

密级 \_\_\_\_\_



**中国科学院大学**  
University of Chinese Academy of Sciences

# 博士学位论文

社交媒体下地理数据的处理与应用

作者姓名 \_\_\_\_\_ 闵巍庆

指导教师 \_\_\_\_\_ 徐常胜 研究员

\_\_\_\_\_ 中国科学院自动化研究所

学位类别 \_\_\_\_\_ 工学博士

学科专业 \_\_\_\_\_ 模式识别与智能系统

培养单位 \_\_\_\_\_ 中国科学院自动化研究所

2015 年 4 月



# Geo-data Processing and Application for Social Media

By  
**Weiqing Min**

A Dissertation Submitted to  
University of Chinese Academy of Sciences  
In partial fulfillment of the requirement  
For the degree of  
Doctor of Engineering

Institute of Automation  
Chinese Academy of Sciences

April, 2015





# 独创性声明

本人声明所递交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确地说明并表示了谢意。

签 名:\_\_\_\_\_日 期: \_\_\_\_\_

# 关于论文使用授权的说明

本人完全了解中国科学院自动化研究所有关保留、使用学位论文的规定，即：中国科学院自动化研究所有权保留送交论文的复印件，允许论文被查阅和借阅；可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。  
(保密的论文在解密后应遵守此规定)

签 名:\_\_\_\_\_导师签名:\_\_\_\_\_日 期: \_\_\_\_\_



## 摘 要

随着Web2.0技术的迅速发展和GPS手持移动智能设备的广泛应用，人们在使用各种社会媒体的同时以各种各样的形式分享自己的地理位置，产生了大量地理数据。例如用户在Flickr上分享某一地点比较有名的地标图片，在Foursquare上分享他们当前所在的商户位置。这类地理数据除了包含地理位置信息外，还通常关联其他各种上下文信息，比如时间和文本标签等。这些内容丰富、形式多样、数量巨大的地理数据能够服务于基于地理位置的各种应用，比如社会媒体的组织和检索、旅游推荐和媒体可视化等，因此如何对来自社会媒体的地理数据进行有效处理就成为开展基于地理位置的各种应用的关键。

来自社会媒体的地理数据和传统的网络多媒体相比有自己更突出的特点，比如这些地理数据通常都包含位置信息，数量巨大和数据异质等。尽管研究人员已经做了大量的工作，但是来自社会媒体的地理数据处理技术仍然存在着一一些关键问题需要解决，比如异质多模态信息的有效融合和统一建模，跨平台的信息关联等。针对以上问题，我们从社会媒体下地理数据的处理和应用两个方面进行了研究和探讨，主要研究内容和贡献如下：

- (1) 基于场景和角度的地标总结。考虑到地标场景和拍摄角度的多样性，我们提出了一个场景角度主题模型对地标进行总结。在角度聚类集合的基础上，该模型能够学习来自不同角度聚类共享的场景主题子空间以及同一场景下不同角度聚类所特有的场景角度主题子空间。我们利用学习的两类主题子空间得到每个地标对应的不同场景和角度的代表性图片。
- (2) 基于时空主题的地标分析。来自社会媒体的地标图片除了地标的视觉信息外，通常还包括和它关联的其他信息，比如文本和时间信息等。本工作充分考虑各种异质信息提出了一种时空主题模型学习地标的三类主题子空间，即所有地标都享有的全局主题子空间，只和某个地标相关的位置主题子空间以及在某个地标的某些时刻对应的时间主题子空间。此外，我们考虑地标和位置主题的关联以及地标-时间和时间主题的关联引入了

基于互信息的正则化的优化目标函数。最后我们利用贝叶斯定理从地标的时间和空间两个方面对发现的主题进行了分析。

- (3) 基于异质元数据的社会事件检测。社交媒体下的地理数据和关联的其他各种信息（比如时间信息）的重要应用之一是检测这类媒体数据中的社会事件实现基于社会事件的媒体组织和搜索。本工作充分考虑社会媒体的各种异质信息包括位置信息，时间信息、文本和视觉信息进行社会事件检测。为此我们提出了一种鲁棒的高阶联合聚类方法。一方面构建星结构的K分图建模社交媒体本身和各种类型的信息之间的依附关系，实现这些异质信息的有效融合；另一方面考虑时间空间内时间之间的关系，并将该关系作为全局正则化项引入到总的目标函数中，进一步提高了社会事件检测的精度。
- (4) 基于位置情景的跨平台个性化推荐。在地理数据处理和分析的基础上，本工作设计了一个基于位置情景的跨平台协同应用：在给定某一位置情景和两个不同的社交媒体平台Flickr和Foursquare的条件下，让Flickr用户能够享受来自Foursquare的当前位置的商户信息服务，同时让Foursquare用户能够享受来自Flickr的和当前位置相关的图片服务。为此，我们提出了一个跨平台多模态主题模型，一方面该模型能够区分和发现两类不同的主题，即平台特有的主题和所有平台共享的主题，以平台共享的主题作为桥梁连接两个平台；另一方面，该模型能够对齐同一主题来自不同平台的多种模态，保证不同平台之间不同模态的特色推荐。

**关键词：** 社交媒体，地理数据，联合聚类，主题模型，跨平台，总结，推荐

## **Abstract**

The fast development of Web2.0 and the widespread use of GPS-equipped mobile smart devices empower people to use location data from different social networks in various ways, which fosters the emergence of geo-data. For example, people can upload landmark images to Flickr and share their present venue information in Foursquare. Besides the location information, these geo-data are associated with other context information, e.g., the time-stamps and textual meta-data. The large-scale geo-data presents rich content with different modalities and thus serves as a handy resource for various location based applications(e.g. social media organization and recommendation, travel recommendation and media visualization, ). Therefore, how to effectively process this kind of the geo-data becomes the key problem of location based application.

Compared with the traditional multimedia, the geo-data has distinctive characteristics: they are generally associated with the geo-location and presents content with the heterogeneous metadata. Although researchers have done a lot of work in recent years, there are still several key technical issues, such as the fusion of the heterogeneous metadata and the information correlation across different platforms. To cope with the above mentioned issues, we conduct the research on the geo-data processing and applications for social media. The main contributions of this dissertation can be summarized as follows:

- (1) Scene and viewpoint based summarization for landmarks. Considering the diversity in both scenes and viewpoints, in order to better visually summarize landmarks, we propose a scene-viewpoint based theme model for modeling both scenes and viewpoints. This model is capable of learning the subspace of both the shared scene themes and viewpoint-specific scene-viewpoint themes. We obtain representative images with different scenes and viewpoint via the two kinds of learned subspace.
- (2) Spatio-temporal theme based landmark analysis. The landmark images from so-

cial networks are generally associated with other information, such as time and text information. We propose a probabilistic topic model to utilize different multimodal information to learn three kinds of theme subspace, i.e., global themes shared by many landmarks, local themes characterizing local characteristics of one landmark and temporal themes happened at a specific moment for one landmark. In addition, we consider the correlation between the local theme and landmarks, the correlation between the temporal theme and landmark-time information and introduce the mutual information-based regularization into the overall objective function. Finally, we analyze the mined themes from the spatio-temporal dimensions according to the Bayes' theorem.

- (3) Heterogeneous metadata based social event detection. In order to utilize different heterogeneous metadata including the location, time-stamp, visual and textual content for social event detection, we propose a robust high-order co-clustering method, which can construct the star-structured  $K$ -partite graph to model the inter-relationships among social media set and the associated  $K-1$  types of metadata. In addition, the intra-relationship between the time-stamps in the time space is introduced as the global regularization, which can further improve the performance.
- (4) Location context based personalized inter-platform recommendation. Based on the geo-data processing and analysis, we design a novel location based cross-platform collaboration application: given the location context and two platforms including Flickr and Foursquare, we recommend location relevant photos from Flickr to Foursquare users and recommend location relevant venues from Foursquare to Flickr users. To solve the problem, we propose a Cross-platform Multi-Modal Topic Model, which is capable of 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. We can correlate these two platforms via the learned shared topics. Furthermore, this model can align multiple modalities from different platforms, which facilitates the characteristic with different modalities recommendation cross different platforms.

**Keywords:** social media, geo-data, co-clustering, topic model, cross-platform, summarization, recommendation





# 目 录

摘要 .....	i
Abstract .....	iii
目录 .....	vii
第一章 绪论 .....	1
1.1 研究背景和意义 .....	1
1.2 研究内容 .....	5
1.2.1 基于地理位置的社会媒体总结 .....	6
1.2.2 基于时空的社会媒体内容分析 .....	6
1.2.3 基于位置情景的跨平台协同应用 .....	7
1.3 论文结构安排 .....	7
第二章 研究现状概述 .....	11
2.1 技术发展综述 .....	11
2.1.1 基于地理位置的社会媒体挖掘 .....	11
2.1.2 基于时空的社会媒体内容分析 .....	13
2.1.3 基于位置情景的跨平台协同应用 .....	14
2.2 研究方法介绍 .....	15
2.2.1 联合聚类 .....	15
2.2.2 概率主题模型 .....	16
第三章 基于场景和角度的地标总结 .....	19
3.1 引言 .....	19
3.2 相关工作 .....	20

3.3	方法框架 .....	21
3.4	角度聚类 .....	21
3.5	基于场景角度的主题建模 .....	22
3.5.1	参数估计 .....	24
3.5.2	总结生成 .....	25
3.6	实验结果与分析 .....	25
3.7	本章小结 .....	28
<b>第四章</b>	<b>基于时空主题的地标分析 .....</b>	<b>29</b>
4.1	引言 .....	29
4.2	相关工作 .....	30
4.3	方法框架 .....	31
4.4	数据准备 .....	32
4.5	多模态时空主题建模 .....	33
4.5.1	文档的生成式过程 .....	33
4.5.2	基于互信息的正则化 .....	34
4.5.3	参数估计 .....	35
4.6	时空主题分析 .....	36
4.6.1	主题的时间分布 .....	36
4.6.2	主题的位置分布 .....	36
4.7	实验结果与分析 .....	37
4.7.1	主题模型评估 .....	37
4.7.2	主题分析 .....	40
4.8	本章小结 .....	42
<b>第五章</b>	<b>基于异质元数据的社会事件检测 .....</b>	<b>43</b>
5.1	引言 .....	43
5.2	相关工作 .....	44
5.2.1	社会事件检测 .....	44

5.2.2	联合聚类	45
5.3	方法框架	46
5.4	基于高阶联合聚类的粗社会事件检测	47
5.4.1	星结构的K分图	47
5.4.2	基于K分图的关系建模	48
5.4.3	时间空间内的关系建模	50
5.4.4	统一的问题描述	50
5.5	聚类 and 样本改进	51
5.6	社会事件检测算法	51
5.7	实验结果与分析	52
5.7.1	数据库	52
5.7.2	评估测度和特征	54
5.7.3	实验结果	54
5.8	本章小结	56
<b>第六章</b>	<b>基于位置情景的跨平台个性化推荐</b>	<b>59</b>
6.1	引言	59
6.2	相关工作	61
6.2.1	跨平台个性化推荐	61
6.2.2	概率主题模型	62
6.3	方法框架	63
6.4	跨平台多模态主题建模	64
6.4.1	问题定义	65
6.4.2	CM <sup>3</sup> TM	65
6.4.3	模型推断	67
6.4.4	参数估计	69
6.5	基于位置情景的跨平台推荐	70
6.5.1	个性化的Foursquare商户推荐给Flickr用户	70
6.5.2	个性化的Flickr图片推荐给Foursquare用户	71

6.6 实验结果与分析 .....	72
6.6.1 实验设置 .....	72
6.6.2 跨平台多模态主题模型评估 .....	74
6.6.3 个性化平台间推荐评估 .....	77
6.7 本章小结 .....	82
<b>第七章 结束语 .....</b>	<b>83</b>
<b>参考文献 .....</b>	<b>85</b>
<b>发表文章目录 .....</b>	<b>103</b>
<b>致谢 .....</b>	<b>105</b>

## 表 格

3.1	地标集合的统计信息 .....	26
3.2	发现的主题的数量统计 .....	27
4.1	地标的统计 .....	37
4.2	发现主题的性能评估 .....	40
4.3	发现主题的性能评估 .....	40
5.1	不同方法的NMI .....	55
5.2	不同方法的NMI比较 .....	56
6.1	关键符号列表 .....	64
6.2	数据统计 .....	73



## 插图

1.1	各种类型的基于地理位置的社会媒体 .....	2
1.2	研究框架图 .....	5
2.1	主题模型 .....	16
3.1	大本钟图片集 .....	20
3.2	可视化地标总结框架 .....	21
3.3	角度聚类例子展示 .....	23
3.4	主题模型的图表示 .....	24
3.5	三个地标(a) 金门大桥(b) 勃兰登堡门和(c) 凯旋门的总结展示。 对于每一个地标，每一行代表一个场景而每一列代表一种拍摄 角度 .....	27
3.6	用户体验比较 .....	28
4.1	清水寺（Kiyomizu-dera）的例子 .....	29
4.2	地标分析框架: (1) 数据准备(2)主题建模(3)主题分析 .....	31
4.3	地标主题模型的图表示 .....	34
4.4	发现的主题展示 .....	39
4.5	金门大桥主题的时间分布 .....	41
4.6	主题“2012 Olympic games in London”的位置分布，时间范围 为2012-07-22 2012-08-14 .....	41
5.1	来自Flickr网站的社会事件例子 .....	44
5.2	星结构的K分图 .....	48
5.3	不同足球事件的例子图片 .....	53
5.4	不同 $\alpha$ 值下的性能比较 .....	56

