06, 2015. Date of publication July 30, 2015; date of current version September 15, 2015. This work was supported in part by the National Program on Key Basic Research Project (973 Program, Project 2012CB316304), by the National Natural Science Foundation of China under Grant 61225009, Grant 61201374, and Grant 61432019, by the Beijing Natural Science Foundation under Grant 4131004 and Grant 4152053, by the open project from the State Key Laboratory for Novel Software Technology of Nanjing University, and by the Deanship of Scientific Research at King Saud University under International Research Group Program IRG 14-18. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Cees Snoek.

W. Min, B.-K. Bao, and C. Xu are with the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: wqmin@nlpr.ia.ac.cn; bingkunbao@gmail.com; csxu@nlpr.ia.ac.cn).

M. S. Hossain is with the SWE Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia (e-mail: mshossain@ksu.edu.sa).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2015.2463226

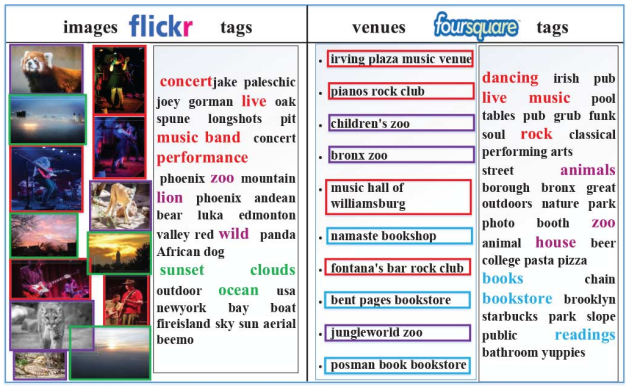


Fig. 1. Different types of knowledge in Flickr and Foursquare, where the shared knowledge is shown in red and purple, the Flickr-specific knowledge is shown in green, and the Foursquare-specific knowledge is shown in blue.

knowledge. Considering that the shared knowledge belongs to all the platforms, we utilize it as a bridge to connect different platforms for inter-platform applications. Specifically, our work is dedicated to extracting the shared knowledge by distinguishing it from the platform-specific knowledge to connect different platforms.

Extracting the shared knowledge from different platforms is not trivial. The challenge is twofold. Firstly, in each platform, the shared knowledge is intertwined with the specific knowledge. If we directly extract the shared knowledge by simply combining content from all platforms, the extracted shared knowledge will probably be quite noisy. Therefore, the first challenge is to differentiate between the shared knowledge and platform-specific knowledge. Secondly, besides the textual content, each platform has its own modality, such as images from Flickr and venues from Foursquare. In order to better realize inter-platform multimedia applications, we need to align different modalities for each kind of topic knowledge across different platforms. Therefore, the second challenge is to align different modalities across different platforms.

In order to solve the above-mentioned problems, we propose a probabilistic topic modeling framework, namely cross-platform multi-modal topic model (CM^{3}TM) to discover topics from different platforms. We define two different types of topics, i.e., shared topics and platform-specific topics. The shared topics characterize the shared knowledge from all platforms and are represented in a shared topic space. In contrast, the platform-specific topics characterize platform-specific knowledge and are represented in a platform-specific topic space. By splitting the topic space into the shared topic space and platform-specific topic space, CM^{3}TM is capable of distinguishing shared topics from platform-specific topics. The learned shared topics are used to bridge different platforms. Furthermore, CM^{3}TM can align multiple modalities for each shared topic across different platforms through the learned shared topic space. We take Flickr and Foursquare as the testbed in our study. As shown in Fig. 2(a), the input of the model includes 1) user's images and annotated tags from Flickr, and 2) user's checked-ins and annotated tags from Foursquare. The output includes 1) the learned shared topic space and platform-specific topic space, and 2) the shared topic distribution of users and platform-specific topic distribution of users.

Based on the proposed CM^{3}TM, we develop a location-context inter-platform recommendation framework, which mainly consists of two components: topic modeling and personalized inter-platform recommendation. CM^{3}TM is employed to learn the topic space and the topic distribution of users. Location-context inter-platform recommendation [Fig. 2(b)] is derived from CM^{3}TM, including: 1) personalized venue recommendation from Foursquare to Flickr users and 2) personalized image recommendation from Flickr to Foursquare users.

The contributions of our work can be summarized as follows.

- We study the problem of inter-platform connection by directly exploring multi-modal content information from different platforms to benefit inter-platform multimedia applications.
- We propose a cross-platform multi-modal topic model (CM^{3}TM), which is able to discover shared topics by distinguishing them from platform-specific topics and align multiple modalities for each shared topic across different platforms.
- We present a location-context inter-platform recommendation framework based on CM^{3}TM and evaluate them through a large-scale real-world dataset from two platforms.

The rest of the paper is organized as follows. Section II reviews the related work. Section III presents the cross-platform multi-modal topic model (CM^{3}TM) in detail. Section IV introduces the location-context personalized inter-platform recommendation application derived from our proposed model. Experimental results are reported in Section V. Finally, we conclude the paper in Section VI.

II. Related Work

Our work is related to two research areas: 1) probabilistic topic models, and 2) personalized recommendation.

A. Probabilistic Topic Models

The Probabilistic Topic Model (PTM) aims to explore a set of topics from a large collection of documents, where a topic is a distribution over a fixed vocabulary and a document is a distribution over topics. A more comprehensive survey of PTM is provided in [5], which has been successfully applied to many domains (e.g., text, images, videos and music) for various tasks such as classification, information retrieval and recommendation [7], [6], [14], [37], [30], [32].

The simplest probabilistic topic model is Latent Dirichlet Allocation (LDA) [7]. In order to extend the LDA model to learn the joint correlations between data of different modalities, such as the text and images, some variants of topic models are developed, such as multimodal-LDA and correspondence LDA [6]. They use a set of shared latent variables to explicitly model images and annotated text to capture correlations between the data of two modalities. Compared with [6], which modeled the multi-modal data in one platform, our method models them across different platforms. This is more challenging due to the alignment

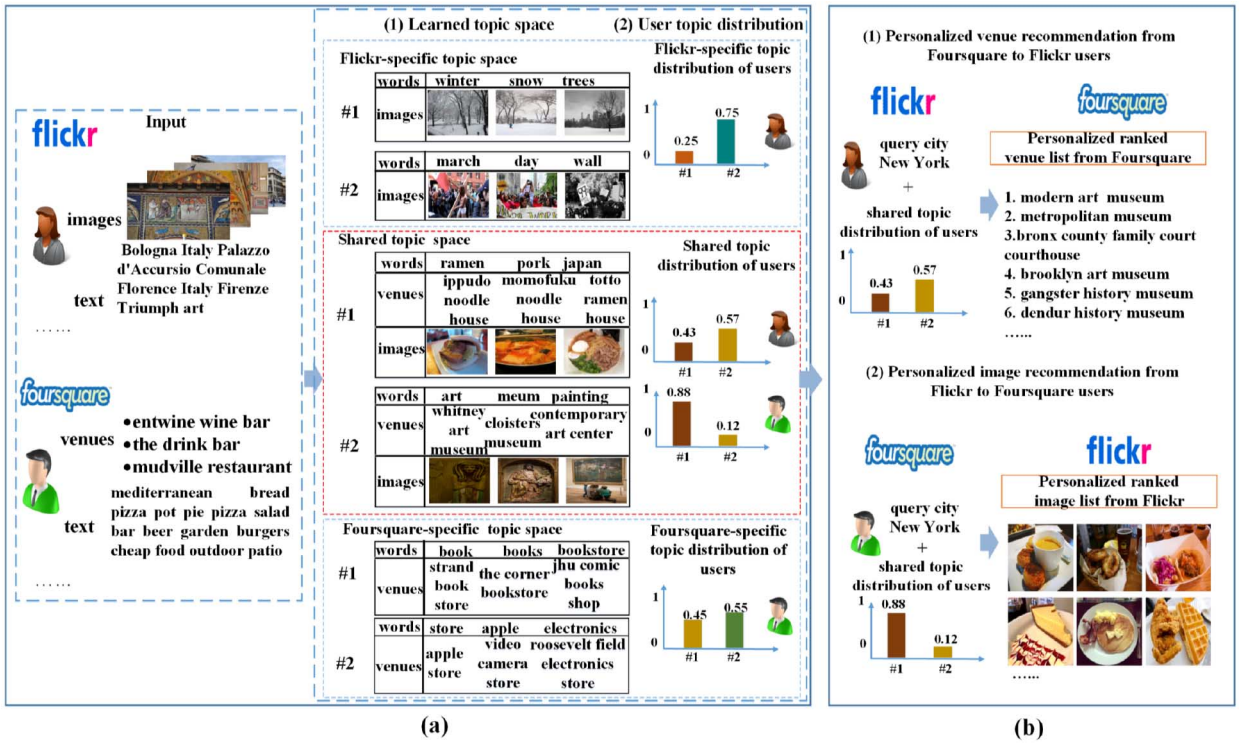


Fig. 2. Proposed personalized recommendation framework. (a) Cross-platform multi-modal topic model. (b) Location-context inter-platform recommendation.

of multiple modalities for each topic across different platforms. The most related work to ours is [20], which divides the topics from different domains into shared topics and domain-specific topics for text classification. Our work is different from [20] in that, we are investigating a novel inter-platform recommendation problem in the social multimedia context. In addition to the textual information, we provide a multi-modal framework, which can simultaneously model visual content and venue information across different platforms. To the best of our knowledge, a model that can discover topics and align different modalities for each topic across different platforms has not been proposed before.

B. Personalized Recommendation

Personalized recommendation in social media can be divided into two types: single-platform based personalized recommendation and cross-platform based personalized recommendation.

1) Single-Platform Based Personalized Recommendation: Most of existing work on personalized recommendation is relevant to the single-platform, such as Point-of-interest (POI) recommendation [47], [35], [22], [38], image recommendation [10], [30], [31] and friend recommendation [34]. The single-platform based personalized recommendation usually utilizes the shared content information and user relationship in this platform for recommendation. For example, some studies [35], [22], [38], [37] explored user preferences, social influence, or geographical influence from Foursquare for POI recommendation. Other work such as [30], [34] utilized the shared content information (i.e., images and tags) and context information from Flickr to conduct personalized image or friend recommendation.

To overcome the drawbacks of solutions in single platform, for example, the sparseness of user data, many researchers start to shift their attention to multiple platform based solutions to exploit more rich data across different platforms.

2) Cross-Platform Based Personalized Recommendation: The basic idea of existing cross-platform based personalized recommendation [44], [43], [49], [50], [33] is to transfer the knowledge from source platforms to the target one to improve the performance of recommendation. Zhang *et al.* [43], [44] proposed to employ the link information transferred from aligned source platforms through overlapping users to predict social links for new users. Different from [43], [44], Zhong *et al.* [49], [50] introduced one transfer algorithm based on topic models to integrate user relationships adaptively from different platforms for item recommendation or friend recommendation. Yan *et al.* [33] proposed a random walk based method to address the cold-start friend recommendation problem through overlapping users. These methods all use the overlapping users as the bridge to connect different platforms. In contrast, Qi *et al.* [28] utilized the biased cross-platform sampling method to do link prediction across platforms via the content attributes of users. Roy *et al.* [29] transferred the knowledge from the tweet stream to improve video recommendation by mining latent features-topics from the tweets. Our work also learns the latent features from the content information as the bridge to connect different platforms for recommendation. However, different from [28], [29], we focus on addressing personalized inter-platform recommendation problems while the work [28], [29] mainly transfer the knowledge from the source platform to the target one to solve the target platform recommendation problems.