6.1	来自Flickr和Foursquare不同类型的知识：红色和紫色标注的代表共享的知识，绿色标注的代表Flickr特有的知识，蓝色标注的代表Foursquare特有的知识 .....	60
6.2	基于位置情景的跨平台个性化推荐框架 .....	63
6.3	CM <sup>3</sup> TM的生成式模型 .....	65
6.4	CM <sup>3</sup> TM挖掘的三类主题例子 .....	75
6.5	Perplexity 比较 .....	77
6.6	Precision 和MAP 在个性化的Foursquare商户推荐给Flickr用户的性能比较 .....	80
6.7	Precision 和MAP 在个性化的Flickr图片推荐给Foursquare用户的性能比较 .....	80
6.8	个性化的跨平台推荐的事例研究 .....	81



# 第一章 绪论

## 1.1 研究背景和意义

Web2.0技术的到来使得各种类型的社交网络得到了迅速的发展，比如新浪微博、街旁网、Flickr、Youtube和Foursquare等。与此同时，全球定位系统及手持智能设备的广泛应用使得人们在使用这些社会媒体的同时以各种各样的形式分享自己的地理位置信息，产生了大量的地理数据。例如，用户在Flickr上分享某一地点比较有名的地标图片，在Youtube上分享带有地理标签标注的视频，在Foursquare或者街旁网上分享他们当前所在的商户位置等。一方面，这些地理数据通常还关联其他各种元数据信息，比如时间信息、文本标签、标题和评论信息等，大量丰富的地理数据为我们基于地理位置的社会媒体组织、总结、搜索和推荐的研究提供了许多机会。另一方面，这些地理数据模态多样，数据异质，并且含有大量噪音，使得无法直接利用用户产生的数据，因此如何对其进行有效处理，以便于有效的组织、搜索和推荐，就成为基于地理位置的社会媒体应用的关键。

社会媒体以多种不同的模态来呈现，包括文本、图像、视频和签到等。大部分社会媒体允许用户在上传这些媒体的同时分享他们的位置信息，图1.1展示了各种各样的地理数据，可以看出，这类来自社会媒体的地理数据通常具有以下特点：

- 位置标注。许多的社会媒体都关联某一种位置信息，这些位置的表示形式是多种多样的，最基本的位置信息比如许多社会媒体都关联GPS地理标签，其他的位置包括地标和签到的商户标签等。如图1.1所示，这些位置信息包括来自Flickr的地标信息，来自新浪微博的城市地理位置信息，来自Foursquare的签到商户等。
- 数据异质。一方面，如图1.1所示，社会媒体下的地理数据类型包括图片（比如Flickr）、视频（比如Youtube）、微博（比如新浪微博）和签到数据（比如Foursquare）等各种模态。另一方面，社会媒体下的地理数据通常还和其他不同类型的实体类型关联在一起，比如对于Flickr的图片，还包括用户、地理位置、时间和文本标签等。这些实体之间存在这不同类型的异质关系。



图 1.1: 各种类型的基于地理位置的社交媒体

而在Foursquare上，每一条签到记录不仅包括地理标签，还包括文本标签、该商户的签到次数和留言等。而且Foursquare的用户、用户所发的留言和用户签到的商户等多种类型的实体之间存在着异质的交互 [134]。

- 数量巨大。据统计，Flickr每月上传的带有地理标签的图片数量约为2,000,000<sup>1</sup>。而Foursquare截止2014年12月已经达到4500万用户和超过50亿次签到<sup>2</sup>。

为了理解和利用这类社交媒体数据，世界范围内的很多高校和科研机构都开展了面向社会媒体的地理数据处理和应用方面的研究。较早的工作主要利用位置信息和其他模态信息实现对某一区域的总结。比如来自yahoo实验室的工作组对某一城市的所有地标图片，充分利用地理标签、时间、文本和视觉信息等，序列使用这些信息通过聚类的方法实现对该城市地标图片的总结。美国康奈尔大学计算机学院的工作组则首先利用图片的地理位置标签对图片进行聚类，找到一些代表性的地标图片，然后充分利用地标的文本，视觉和时间信息突出解决大规模地标图像分类的问题 [29, 67]。希腊雅典国家技术大学的图像和视频分析组 [4]在利用地理标签进行图像的聚类后解决地标的检索问题。除

<sup>1</sup><https://www.flickr.com/map>

<sup>2</sup><http://36kr.com/p/64076.html>

了基于地理位置的总结外，地理位置作为一种情景信息还可以用于社会事件检测。汉诺威莱布尼兹大学的L3S研究中心 [22]利用Flickr图片集的地理标签、时间和文本标签信息通过基于小波的空间分析方法检测社会事件。以上大都基于单个平台的研究，为了解决单个平台面临的问题比如数据的稀疏性等，现在的一些研究开始转移到基于位置的跨平台社交媒体挖掘技术中。例如，伊利诺伊芝加哥分校的小组 [132] 利用Foursquare和Twitter两个平台的地理标签和文本信息等关联两个平台进行异质链接推荐。此外，在国内外知名会议和期刊上，很多和社交媒体下地理数据处理和应用相关的工作，陆续被发表出来，国际会议如IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision (ICCV), ACM International Conference on Multimedia (MM), ACM International Conference on World Wide Web (WWW), ACM International Conference on Web Search and data mining (WSDM)等，当前的一些权威期刊包括IEEE transactions on Pattern Analysis and Machine Intelligence(PAMI), IEEE transactions on Multimedia (TMM)等，为该领域的研究人员提供了更多的交流机会。

近些年来，虽然研究者对社交媒体下地理数据处理和应用方面的研究不断深入，但仍然存在一些关键的技术问题需要亟待解决：如何将位置信息和该社交媒体的其他模态和类型的信息进行有效的融合，如何对某一位置进行合理有效的总结和分析等。对这些关键技术和科学问题的研究进行深入探讨，具有重要的理论研究意义，具体体现在以下几个方面：

- 解决社交媒体的异质信息表示和统一建模问题。社交媒体下的地理数据相比于单一媒体，不仅包括媒体内容和位置信息，还包括与之关联的丰富的异质元数据信息，比如时间和文本标签等。而且社交媒体和这些元数据之间存在着丰富的异质关系，利用现有的数据挖掘技术充分考虑各种模态信息以及上下文信息实现这些信息的有效融合和统一建模有助于挖掘社交媒体更为丰富和准确的知识，进而为社交媒体下地理数据的处理与应用提供基础。
- 促进语义多媒体理解技术的发展。语义多媒体理解旨在对多媒体包括图像和视频等内容的语义解释，始终是计算机视觉和多媒体领域研究中具有挑战性的问题。社交媒体下的地理数据由于本身引入新的位置维度信息有利于缩小语义鸿沟，加强多媒体的语义理解。此外，通过数据挖掘技术充分利用位置信息和其他元数据信息结合，比如时间可以进行社交媒体事件检测，而检测

到的事件也同样有利于多媒体语义的理解。

- 促进和地理位置研究相关的学科发展。社交媒体下的地理数据处理和推荐技术的研究涉及的很多问题不能直接采用传统的多媒体计算模型及其理论来描述。通过研究新的挖掘技术，充分考虑社会媒体的位置信息和其他各种信息和关系可以帮助建立适合基于地理位置的社交媒体的新型计算框架。这些新的方法可以促进和带动和地理位置研究相关学科的发展，比如地理信息服务（GIS）和基于地理位置的服务（LBS）等，丰富相关学科的研究内容和研究方法。

另外，社交媒体下的地理数据处理技术已经广泛应用到许多生产生活领域，并渐渐展现出更为广阔的应用前景，如在以下几个方面的应用：

- 社交媒体组织、标注和检索：多媒体的语义组织已经成为多媒体搜索领域的重要任务。来自社交媒体的地理数据可以允许我们依据位置信息进行社交媒体组织，实现媒体的语义标注增强，甚至进一步通过位置信息和其他信息比如时间信息结合产生的事件信息进行社交媒体的组织，索引和检索。
- 推荐：带有位置信息的社交媒体比如图片、视频、日志和博客等经常反应出许多和位置相关的服务和产品所需要的重要信息。例如地标图片是某地方具有独特地理特色的建筑物（例如摩天大楼、教堂和寺庙等）。这些地标是旅游者旅游的必经之地。因此通过相应的地理数据处理技术对这些地标图片进行分类和总结可以应用到旅游推荐中去。此外，通过对用户上传的带有地理信息的社交媒体数据进行处理和分析发现用户的旅游路线，可以进行个性化的路线推荐；我们还可以综合分析用户在各个网站的信息挖掘感兴趣的旅游区域（ROI）进而可以实现ROI的推荐等。
- 可视化：由于位置信息的获得，图片和其他媒体能够以一种新的维度信息位置进行可视化。这种实现可以通过以下几种途径，比如根据地理标签信息对某一地区的地标图片进行总结得到该地区多样性和全面的图片集从而实现对某一地区的可视化总结，当然也可以对用户旅游轨迹和路线进行可视化总结。可视化可以直接应用到旅游相关的行业和其他应用中。
- 位置识别：根据社交媒体或者社交媒体和其他模态结合的信息识别该媒体的地理位置信息。该位置可以是GPS信息或者地标信息。通过现有的带有位置信息的社交媒体数据，我们可以比较容易得到大量的训练数据，利用这些训练数据建立相应的位置识别模型已达到位置识别的目的。

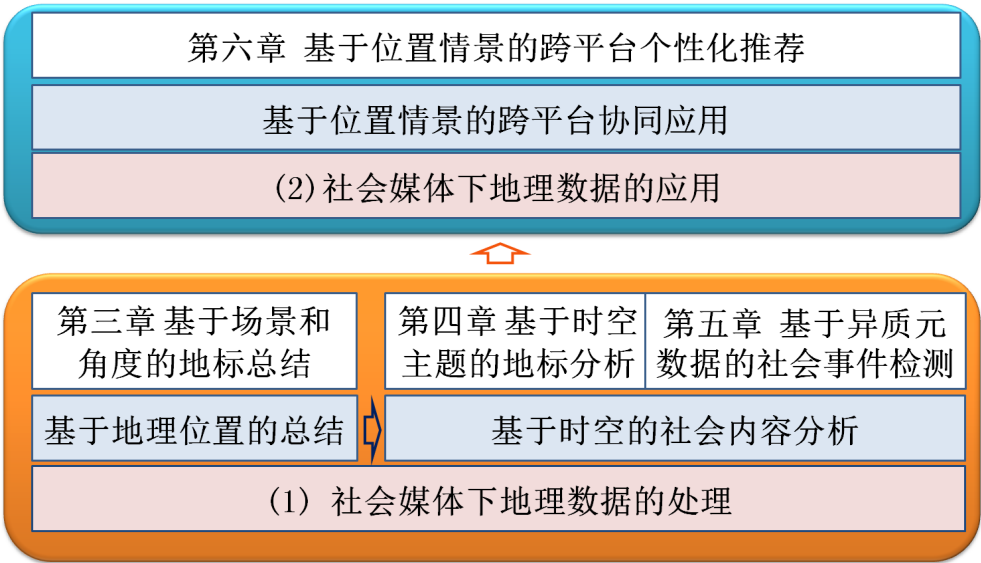


图 1.2: 研究框架图

1.2 研究内容

社交媒体下的地理数据处理和应用研究旨在对来自社会媒体的地理数据进行处理和分析以便于基于地理位置的各种应用。因此我们的研究主要包括以下两个方面（1）社交媒体下地理数据的处理和（2）社交媒体下地理数据的应用。其中，社交媒体下地理数据的处理是基于地理位置的应用的基础。第一个研究方面主要包括基于地理位置的社交媒体总结和基于时空的社交媒体内容分析。基于地理位置的社交媒体总结主要针对一些重要位置比如地标或者城市等，充分利用文本和视觉等多模态的社交媒体数据挖掘各种各样的地理位置的主题实现该地理位置的总结；在基于地理位置的社交媒体总结的基础上，基于时空的社交媒体内容分析则进一步引入时间维度，从时空两个维度分析媒体内容，比如地标的时空主题分析和社会事件监测。我们对以上来自社会媒体的地理数据处理均是基于单平台的，相比之下，对于第二个研究方面，我们考虑到社交媒体信息的多平台性，研究基于位置情景的跨平台个性化推荐，从一个平台挖掘和当前位置相关的知识推荐给另一个平台的用户。具体来说我们的研究内容包括（1）基于地理位置的社交媒体总结；（2）基于时空的社交媒体内容分析和（3）基于位置情景的跨平台协同应用。图1.2的研究框架显示了这几个方面工作的关系。



### 1.2.1 基于地理位置的社交媒体总结

基于地理位置的社交媒体总结是基于地理位置的应用的基础。比如对某一重要地标或者热门城市的总结便于用户理解该地理区域的特色从而有利于用户的旅游规划。基于地理位置的社交媒体总结要求地理位置总结的结果具有代表性和多样性,而且要求对于不同的地理位置需要考虑不同的总结角度。我们对该部分研究内容主要针对地标进行总结。用户上传的大量的地标图片不仅拍摄角度多种多样(例如正面拍摄、自底部向上拍摄和近距离拍摄),而且拍摄的场景由于不同的拍摄的时间也会呈现不同的外观(比如某一地标的阴天场景和晴天场景)。基于场景和拍摄角度的地标总结能够无冗余的展现地标多样性的结果,这为我们基于该地理位置的总结提供了解决思路。基于此,我们主要的研究内容是如何利用各种模态的信息建模场景和角度实现该地理位置的总结。由于场景是由底层表观特征表示而拍摄角度则是通过几何信息表示,因此如何通过社交媒体提供的信息解决场景和拍摄角度表示的不一致性是我们研究基于该类地理位置的社交媒体总结的关键。

### 1.2.2 基于时空的社交媒体内容分析

基于时空的社交媒体内容分析在各种领域中有着广泛的应用。来自社交媒体比如Flickr和Youtube的地理数据通常还关联时间信息和文本信息等,这些丰富的信息为我们基于时空的社交媒体内容分析提供了大量的数据。基于此,我们主要的研究内容通过从各种异质的社交媒体数据中有效挖掘和分析社会媒体的时空模式。一方面,基于地理位置的社交媒体包含位置、时间、文本和视觉信息等,为了充分利用这些信息,我们需要有效的融合各种社交媒体信息实现对社会媒体的统一建模;另一方面,对于每一种时空模式都有相应的多模态信息来表示,我们需要对齐来自同一模式各种多模态信息的表示。对于第一个问题,我们通过建模社交媒体和各类元数据之间的依附关系实现这些信息的联合聚类或者主题建模。对于第二个问题,我们通过学习共享的子空间去直接建模视觉和文本信息捕获不同模态数据之间的相关性。针对该研究内容,我们的工作主要集中在基于时空主题的地标分析和社会事件检测。

### 1.2.3 基于位置情景的跨平台协同应用

不同的社交媒体平台都有自己独特的服务，例如Flickr是一个图像分享网站用来分享带有地理标签的图片而Foursquare则提供一个个性化的位置搜索服务。充分利用各个平台之间的互补信息实现基于位置情景的跨平台协同应用受到越来越多的重视。由于同一用户在不同平台的账号在许多情况下很难获取，而且不同的平台包含不同的异质媒体数据，因此如何关联多个社交媒体平台成为跨平台协同应用的关键。此外，不同的平台都有自己的多模态信息，在关联的基础上，如何对齐来自同一类主题不同的模态是我们要解决的另外一个问题。针对第一个问题，我们充分利用从各个平台挖掘的共享的知识作为桥梁连接不同的平台；而对于第二个问题，我们则通过学习共享的子空间去建模不同模态之间的相关性。具体来说，本文解决如下一个新奇的跨平台问题：给定特定的位置情景和两个平台Flickr和Foursquare，我们推荐来自Flickr的和当前位置相关的图片给Foursquare用户，同时推荐来自Foursquare的当前位置的商户信息给Flickr用户，以实现基于位置情景的跨平台个性化推荐应用。

## 1.3 论文结构安排

针对以上三个方面的研究，如图1.2所示，论文具体研究工作包括从第三章到第六章共四章分别进行阐述，其各个章节的具体安排如下：

- 第二章首先回顾了基于地理位置的社交媒体总结，基于时空的社交媒体内容分析和基于位置情景的跨平台协同应用的研究现状；然后介绍了这些研究工作中常用的两种数据挖掘方法：（1）联合聚类和（2）概率主题模型。
- 第三章研究基于场景和角度的地标总结。来自社会媒体的地标图片呈现出两个特点：地标场景的多样性和拍摄角度的多样性。基于以上两个特点，我们提出了一个场景角度主题模型同时建模场景和角度，该模型能够学习出来自不同角度聚类共有的场景主题子空间以及同一场景下不同角度所特有的场景角度主题子空间。利用学习的这两类主题子空间可以得到场景角度主题的高层语义表示。基于该方法，我们提出了一个地标总结框架：首先通过基于GIST特征的角度聚类得到每个地标不同角度的聚类，然后以每个角度聚类作为一个集合，通过提出的场景角度主题模型得到每个角度聚类中不同场景的主题。基于学习的场景角度主题，每个地标最终由不同场景和不同角度的代表性的图片表示。

- 第四章基于时空主题的地标分析。来自图片分享网站的地标图片除了包含地标的视觉信息，通常还包括和它关联的其他信息，比如时间和文本标签信息等。这些数据包含着三类主题：全局主题、位置主题和时间主题。其中全局主题是所有地标共有的主题，比如天气，交通等；位置主题特指某一个地标所特有的风格或外观属性；时间主题是指某个地标在某个时间点上特有的风格或外观属性。为了充分利用各种异质信息，本工作提出了一种时空主题模型去挖掘该地标的三类主题，每类主题对应不同的主题子空间。通过挖掘的主题，我们以图文并茂的方式的可视化发现的主体。基于该模型，我们提出了一个地标主题分析框架。该框架包括三个部分：首先，数据收集主要是从图片分享网站中下载地标图片并进行相应的预处理；其次，主题建模则是利用提出的时空主题模型从地标图片中挖掘三类主题；最后从地标的时间和空间两个方面对发现的主体进行分析。
- 第五章研究基于异质元数据的社会事件检测。基于地理位置的社交媒体和关联的时间信息的重要应用之一是社会事件检测。本工作考虑社交媒体和关联的位置信息外，同时考虑其它上下文信息包括时间和文本标签进行社会事件检测，我们提出了一种高阶的联合聚类方法。该方法一方面采用星结构的K分图建模该社交媒体和位置、时间、文本和视觉信息之间的关系，另一方面考虑时间空间内容时间与时间的关系，并将该关系作为全局正则化项引入到总的优化目标函数中。在该方法的基础上，我们进一步提出了一个社会事件检测框架，该框架包括粗事件分类和聚类样本改进。首先在粗聚类中，我们考虑位置、时间、文本和视觉信息四种异质元数据和该媒体对象之间的关系设计了高阶的联合聚类的方法；其次在得到的粗聚类中，我们采用一种启发式的方法实现聚类的过滤和融合，同时去掉每个聚类中有噪音的图像。最终我们得到了每个事件的聚类达到了事件检测的目的。
- 第六章研究基于位置情景的跨平台个性化推荐。随着Web2.0的迅速发展，各种各样的社会平台迅速发展，例如，Flickr是一个图像分享网站用来分享个人的带有地理标签的图片而Foursquare则提供一个个性化的位置搜索服务。为了能够让来自Flickr用户能够享受来自Foursquare的商户信息，同时也为了让来自Foursquare用户能够享受来自Flickr的带有地理标签的图片服务。本文解决如下一个跨平台协同推荐问题：给定特定的位置情景和两个平台Flickr和Foursquare，我们推荐来自Flickr的图片给Foursquare用户，同时推



荐来自Foursquare 的商户信息给Flickr用户。为此我们提出了一个跨平台多模态主题模型，该模型能够（1）区分两类不同的主题，也就是平台特有的主题和所有平台都共享的主题；（2）对于同一主题，对齐来自不同平台的多种模态。我们最终将提出的模型应用到基于位置情景的平台间推荐问题中。

- 第七章对全文进行总结，并对未来的研究工作进行了展望。



## 第二章 研究现状概述

随着社会媒体的快速发展和移动手持设备的广泛应用，基于位置的社会媒体资源与日俱增，如何对这类社会媒体数据进行有效的挖掘已成为一个日益重要的研究课题。近年来，广大研究者在该领域中展开了广泛的研究，并取得了许多成果。本章将对近年来基于地理位置的社会媒体挖掘技术进行介绍，具体来说，本章第一节就基于地理位置的社会媒体挖掘、基于时空的社会媒体内容分析和基于位置情景的跨平台协同应用三个方面来综述目前的技术发展现状。针对各研究工作更具体的研究现状介绍，可参考各章的相关工作部分。为了实现这些技术，本章第二节主要就其中的两类数据挖掘方法即联合聚类和概率主题模型方法进行介绍。

### 2.1 技术发展综述

#### 2.1.1 基于地理位置的社会媒体挖掘

基于地理位置的社会媒体挖掘研究由于其潜在的应用前景而得到广泛的重视。文献 [72, 139] 全面综述了基于地理位置的社会媒体研究和应用近年来的进展。由于本工作针对某一类重要的地理位置即地标相关的挖掘工作进行研究，因此我们主要对和地标挖掘相关的工作进行综述。地标挖掘主要是从大量的地标数据中挖掘出和地标相关的有意义的模式以便对地标进一步的总结、识别、分类和检索。本小节具体从地标组织和总结，地标识别和分类，地标检索和基于地标的旅游应用这四个方面的工作逐一展开介绍。

(1) 地标组织和总结 Xue等人 [115] 直接利用图片的视觉信息获取地标的拍摄角度（例如正面拍摄和远距离拍摄），然后通过基于拍摄角度的聚类方法实现地标的可视化总结。文献 [102, 136] 则是从场景（比如夜景和雪景）的角度通过基于视觉信息的聚类或者主题建模的方法实现基于视觉特征的地标总结。社交媒体中除了图片的视觉信息，还包括来自社会媒体的各种元数据信息，比如文本标签和时间信息等。为了充分利用这些信息，Kennedy等人 [60, 61] 在基于视觉信息聚类的基础上，利用地理标签和时间信息对聚类进行排序，选出代表性的聚类用以总结某一地区。和 [60, 61] 的工作相似，Rudinac 等人 [96] 首先

利用多种模态信息构建一个多模态的图,然后基于多模态图采用随机游走的方法选出代表性的图片。Ji等人[55,56]也采用图模型的方法从博客中挖掘和总结地标。Chen等人[24]在总结的基础上进一步采用卡通技术为每一个地标在地图上生成一个地标图标。除了图片分享网站外,Pang等人[81]利用主题模型从旅游日志中挖掘出某些地区相关的本地主题实现对某一城市的总结。

**(2) 地标识别和分类** 最早基于地标识别的工作[138]主要从互联网中收集大量的地标图片进行聚类建立地标可视化模型进而实现地标识别。相比[138],Li等人[29,67]收集更大规模的地标数据集包括500个地标和30,000,000地标图片,然后用多类的支持向量机训练模型实现大规模的地标分类。以上方法由于没有使用地标的几何信息无法克服遮挡和多角度问题,为了解决这个问题,Xiao等人[112-114]则是将每个地标进行3D模型的重构,基于每个地标的3D点云模型实现地标鲁棒的分类。和[114]的思想相似,Hao等人[47]从地标图片中提取出3D的“视觉短语”进行地标的识别。Raguram等人[91]则同时考虑地标图片的2D表现特征和3D几何特征进行地标的3D建模和识别。

**(3) 地标检索** Gavves等人[44,45]通过提取可视化同义词(visual synonyms)特征进行地标的图像检索。不同于[44,45],Avrithis等人[5]利用基于向量量化的聚类方法得到每个地标代表性的图片,每个地标最终都可以用一些代表性图片生成的场景地图(scene map)来表示实现地标的检索。此外,随着移动互联网的发展和手持设备的广泛应用,移动地标搜索的相关的工作也相继展开,工业界已经成熟的产品比如Google的Goggle<sup>1</sup>和Baidu搜索APP<sup>2</sup>。而在科研界,Chen等人[21]通过采用直方图均衡化增强所有图片的对比度以提取更为鲁棒的特征,同时融入了GPS信息实现移动地标识别。而Ji等人[54]则是通过判别性词典编码的方式得到判别性的地标描述以实现移动地标的检索。

**(4) 基于地标的旅游应用** 近年来各种各样基于地标的旅游推荐系统在工业界和科研界都得到了极大的发展。在工业界,第一个面向旅游的交互式网站WikiTravel<sup>3</sup>能够为用户提供及时的和地标相关的旅游信息。之后其他一些基于地标级别的旅游系统也得到广泛的传播和发展,比如YahooTravel<sup>4</sup>、国内的

<sup>1</sup><https://support.google.com/websearch/answer/166331?hl=en>

<sup>2</sup><http://lvyoud.baidu.com/>

<sup>3</sup><http://wikitravel.org/en/Main-Page>

<sup>4</sup><https://www.yahoo.com/travel>

百度旅游<sup>5</sup>和蚂蜂窝<sup>6</sup>等。在科研界,比较早的旅游系统是Snively等人 [103]设计的PhotoTourism,该系统提供一个新奇的3D接口能够交互式的浏览和挖掘大规模的无结构的地标图片集。Gao 等人 [43]提出了W2Go 旅游推荐系统,以给定的城市作为查询,W2Go能够使用自动的地标排序方法返回该城市中比较流行的地标。此外,Xin等人 [71]提出了旅游路线推荐系统Photo2Trip。该系统能够利用大规模图片的地理标签信息生成旅游路径实现旅游路径的推荐。

### 2.1.2 基于时空的社会媒体内容分析

基于时空的社会媒体内容分析主要是挖掘和分析社会媒体的时空模式以进一步应用到各种领域中。相关的工作主要包括基于时空主题的分析和社会事件检测。对于时空主题分析,Mei等人 [74]提出一个概率主题模型挖掘博客内容的时空主题和分析不同主题在不同位置或不同时间的分布应用到市场营销等领域。Hu等人 [53]则基于Foursquare的签到数据和时间信息,通过时空主题建模和分析同时考虑时间和商户位置信息为用户做出更为合理的推荐。和Hu工作相似,Yuan [125]充分分析地理和时间方面的影响力实现时间意识的感兴趣商户的推荐。

对于社会事件检测,社会事件<sup>7</sup>定义为由人计划和参与的事件。通常一个社会事件可以由和事件相关的媒体元数据信息指定,包括位置、时间和人等。社会事件可以用于社会媒体的浏览 [46]、检索和推荐,因而社会事件检测也成为科研界关注的焦点。许多科研机构比如MediaEval等已多次举办关于社会事件检测的挑战,对社会事件检测技术的发展起到了推动作用。较早的工作 [92,93]充分利用来自Flickr图片的地理标签和时间信息用突发性检测方法从Flickr的文本标签中提取带有事件和地理位置语义的文本标签。Chen 等人 [23]则是基于地理位置和时间信息采用小波变换的方法从Flickr的文本标签中检测社会事件,并在此基础上进一步将事件区分为周期性事件和非周期性事件。Becker等人 [9,10]充分利用文本信息和丰富的情景信息包括地理位置和时间通过集成的方法进行相似度的学习,基于学习的相似度,使用在线的聚类方法检测事件。和 [9,10]相似,Zaharieva等人 [127]也充分利用了文本、时间和空间的特征,分别基于这些特征进行聚类以检测社会事件。而文献 [86,111]融合各种信