In addition to the cross-platform based recommendation, another relevant area is the cross-domain recommendation. A more comprehensive survey of the cross-domain recommendation is provided in [17], [11]. The core concept of cross-domain recommendation is to improve the recommendation performance of one domain by exploiting the information from other domains. For example, Li *et al.* [18], [19] used the latent factor model to learn an implicit cluster-level rating pattern that can be shared with a target domain for transferring the knowledge from an auxiliary domain. Zhang *et al.* [45] extended the latent factor model to multiple domains to learn an implicit correlation matrix, which links different domains for knowledge transfer by the implicit domain correlations. In addition, Pan *et al.* [27], [26] introduced a method named coordinate system transfer over multiple data domains to learn shared latent features. Similar to [27], [26], we also learn the shared space to transfer the knowledge among different platforms. However, unlike the above mentioned work that is only to serve the target domain by virtue of the auxiliary domains, we focus on the inter-platform recommendation application to benefit each other simultaneously. Furthermore, in addition to the textual information from both platforms, we allow different modalities from each platform and thus provide a generic multi-modal framework in the multimedia context for cross-platform applications.

Recently, some work focused on collecting rich information from different platforms for topic detection or recommendation [1], [2], [41], [23], [39], [21]. For example, Bao *et al.* [1], [2] utilized the co-clustering method to detect emerging topics from multimedia streams from different platforms, including Twitter, Flickr and New York Times. Zahlka *et al.* [41] extracted the concepts from venue images and associated annotations collected using Foursquare, Picasa and Flickr and then learned the user taste through the designed interactive map interface for personalized recommendation. Liu *et al.* [23] analyzed both geo-tagged images from Flickr and check-ins from Foursquare to exploit the travelers' flavors and the preferences of daily-life activities of local residents for discovering and recommending the area of interest in a city. Unlike the above mentioned work which focused on collecting the information from different platforms for recommendation, we transfer the knowledge between different platforms for recommendation to benefit users from these platforms simultaneously.

III. The Cross-Platform Multi-Modal Topic Model

This section introduces the cross-platform multi-modal topic model ($\mathrm{CM}^3\mathrm{TM}$), which is to discover both shared and platform-specific topics from users' documents[1] in Flickr and Foursquare. Table I lists relevant notations used in this paper.

A. Problem Statement

As discussed in Section I, we assume that there are three kinds of topics in user documents in Flickr and Foursquare:

[1]Here, we aggregate the records using the user as the unit, which is more intuitive for user-oriented applications (e.g., personalized search and recommendation). In Flickr, one user document includes the shared images and associated text information (e.g., tags and the title) while one Foursquare user document include all the venues where this user has checked in and the associated text information, such as tags and tips this user left.

TABLE I
List of Key Notations

Notations	Description
c_m	the m-th social platform
U, U^{c_1}, U^{c_2}	the user set, Flickr user set, Foursquare user set
I_u, T_u	a collection of items (i.e., images in Flickr or venues in Foursquare), associated text for user document u
W, V, L	textual word vocabulary, visual word vocabulary, venue vocabulary
K, K^{c_1}, K^{c_2}	the number of shared topics, Flickr-specific topics, Foursquare-specific topics
n_u^w, n_u^v, n_u^l	the number of textual words, visual words and venues in user document u
ψ, ϕ, φ	the multinomial distributions over textual words, visual words and venue words for shared topics
ψ^{c_1}, ϕ^{c_1}	the multinomial distributions over textual words and visual words for Flickr-specific topics
ψ^{c_2}, φ^{c_2}	the multinomial distributions over textual words and venue words for Foursquare-specific topics
θ'^{c_m}, θ	the multinomial distributions over platform c_m-specific topics and shared topics for users
s^w, s^v , s^l	binary labels denoting whether the generation of the textual word, visual word and venue is drawn from θ or θ'^{c_m}
z^w, z^v, z^l	topic assignment for the textual word, visual word and venue
ρ^{c_m}	the parameter for sampling binary variable s^w, s^v or s^l from the platform c_m
$\lambda_1^{c_m}$, $\lambda_2^{c_m}$	Beta priors to generate ρ^{c_m}
$\alpha_1^{c_m}$, $\alpha_2^{c_m}$	Dirichlet priors to multinomial distribution θ^{c_m}, θ'^{c_m}
β, β^{c_m}	Dirichlet priors to multinomial distribution ψ, ψ^{c_m}
γ, γ^{c_1}	Dirichlet priors to multinomial distribution ϕ, ϕ^{c_1}
η, η^{c_2}	Dirichlet priors to multinomial distribution φ, φ^{c_2}

1) shared topics, which are shared by both social platforms; 2) Flickr- specific topics, which are related only to the Flickr platform; and 3) Foursquare-specific topics, which correspond to the characteristics of the Foursquare platform. In addition, each kind of topics is represented by different modalities. Based on the above concepts,[2] the problem of $\mathrm{CM}^3\mathrm{TM}$ can be defined as follows.

Definition 1: ($\mathrm{CM}^3\mathrm{TM}$). Given a collection of users from the Flickr platform c_1 and the Foursquare platform c_2, $U = \{u_1^{c_1}, u_2^{c_1}, \ldots, u_{|U^{c_1}|}^{c_1}, u_1^{c_2}, u_2^{c_2}, \ldots, u_{|U^{c_2}|}^{c_2}\}$, where $u_i^{c_m}$ corresponds to a two-dimensional tuple $[I_{u_i^{c_m}}, T_{u_i^{c_m}}]$ from social platform c_m, the goal of $\mathrm{CM}^3\mathrm{TM}$ is to learn: 1) three kinds of topic spaces: the shared topic space ψ, ϕ, φ; the Flickr-specific topic space ψ^{c_1}, ϕ^{c_1} and the Foursquare-specific topic space ψ^{c_2}, φ^{c_2}; and 2) the distributions over shared topics, Flickr-specific topics and Foursquare-specific topics for users: θ, θ'^{c_1}, θ'^{c_2}.

B. Cross-Platform Multi-Modal Topic Model

Two key ideas are exploited in our topic model. Firstly, our model should distinguish between shared topics and platform-specific topics. In order to realize it, we can split the topics into three groups: K shared topics with parameters $\psi = \{\psi_1, \ldots, \psi_K\}$, K^{c_1} Flickr-specific topics with parameters $\psi^{c_1} = \{\psi_1^{c_1}, \ldots, \psi_{K^{c_1}}^{c_1}\}$ and K^{c_2} Foursquare-specific topics with parameters $\psi^{c_2} = \{\psi_1^{c_2}, \ldots, \psi_{K^{c_2}}^{c_2}\}$. Each topic is represented by a multinomial distribution over words with different parameters. Our model can distinguish and learn three kinds of distributions to discover these topics. Secondly, in order to facilitate the inter-platform applications, our model

[2]Generally,the shared topic comprises of the keywords belonging to both platforms, while the Flickr-specific and Foursquare-specific topics are composed by the keywords from Flickr and Foursquare, respectively.

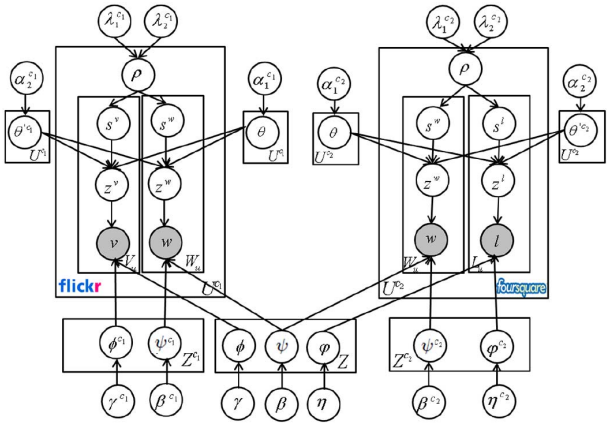


Fig. 3. Probabilistic generative model of CM^{3}TM.

should align different modalities for each topic, especially across different platforms. To realize this idea, we correlate the topics from different modalities by sampling them from the same distribution over topics. Since the distribution over textual words for shared topics ψ is learned from all platforms, the textual modality, visual modality and venue modality for each shared topic can be aligned. Specifically, in Flickr, besides the distributions over textual words for shared topics ψ and Flickr-specific topics ψ^{c_1}, there are corresponding distributions over visual words for shared topics ϕ and Flickr-specific topics ϕ^{c_1}. Similarly, in Foursquare, the distributions over venue words for shared topics φ and Foursquare-specific topics φ^{c_2} correspond to ψ and ψ^{c_2}, respectively. The shared topic-image distribution ϕ and shared topic-venue distribution φ will facilitate the following inter-platform recommendation. ϕ is used for personalized image recommendation from Flickr to Foursquare users while φ for personalized venue recommendation from Foursquare to Flickr users. In addition, for each platform m, a latent switch variable s^x, $x = \{w, v, l\}$ is introduced to control whether the generation of the textual word, the visual word and the venue is drawn from θ or θ'^{c_m}. Fig. 3 illustrates the graphical representation of the generation process.

The details of the generative process are as follows.