---

<sup>5</sup><http://lvyou.baidu.com/>

<sup>6</sup><http://www.mafengwo.cn/>

<sup>7</sup><http://www.multimediaeval.org/mediaeval2013/sed2013/index.html>

信息的特征通过聚类的方法检测社会事件。不同于以上工作, Yuan等人 [122,123] 充分利用gps轨迹信息和时间进行事件的识别和分类。以上工作均未考虑丰富的图片视觉信息, 为了充分利用该类信息, Liu等人 [69]通过三步实现社会事件检测。首先, 基于某些条件得到一些事件实例, 然后通过提出的事件鉴别模型得到事件和图片之间的关系, 最后通过可视化的删减和基于启发式的修正改进聚类的结果。Quack等人 [90] 通过视觉、文本和空间特征的相似度进行聚类, 然后通过模型训练实现事件和物体的分类。而Papadopoulos等人 [84] 通过基于视觉和文本信息的图聚类以及地理位置信息的聚类融合得到训练集, 通过训练SVM分类器实现地标和事件的分类。Reuter等人 [94] 则实现了在线的社会事件检测。此外, Brenner [16] 进一步借助外部信息源比如Playerhistory.com等网站作为辅助信息实现社会事件的检测。

### 2.1.3 基于位置情景的跨平台协同应用

基于跨平台的协同应用主要包括两个方面的研究内容, (1) 融合来自不同平台的信息进行用户的建模。Abel等人 [1] 通过聚合来自Flickr、Twitter和Delicious的用户信息解决冷启动的推荐问题。Zahalka等人 [126]从Picasa、Foursquare和Flickr的网站中收集带有文本标签的商户图片然后通过一个交互式的接口学习用户的兴趣实现个性化的推荐。Zhong等人 [140], [141] 提出了一个主题模型方法自适应的融合来自不同平台的信息进行知识的转移实现商品或者朋友推荐。Liu等人 [68] 同时分析来自Flickr的带有地理标注的图片和来自Foursquare的签到数据从中挖掘旅游者的兴趣和当地居民活动的偏好, 进而发现和推荐一个城市中感兴趣的地理位置。(2) 充分利用不同平台各自的特色实现协同应用。Yan等人 [116] 以不同平台共有的用户作为桥梁提出了基于随机游走的方法用于解决朋友推荐中的冷启动问题。Zhang等人 [131,132] 也以两个平台共有的用户作为桥梁借助两个平台的位置、时间和文本信息转移来自源平台的链接信息用于预测新用户的社会链接信息。Qi等人 [88,89] 以用户的内容属性作为桥梁充分使用有偏的跨平台采样方法进行跨平台的链接预测。Roy等人 [95]通过挖掘博文的主题将Twitter平台的知识转移到Youtube, 从而进行视频的个性化推荐。Deng等人 [31] 也是通过融入来自Twitter用户的信息实现个性化的Youtube视频推荐。不同于 [31,95] Yan等人 [117], 则是在给定来自Youtube的视频的情况下, 在Twitter上找到能够进行最大化该视频推广的Twitter追随者。

除此之外, 和跨平台推荐比较相关的一个领域是跨域推荐。文献 [38,64]

较为全面的描述了近几年跨域推荐的研究进展。跨域推荐的核心概念是充分挖掘来自多个领域的信息以便于目标平台的推荐。例如Li等人 [38, 64]用潜在因子模型学习间接的聚类级别的打分模式用于转移来自源领域的知识。Zhang等人 [133]将潜在因子模型扩展到多领域去学习一个间接的关联矩阵, 该矩阵通过间接的领域关联链接不同领域的知识。而Pan等人 [79, 80] 引入了一个坐标系转移的方法去学习多个领域的潜在特征进行转移。

## 2.2 研究方法介绍

为了实现基于地理位置的社会媒体内容分析和应用, 我们需要考虑基于地理位置的社会媒体包含的各类信息, 比如位置信息、时间信息、文本信息和视觉信息等, 为了充分利用这些信息, 关键是要有效的融合各种社会媒体信息实现对社会媒体的统一建模以便有效的挖掘社会媒体的主题。而联合聚类和概率主题模型则为解决这些关键问题提供了方法上的支持, 因此本小节主要针对联合聚类和概率主题模型两类方法分别给予相关的介绍, 在第三章到第七章具体的研究工作中, 将对这些方法进行利用和改进。

### 2.2.1 联合聚类

相比于基于一维的传统聚类方法, 联合聚类能够在两个维度上同时进行聚类。令 $X$ 和 $Y$ 为两个离散的随机变量, 且对应的取值集合分别为 $\{x_1, \dots, x_m\}$ 和 $\{y_1, \dots, y_n\}$ 。令 $p(X, Y)$ 表示随机变量 $X$ 和 $Y$ 的联合概率分布, 可以看做一个 $m \times n$ 的矩阵。我们的问题是同时将 $X$ 和 $Y$ 分别聚类成 $k$ 和 $l$ 个硬聚类。令 $X$ 的 $k$ 个聚类表示成 $\{\hat{x}_1, \dots, \hat{x}_k\}$ , 令 $Y$ 的 $l$ 个聚类表示成 $\{\hat{y}_1, \dots, \hat{y}_l\}$ , 则我们要找到以下两个映射 $C_X$ 和 $C_Y$

$$C_X : \{x_1, \dots, x_m\} \rightarrow \{\hat{x}_1, \dots, \hat{x}_k\} \quad C_Y : \{y_1, \dots, y_n\} \rightarrow \{\hat{y}_1, \dots, \hat{y}_l\} \quad (2.1)$$

也即最终的聚类表示成 $\hat{X} = C_X(X)$ ,  $\hat{Y} = C_Y(Y)$ 。

Charrad等人 [20]综述了联合聚类的方法近年来的进展。Dhillon等人 [33] 将联合聚类看做是一个二分图分割问题, 利用提出的谱聚类方法进行联合聚类。该方法的局限在于每一种源的数据都要至少和另一个源的数据结合在一起。为了克服这个问题, Dhillon等人 [34] 提出一个基于信息论的联合聚类方法以最小互



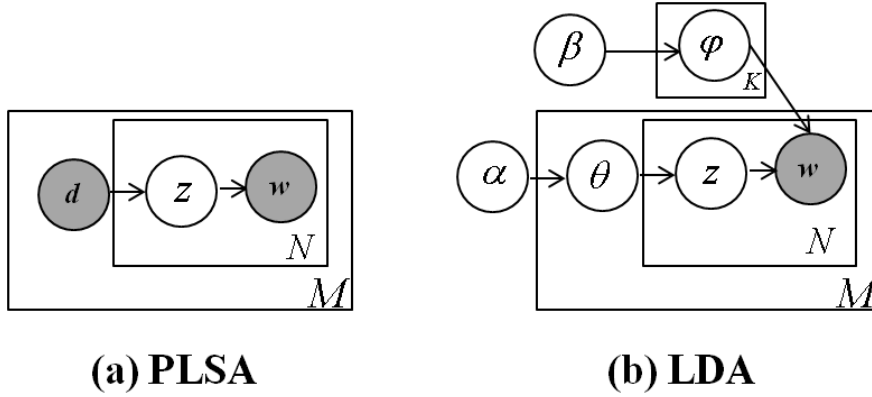


图 2.1: 主题模型

信息损失的形式求解对应的映射，包括基于行的聚类 and 基于列的聚类，最终要优化以下损失函数

$$\min_{C_X^*, C_Y^*} (I(X; Y) - I(\hat{X}; \hat{Y})) \quad (2.2)$$

其中  $I(X; Y)$  表示  $X$  和  $Y$  的互信息； $I(\hat{X}; \hat{Y})$  表示  $\hat{X}$  和  $\hat{Y}$  的互信息。 $C_X^*, C_Y^*$  是最终要求的函数

Banerjee 等人 [6] 则建议了一个泛化的最大熵联合聚类方法通过 Bregman 信息原则。除了成对的联合聚类的方法，Gao 等人 [41] 通过使用将基于一致的信息论的联合聚类方法扩展到文本、视觉等多种类型的联合聚类。为了处理真实世界中的噪音数据，Deodhar 等人 [32] 提出了一种鲁棒重叠的联合聚类方法。

### 2.2.2 概率主题模型

概率主题模型已经成功应用到各个领域比如文本、图片、视频和音乐等 [13, 14, 50, 97, 109, 120]。文献 [12] 比较全面的介绍了概率主题模型的相关知识。基本的概率主题模型包括潜在语义主题分析 (Probabilistic Latent Semantic Analysis, PLSA) [51] 和潜在狄利克雷分配 (Latent Dirichlet Allocation, LDA) [14] (如图 2.1)。对于潜在语义分析 (PLSA)，它假设整个文档集合对应  $K$  个主题，每个文档对应一个特定的  $K$  维主题分布。文档中每个词的生成分成两步 (1) 根据文档  $d$  的主题分布选择一个主题  $z$ ；(2) 由该主题对应生成一个词  $w$ 。为了解决潜在语义分析可能存在的过拟合问题，潜在狄利克雷分配引入狄利克雷先验。对于文档集合汇总的每篇文档，LDA 定义了如下的生成过程：(1) 对每



篇文档，从狄利克雷分布中抽取一个主题分布；（2）从主体分布抽样一个主题；（3）从所抽样的主题根据主体-词分布抽样一个单词。该模型常用吉布斯采样 [48] 的方法进行模型推断和参数估计。

由于本工作重点是基于地理位置的社会媒体的数据挖掘方法的研究，根据该类型社会媒体的特点即带有位置信息、多模态异质和数量巨大。针对这三个特点，我们对主题模型现有方法的介绍包括基于地理位置的主题模型、基于多模态的主题模型和基于大规模数据的主题模型。

**（1）基于地理位置的主题模型** 基于地理位置的主题模型通过融入位置信息将原有的主题模型进行变形 [52, 81, 110, 121]。例如，Yin等人 [121] 组合地理位置和文本信息用于发现地理主题同时比较不同地理位置下的主题。这里地理位置的信息的分布假定是高斯分布，为了支持检测更为复杂的非高斯分布的空间结构，Kling等人 [62] 提出了一个基于多狄利克雷过程的地理主题模型检测非高斯的地理位置主题。Pang等人 [81] 利用主题模型挖掘基于地理位置的两类主题，即全局主题和本地主题。这里全局主题是所有区域都有的主题而本地主题则是针对某个区域特有的主题。而Mei等人 [74] 进一步考虑时间因素进行主题的时空分析。和 [74] 相似，Hu等人 [53] 通过时空主题建模和分析同时考虑时间和商户位置信息为用户做出更为合理的推荐。

**（2）基于多模态的主题模型** 为了建立不同模态比如图片和文本之间的主题关联，基于多模态的主题模型得到了广泛的重视。Blei等人 [13] 提出了两类方法：多模态LDA和对应性LDA。这些方法用一组共享的潜在变量去直接建模图像和文本捕获不同模态之间的相关性。Putthividhy等人 [87] 提出了一个基于回归的主题模型，该模型允许不同模态的主题数量可以是不同的因而能够捕获更一般的关联结构。但以上工作均假设文本都是和相应的图片成对关联，但是现实情景下，文本和图片的关系是并不是完全对应的，为了能够更一般的建模多模态主题，Jia等人 [57] 提出了一个基于马尔科夫场的主题模型，该模型既能编码不同模态之间的关系，还可以建模单个模态的主题。Virtanen 等人 [108] 则将主题空间分成三类：文本特有的主题空间，文本和图片共享的主题空间和图片特有的主题空间，提出可分解的主题模型分别学习这三类主题空间。

**（3）基于大规模数据的主题模型** 为了将主题模型成功应用到大规模数据中，国内外一些学者提出了各种各样的在线的主题模型，主要包括基于PLSA的方法和基于LDA的方法。对于基于PLSA的在线方法，Hofmann等

人 [51]首先提出了一个在线模型, 该模型只能对新的文档进行模型的更新, 无法处理新文档中的新词。为了解决这个问题, Chou等人 [28]通过调整前一时间段数据流的模型推导出能够处理当前数据流的模型。该方法首先将新来的文档融入到训练的模型中进行模型更新然后基于新词进行模型的更新。Gohr等人 [35]也采用了相似的方法。相比 [28, 35], Brants等人 [15]提出了一个统一的模型能够同时处理新的文档和新词。以上提及的在线模型均是采用最大似然估计的方法。Chien 等人 [26]则提出了一个新的增量学习算法, 采用最大后验估计的方法更新模型参数。除了基于PLSA的在线方法, 基于LDA的在线主题方法也得到广泛的研究 [3, 19, 49]。Alsumait等人 [3] 提出了一个在线的LDA方法, 通过利用扩展的吉布斯采样方法推导出下一个时间段的主题-词分布。该模型能够通过将来自新文档的新词加入到词典中更新词典去检测新兴主题。Hoffman等人 [49]开发了一个在线的变分贝叶斯方法, 但是他们假定主题词分布是从有限的狄利克雷分布采样得到因而无法使用新词进行模型的更新。Canini等人 [19]通过从狄利克雷过程而不是狄利克雷分布采样主题解决了该问题。

此外, 随着社交媒体各种平台的发展, 国内外一些学者提出了基于跨域的主题模型, 比如Li等人 [65]将不同域的主题类型分成共享的主题和基于域特有的主题两类建立主题模型挖掘这两类主题模型进行文本分类。Paul等人 [85]提出了一个跨集合的主题模型, 该模型能够区分所有集合共享的主题以及共享主题中每个集合中的特色信息。Bao等人 [8] 则利用跨集合的主题模型实现了跨域分类。

## 第三章 基于场景和角度的地标总结

### 3.1 引言

地标是一个城市独特的建筑物，具有显著的物理、文化和历史特征，比如纪念碑、教堂和广场等。在线图片分享网站的快速发展和数字图片捕获设备的广泛应用为我们提供了大规模的地标图片集。如图3.1所示，这些地标图片集呈现出不同的场景（比如阴天和晴天场景）和不同的拍摄角度（比如正面拍摄、自底部向上拍摄、近距离拍摄和远距离拍摄等）。带有多场景和多角度的地标图片能够低冗余度的对地标进行全面的总结，这样的总结可以帮助更好的组织和浏览地标。而且地标总结的结果能够支持更多的应用，比如地标识别和分类、地标检索和自动的3D重构等。尽管现有的很多工作对基于视觉特征的地标总结 [102, 115, 136]有了深入的研究，但这些工作并未能同时考虑场景和角度的多样性进行总结因而可能无法为用户提供满意的地标总结结果。

我们认为一个地标的总结应该考虑以下两个方面（1）**场景** 一个地标在一天中不同的时间段内或者不同的天气情况下会展示不同的场景，比如大本钟在晴天时的场景和埃菲尔铁塔在大雪时的场景。不同类别的地标场景意味着不同的视觉感知；（2）**拍摄角度** 图片的拍摄角度具有多样性，包括正面拍摄、自底部向上拍摄、近距离拍摄和远距离拍摄等。图片由于不同的拍摄角度而反映出很大的内容不同 [115]。比如，近距离拍摄经常要描述图片的一些细节而远距离拍摄要给出一个大概的宏观描述。自底向上的拍摄角度可能要展示一些不同寻常的景色。本文的目的就是基于以上两个方面从地标的图片集中挖掘潜在的场景-角度主题更好的总结地标。

场景通常是由图片的表观信息表示而拍摄角度则是由几何信息表示。如何克服场景和角度表示的不一致性是我们解决基于场景和角度的地标总结的关键。为此我们提出了一个可视化的地标总结框架。首先通过角度聚类将相似角度的图片聚在一起。其次我们发现每一个角度聚类中不同的场景。一张图片可以通过视觉词袋表示，相同场景的地标应该由相同的视觉词表示。因此，为了发现所有角度聚类中的场景，我们提出了一个场景角度主题模型，该模型能够发现来自所有的角度聚类共享的场景主题以及每个角度聚类中特有的场景角度



图 3.1: 大本钟图片集

主题，同时能够按照角度区分这些主题。基于这些发现的主题，我们用相应的代表性的地标图片实现地标的总结。

### 3.2 相关工作

相关工作主要包括两个方面的内容：基于视觉特征的地标总结和跨集合主题模型。接下来我们分别给予介绍。

基于视觉特征的地标总结主要是选择代表性和多样性的地标图片以便于地标的浏览、组织和检索。Simon等人 [102]提出一个优化算法，考虑似然性、覆盖率和正交性三个准则发现图片集合中典型的场景图片总结地标。Zhao 等人 [136]通过主题建模的方法发现代表性的场景主题总结地标。不同于以上方法，Xue 等人 [115]根据图片的拍摄角度提出了一个角度建模的方法。此外，还有一类工作比如 [60, 61, 96]不仅利用视觉信息，还充分考虑一些元数据信息比如地理标签和时间信息辅助进行地标的总结。本工作与上述方法的不同在于我们的工作利用视觉信息同时考虑场景和角度两个方面实现地标总结。

由于本工作的重点是基于多个集合的主题建模，因此我们主要介绍主题模型中的跨集合主题模型的相关工作。Zhai等人 [129]提出了一个跨集合混合模型，该模型扩展原有的PLSA模型使之能够区分不同文档集中共享的主题信息和每一个集合特有的主题信息。为了防止过拟合，Paul 等人 [85]通过引入狄利

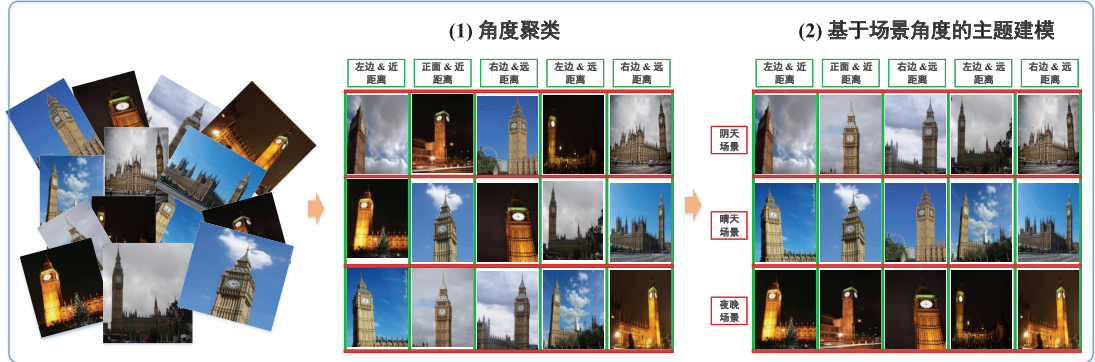


图 3.2: 可视化地标总结框架

克雷先验提出了跨集合的潜在狄利克雷分配模型。该模型已经成功应用到文本的分析比较中 [85, 129]。Bao等人 [8] 进一步扩展跨集合主题模型到有监督的模型。Gao等人 [42] 使用有监督的跨集合模型分别进行新闻和图像分类。不同于以上方法，我们跨集合的主题模型基于视觉信息同时考虑建模场景和角度实现地标的总结。

### 3.3 方法框架

本章提出了一种基于场景角度主题建模的方法，基于该模型，我们提出了一个可视化的地标总结框架。方法框架如图3.2所示，该框架包括两部分：角度聚类和基于场景角度的主题建模。首先利用谱聚类将同一地标图片按照拍摄角度分别聚类成不同的组；然后利用提出的基于场景角度的主题建模，从角度聚类中挖掘出共享的场景主题和每个角度聚类特有的场景角度主题；利用发现的主题找到代表性的图片总结地标。图3.2展示了基于场景角度主题的大本钟的总结结果，每一个主题是由带有某种场景和某种角度组合的一张图片表示。

### 3.4 角度聚类

Raguram等人 [66] 考虑到一个地标的拍摄角度是由图片中地标的空间结构决定的，同时证明全局的GIST描述子 [78] 能够根据感知相似度聚类空间结构相似的地标图片。受此工作的启发，我们同样采用GIST特征描述子，通过GIST特征构建特征的相似度矩阵进行聚类。我们通过以下的高斯核计算两张图片的相



似度 $W_{p,q}$ :

$$W_{p,q} = \exp\left(-\frac{\|\mathbf{v}_p - \mathbf{v}_q\|^2}{\sigma^2}\right) \quad (3.1)$$

其中 $\mathbf{v}_p$ 和 $\mathbf{v}_q$ 分别表示图片 $p$ 和 $q$ 的特征向量。

然后我们用构建的相似度矩阵通过谱聚类 [107] 聚类具有相似角度的图片。通过谱聚类，我们获得每个地标 $l$ 聚类的集合 $\{G_{l,1}, \dots, G_{l,k}, \dots, G_{l,|K|}\}$ 。为了选择代表性的角度聚类 and 过滤不一致的聚类，我们将聚类按照代表性进行排序。具体来说，对于每个聚类 $k$ ，我们通过如下的相似度公式计算类内的相似度 $S_{intra}$ :

$$S_{intra} = \frac{1}{|G_{l,k}|} \sum_{p \in G_{l,k}} \frac{1}{|G_{l,k}| - 1} \sum_{q \in G_{l,k} \setminus p} W_{p,q} \quad (3.2)$$

我们可以看出类内相似度 $S_{intra}$ 越高，则表明该类更加的紧凑和具有代表性。最终，我们选择类内相似度高的聚类进行接下来的主题建模。图3.3给出了两个地标大本钟和凯旋门的例子。每一列表示一个角度聚类，比如大本钟，每一列对应的拍摄角度包括近距离拍摄、正面拍摄和远距离拍摄等。每一个角度聚类 $k$ 的每一张图片 $p$ 通过如下的度 $D_p$ 来计算排序。

$$D_p = \sum_{q \in G_{l,k}} W_{p,q} \quad (3.3)$$

### 3.5 基于场景角度的主题建模

如图3.3所示，每个角度聚类包括不同的场景，比如法国凯旋门第二行包括夕阳、晚景、乌云和晴天四种场景。也即如果一张图片由相应的视觉词袋模型表示，不同场景的地标图片应该有相对应的视觉词表示。因此类比文档的主题模型，我们可以充分利用主题模型挖掘每个角度聚类中不同类别的场景主题。同时，每一种场景主题的表达还和对应的所在的角度聚类有关。为了挖掘这些主题，我们设计一种新的主题模型，即场景角度主题模型，该模型包括父层和子层两层。在父层，我们发现来自所有角度聚类共享的场景主题。每一个场景主题的词是由所有的角度聚类的图片共享。而在子层，我们挖掘来自同一角度聚类的场景角度主题，这些贡献于某一场景角度主题的词是针对某一角度聚类的。也即每一种角度的场景主题的词有两部分组成：一部分是所有角度聚类都



图 3.3: 角度聚类例子展示

共享的视觉词，另一部分则是针对每个角度所特有的视觉词。此外还包括背景主题。该类主题由无意义和无判别性的词组成。综上所述，图片的视觉词是由这三类主题共同生成的。为了完整的表示每一个场景主题，我们采用相同的主题索引 $z$ 。当从一个角度聚类 $g$ 产生一张图片的时候，我们首先选择主题索引 $z$ ，然后从这三类主题-词分布中采样视觉词，也即共享主题-词分布 $\phi_z^C$ ，场景角度主题-词分布 $\phi_{g,z}^S$ 和背景主题-词分布 $\phi^B$ 。我们引入变量 $x$ 控制选择不同类型的主题-词分布。图3.4给出了生成式模型的图表示，具体生成式过程如下：

1. 采样 $\phi^B \sim Dir(\beta^B)$ ， $\phi_z^C \sim Dir(\beta^C)$ ， $\phi_{g,z}^S \sim Dir(\beta^S)$
2. 采样多项式分布 $\psi_{g,z} \sim Dir(\gamma)$
3. 对于每张图片 $I$ 
  - (a) 采样一个主题混合 $\theta_I \sim Dir(\alpha)$
  - (b) 对于图片 $I$ 的第 $i$ 个词
    - i. 采样主题 $z_i \sim Mult(\theta_I)$
    - ii. 采样 $x_i \sim Mult(\psi_{g,z_i})$

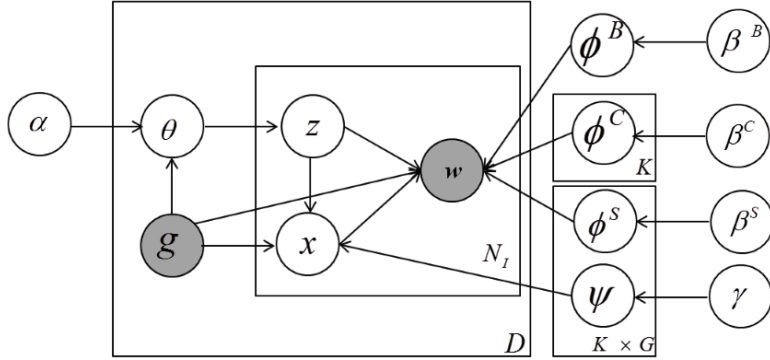


图 3.4: 主题模型的图表示

- iii. 如果  $x_i = 0$ , 采样一个词  $w_i \sim Mult(\phi^B)$
- 如果  $x_i = 1$ , 采样一个词  $w_i \sim Mult(\phi_{z_i}^C)$
- 如果  $x_i = 2$ , 采样一个词  $w_i \sim Mult(\phi_{g, z_i}^S)$

其中  $\alpha, \beta^B, \beta^C, \beta^S$  和  $\gamma$  是对称狄利克雷的超参数。

### 3.5.1 参数估计

在给定观测词  $\mathbf{w}$  的条件下, 我们的目标是估计新的赋值  $\mathbf{z}$  和  $\mathbf{x}$ 。具体来说, 我们使用吉布斯采样通过迭代的方式轮流估计  $\mathbf{z}$  和  $\mathbf{x}$ 。

$$p(z_i | \mathbf{x}, \mathbf{z}_{-i}, \mathbf{w}, \Phi) \propto \begin{cases} (n_I^{z_i, -i} + \alpha_{g_I z_i}) \times \frac{n_{G, x_i}^{w_i, -i} + \beta^B}{n_{G, x_i}^{(\cdot), -i} + W\beta^B}, & x_i = 0 \\ (n_I^{z_i, -i} + \alpha_{g_I z_i}) \times \frac{n_{z_i, x_i}^{w_i, -i} + \beta^C}{n_{z_i, x_i}^{(\cdot), -i} + W\beta^C}, & x_i = 1 \\ (n_I^{z_i, -i} + \alpha_{g_I z_i}) \times \frac{n_{g_I, z_i, x_i}^{w_i, -i} + \beta^S}{n_{g_I, z_i, x_i}^{(\cdot), -i} + W\beta^S}, & x_i = 2 \end{cases} \quad (3.4)$$

$$p(x_i | \mathbf{x}_{-i}, \mathbf{z}, \mathbf{w}, \Phi) \propto$$

$$\begin{cases} (n_{g_I, z_i}^{x_i, -i} + \gamma) \times \frac{n_{G, x_i}^{w_i, -i} + \beta^B}{n_{G, x_i}^{(\cdot), -i} + W\beta^B}, & x_i = 0 \\ (n_{g_I, z_i}^{x_i, -i} + \gamma) \times \frac{n_{z_i, x_i}^{w_i, -i} + \beta^C}{n_{z_i, x_i}^{(\cdot), -i} + W\beta^C}, & x_i = 1 \\ (n_{g_I, z_i}^{x_i, -i} + \gamma) \times \frac{n_{g_I, z_i, x_i}^{w_i, -i} + \beta^S}{n_{g_I, z_i, x_i}^{(\cdot), -i} + W\beta^S}, & x_i = 2 \end{cases} \quad (3.5)$$