1) For each shared topic $z \in \{1, \ldots, K\}$, draw $\psi_z \sim Dir(\beta)$, draw $\phi_z \sim Dir(\gamma)$, draw $\varphi_z \sim Dir(\eta)$
2) For Flickr platform c_1-specific topic $z \in \{1, \ldots, K^{c_1}\}$, draw $\psi_z^{c_1} \sim Dir(\beta^{c_1})$, draw $\phi_z^{c_1} \sim Dir(\gamma^{c_1})$
3) For Foursquare platform c_2-specific topic $z \in \{1, \ldots, K^{c_2}\}$, draw $\psi_z^{c_2} \sim Dir(\beta^{c_2})$, draw $\varphi_z^{c_2} \sim Dir(\eta^{c_2})$
4) For each platform $m \in \{c_1, c_2\}$
   a) if $m = c_1$, for each user $u \in U^{c_1}$
      i) draw $\theta_u^{c_1} \sim Dir(\alpha_1^{c_1})$,
      ii) draw $\theta_u'^{c_1} \sim Dir(\alpha_2^{c_1})$,
      iii) draw $\rho_u^{c_1} \sim Beta(\lambda_1^{c_1}, \lambda_2^{c_1})$,
      iv) for each textual word $w_{u,j} \in \mathbf{w}_u$
         A) Draw a switch variable $s_{u,j}^w \sim Binomial(\rho_u^{c_1})$
         B) if $s_{u,j}^w = 0$, draw a topic $z_{u,j}^w \sim Multi(\theta_u^{c_1})$ and then draw a word $w_{u,j} \sim Multi(\psi_{z_{u,j}^w}^{c_1})$
         C) if $s_{u,j}^w = 1$, draw a topic $z_{u,j}^w \sim Multi(\theta_u'^{c_1})$ and then draw a word $w_{u,j} \sim Multi(\psi_{z_{u,j}^w}^{c_1})$
      iv) for each visual word $v_{u,j} \in \mathbf{v}_u$
         A) Draw a binary switch $s_{u,j}^v \sim Binomial(\rho_u^{c_1})$
         B) if $s_{u,j}^v = 0$, draw a topic $z_{u,j}^v \sim Multi(\theta_u^{c_1})$ and then draw a word $v_{u,j} \sim Multi(\phi_{z_{u,j}^v})$
         C) if $s_{u,j}^v = 1$, draw a topic $z_{u,j}^v \sim Multi(\theta_u'^{c_1})$ and then draw a word $v_{u,j} \sim Multi(\phi_{z_{u,j}^v}^{c_1})$
   b) if $m = c_2$, for each user $u \in U^{c_2}$
      i) draw $\theta_u^{c_2} \sim Dir(\alpha_1^{c_2})$,
      ii) draw $\theta_u'^{c_2} \sim Dir(\alpha_2^{c_2})$,
      iii) draw $\rho_u^{c_2} \sim Beta(\lambda_1^{c_2}, \lambda_2^{c_2})$,
      iv) for each textual word $w_{u,j} \in \mathbf{w}_u$
         A) Draw a binary switch $s_{u,j}^w \sim Binomial(\rho_u^{c_2})$
         B) if $s_{u,j}^w = 0$, draw a topic $z_{u,j}^w \sim Multi(\theta_u^{c_2})$ and then draw a word $w_{u,j} \sim Multi(\psi_{z_{u,j}^w})$
         C) if $s_{u,j}^w = 1$, draw a topic $z_{u,j}^w \sim Multi(\theta_u'^{c_2})$ and then draw a word $w_{u,j} \sim Multi(\psi_{z_{u,j}^w}^{c_2})$
      iv) for each venue $l_{u,j} \in \mathbf{l}_u$
         A) A) Draw a binary switch $s_{u,j}^l \sim Binomial(\rho_u^{c_2})$
         B) if $s_{u,j}^l = 0$, draw a topic $z_{u,j}^l \sim Multi(\theta_u^{c_2})$ and then draw a venue $l_{u,j} \sim Multi(\varphi_{z_{u,j}^l})$
         C) if $s_{u,j}^l = 1$, draw a topic $z_{u,j}^l \sim Multi(\theta_u'^{c_2})$ and then draw a venue $l_{u,j} \sim Multi(\varphi_{z_{u,j}^l}^{c_2})$

Note that our proposed model is different from other simple methods in discovering three kinds of topics. For example, one simple method first separates the items from Flickr and Foursquare into three sets based on the text similarity, where one is shared by two platforms and the other two sets are specific to one platform, respectively. The traditional topic model (e.g., LDA [7]) is then applied to discover topics from each set. In contrast, our proposed model can discover three kinds of topics simultaneously in a unified framework. Furthermore, compared with the simple method based on the text similarity, our proposed topic model is capable of distinguishing among three kinds of topics based on more accurate semantic relations.

C. Model Inference

The CM^{3}TM model includes four sets of variables: the binary switch labels $\mathbf{s}^w, \mathbf{s}^v$, the topic assignment $\mathbf{z}^w, \mathbf{z}^v$ from Flickr and the binary switch labels $\mathbf{s}^w, \mathbf{s}^l$, the topic assignment $\mathbf{z}^w, \mathbf{z}^l$ from Foursquare. We use Gibbs sampling [13] for the model inference. The update rules for latent variables are as follows.

For latent variable $\mathbf{z}^w$ and $\mathbf{s}^w$ from Flickr c_1, we employ a two-step Gibbs sampling procedure:

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

$$p(s_i^w = 0|\mathbf{s}_{\neg i}^w, \mathbf{z}^w, \mathbf{w}) \propto \frac{n_{u,z_i,\neg i}^{w,s_i^w=0,c_1} + \alpha_1^{c_1}}{\sum_z n_{u,z,\neg i}^{w,s_i^w=0,c_1} + K\alpha_1^{c_1}} (n_{u,s_i^w=0,\neg i}^{w,c_1} + \lambda_1^{c_1}) \quad (1)$$

$$p(s_i^w = 1|\mathbf{s}_{\neg i}^w, \mathbf{z}^w, \mathbf{w}) \propto \frac{n_{u,z_i,\neg i}^{w,s_i^w=1,c_1} + \alpha_2^{c_1}}{\sum_z n_{u,z,\neg i}^{w,s_i^w=1,c_1} + K^{c_1}\alpha_2^{c_1}} (n_{u,s_i^w=1,\neg i}^{w,c_1} + \lambda_2^{c_1}) \quad (2)$$

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

$$p(z_i^w|s_i^w = 0, \mathbf{z}_{\neg i}^w, \mathbf{s}_{\neg i}^w, \mathbf{w}) \propto \frac{n_{z_i,w_i,\neg i} + \beta}{\sum_{w'} n_{z_i,w',\neg i} + W\beta} (n_{u,z_i,\neg i}^{w,s_i^w=0,c_1} + \alpha_1^{c_1}) \quad (3)$$

$$p(z_i^w|s_i^w = 1, \mathbf{z}_{\neg i}^w, \mathbf{s}_{\neg i}^w, \mathbf{w}) \propto \frac{n_{z_i,w_i,\neg i}^{c_1} + \beta^{c_1}}{\sum_{w'} n_{z_i,w',\neg i}^{c_1} + W\beta^{c_1}} (n_{u,z_i,\neg i}^{w,s_i^w=1,c_1} + \alpha_2^{c_1}) \quad (4)$$

where $i = (u, j)$ is the current index. The superscript $\neg i$ denotes a counting variable that excludes the i-th word index in the corpus. $n_{u,z_i,\neg i}^{w,s_i^w=0,c_1}$ is the number of times that the shared topic $z_i(s_i^w = 0)$ represented by the words w is assigned to the user u in Flickr. $n_{u,z_i,\neg i}^{w,s_i^w=1,c_1}$ is the number of times that Flickr-specific topic $z_i(s_i^w = 1)$ represented by the words is assigned to the user u in Flickr. $n_{z_i,w_i,\neg i}$ is the number of times that the word w_i is assigned to the shared topic $z_i(s_i^w = 0)$. $n_{z_i,w_i,\neg i}^{c_1}$ is the number of times that the word w_i is assigned to the Flickr-specific topic $z_i(s_i^w = 1)$. $n_{u,s_i^w=0,\neg i}^{w,c_1}$ is the number of times that $s_i^w = 0$ is assigned to the user u from Flickr. $n_{u,s_i^w=1,\neg i}^{w,c_1}$ is the number of times that $s_i^w = 1$ is assigned to the user u from Flickr. The update rules for variables concerning visual words $\mathbf{z}^v$ and $\mathbf{s}^v$ from the Flickr social platform can be derived analogously.

For $\mathbf{z}^w$ and $\mathbf{s}^w$ from social platform Foursquare c_2, the update rules are as follows:

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

$$p(s_i^w = 0|\mathbf{s}_{\neg i}^w, \mathbf{z}^w, \mathbf{w}) \propto \frac{n_{u,z_i,\neg i}^{w,s_i^w=0,c_2} + \alpha_1^{c_2}}{\sum_z n_{u,z,\neg i}^{w,s_i^w=0,c_2} + K\alpha_1^{c_2}} (n_{u,s_i^w=0,\neg i}^{w,c_2} + \lambda_1^{c_2}) \quad (5)$$

$$p(s_i^w = 1|\mathbf{s}_{\neg i}^w, \mathbf{z}^w, \mathbf{w}) \propto \frac{n_{u,z_i,\neg i}^{w,s_i^w=1,c_2} + \alpha_2^{c_2}}{\sum_z n_{u,z,\neg i}^{w,s_i^w=1,c_2} + K^{c_2}\alpha_2^{c_2}} (n_{u,s_i^w=1,\neg i}^{w,c_2} + \lambda_2^{c_2}) \quad (6)$$

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

$$p(z_i^w|s_i^w = 0, \mathbf{z}_{\neg i}^w, \mathbf{s}_{\neg i}^w, \mathbf{w}) \propto \frac{n_{z_i,w_i,\neg i} + \beta}{\sum_{w'} n_{z_i,w',\neg i} + W\beta} (n_{u,z_i,\neg i}^{w,s_i^w=0,c_2} + \alpha_1^{c_2}) \quad (7)$$

$$p(z_i^w|s_i^w = 1, \mathbf{z}_{\neg i}^w, \mathbf{s}_{\neg i}^w, \mathbf{w}) \propto \frac{n_{z_i,w_i,\neg i}^{c_2} + \beta^{c_2}}{\sum_{w'} n_{z_i,w',\neg i}^{c_2} + W\beta^{c_2}} (n_{u,z_i,\neg i}^{w,s_i^w=1,c_2} + \alpha_2^{c_2}) \quad (8)$$

where the meaning of parameters can be obtained by analogy with the parameters from the Flickr platform. Similarly, the update rules for variables concerning venues $\mathbf{z}^l$ and $\mathbf{s}^l$ from social platform Foursquare c_2 are derived analogously.

D. Parameter Estimation

After a sufficient number of sampling iterations, we can estimate the parameters as follows:

$$\begin{aligned}
\hat{\rho}_{u,s=0}^{c_1} &= \frac{n_{u,s=0}^{w,c_1} + n_{u,s=0}^{v,c_1} + \lambda_1^{c_1}}{\sum_{s'} n_{u,s'}^{w,c_1} + \sum_{s'} n_{u,s'}^{v,c_1} + \lambda_1^{c_1} + \lambda_2^{c_1}} \\
\hat{\rho}_{u,s=1}^{c_1} &= \frac{n_{u,s=1}^{w,c_1} + n_{u,s=1}^{v,c_1} + \lambda_2^{c_1}}{\sum_{s'} n_{u,s'}^{w,c_1} + \sum_{s'} n_{u,s'}^{v,c_1} + \lambda_1^{c_1} + \lambda_2^{c_1}} \\
\hat{\rho}_{u,s=0}^{c_2} &= \frac{n_{u,s=0}^{w,c_2} + n_{u,s=0}^{l,c_2} + \lambda_1^{c_2}}{\sum_{s'} n_{u,s'}^{w,c_2} + \sum_{s'} n_{u,s'}^{l,c_2} + \lambda_1^{c_2} + \lambda_2^{c_2}} \\
\hat{\rho}_{u,s=1}^{c_2} &= \frac{n_{u,s=1}^{w,c_2} + n_{u,s=1}^{l,c_2} + \lambda_2^{c_2}}{\sum_{s'} n_{u,s'}^{w,c_2} + \sum_{s'} n_{u,s'}^{l,c_2} + \lambda_1^{c_2} + \lambda_2^{c_2}} \\
\hat{\theta}_{u,z}^{c_1} &= \frac{n_{u,z}^{w,s^w=0,c_1} + n_{u,z}^{v,s^v=0,c_1} + \alpha_1^{c_1}}{\sum_{z'} n_{u,z'}^{w,s^w=0,c_1} + \sum_{z'} n_{u,z'}^{v,s^v=0,c_1} + K\alpha_1^{c_1}} \\
\hat{\theta}'^{c_1}_{u,z} &= \frac{n_{u,z}^{w,s^w=1,c_1} + n_{u,z}^{v,s^v=1,c_1} + \alpha_2^{c_1}}{\sum_{z'} n_{u,z'}^{w,s^w=1,c_1} + \sum_{z'} n_{u,z'}^{v,s^v=1,c_1} + K^{c_1}\alpha_2^{c_1}} \\
\hat{\theta}_{u,z}^{c_2} &= \frac{n_{u,z}^{w,s^w=0,c_2} + n_{u,z}^{l,s^l=0,c_2} + \alpha_1^{c_2}}{\sum_{z'} n_{u,z'}^{w,s^w=0,c_2} + \sum_{z'} n_{u,z'}^{l,s^l=0,c_2} + K\alpha_1^{c_2}} \\
\hat{\theta}'^{c_2}_{u,z} &= \frac{n_{u,z}^{w,s^w=1,c_2} + n_{u,z}^{l,s^l=1,c_2} + \alpha_2^{c_2}}{\sum_{z'} n_{u,z'}^{w,s^w=1,c_2} + \sum_{z'} n_{u,z'}^{l,s^l=1,c_2} + K^{c_2}\alpha_2^{c_2}} \\
\hat{\psi}_{z,w} &= \frac{n_{z,w} + \beta}{\sum_{w'} n_{z,w'} + W\beta} \\
\hat{\psi}'^{c_m}_{z,w} &= \frac{n_{z,w}^{c_m} + \beta^{c_m}}{\sum_{w'} n_{z,w'}^{c_m} + W\beta^{c_m}} \\
\hat{\phi}_{z,v} &= \frac{n_{z,v} + \gamma}{\sum_{v'} n_{z,v'} + V\gamma} \\
\hat{\phi}'^{c_1}_{z,v} &= \frac{n_{z,v}^{c_1} + \gamma^{c_1}}{\sum_{v'} n_{z,v'}^{c_1} + V\gamma^{c_1}} \\
\hat{\varphi}_{z,l} &= \frac{n_{z,l} + \eta}{\sum_{l'} n_{z,l'} + L\eta} \\
\hat{\varphi}'^{c_2}_{z,l} &= \frac{n_{z,l}^{c_2} + \eta^{c_2}}{\sum_{l'}^{c_2} n_{z,l'} + V\eta^{c_2}}
\end{aligned} \quad (9)$$