表 3.1: 地标集合的统计信息

地标	#图片	地标	#图片
大本钟	13136	巴黎圣母院	8822
自由女神像	7575	凯旋门	8487
金门大桥	10745	勃兰登堡门	7582
伦敦塔桥	10814	万神庙	7153
埃菲尔铁塔	12419	罗勒大教堂	8458

题数量  $K = 8$ ,  $\alpha = 1.0$ ,  $\beta^B = \beta^C = \beta^S = 0.01$ ,  $\gamma = 1.0$ 。吉布斯采样迭代的次数为2000。

为了评估我们的方法，我们考虑比较以下两种基准方法：

- VC: 角度聚类 (Viewpoint Clustering, VC)
- LDA: 潜在狄利克雷分配 [13]

我们的方法为SV (Scene-Viewpoint)。对于所有的方法，我们为每个地标选择相同数量的图片以确保公平比较。具体来说，对于我们的方法，既然对于每一个角度聚类，场景主题的数量是8，那么来自所有角度聚类的场景主题的数量是  $8 \times 5 = 40$ 。同时为每个角度聚类的每个场景选择5张图片；对于VC，我们选择排名靠前的5个角度聚类然后从每一个角度聚类中选择40张图片。对于LDA，我们将排名前5的角度聚类聚在一起然后直接应用LDA。主题的数量和我们的方法一样为40。对每一个主题，我们按照  $p(z|I)$  选择前5张图片。

我们首先进行模型的定量评估。我们随机选择4个地标然后让20个用户为每一个地标标定有意义的主题。注意对于VC，我们认为一个角度聚类就是一个主题。表格3.2统计了发现的主题的数量。我们发现我们的方法发现更多有意义的主题。原因在于基准方法仅仅基于某一个维度角度或者场景聚类，而我们的方法同时考虑中这两个维度总结地标。

对于地标总结，我们选择每一个主题的排序第一的图片作为地标总结结果的展示。图3.5展示了三个地标金门大桥、勃兰登堡门和凯旋门的地标总结结果。图片的每一行对应一个场景。例如，在金门大桥(图3.5 (a))倒数第二行是夕阳景色而在勃兰登堡门(图3.5 (b))的最后一行是夜景。图片的每一列代表某一个角度。例如，金门大桥的第二列是正面近景而凯旋门(图3.5 (c)) 的最后一列

表 3.2: 发现的主题的数量统计

方法	大本钟	埃菲尔铁塔	巴黎圣母院	伦敦塔桥	均值
VC	5	5	5	5	5
LDA	22	24	26	21	23
SV	25	30	30	25	27



图 3.5: 三个地标(a) 金门大桥(b) 勃兰登堡门和(c) 凯旋门的总结展示。对于每一个地标，每一行代表一个场景而每一列代表一种拍摄角度

是正面远景。从结果展示可以看出基于场景和角度的地标总结达到了比较好的总结效果。

为了评估我们的方法，我们让20个用户从四个方面判断总结的结果：代表性、全面性、冗余度和满意度。用户被要求从这个四方面分别进行打分，分数从0到10（10代表最高分）。图3.6展示了所有用户得分的平均值。从中我们能够看到我们的方法在全面性、冗余度和满意度上得到了更为显著的增益。具体来说，我们的方法SV在全面性的得分是最高的。这是因为SV基于角度和场景进行地标的总结而另外两个基准方法仅仅针对某一个维度进行总结。此外，SV在冗余度上的得分要高于LDA，而LDA则要高于VC。SV在表示上的得分可以和LDA比较。

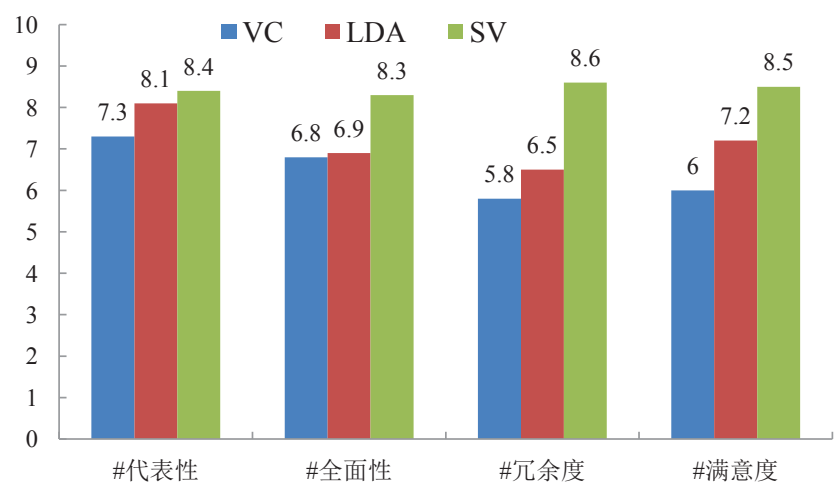


图 3.6: 用户体验比较

### 3.7 本章小结

本章提出了一个基于场景和角度的地标总结框架。在该框架中，我们首先将不同的图片按照角度进行聚类。然后通过提出的场景角度主题模型发现每个角度聚类中的场景。该模型能够从所有的角度聚类共享的场景主题以及每个角度聚类中特有的场景角度主题，同时能够按照角度区分这些主题。基于这些发现的主题，我们用相应的代表性的图片实现地标的可视化总结。实验结果展示我们方法的有效性。

## 第四章 基于时空主题的地标分析

### 4.1 引言

地标由于其独特的物理和文化特色而成为旅游者感兴趣的地方。这些独特的物理外观和文化特色构成了地标的本地特色。此外地标在不同的时刻会展示不同的景色或者发生某些重大活动，也即地标具有时间性。我们以日本的清水寺（Kiyomizu-dera）作为例子，从图4.3<sup>1</sup>中我们可以看到除了清水寺本身独特的位置特色“Temple”，还包括两类不同季节对应的场景包括春季的樱花和秋天的枫叶。如果我们能够同时提供清水寺的本地特色和不同时间段的景色给旅游者，那么旅游者就会很容易根据他们的兴趣爱好决定是否去清水寺旅游以及旅游的时间。因此挖掘和地标的本地特色和时间特色相关的知识有助于用户的旅游规划。



图 4.1: 清水寺（Kiyomizu-dera）的例子

图片分享网站比如Flickr和Picasa为我们提供了大量地标图片资源。并且每张图片还关联其他一些元数据信息，比如文本标签和时间信息等。现有的工作主要利用这些信息进行地标的总结 [60,61]、地标识别 [138]和检索 [5]。但是大部分工作均集中挖掘和地标本地特色相关的知识而忽略了地标的时特色。挖掘和地标的时特色相关的知识能够帮助旅游者根据自己的兴趣选择某一地标

<sup>1</sup>图4.3展示的文本文、图片和时间信息均来自Flickr



旅游的合适时间和在某一时间段内可以去旅游的地标,并且这些信息应该是旅游者需要了解的基本信息。因此本章考虑地标的两类主题,即和本地特色相关的位置主题以及和时间特色相关的时间主题进行地标的总结和分析。位置主题是特指某一个地标所特有的风格或外观属性,而时间主题则是在某个地标的某些时刻发生的主题,比如金门大桥的航空展和清水寺的红叶季等。此外地标还存在一类全局主题 [81],即所有地标所共有的主题,比如用户在上传地标图片时标注的天气或者交通信息等。位置主题作为代表性的属性已经被证明他的有用性。时间主题在地标的挖掘也同样是不可缺少的,可以作为位置主题的补充因而便于对地标的全面理解。并且时间主题能够为地标历史浏览和旅游者的旅游决策提供更加全面和结构化的信息。因此本章讨论发现地标的这三类主题并进行相关的分析。

为了发现这三类主题,本章提出了一个多模态时空主题模型充分利用多模态信息和地标的时空信息学习全局主题、位置主题和时间主题三类主题的子空间。同时我们考虑地标和位置主题的关联以及地标-时间和时间主题的关联引入了基于互信息的正则化的目标函数。基于这个模型,我们提出了一个地标时空主题框架,利用贝叶斯定理从地标的时间和空间两个方面对挖掘的主题进行分析。

## 4.2 相关工作

和本工作相关的两个领域包括基于多模态的地标总结和地理主题建模。

较早的基于多模态的地标总结的工作 [60]利用多模态信息和上下文信息,包括文本标签、视觉内容、地理标签和时间信息通过聚类的方法总结地标。Ji等人 [56]融合来自Flickr社区的内容和社区信息通过图建模的方法挖掘城市比较流行的地标。Li等人 [29, 67]收集大规模的地标数据集通过训练多类的支持向量机实现大规模的地标分类。此外一些工作 [43, 71]充分利用这些丰富的异质多模态信息开发旅游推荐系统。比如Gao等人 [43]提出了一个旅游推荐系统W2Go。W2Go以给定的城市名称作为查询,该系统能够使用自动的地标排序方法返回该城市中比较流行的地标。Xin等人 [71]提出了旅游路线推荐系统Photo2Trip,利用大规模图片的地理标签信息生成旅游路径便于旅游路径的推荐。以上工作主要集中挖掘和地标的本地特色相关的知识而忽略了地标的时



图 4.2: 地标分析框架: (1) 数据准备(2)主题建模(3)主题分析

对于地理主题发现，Yin等人 [121]组合地理位置和文本信息发现地理位置主题同时对在不同地理位置下的同一主题进行比较。为了支持检测更为复杂的非高斯分布的空间结构，Kling等人 [62] 提出了一个基于多狄利克雷过程的地理主题模型发现非高斯的地理位置主题。Pang等人 [81] 利用主题模型发现基于地理位置的两类主题，即全局主题和位置主题。这里全局主题是所有城市都有的主题而本地主题则是针对某个城市特有的主题。然后利用发现的位置主题的关键词检索Flickr的图片，实现城市总结的可视化。Mei等人 [74]进一步考虑时间因素对主题进行了时空分析。不同于 [74]，一方面我们同时发现三类主题也即全局主题、位置主题以及时间主题而 [74]主要集中于发现和分析时间主题。另一方面，我们的框架能够同时融入多种模态信息提供了一个多模态的主题发现和分析框架，而 [74] 仅仅是基于文本内容的主题分析。

### 4.3 方法框架

如图4.2，我们的框架包括三个模块：

- (1) **数据准备** 我们从Flickr爬取大量的地标图片同时进行预处理；
- (2) **主题建模** 我们提出了一个多模态时空主题模型发现三类主题，即全局主题、位置主题和时间主题；一方面我们的模型通过学习三类主题的子空间发现三类主题；另一方面我们的模型充分利用文本和视觉信息，使得

同一主题同时由相关的文本和图片共同表示。此外为了准确的学习三类子空间，我们考虑地标和位置主题的关联以及地标-时间和时间主题的关联引入了基于互信息的正则化的目标函数，最终通过泛化的期望-最大值算法进行参数估计。

- (3) **主题分析** 我们从地标的的时间和空间两个方面利用贝叶斯定理对发现的主题进行分析，包括基于某一个地标的主题分布和基于特定时间的主题的位置分布。

## 4.4 数据准备

为了发现和分析地标主题，我们需要建立地标数据库，主要包括（1）地标列表选择和（2）地标图片收集。对于地标列表选择，我们首先从维基百科旅游网页<sup>2</sup>中选择参观人次比较多的国家，对于每个国家，我们利用一些旅游网站比如Yahoo旅游指导等选择一些著名的地标。考虑到如果用户大量上传某一个地标的图片到图片分享网站比如Flickr，那么该地标就可能非常流行。因此我们进一步用Flickr API<sup>3</sup>过滤和选择地标图片。在给定一个地标名称作为输入，我们可以直接通过这个API获得返回图片的数量。我们选择返回图片数量超过200,000的地标作为最终的地标列表。接下来我们从Flickr上下载每个地标的图片集。对于每个地标，我们下载图片和关联的元数据，包括标签、标题、描述、拍摄时间和GPS信息。但是如果我们简单的选择地标名称作为输入收集图片，我们可能会得到比较多的噪音图片，原因在于（1）一些地标的名称是多义词。比如地标“White House”不仅可以表示美国总统的官邸还表示喷着白色油漆的房子；（2）许多图片同时标注不止一个地标。例如虽然一张图片标有大本钟和埃菲尔铁塔，但图片本身表示的是埃菲尔铁塔。为了解决这些问题，我们组合地标名称和相应的城市名称作为查询，例如为了爬取大本钟“Big Ben”的图片，对应的查询为“Big Ben, London”。这种约束能够极大的减少噪音图片。其次我们使用可以获得的地理标签信息进行去噪。具体来说，我们使用来自维基百科的每一个地标的经纬度信息作为聚类中心，进行基于地理位置的图片聚类。距离超过一定的阈值的图片将被过滤。对于没有地理信息标注的图片，如果他们的标注超过一个地标，我们就直接去除他们。