where $m \in \{1, 2\}$. Let $s \in \{0, 1\}$, $x \in \{w, v, l\}$. $n_{u,s_i^w=s}^{w,c_m}$ is the number of times that $s_i^w = s$ is assigned to the user u from the social platform c_m. $n_{u,s_i^v=s}^{v,c_1}$ is the number of times that $s_i^v = s$ is assigned to the user u from the Flickr platform. $n_{u,s_i^l=s}^{l,c_2}$ is the number of times that $s_i^l = s$ is assigned to the user u from the Foursquare platform. $n_{u,z}^{x,s_i^x=s,c_m}$ is the number of times that the topic z represented by x is assigned to the user u in platform c_m. $n_{z,w}$ is the number of times that the word w is assigned to the shared topic z. $n_{z,w}^{c_m}$ is the number of times that the word w is assigned to the platform c_m-specific topic z. $n_{z,v}$ is the number of times that the visual word v is assigned to the shared topic z. $n_{z,v}^{c_1}$ is the number of times that the visual word v is assigned to the platform c_1-specific topic z. $n_{z,l}$ is the number of times that the venue l is assigned to the shared topic z. $n_{z,l}^{c_2}$ is the number of times that the venue l is assigned to the platform c_2-specific topic z.

IV. Location-Context Inter-Platform Recommendation

CM^{3}TM can be potentially applied to many new inter-platform applications based on the shared topics and their corresponding topic distributions of users learned from CM^{3}TM. In this section, we present the inter-platform recommendation applications by incorporating the location context. The task is described as follows: given a user u from a platform and a query city r, the goal is to recommend a ranked item list from the other platform to this user according to his/her interest. It includes the following two cases: 1) personalized venue recommendation from Foursquare to Flickr users, and 2) personalized image recommendation from Flickr to Foursquare users.

A. Personalized Venue Recommendation From Foursquare to Flickr Users

For a Flickr user u_{c_1} and a query city r, we introduce the shared topic variable z_k to rank the recommended venues from Foursquare by the following equation:

$$\begin{aligned} p(l_{c_2}|u_{c_1}, r) &= \sum\nolimits_{k\in K} p(l_{c_2}|z_k)p(z_k|u_{c_1}) \\ &= \begin{cases} \sum_{k\in K} \hat{\varphi}_{z_k,l_{c_2}} \hat{\theta}^{c_1}_{u,z_k}, & l_{c_2} \in L_r \\ 0, & l_{c_2} L_r \end{cases} \end{aligned}$$

where L_r is the set of venues in the city r.

B. Personalized Image Recommendation From Flickr to Foursquare Users

For a Foursquare user u_{c_2} and a query city r, the similarity between the user-topic distribution and the image-topic distribution is calculated by the following equation:

$$sim(u_{c_2}, I_{c_1}) = \begin{cases} \omega sim(\hat{\boldsymbol{\theta}}^{c_2}_{u_{c_2}}, \mathbf{d}^w_{I_{c_1}}) \\ \quad +(1-\omega) sim(\hat{\boldsymbol{\theta}}^{c_2}_{u_{c_2}}, \mathbf{d}^v_{I_{c_1}}), & geo_{I_{c_1}} \in Geo_r \\ 0, & geo_{I_{c_1}} \notin Geo_r \end{cases} \quad (11)$$

where ω is the weight parameter controlling the strength of the visual content and text content for the image I_{c_1}. $geo_{I_{c_1}}$ represents the geo-tag information of the image I_{c_1}. Geo_r represents the geographical area of the city r. $\hat{\theta}^{c_2}_{u_{c_2}}$ is the learned shared topic distribution of user u_{c_2}. $\mathbf{d}^w_{I_{c_1}}$ is the topic distribution of the image textual content $\mathbf{d}^w_{I_{c_1}} =< d^w_{I_{c_1},z_1}, \dots d^w_{I_{c_1},z_k}, \dots, d^w_{I_{c_1},z_K} >$. $\mathbf{d}^v_{I_{c_1}}$ is the topic distribution of the image visual content. $\mathbf{d}^v_{I_{c_1}} =< d^v_{I_{c_1},z_1}, \dots d^v_{I_{c_1},z_k}, \dots, d^v_{I_{c_1},z_K} >$

$$\begin{aligned} d^w_{I_{c_1},z_k} &= \frac{1}{n^w_{I_{c_1}}} \sum_{i=1}^{n^w_{I_{c_1}}} p(z_k|w_{I_{c_1},i}) \\ &= \frac{p(z_k)}{n^w_{I_{c_1}}} \sum_{i=1}^{n^w_{I_{c_1}}} \frac{p(w_{I_{c_1},i}|z_k)}{p(w_{I_{c_1},i})} \\ &= \frac{p(z_k)}{n^w_{I_{c_1}}} \sum_{i=1}^{n^w_{I_{c_1}}} \frac{\hat{\psi}_{z_k,w_{I_{c_1},i}}}{p(w_{I_{c_1},i})}. \end{aligned} \quad (12)$$

Similarly,

$$d^v_{I_{c_1},z_k} = \frac{p(z_k)}{n^v_{I_{c_1}}} \sum_{i=1}^{n^v_{I_{c_1}}} \frac{\hat{\phi}_{z_k,v_{I_{c_1},i}}}{p(v_{I_{c_1},i})} \quad (13)$$

where $n^w_{I_{c_1}}$ and $n^v_{I_{c_1}}$ indicate the number of textual words and visual words in image I_{c_1} respectively. $p(w_{I_{c_1},i})$ and $p(v_{I_{c_1},i})$ are the word prior distribution and visual word prior distribution. We assume that the visual word prior and tag word prior all follow the uniform distribution, i.e., $p(w_{I_{c_1},i}) = \frac{1}{W}$ and $p(v_{I_{c_1},i}) = \frac{1}{V}$. $p(z_k)$ is the topic prior distribution and is calculated as

$$\begin{aligned} p(z_k) &= \sum_u p(z_k|u)p(u) \\ &= \frac{\sum_{u\in U^{c_1}} \hat{\theta}^{c_1}_{u,z_k}(n^w_u + n^v_u) + \sum_{u\in U^{c_2}} \hat{\theta}^{c_2}_{u,z_k}(n^w_u + n^l_u)}{\sum_{u\in U^{c_1}}(n^w_u + n^v_u) + \sum_{u\in U^{c_2}}(n^w_u + n^l_u)} \end{aligned} \quad (14)$$

where n^w_u, n^v_u and n^l_u indicate the number of textual words, visual words and venues for user u, respectively.

V. Evaluation

In this section, we firstly describe the experimental setting including the dataset and implementation details. We then evaluate the performance of the proposed CM^{3}TM qualitatively and quantitatively. Finally, we verify the effectiveness of the proposed personalized inter-platform recommendation derived from CM^{3}TM.

A. Experimental Settings

1) Dataset: Two representative platforms are selected as the testbed: Foursquare, a famous location based social media platform and Flickr, a popular image-sharing social media platform. In order to conduct the evaluation of the location-context inter-platform recommendation, we select New York city as the test city and collect the data from the two platforms. The detailed procedure of data crawling is as follows.

- *Flickr*: To download enough images, we firstly use two kinds of queries: the name of the city New York and its corresponding GPS information to retrieve seed images respectively. Then, the owner's IDs of the seed images are retrieved. Based on the owner IDs, we download all the geotagged images through Flickr API.[3] For each image, we also crawl the associated metadata, including the tags, title, description, geo-tag information and favorite count. Note that we only focus on images taken in New York City and there is no requirement of the hometown of user IDs.
- *Foursquare*: Since the personal check-in information in Foursquare cannot be directly accessed, we turn to the Twitter streams,[4] where tweets containing check-in information are shared to the public. Following the work [46], we monitor the Twitter's streams and capture check-ins by crawling tweets with the keyword "4sq". Each tweet contains a short check-in message and a link pointing to the Foursquare check-in page, where we are able to retrieve more information related to the venues and users. In

[3][Online]. Available: https://www.flickr.com/services/api/

[4][Online]. Available: https://dev.twitter.com/docs/streaming-api

TABLE II
STATISTICS OF OUR COLLECTED DATA

♯ platform	♯ user	♯ image	♯ venue	♯ check-in	♯ tip
Flickr	11,996	856,795	-	-	-
Foursquare	13,143	-	40,166	144,805	516,655

total, our dataset includes 7,383,865 check-ins performed by 1,224,899 users at 2,509,641 venues from April 1 to July 1 in 2014. We firstly extract the check-in data from New York based on the check-in latitude and longitude. For Foursquare users corresponding to these check-ins in New York, we then select their check-in data in other cities. In addition, since leaving a tip in a venue can be considered as one check-in [3], we crawl their tips using Foursquare API,[5] which is open to the public. For each venue, we also download the venue information including the tags, latitude and longitude, the count of check-ins and the count of tips.

For each platform, we remove the users who have less than 8 records. The reason is that we focus on selecting the users with more information for common recommendation applications. The core of our inter-platform recommendation is to correlate different platforms based on shared topics from the user documents of all the platforms. User interests can be represented by the shared topics. For users with fewer records, their interests are hard to determine. Therefore, similar to [4], we do not consider them in our work. For the text information from Foursquare (venue tags and tips) and Flickr (tags, the title and the description), we clean them by removing the stopwords, html tags and camera related words such as "Cannon" and "35 mm" [12]. After the data preprocessing, the number of unique words from two platforms is 22, 532. The statistics of the resulting dataset are shown in Table II.

2) Implementation details: For each image from Flickr, we choose to represent the visual content by region-level Maximally Stable Extremal Region (MSER) features. Compared with the keypoint based descriptors, MSER regions indicate local homogeneous parts in objects and show higher distinctness [24]. For this reason, a lot of work, such as [9], [30] used MSER for image feature extraction. In our implementation, we compute one SIFT descriptor in each detected elliptical region from MSER. About 447,030 MSER descriptors are extracted from 20,000 sample images, which are further quantized to constitute a dictionary of 1,024 visual words. As for venues from Foursquare, similar to [16], each user's venues are aggregated as the venue document, where each venue is considered as one "word".

As for the hyper-paramerters of the model, without any prior knowledge, we empirically set the fixed values, i.e., $\alpha_1^{c_1} = \alpha_2^{c_1} = \alpha_1^{c_2} = \alpha_2^{c_2} = 1.0$, $\lambda_1^{c_1} = \lambda_2^{c_1} = \lambda_1^{c_2} = \lambda_2^{c_2} = 1.0$, $\beta = \beta^{c_1} = \beta^{c_2} = 0.01$, $\gamma = \gamma^{c_1} = 0.01$, $\eta = \eta^{c_2} = 0.01$.

B. Evaluation of CM^{3}TM

1) Qualitative Evaluation: We demonstrate the effectiveness of CM^{3}TM by examining the discovered topics. We visualize them by providing the five top-ranked words, the top five related images and the top five relevant venues. Each of the shared

[5][Online]. Available: https://developer.foursquare.com/docs/users/tips

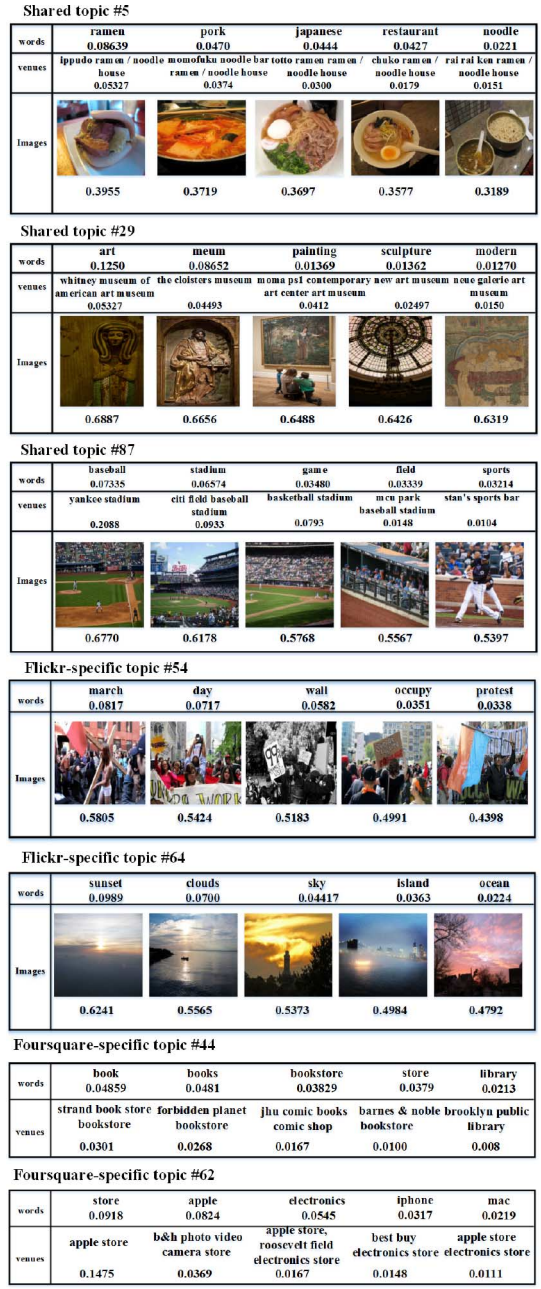


Fig. 4. Illustration of three kinds of discovered topics by CM^{3}TM.

topics is represented by relevant words, images and venues. The Flickr-specific topic is represented by relevant words and images. The Foursquare-specific topic is represented by relevant words and venues. Fig. 4 illustrates some discovered topics including three shared topics, two Flickr-specific topics and two Foursquare-specific topics. The textual words are sorted by their probabilities generated from the corresponding topic $p(w_i|z_k)$, while the venues are sorted by $p(l_i|z_k)$. The images are sorted by the weighted sum of cosine similarity between the topic Z_i and the image I_j

$$sim(Z_i, I_j) = \nu \frac{\mathbf{w}_{z_i}\mathbf{w}_{I_j}}{\|\mathbf{w}_{z_i}\|\|\mathbf{w}_{I_j}\|} + (1-\nu)\frac{\mathbf{v}_{z_i}\mathbf{v}_{I_j}}{\|\mathbf{v}_{z_i}\|\|\mathbf{v}_{I_j}\|} \quad (15)$$

where ν is the weight parameter (in our experiment, $\nu = 0.6$). $\mathbf{w}_{z_i}$ and $\mathbf{v}_{z_i}$ represent the learned topic-textual word distribu-

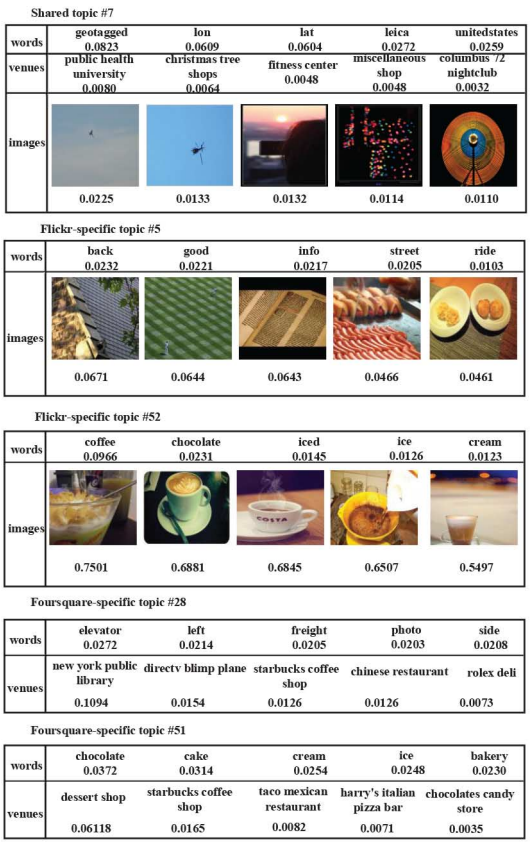


Fig. 5. Illustration of negative examples from CM^{3}TM.

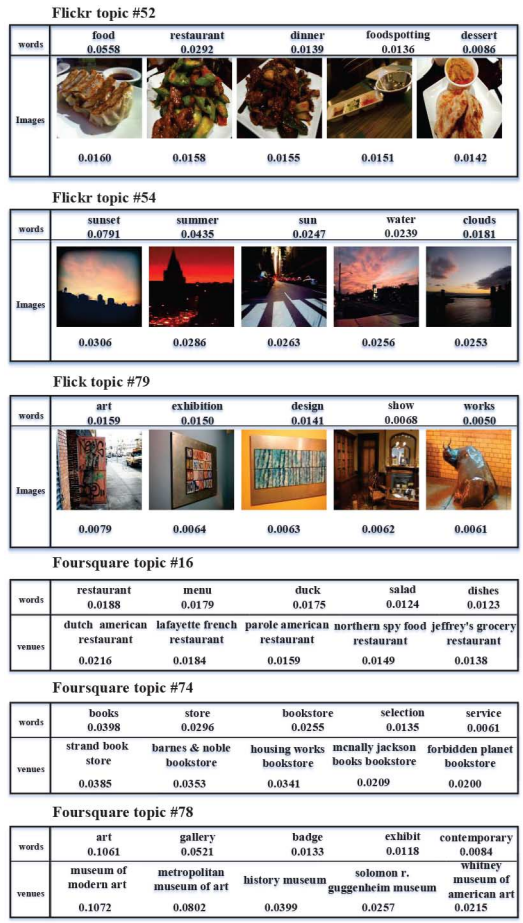


Fig. 6. Illustration of discovered topics from Flickr and Foursquare.

tion and topic-visual word distribution, respectively. $\mathbf{w}_{I_j}$ and $\mathbf{v}_{I_j}$ represent the normalized vector with Term Frequency (TF) on textual words and visual words of image I_j, respectively. As shown in Fig. 4, we can see that some shared topics are about food (shared topic ♯5), art (shared topic ♯29) and sports (shared topic ♯87). These topics are shared by all platforms. In contrast, the Flickr-specific topics include political events (Flickr-specific topic ♯54) and the nature (Flickr-specific topic ♯64), which generally are not in Foursquare. The Foursquare-specific topics are referred to ones such as reading (Foursquare-specific topic ♯44) and gadget freaks (Foursquare-specific topic ♯62). By providing a combination of representative words, images and venues, it becomes very easy to interpret the domain knowledge associated with each topic. We can see that (1) there exist three kinds of topics: shared topics, Flickr-specific topics and Foursquare-specific topics in two platforms and (2) by considering the textual words, visual image content and venues, the discovered topics show high consistency among the semantic concepts, visual themes and venue themes.

In addition, we illustrate some negative examples in Fig. 5. We observe that (1) some topics are meaningless or lack interpretability, such as shared topic ♯7, Flickr-specific topic ♯5 and Foursquare-specific topic ♯28; and (2) there are relevant topics between Flickr-specific topics and Foursquare-specific topics, e.g., Flickr-specific topic ♯52 and Foursquare-specific topic ♯51. These topics are about food and actually are shared topics. The possible reasons are as follows: (1) the topic model does not discover topics based on the semantics but the statistical patterns and the mined patterns are not always meaningful; (2) since the topic model is a probabilistic model, the three kinds of topics are split based on the probability, and thus are not fully independent; and (3) the number of topics is predefined and is generally not consistent with the number of actual topics. In practice, since our cross-platform recommendation is based on the shared topics, we generally set the number of topics larger so that all the shared topics are discovered from two platforms. In this case, there are possibly more relevant topics between two platform-specific topics. However, they do not affect our recommendation performance.

In order to further verify the effectiveness of our model, we have conducted an experiment of topic modeling for each platform, respectively. We use the conventional multi-modal topic model [6] for topic modeling on the multi-modal data from each platform. The number of topics is set as 100 for each platform. For the discovered topics, similar visualization and sorting strategies are adopted. Fig. 6 illustrates three discovered topics for each platform. We observe that there are some topics with similar semantics between Flickr and Foursquare. For example, Flickr topic ♯52 and Foursquare topic ♯16 are about food while Flickr topic ♯79 and Foursquare topic ♯78 are relevant to art. Compared with the results from Fig. 4, we can see that our model is indeed capable of correlating these similar topics between Flickr and Foursquare and discovering these shared topics from two platforms.