<sup>2</sup><http://en.wikipedia.org/wiki/Tourism>

<sup>3</sup><http://www.flickr.com/services/api/explore/flickr.photos.search>



## 4.5 多模态时空主题建模

多模态时空主题建模通过建模文档的生成过程发现多类主题。这里来自Flickr的每一张图片和关联的文本、时间信息被看做是一个文档。该模型需要发现文档中的三类主题：（1）**时间主题** 该类主题主要描述在某些特定时刻发生的一些活动或场景；（2）**位置主题** 这类主题和某一个地标关联，描述某个地标特有的物理外观；（3）**全局主题** 该类主题则是对应所有的地标都共享的主题。

基于以上概念，多模态时空主题建模的问题定义如下：给定文档集合 $D = \{d_1, d_2, \dots, d_{|D|}\}$ ，这里 $d_i$  包括标注的时间戳 $t_{d_i}$  和地标 $l_{d_i}$ ，该模型学习以下三类主题空间：（1）全局主题空间 $\varphi^{gl}$ ；（2）位置主题空间 $\varphi^{loc}$  和（3）时间主题空间 $\varphi^{tl}$ 。

### 4.5.1 文档的生成式过程

给定Flickr的文档集合 $D$ ，该集合的词典 $Y$ 包括文本词典 $W$ 和视觉词典 $V$ ，地标集合 $L$ 和时间间隔集合 $t_d \in T$ 。基于全局主题 $Z^{gl}$ 、位置主题 $Z^{loc}$  和时间主题 $Z^{tl}$ 的集合，文档中的每一个词 $d$ 是通过以下三种方式之一生成：（I）来自特定文档分布 $\theta_d$ 的全局主题 $Z^{gl}$ ；（II）来自某一特定地标的分布 $\psi_{l_d}$ 的位置主题；和（III）来自特定地标和时间分布 $\phi_{(l_d, t_d)}$ 的时间主题 $z \in Z^{tl}$ 。变量 $x$ 是从特定文档的多项式分布 $\pi_d$ 中采样得到，其中 $\pi_d \triangleq \{p(x|d) | x \in \{gl, loc, tl\}\}$ 表示三类主题的比例分布。图4.3表示生成式过程的图表示。其中 $\psi_l$ 是一个 $|L| \times |Z^{loc}|$ 的矩阵。 $\phi_{(l, t)}$ 是一个 $|LT| \times |Z^{tl}|$ 矩阵。 $|LT| = |L| \times |T|$ 。在我们的实验中，地标的数量 $|L| = 20$ ，时间范围从2010-01-01到2012-12-31。我们简单的将一个月分成四个时间间隔， $|T| = 3 \times 48 = 144$ 。来自文档集合的每个文档 $d$ 的生成式过程如下表示：对于来自文档 $d$ 中的每个词 $y_{d,n} \in Y$

1. 采样 $x_{d,n} \sim Multi(\pi_d)$
2. 如果 $x_{d,n} = gl$ , 则采样一个全局主题 $z_{d,n} \sim Multi(\theta_d)$
3. 如果 $x_{d,n} = loc$ , 则采样一个位置主题 $z_{d,n} \sim Multi(\psi_{l_d})$
4. 如果 $x_{d,n} = tl$ , 则采样一个时间主题 $z_{d,n} \sim Multi(\psi_{(l_d, t_d)})$
5. 采样 $y_{d,n} \sim Multi(\varphi_{z_{d,n}}^{x_{d,n}})$

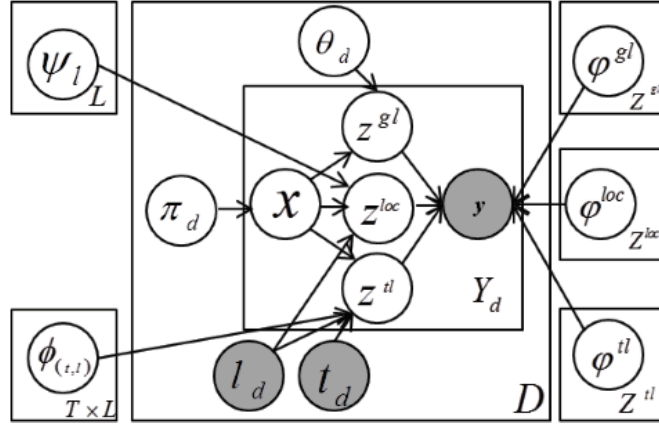


图 4.3: 地标主题模型的图表示

根据文档集合 $D$ 的生成式过程，基于文档集 $D$ 的似然估计可以表示成

$$\begin{aligned}
 L(D) = & \sum_{d \in D} \sum_{y \in Y} n(d, y) \\
 & \times \log[p(x = gl|d) \sum_{z \in Z^{gl}} \theta_{d,z} \varphi_{z,y}^{gl}] \\
 & + p(x = loc|d) \sum_{z \in Z^{loc}} \psi_{l_d,z} \varphi_{z,y}^{loc} \\
 & + p(x = tl|d) \sum_{z \in Z^{tl}} \phi_{(t_d, l_d), z} \varphi_{z,y}^{tl}
 \end{aligned} \tag{4.1}$$

#### 4.5.2 基于互信息的正则化

和 [81]相似，文档中每个位置主题的发生应该和这些对应的地标位置相关。我们用互信息测量位置集合 $L$ 和位置主题集合 $Z^{loc}$ 的相关性，定义一个基于位置的互信息如下：

$$I_l(L; Z^{loc}) \triangleq \sum_{l \in L} \sum_{z \in Z^{loc}} p(l, z) \log \frac{p(l, z)}{p(l)p(z)} \tag{4.2}$$

其中概率分布 $\{p(l)\}_{l \in L}$ 在没有位置的先验知识的情况下设为均匀分布。

同样给定一个地标，文档集合中的每个时间主题发生的概率应该和相应的时间-位置相关。基于时间-位置的互信息定义为

$$I_{(l,t)}((L, T); Z^{tl}) \triangleq \sum_{(t,l) \in (L, T)} \sum_{z \in Z^{tl}} p((l, t), z) \log \frac{p((l, t), z)}{p((l, t))p(z)} \tag{4.3}$$

其中概率分布  $\{p(l, t)\}_{(l,t) \in (L,T)}$  在没有先验知识的情况下同样被设为均匀分布。

### 4.5.3 参数估计

由于在目标函数中引入互信息的正则化项，因此我们利用泛化的期待-最大化算法（Generalized EM, GEM）[77]通过求解下面的优化问题估计参数。

$$\max(L(D) + \lambda_1 I_l(L; Z^{loc}) + \lambda_2 I_{(l,t)}((L, T); Z^{lt})) \quad (4.4)$$

这里  $\lambda_1$  and  $\lambda_2$  是正则化的参数。

具体的估计过程如下：

**E-Step:**

$$p(z, x|d, w) \propto \begin{cases} p^{(n)}(x|d)(\theta_{d,z})^{(n)}(\varphi_{z,w}^{gl})^{(n)} & x = gl, z \in Z^{gl} \\ p^{(n)}(x|d)(\psi_{l_d,z})^{(n)}(\varphi_{z,w}^{loc})^{(n)} & x = loc, z \in Z^{loc} \\ p^{(n)}x|d(\psi_{(t_d,l_d),z})^{(n)}(\varphi_{z,w}^{lt})^{(n)} & x = tl, z \in Z^{tl} \end{cases} \quad (4.5)$$

**M-Step:**

$$p^{(n+1)} = \frac{\sum_{z \in Z^x} \sum_{y \in Y} n(d, y) p(z, x|d, y)}{\sum_{y \in Y} n(d, y)}, \quad (4.6)$$

$$x \in \{gl, loc, tl\}$$

$$(\varphi_{z,w}^x)^{(n+1)} = \frac{\sum_{d \in C} n(d, y) p(z, x|d, y)}{\sum_{Y' \in Y} \sum_{d \in C} n(d, y) p(z, x|d, Y')}, \quad (4.7)$$

$$x \in \{gl, loc, tl\}, z \in Z^x$$

$$(\theta_{d,z})^{(n+1)} = \frac{\sum_{y \in Y} n(d, y) p(z, x = gl|d, y)}{\sum_{z' \in Z^{gl}} \sum_{y \in Y} n(d, y) p(z', x = gl|d, Y')}, \quad (4.8)$$

$$z \in Z^{gl}$$

$$(\psi_{l,z})^{(n+1)} = \frac{\sum_{d|l_d=l} \sum_{y \in Y} n(d, y) p(z, x = loc|d, y)}{\sum_{z' \in Z^{loc}} \sum_{d|l_d=l} \sum_{y \in Y} n(d, y) p(z', x = loc|d, y)}, \quad (4.9)$$

$$z \in Z^{loc}$$

$$\begin{aligned}
(\psi_{(t,l),z})^{(n+1)} = & \frac{\sum_{d|t_d=t, l_d=l} \sum_{y \in Y} n(d, y) p(z, x = tl | d, Y)}{\sum_{z' \in Z^{loc}} \sum_{d|t_d=t, l_d=l} \sum_{y \in Y} n(d, y) p(z', x = tl | d, y)}, \quad (4.10) \\
& z \in Z^{tl}
\end{aligned}$$

## 4.6 时空主题分析

在GEM参数估计之后，我们利用这些参数进行主题分析，包括主题的时间分布和主题的位置分布。

### 4.6.1 主题的时间分布

一个地标的时间主题包括周期性的主题和非周期性的主题。周期性的主题是在某些固定时刻经常发生的主题而非周期性的主题则是仅在某些时刻发生的主题。在给定一个地标 $l$ 的情况下，我们通过计算在时间主题 $z$ 和地标 $l$ 的条件下的时间 $t$ 的分布区分这两类主题，也即 $p(t|z, l)$ 。 $p(t|z, l)$ 通过如下的贝叶斯定律计算。

$$p(t|z, l) = \frac{\phi_{(l,t)} p(t, l)}{\sum_{t' \in T} \phi_{(l,t')} p(t', l)} \quad (4.11)$$

其中 $p(t, l)$ 是在位置 $l$ 和时间间隔 $t$ 对应的词的数量和来自所有时间段的词的数量之比。

### 4.6.2 主题的位置分布

除此之外，我们也可以分析在给定的特定时间间隔的条件下的主题的位置分布。同样我们也可以通过贝叶斯定理进行推导

$$p(l|z, t) = \frac{\phi_{(l,t)} p(t, l)}{\sum_{l' \in L} \phi_{(t,l')} p(t, l')} \quad (4.12)$$

同样 $p(t, l)$ 是在时间间隔 $t$ 和地标 $l$ 对应的词的数量和来自所有地标下的词的数量。

表 4.1: 地标的统计

地标	地标
凯旋门	大本钟
布鲁克林大桥	白金汉宫
埃菲尔铁塔	华盛顿纪念碑
紫禁城	金门大桥
长城	清水寺
伦敦眼	林肯纪念堂
自由女神像	巴黎圣母院
颐和园	悉尼歌剧院
东京塔	伦敦塔桥
特拉法尔加广场	白宫

4.7 实验结果与分析

我们在来自6个国家的20个地标（表4.1）上进行实验评估验证我们模型的有效性。

这些地标图片拍摄的时间范围从2010-01-01到2012-12-31。每张图片的标签、标题和描述连在一起作为图片的文本信息。我们通过预处理去除文本中的停用词、html和摄像机相关的词比如“Cannon”和“35mm”。此外，既然地标名字和相关的城市名称对于发现地标主题是没有太大意义的，我们也全部将他们去除。然后我们去掉频率低于50的词，选择超过8个词的图片作为我们最后的数据集。最终的图片集包括435,810图片和由22,703个词组成的文本词典。对于图片特征，我们从每张图片中提取1024维的HOG [30]特征。

4.7.1 主题模型评估

我们通过训练提出的主题模型学习三类主题，即全局主题、位置主题和时间主题。对应的主题数量经验性的分别设置为50、50和150。此外，我们设置(4.4)的正则化参数 $\lambda_1 = \lambda_2 = 1e + 6$ 。位置的粒度被设置为一个地标。对时间的粒度，我们简单的将一个月分成4个时间间隔，前三个时间间隔以7天作为一个单元而每个月的剩余部分作为第四个时间间隔。这里要注意的是，我们的框架可以很容易的扩展到不同的位置和时间粒度去。

#### 4.7.1.1 定性的事例分析

我们通过提供代表性的关键词和图片解释发现的主题。文本标签可以根据 $p(w|z)$ 排序得到而图片则通过如下定义的主题 $Z_i$ 和文档 $d_j$ 之间的余弦相似度来排序：

$$\text{sim}(Z_i, d_j) = \frac{(\mathbf{w}_{z_i}, \mathbf{v}_{z_i})(\mathbf{w}_{d_j}, \mathbf{v}_{d_j})}{|(\mathbf{w}_{z_i}, \mathbf{v}_{z_i})||(\mathbf{w}_{d_j}, \mathbf{v}_{d_j})|} \quad (4.13)$$

其中 $(\mathbf{w}, \mathbf{v})$ 分别表示连接的文本词和可视化词的向量。图4.4展示了一些发现的主题的例子。我们根据 $\theta_d$ 、 $\psi_l$ 和 $\phi_{(l,t)}$ 标注发现的全局主题、位置主题和时间主题。为了便于解释发现的主题，我们用地标名标注位置主题。而对于时间主题，则标注主题的标题和时间。从图4.4中我们可以看到我们提出的主题模型能够发现一些时间主题，包括某些时刻显著性的景色（比如时间主题#114）和一些重要活动（比如时间主题#67 布鲁克林大桥的占领华尔街运动和时间主题#70 伦敦塔桥的伦敦马拉松）。通过这些时间主题，我们能够加深对地标社会功能的理解。对于位置主题，他们主要描述某些地标的风格，例如位置主题#27 描述自然风格“harbor”。相比之下，全局主题比如#12和#50 倾向于和许多地标相关联。