2) Quantitative Evaluation: We resort to the perplexity as the metric, which is a standard quantitative measure in topic

modeling literature to compare the performance of various topic models [7]. The lower the perplexity score is, the better the generalizability of the topic model is. In our evaluation, we adopt the caption perplexity [6] as the performance measure, which is suitable for evaluating the topic model with the multi-modal information and is defined as

$$perplexity(D_{test}) = \exp\left(-\frac{\sum_{d\in D_{test}} \ln p(\mathbf{w}_d|\mathbf{v}_d)}{\sum_{d\in D_{test}} n_d^w}\right) \tag{16}$$

where D_{test} is the test set, $\mathbf{w}_d$ is the textual word vector of the text document and $\mathbf{v}_d$ is the visual word vector of the image.

We define the variant of the caption perplexity for CM^{3}TM as

$$perplexity(D_{test}) = \exp\left(-\frac{\sum_{u\in U_{test}^{c_1}} \ln p(\mathbf{w}_u|\mathbf{v}_u) + \sum_{u\in U_{test}^{c_2}} \ln p(\mathbf{w}_u|\mathbf{l}_u)}{\sum_{u\in U_{test}} n_u^w}\right) \tag{17}$$

$$p(\mathbf{w}_u|\mathbf{v}_u) = \prod_{w_u\in\mathbf{w}_u} \sum_{k\in K_1} p(w_u|z_k)p(z_k|\mathbf{v}_u)$$
$$p(\mathbf{w}_u|\mathbf{l}_u) = \prod_{w_u\in\mathbf{w}_u} \sum_{k\in K_2} p(w_u|z_k)p(z_k|\mathbf{l}_u) \tag{18}$$

where $\mathbf{w}_u$, $\mathbf{v}_u$, $\mathbf{l}_u$ represent the textual word vector, visual word vector and venue vector of user u, respectively. After model training, we obtain the topic-textual word distribution $p(w_u|z_k)$, topic-visual word distribution $p(v_u|z_k)$ and topic-venue distribution $p(l_u|z_k)$. Similar to [8], the estimation of $p(z_k|\mathbf{v}_u)$ can be approximated by running Gibbs sampling over all the extracted visual words from the images of the test user (no words used) using $p(v_u|z_k)$. Similarly, the estimation of $p(z_k|\mathbf{l}_u)$ can be approximated by running Gibbs sampling over all the extracted venues from the venue documents of the test user (no words used) using $p(l_u|z_k)$. K_1 and K_2 are the sum of topics from Flickr and Foursquare, respectively. In CM^{3}TM, $K_1 = K + K^{c_1}$, $K_2 = K + K^{c_2}$.

We design the following baseline for comparison: *Basic Version*: **CM^{3}TM_BV**. This model does not distinguish between shared topics and platform-specific topics and considers that the two platforms share the same topic space.

We randomly divide the dataset into two parts: the training set and the test set. We randomly select 80% of the users from each platform and their data as the training data. The remaining 20% users and their data are used as the test data. Without any prior knowledge, we simply set the proportion of the shared topics in each platform to 0.5, that is $K = K^{c_1} = K^{c_2}$. Fig. 7 shows the perplexity of the test set for different values of topic number $K^{all} = K + K^{c_1} + K^{c_2}$. We can see that the proposed method, which is more flexible than the baseline CM^{3}TM_BV, reaches better (lower) perplexity on the test data due to being able to describe both shared topics and platform-specific topics without needing to introduce noisy topics. Distinguishing shared topics from platform-specific topics enables to better separate irrelevant or inconsistent topic knowledge in different platforms. In addition, the better caption perplexity also means our model

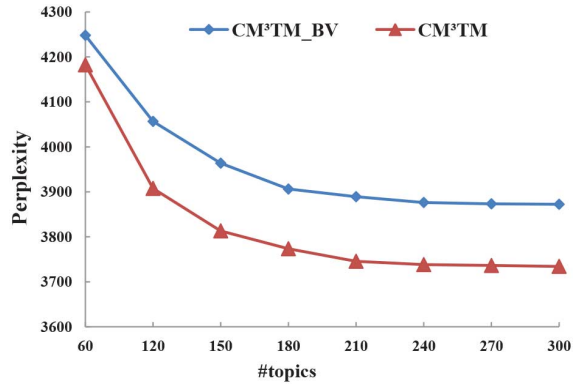


Fig. 7. Perplexity comparison. (a) Precision@K. (b) MAP@K.

achieves higher consistency among topics of different modalities than the baseline. It is clear that the perplexity is stable when $K^{all} = 240$ (that is $K = K^{c_1} = K^{c_2} = 80$), and therefore we choose the desired topic number $K = K^{c_1} = K^{c_2} = 80$ for CM^{3}TM in the following experiments.

3) Computational Complexity Analysis: Without loss of generality, let $K = K^{c_1} = K^{c_2}$ and the iteration number of Gibbs sampling is I. In each iteration, it requires to go through all users U from both platforms. For each user $u \in U^{c_1}$, it needs $O((W_u + V_u)K)$ operations to calculate (1)–(8). Similarly, for each user $u \in U^{c_2}$, it needs $O((W_u + L_u)K)$ for one update. V_u, L_u and W_u are the total count of words, venues and visual words in user document u, respectively. Thus, the whole time complexity of CM^{3}TM is $O(U^{c_1}(W_u + V_u)KI + U^{c_2}(W_u + L_u)KI) = O((UW_u + U^{c_1}V_u + U^{c_2}L_u)KI)$. In our experiment, the training procedure on the whole dataset includes 1,000 iterations of Gibbs sampling and lasts for approximately 462,000 seconds (about 5 days) on a PC with Intel Core 3.40 GHz processor. Thus the average time cost of each iteration is about 462 seconds.

For the scalability of our model, when this model scales to another dataset, similar to the conventional topic models, the number and the value of prior parameters are not changed. We need to enlarge the number of topics to discover more topics from the large dataset. When the number of topics increases, training our model takes more time. In this case, we possibly resort to fast samplings methods (e.g., [25], [48]) for speeding up the model training.

C. Evaluation of Personalized Inter-Platform Recommendation

1) Evaluation Strategies: The purpose of personalized inter-platform recommendation is to recommend users in one platform with items from the other platform according to users' interests. The ideal evaluation strategy is to compare the recommended item list with the actual information left by users in this platform. However, it is hard to verify the accounts in two platforms from the same user, hence we resort to an approximate evaluation.

For personalized venue recommendation to Flickr users, we asked 20 users to label the returned venue list. The 20 users are all students, including 10 males and 10 females. Their ages are from 20 to 30. For guaranteeing the fairness of the experiment, there are no authors in the user group. They can consult

the uploaded image set of this Flickr user to help make judgement. For one venue, if more than 10 users thought it is relevant to this Flickr user, the venue is annotated with label 1, and 0 otherwise. Similarly, for personalized image recommendation to Foursquare users, we also asked these 20 users to label the returned image list by consulting the check-in records of this Foursquare user. For one image, if more than 10 users thought it is relevant to this Foursquare user, the image is annotated with label 1, and 0 otherwise. Note that when crawling data, we found that some venue images exist in Flickr. In order to avoid the bias on the groundtruth generation, we do not show any venue image in Flickr to the annotators. There are two steps to find the venue images in Flickr: (1) all the Flickr images tagged with "Foursquare"; (2) all the Flickr images tagged with the venue name listed on our built venue vocabulary.

2) Evaluation Metrics: The aim of the personalized inter-platform recommendation is to provide each user a ranked list of items. Similar to traditional information retrieval tasks, we use Precision@*K*, MAP@*K* to measure the quality of the ranked list of recommended items.

For a given query $u \in U$, Precision@K is defined as

$$Precision@K = \frac{\sum_{k=1}^{K} r(k)}{K}. \tag{19}$$

MAP@*K* is the mean of average precision scores over test users U and is defined as

$$MAP@K = \frac{1}{U}\sum_{u=1}^{U} \frac{\sum_{k=1}^{K} Precision@uk * r(uk)}{\sum_{k=1}^{K} r(uk)} \tag{20}$$

where r_k is the relevance level at position k, which is 0 for "Not Relevant" and 1 for "Relevant" in our experiment. r_{uk} is the relevance level at position k for user u. Precision@uk is the precision at position k for user u, and K is the truncation level. In our study, we set $K \in \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$.

After model learning, we randomly select 1,000 users as the test queries for each platform, where each query includes the topic distribution of the query user and the query city New York. We rank venues and images using (10) and (11), respectively. In personalized image recommendation to Foursquare users, we set the weight $\omega = 0.6$.

3) Baselines: Since it is hard to obtain common users across these two platforms, we cannot compare our method with existing methods, which use the overlapping users as the bridge. Therefore, we consider the following baselines for comparison:

- *Popularity (POP)*: This approach provides the same recommendation list of items to all users according to this item's popularity. In Foursquare, let $score\_l_j$ be the popularity of venue j, then

$$score\_l_j = \tau \sharp check\_in_j + (1-\tau)\sharp user\_tip_j \tag{21}$$

  where $\sharp check\_in_j$ is the number of check-ins at venue j and $\sharp user\_tip_j$ is the number of tips left at venue j. $\tau = 0.5$.

  In Flickr, the number of favorites explicitly shows how many people like the image. Hence it is a straightforward reflection of the photo's degree of popularity. Let $score\_p_j$ be the popularity score of image p_j, then

$$score\_p_j = \sharp favorite_j \tag{22}$$

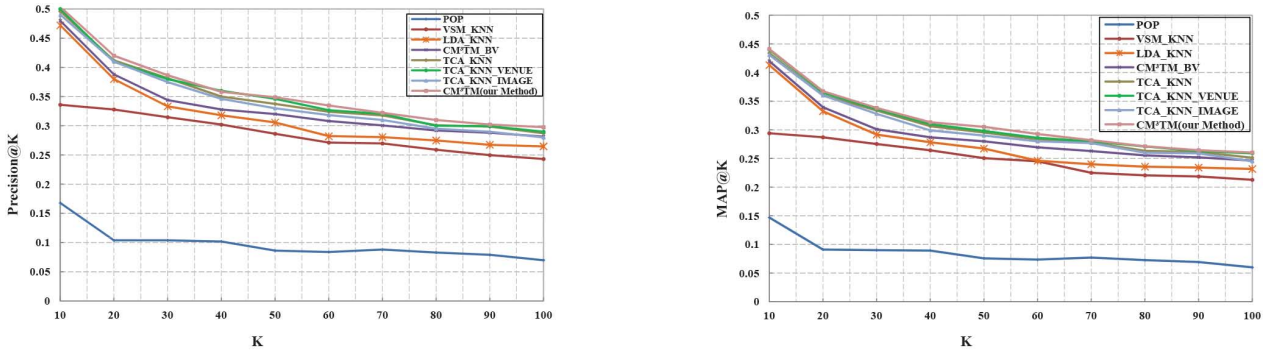
  where $\sharp favorite_j$ is the number of favorites for this image j.
- *Vector Space Model-Based K-Nearest Neighbors Algorithm (VSM_KNN)*: This method utilizes the user activity history and item's text content to create a text vector with TF-IDF. Particularly, for each Flickr user, we aggregate the text of uploaded images to one user document. Similarly, the tags and user tips of check-in venues by one Foursquare user are aggregated to one user document. Each image and each venue are also represented by relevant text information. For Flickr users, VSM_KNN retrieves all venues to find k nearest neighbors by computing the cosine similarity between the user's text vector and venue's text vector as follows:

$$sim(u_{c_1}, l_{c_2}) = \frac{\mathbf{w}_{u_{c_1}}\mathbf{w}_{l_{c_2}}}{\|\mathbf{w}_{u_{c_1}}\|\|\mathbf{w}_{l_{c_2}}\|}. \tag{23}$$

  Similarly, for Foursquare users, VSM_KNN retrieves all images to find *k* nearest neighbors by computing the cosine similarity between the user's text vector and image's text vector as follows:

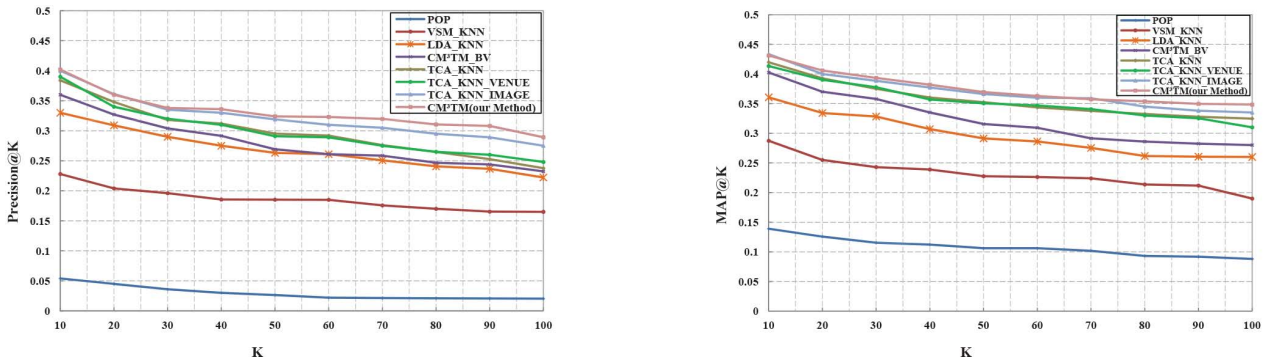
$$sim(u_{c_2}, I_{c_1}) = \frac{\mathbf{w}_{u_{c_2}}\mathbf{w}_{I_{c_1}}}{\|\mathbf{w}_{u_{c_2}}\|\|\mathbf{w}_{I_{c_1}}\|} \tag{24}$$

  where $\mathbf{w}_{u_{c_1}}$, $\mathbf{w}_{I_{c_1}}$, $\mathbf{w}_{u_{c_2}}$ and $\mathbf{w}_{l_{c_2}}$ represent the feature vector of the Flickr user, Flickr image, Foursquare user and Foursquare venue respectively.
- *LDA-Based K-Nearest Neighbors Algorithm (LDA_KNN)*: LDA [7] is used to obtain the topic distribution of each user and item. This model only considers the text information and does not distinguish between shared topics and platform-specific topics. The cosine similarity is computed similar to (23) and (24).
- $\mathbf{CM^3TM\_BV}$: Compared with *LDA_KNN*, this model incorporates the multi-modal information. The recommendation strategy is similar to CM^{3}TM and the similarity is computed according to (10) and (11).
- *TCA-Based K-Nearest Neighbors Algorithm (TCA_KNN)* [20]: Compared with *LDA_KNN*, this model can distinguish between shared topics and platform-specific topics. Each user and each item is represented by their corresponding shared topic distributions. The cosine similarity is computed similar to (23) and (24).
- *TCA_KNN_VENUE*: Compared with TCA_KNN, we introduce the venue modality from Foursquare in this model. Therefore, the input data from Flickr in this model are the same as TCA_KNN while the input data from Foursquare in this model are the same as CM^{3}TM. For image recommendation from Flickr to Foursquare users, each user and each image item is represented by their corresponding shared topic distributions and the cosine similarity between their topic distributions is computed via (23). As for venue recommendation from Foursquare to Flickr users, we can use (10) to obtain the ranked venue list.



(a) (b)

Fig. 8. Precision and MAP of personalized venue recommendation from Foursquare to Flickr users. (a) Precision@K. (b) MAP@K.



(a) (b)

Fig. 9. Precision and MAP of personalized image recommendation from Flickr to Foursquare users. (a) Precision@K. (b) MAP@K.

- *TCA_KNN_IMAGE*: Compared with TCA_KNN, we introduce the image modality from Flickr in this model. Therefore, the input data from Foursquare in this model are the same as TCA_KNN while the input data from Flickr in this model are the same as our proposed CM^{3}TM. For image recommendation from Flickr to Foursquare users, we can use (11) to obtain the ranked image list for each user. As for venue recommendation from Foursquare to Flickr users, each user and each venue item is represented by their corresponding shared topic distributions and the cosine similarity between their topic distributions is computed via (24).

4) Results Analysis: Here, we report the performance comparison for both personalized venue recommendation from Foursquare to Flickr users and personalized image recommendation from Flickr to Foursquare users. Figs. 8 and 9 show Precision@K and MAP@K for two settings. As shown in Fig. 8 in the first setting, firstly, POP and VSM_KNN drop behind four other model-based methods, showing the advantage of using latent topic models to model user's interest and produce recommendations. Secondly, CM^{3}TM_BV outperforms LDA_KNN, justifying the benefit brought by the consistency between textual topics and venue topics. TCA_KNN further improves the performance compared with CM^{3}TM_BV and LDA_KNN. This is because TCA_KNN can distinguish shared topics from platform-specific topics and filters platform-specific topics as the irrelevant information. This guarantees the consistency of the transferred knowledge across two platforms. Similar results can be observed in personalized image recommendation from Flickr to Foursquare users in Fig. 9. Compared with TCA_KNN, TCA_KNN_VENUE further improves the venue recommendation performance and its performance is comparable to our proposed CM^{3}TM. This is because TCA_KNN_VENUE incorporates the venue modality into the topic model and can align both the venue modality and textual modality for each shared topic. The ranked venue list is derived directly based on the shared topics without additional similarity computation. As for image recommendation, the TCA_KNN_VENUE does not introduce the visual modality into the model. Therefore, there is no essential improvement on the performance of image recommendation, which is similar to the the performance of TCA_KNN. Similar analysis is conducted on TCA_KNN_IMAGE. On one hand, TCA_KNN_IMAGE outperforms TCA_KNN in the image recommendation and is comparable to our proposed model. On the other hand, there is no essential improvement on the venue recommendation from Foursquare to Flickr users. Finally, in both recommendation settings, our proposed CM^{3}TM yields better performance than other baselines in terms of precision and MAP. The reason is that our method is capable of aligning the textual modality, visual modality and venue modality for each shared topic from two platforms. This guarantees the consistency among the textual topics, visual topics and venue topics.

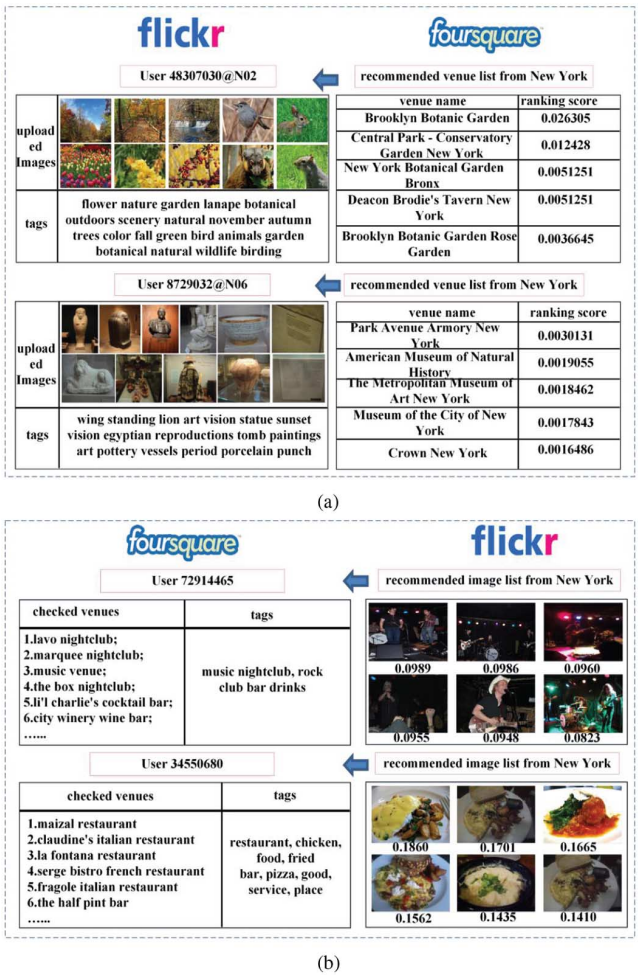


Fig. 10. Case study on personalized inter-platform recommendation. (a) Two examples on personalized venue recommendation from Foursquare to Flickr users. (b) Two examples on personalized image recommendation from Flickr to Foursquare users.

5) Qualitative Case Study: We demonstrate the effectiveness of the proposed inter-platform recommendation by providing some users' recommended results, including the following two settings: (1) personalized venue recommendation from Foursquare to Flickr users [Fig. 10(a)] and (2) personalized image recommendation from Flickr to Foursquare users [Fig. 10(b)]. In (1), Fig. 10(a) shows two test Flickr users with their shared images and corresponding recommended venue list from Foursquare in New York. Since users can express their interests by uploading their images, we show corresponding uploaded images and tags. Take the Flickr user "48307030@N02" as an example, we can see that this Flickr user likes the scenic views and animals. The corresponding recommended venue list from Foursquare includes some parks, which better satisfies the interest of the Flickr user. Similarly, in (2), Fig. 10(b) also shows two test Foursquare users and corresponding recommended image list from Flickr. Take Foursquare user "72914465" as an example, this Foursquare user often checks in at venues of nightclub and music venues. We roughly conclude that he/she is more interested in music. Our method correspondingly recommends this Foursquare user some images about the music concert.

VI. Conclusion

In this paper, we have proposed a cross-platform multi-modal probabilistic model (CM^{3}TM) to address the inter-platform recommendation problem. CM^{3}TM is capable of differentiating between shared topics and platform-specific topics, and aligning topics on different modalities across different platforms. CM^{3}TM is very flexible and can be generalized to other cross-platform recommendation problems in the social multimedia research. The evaluation has demonstrated its effectiveness in connecting different platforms with different modalities through our introduced location-context personalized inter-platform recommendation applications.

In the future, we will be working towards two directions: 1) based on the proposed generative model, more real applications could be designed, such as inter-platform friend suggestion and 2) we plan to extend our model by introducing the social relationships (e.g., friendship) [40], [36]. The content information and social relationship information from different platforms are complementary and can together enhance the performance of the model.

References

[1] B.-K. Bao, W. Min, J. Sang, and C. Xu, "Multimedia news digger on emerging topics from social streams," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 1357–1358.

[2] B.-K. Bao, C. Xu, W. Min, and M. S. Hossain, "Cross-platform emerging topic detection and elaboration from multimedia streams," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 11, no. 4, pp. 54:1–54:21, 2015.

[3] J. Bao, Y. Zheng, and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in *Proc. 20th Int. Conf. Adv. Geographic Inf. Syst.*, 2012, pp. 199–208.

[4] J. Bao, Y. Zheng, and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in *Proc. 20th Int. Conf. Adv. Geographic Inf. Syst.*, 2012, pp. 199–208.

[5] D. Blei, L. Carin, and D. Dunson, "Probabilistic topic models," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 55–65, Nov. 2010.

[6] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2003, pp. 127–134.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[8] X. Chen, C. Lu, Y. An, and P. Achananuparp, "Probabilistic models for topic learning from images and captions in online biomedical literatures," in *Proc. 18th ACM Conf. Inf. Knowl. Manage.*, 2009, pp. 495–504.

[9] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (MSER) tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 2006, vol. 1, pp. 553–560.

[10] J. Fan, D. A. Keim, Y. Gao, H. Luo, and Z. Li, "JustClick: Personalized image recommendation via exploratory search from large-scale Flickr images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 273–288, Feb. 2009.

[11] I. Fernández-Tobas, I. Cantador, M. Kaminskas, and F. Ricci, "Cross-domain recommender systems: A survey of the state of the art," in *Proc. Spanish Conf. Inf. Retrieval*, 2012, pp. 187–198.

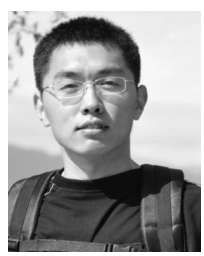
[12] Y. Gao *et al.*, "W2go: A travel guidance system by automatic landmark ranking," in *Proc. Int. Conf. Multimedia*, 2010, pp. 123–132.

[13] T. L. Griffiths and M. Steyvers, "Finding scientific topics," in *Proc. Nat. Academy Sci. USA*, 2004, vol. 101, no. Suppl 1, pp. 5228–5235.

[14] M. D. Hoffman, D. M. Blei, and P. R. Cook, "Content-based musical similarity computation using the hierarchical Dirichlet process," in *Proc. ISMIR*, 2008, pp. 349–354.

[15] M. Jiang *et al.*, "Social contextual recommendation," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, 2012, pp. 45–54.

[16] K. Joseph, C. H. Tan, and K. M. Carley, "Beyond local, categories and friends: Clustering foursquare users with latent topics," in *Proc. 2012 ACM Conf. Ubiquitous Comput.*, 2012, pp. 919–926.
[17] B. Li, "Cross-domain collaborative filtering: A brief survey," in *Proc. 23rd IEEE Int. Conf. Tools Artificial Intell.*, Nov. 2011, pp. 1085–1086.
[18] B. Li, Q. Yang, and X. Xue, "Can movies and books collaborate? Cross-domain collaborative filtering for sparsity reduction," in *Proc. 21st Int. Joint Conf. Artificial Intell.*, 2009, pp. 2052–2057.
[19] B. Li, Q. Yang, and X. Xue, "Transfer learning for collaborative filtering via a rating-matrix generative model," in *Proc. 26th Annu. Int. Conf. Machine Learn.*, 2009, pp. 617–624.
[20] L. Li, X. Jin, and M. Long, "Topic correlation analysis for cross-domain text classification," in *Proc. AAAI*, 2012, pp. 998–1004.
[21] T. Li, S. Yan, T. Mei, X.-S. Hua, and I.-S. Kweon, "Image decomposition with multilabel context: Algorithms and applications," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2301–2314, Aug. 2011.
[22] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1043–1051.
[23] J. Liu, Z. Huang, L. Chen, H. T. Shen, and Z. Yan, "Discovering areas of interest with geo-tagged images and check-ins," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 589–598.
[24] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[25] D. Newman, A. Asuncion, P. Smyth, and M. Welling, "Distributed algorithms for topic models," *J. Mach. Learn. Res.*, vol. 10, pp. 1801–1828, 2009.
[26] W. Pan, N. N. Liu, E. W. Xiang, and Q. Yang, "Transfer learning to predict missing ratings via heterogeneous user feedbacks," in *Proc. Int. Joint Conf. Artificial Intell.*, 2011, vol. 22, p. 2318.
[27] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang, "Transfer learning in collaborative filtering for sparsity reduction," in *Proc. 24th Conf. Artificial Intell.*, 2010, vol. 10, pp. 230–235.
[28] G.-J. Qi, C. C. Aggarwal, and T. Huang, "Link prediction across networks by biased cross-network sampling," in *Proc. IEEE 29th Int. Conf. Data Eng.*, Apr. 2013, pp. 793–804.
[29] S. D. Roy, T. Mei, W. Zeng, and S. Li, "Socialtransfer: Cross-domain transfer learning from social streams for media applications," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 649–658.
[30] J. Sang and C. Xu, "Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 19–28.
[31] Y. Shi, P. Serdyukov, A. Hanjalic, and M. Larson, "Nontrivial landmark recommendation using geotagged photos," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 3, pp. 47:1–47:27, 2013.
[32] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 448–456.
[33] M. Yan, J. Sang, T. Mei, and C. Xu, "Friend transfer: Cold-start friend recommendation with cross-platform transfer learning of social knowledge," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2013, pp. 1–6.
[34] T. Yao, C.-W. Ngo, and T. Mei, "Context-based friend suggestion in online photo-sharing community," in *Proc. 19th ACM Int. Conf. Multimedia*, 2011, pp. 945–948.
[35] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2011, pp. 325–334.
[36] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2011, pp. 325–334.
[37] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen, "Lcars: A location-content-aware recommender system," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery data Mining*, 2013, pp. 221–229.
[38] J. J.-C. Ying, E. H.-C. Lu, W.-N. Kuo, and V. S. Tseng, "Urban point-of-interest recommendation by mining user check-in behaviors," in *Proc. ACM SIGKDD Int. Workshop Urban Comput.*, 2012, pp. 63–70.
[39] N. J. Yuan *et al.*, "We know how you live: Exploring the spectrum of urban lifestyles," in *Proc. 1st ACM Conf. Online Social Netw.*, 2013, pp. 3–14.
[40] Q. Yuan, L. Chen, and S. Zhao, "Factorization vs. regularization: Fusing heterogeneous social relationships in top-n recommendation," in *Proc. 5th ACM Conf. Recommender Syst.*, 2011, pp. 245–252.
[41] J. Zahálka, S. Rudinac, and M. Worring, "New Yorker melange: Interactive brew of personalized venue recommendations," in *Proc. ACM Int. Conf. Multimedia*, 2014, pp. 205–208.
[42] F. Zhang, N. J. Yuan, D. Lian, and X. Xie, "Mining novelty-seeking trait across heterogeneous domains," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 373–384.
[43] J. Zhang, X. Kong, and P. S. Yu, "Predicting social links for new users across aligned heterogeneous social networks," in *Proc. IEEE 13th Int. Conf. Data Mining*, 2013, pp. 1289–1294.
[44] J. Zhang, X. Kong, and P. S. Yu, "Transferring heterogeneous links across location-based social networks," in *Proc. 7th ACM Int. Conf. Web Search Data Mining*, 2014, pp. 303–312.
[45] Y. Zhang, B. Cao, and D.-Y. Yeung, "Multi-domain collaborative filtering," *CoRR*, vol. abs/1203.3535, 2012 [Online]. Available: http://arxiv.org/abs/1203.3535
[46] Y.-L. Zhao, L. Nie, X. Wang, and T.-S. Chua, "Personalized recommendations of locally interesting venues to tourists via cross-region community matching," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 50:1–50:26, 2014.
[47] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with GPS history data," in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 1029–1038.
[48] X. Zheng, J. K. Kim, Q. Ho, and E. P. Xing, "Model-parallel inference for big topic models," *CoRR*, vol. abs/1411.2305, 2014 [Online]. Available: http://arxiv.org/abs/1411.2305
[49] E. Zhong, W. Fan, J. Wang, L. Xiao, and Y. Li, "Comsoc: Adaptive transfer of user behaviors over composite social network," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 696–704.
[50] E. Zhong, E. W. Xiang, W. Fan, N. N. Liu, and Q. Yang, "Friendship prediction in composite social networks," *CoRR*, vol. abs/1402.4033, 2014 [Online]. Available: http://arxiv.org/abs/1402.4033



Weiqing Min received the B.E. degree in communication engineering from Shandong Normal University, Jinan, China, in 2008, the M.E. degree in communication and information systems 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, Beijing, China, in 2015.

He is currently a Postdoc with the Institute of Computing Technology, Chinese Academy of Sciences. His current research interests include landmark-based multimedia search and mining and visual recognition in social media.

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 Assistant Researcher with the Institute of Automation, Chinese Academy of Sciences, Beijing, China, and a Researcher at the China–Singapore Institute of Digital Media, Singapore. Her current research interests include cross-media cross-modal image search, social event detection, image classification and annotation, and sparse/low rank representation.

Dr. Bao served as a Technical Program Committee Member for several international conferences, such as MMM2013, ICME2014, and PCM2013. She also served as a Special Session Organizer for MMM2013 and PCM2013. She was a Guest Editor for the *Multimedia System Journal* and *Multimedia Tools and Applications*. She was the recipient of the Best Paper Award from ICIMCS'09.

Changsheng Xu (M'97–SM'99–F'14) is a Professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, and Executive Director of the China–Singapore Institute of Digital Media, Singapore. He has authored or coauthored over 200 refereed research papers. He holds 30 granted/pending patents. His research interests include multimedia content analysis/indexing/retrieval, pattern recognition, and computer vision.

Dr. Xu is an IAPR Fellow and an ACM Distinguished Scientist. He is an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, the *ACM Transactions on Multimedia Computing, Communications and Applications*, and the *ACM/Springer Multimedia Systems Journal*. He served as Program Chair of ACM Multimedia 2009. He has served as Associate Editor, Guest Editor, General Chair, Program Chair, Area/Track Chair, Special Session Organizer, Session Chair, and TPC Member for over 20 IEEE and ACM multimedia journals, conferences, and workshops. He was the recipient of the Best Associate Editor Award of the *ACM Transactions on Multimedia Computing, Communications and Applications* in 2012 and the Best Editorial Member Award of the *ACM/Springer Multimedia Systems Journal* in 2008.

M. Shamim Hossain (S'03–M'07–SM'09) received the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Ottawa, ON, Canada, in 2009.

He is an Associate Professor with King Saud University, Riyadh, Saudi Arabia. He has authored or coauthored more than 70 publications, including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. His research interests include serious games, cloud and multimedia for health care, resource provisioning for big data processing on media clouds, and biologically inspired approaches for multimedia and software system.

Dr. Hossain is a Member of the ACM and ACM SIGMM. He has served as a Member of the Organizing and Technical Committees of several international conferences and workshops. He served as a Co-Chair of the 1st, 2nd, 3rd, 4th, and 5th IEEE ICME Workshop on Multimedia Services and Tools for E-Health. He served as a Co-Chair of the 1st Cloud-Based Multimedia Services and Tools for E-Health Workshop 2012 with ACM Multimedia. He currently serves as a Co-Chair of the 4th IEEE ICME Workshop on Multimedia Services and Tools for E-Health. He is on the Editorial Board of the *Springer International Journal of Multimedia Tools and Applications*. He was on the Editorial Board of a number of journals, including the *International Journal of Advanced Media and Communication*. Previously, he served as a Guest Editor of the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE and *Springer Multimedia Tools and Applications*. Currently, he serves as a Lead Guest Editor of the *Elsevier Future Generation Computer Systems*, *International Journal of Distributed Sensor Networks*, and *Springer Cluster Computing*.