#### 4.7.1.2 定量估计

为了定量评估我们提出的模型，我们考虑以下三种基准方法作为比较：

- PLSA [51]: 该方法直接学习250个主题，没有考虑视觉模态。
- 日志建模（Travelogue Model, TM [81]): 该方法直接学习全局主题和位置主题，没有考虑视觉模态。我们设定全局主题和位置主题的数量分别为50和200. 我们经验的设定正则化参数的值 $\lambda = 1e + 6$
- mmSTTM\_Text: 这种方法和mmSTTM模型类似，只是没有考虑视觉模态。三类主题设置的数量和我们的方法一致。

既然我们的目标是用发现的位置主题和时间主题分析地标，我们首先评估有意义的位置主题和时间主题的数量。具体说来，我们用MAP@K评估。

我们让10个用户标注语义有意义的主题。他们能够依靠Flickr的数据库和外部资源帮助判断。表格4.2统计了各种方法发现的主题数量。我们能够发现我们提出的mmSTTM模型超出了其他三类基准方法。同时我们通过实验结果发现：（1）既然TM方法能够区分全局和位置主题，因而能够过滤掉代表全

主题类型	前5个词	前5张相关的图片				
全局主题#12	blue 0.2531 sky 0.23845 world 0.063868 nature 0.023623 cloud 0.02143	 0.312264	 0.228972	 0.228302	 0.215036	 0.203640
全局主题#17	night 0.38077 lights 0.13088 dark 0.020122 nightshot 0.017241 nightphotography 0.01534	 0.311728	 0.307387	 0.297862	 0.282610	 0.262326
全局主题#50	architecture 0.11974 building 0.097822 monument 0.033438 historic 0.032848 famous 0.031878	 0.146580	 0.143253	 0.122393	 0.103129	 0.099448
位置主题#5 颐和园	kunming 0.022126 lake 0.016091 longevity 0.015694 hill 0.0097054 bridge 0.0095046	 0.157004	 0.152931	 0.152222	 0.150788	 0.150391
位置主题#26 大本钟	thames 0.12577 river 0.055631 southbank 0.044495 riverthames 0.031256 bank 0.024509	 0.135706	 0.135705	 0.131267	 0.130258	 0.126970
位置主题#27 悉尼歌剧院	harbour 0.057095 vivid 0.038343 bridge 0.031438 harbourbridge 0.0206 sydneyharbour 0.0195	 0.078688	 0.072973	 0.071117	 0.067875	 0.066036
时间主题#67 10/01/2011-10/07/2011 布鲁克林大桥 占领华尔街	occupy 0.094204 occupywallstreet 0.079832 ows 0.058307 march 0.032343 wall 0.029606	 0.101686	 0.101680	 0.101607	 0.101482	 0.101471
时间主题#70 08/08/2012-08/14/2012 伦敦塔桥 伦敦马拉松	marathon 0.22029 philippound 0.040515 runner 0.025902 philip 0.023851 pound 0.023138	 0.203266	 0.203159	 0.202948	 0.185978	 0.183135
时间主题#114 04/01/2010-04/07/2010 华盛顿纪念碑 樱花季	spring 0.12389 cherry 0.08374 blossom 0.060646 april 0.047882 sakura 0.043024	 0.148427	 0.147661	 0.145635	 0.142938	 0.135086

图 4.4: 发现的主题展示



表 4.2: 发现主题的性能评估

# 方法	# 位置主题	# 时间主题
PLSA	17	22
TM	20	25
mmSTTM_Text	27	39
mmSTTM	30	45

局主题的词，所以性能超过了PLSA，这里得到的统计结果和 [81]一致；(2) mmSTTM\_Text方法超出了TM方法，这是因为该方法能够区分三类主题；(3) 我们的模型进一步提高了性能，这是因为多模态信息之间能够相互补充增强。

表 4.3: 发现主题的性能评估

# 方法	文本词MAP@10	图片词MAP@10
PLSA	0.6188	0.5887
TM	0.6867	0.6368
mmSTTM_Text	0.7837	0.7411
mmSTTM	0.8399	0.8025

为了进一步评估有意义的主题的语意一致性，我们让10个用户去标注每一个标注的主题前10个关键词是否和该主题相关。对于每一个词，如果超过6个用户认为是相关的则标注为1，否则为0。标签排序通过MAP@10评估。相似的方法被用于标注的主题的图片。表格4.3给出了实验结果，再一次验证了我们方法的优越性。

#### 4.7.2 主题分析

对于主题的时间分布，我们按照(4.11)画出了同一个地标主题的时间分布。不失为一般性，我们随机选择地标金门大桥作为一个例子进行主题的时间分布的分析。我们首先根据 $\phi_{(l,t)}$ 人工标注每一个时间主题。图4.5显示了来自金门大桥不同主题的分布。我们能够看到主题“Fleet Week”有明显的周期性。他发生在每年的10月6号到10月9号。相比之下，其他时间主题尤其是“75 anniversary celebration”属于非周期主题。从中我们很容易分析其他地标的主题时间分布。通过这种分析，我们能够区分周期性和非周期性的主题。



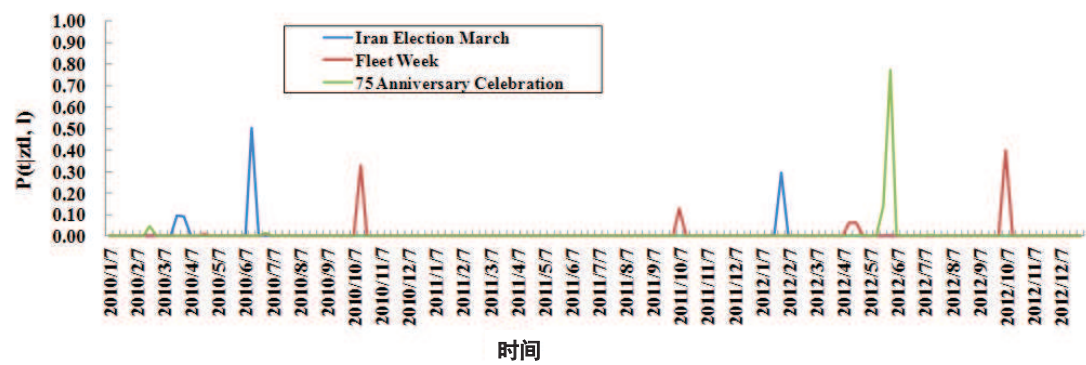


图 4.5: 金门大桥主题的时间分布

地标	07/22/2012-07/31/2012		08/01/2012-08/07/2012		08/08/2012-08/14/2012	
	$P(l z_t, t)$	相关的图片	$P(l z_t, t)$	相关的图片	$P(l z_t, t)$	相关的图片
伦敦塔桥	0.64		0.58		0.30	
大本钟	0.30		0.11		0.10	
伦敦眼	0.05		0.01		0.04	
白金汉宫	0.003		0.07		0.33	
特拉法加广场	0.0001		0.22		0.219	

图 4.6: 主题“2012 Olympic games in London”的位置分布，时间范围为2012-07-22 2012-08-14

而对于在给定时间间隔某个主题的位置分布，我们以发生在2012年07月22日到2012年08月14日的“2012 Olympic games in London”主题作为例子进行分析。图4.6展示了相应的分析结果。既然2012年的奥林匹克运动发生在伦敦，该主题主要是基于来自伦敦的几个地标的分布，比如伦敦塔桥和大本钟。我们能够看到主题主要分布在伦敦塔桥。这是因为伦敦塔桥带有欧林匹克环因而变成欧林匹克运动中最重要旅游景点之一。除此之外，在白金汉宫，主题在第三周变得最强，原因可能是更多的人喜欢观看男子马拉松。通过这种分析，我们展示了重要的事件的是如何影响地标的流行度的，因而这样的分析也便于进一步的地标推荐。

## 4.8 本章小结

本章提出了一个多模态时空主题模型挖掘地标的三类主题，即所有地标都享有的全局主题，只和某个地标相关的位置主题以及在某个地标的某些时刻发生的时间主题。基于该模型，我们提出了一个地标主题分析框架，该框架包括三个部分：数据准备、主题建模和主题分析。该框架从地标的时间和空间两个方面分析了挖掘的主题。实验的主观和客观评价表明了我们模型和框架的有效性。未来我们的工作包括如下两个方面：一方面基于实践的目的我们将提出的模型转为在线的模型以应对大规模地标数据；另一方面，我们将提出的模型应用到其他方面，比如基于事件的地标浏览和地标流行度分析等。

## 第五章 基于异质元数据的社会事件检测

### 5.1 引言

社会媒体的迅速发展和GPS手持移动设备的广泛应用使得用户上传和分享大量的带有地理位置标注的图片或者视频，这些地理数据通常关联时间等各种异质元数据信息，为我们捕获各种各样真实世界的社会事件提供了信息源（如图5.1）。检测来自社会媒体的事件一方面可以应用于社会媒体的组织和检索；另一方面能够促进社会媒体的语义理解。因而社会事件检测受到了越来越多的关注。例如MediaEval组织<sup>1</sup>近年来组织各种各样的社会事件检测的挑战，推动了社会事件检测技术的发展，但社会事件检测的性能仍然有巨大的改进空间。

传统的事件检测通常只涉及到一种模态，比如新闻文档或者图片。相比之下，地理数据包括各种各样的异质元数据，比如Flickr的图片除了包含位置信息外，还包括时间、视觉和文本信息等。因此如何有效融合多种异质元数据信息成为社会媒体下基于地理数据的事件检测的一个重要问题 [86] [10] [23] [111] [142] [84]。一些工作比如文献 [10] 利用包括时间戳、位置和文本信息在内的元数据信息，训练一个分类器学习媒体之间的相似度。基于学习的相似度实现社会媒体事件的聚类。Petkos等人首先使用来自所有模态成对相似的集合训练分类器预测同一类别的关系，然后利用预测的关系进行多模态聚类。这类方法的局限性在于过度依赖社会事件集合现有的知识。另一些工作则是序列的处理这些异质元数据。例如Wang等人 [111] 的社会事件检测框架包括两个聚类过程，第一个聚类是基于时间的聚类而第二个聚类则是基于位置和文本信息。Papadopoulos等人 [84] 将社会事件检测分为三步：基于城市级分类器的图片过滤，基于时间信息的事件分割和基于关联元数据的事件扩展。以上方法没有有效的融合所有异质元数据的信息进行统一建模。

不同于以上工作，本章工作将社会事件检测看做是一个聚类问题，通过考虑社会媒体和其他类型元数据之间的关系联合聚类社会媒体和关联的多种异质数据，实现对所有类型信息的有效融合和统一建模。本章的主要贡献如下

- 提出了一个高阶的联合聚类方法。一方面采用星结构的K分图建模社会媒体

---

<sup>1</sup><http://www.multimediaeval.org/>



图 5.1: 来自Flickr网站的社会事件例子

和关联的位置、时间、文本和视觉信息之间的依附关系，另一方面考虑时间空间内时间之间的关系，将该关系作为全局正则化项引入到总的目标函数中进行优化。

- 提出了一个社会事件监测框架。该框架包括基于高阶联合聚类的粗事件检测和聚类样本改进两部分，一方面我们通过粗的社会事件检测得到事件聚类的初步结果，另一方面我们基于启发式的方法过滤和融合聚类，同时去除聚类中有噪音的图片，进一步提高了社会事件检测的精度。

## 5.2 相关工作

本节针对和我们工作相关的两个领域社会事件检测和联合聚类逐一介绍。

### 5.2.1 社会事件检测

基于社会媒体的事件检测因其巨大的应用前景而受到了广泛的关注。Papadopoulos等人 [84]提出了一个图像分析框架用于自动从大规模的图片集

合中检测地标和事件以改进用户的浏览和搜索体验。Chen等人 [23]使用来自Flickr图片的文本标签和关联的时间和经纬度信息通过小波变换的方法发现社会事件。Firan等人 [39]充分挖掘用户产生的社会信息包括图片的标签、标题和图片描述信息将图片按事件进行分类。此外, MediaEval 每年组织社会事件检测的挑战 [82] [83]以推进社会事件检测技术的不断发展。许多研究者提出了各种有效的方法解决这个挑战。在2011年的挑战中, Brenner等人 [16] 组合各种各样的图片的信息和外部的数据源训练一个分类器模型。Liu等人 [69]通过以下三步解决事件检测问题, 首先在给定的条件下通过查询获得所有的事件实例。然后使用提出的事件鉴别模型预测事件和图片之间的关系。最终采用一些启发式的方法进一步改进事件检测的结果。在2012年的挑战中, Schinas等人 [99]提出了一个基于同类模型的方法, 该方法使用来自SED2011挑战的数据训练模型用于预测两张图片是否属于同一个事件。Brenner [17]提出一个框架融合来自各种各样的模态信息和外部数据信息通过分类的方法检测社会事件和检索关联的图片。

聚类方法已经广泛应用在基于文档和社会媒体的事件检测中。一些工作利用决定性的聚类方法将样本聚到不同的主题中, 比如K均值聚类、单链接和单通道聚类方法 [2]。由于社会媒体包括多种异质元数据, 如何恰当的融合这些信息是社会事件检测的重要挑战之一。Quack等人 [90]提出一个方法从社区图片集合中检测事件和物体。首先利用各种模态信息包括视觉、文本和空间位置关系进行图片聚类。获得的聚类进一步通过训练分类器分成事件和物体。Makkonen等人 [73]首先提取包括名字、时间和位置在内的语义特征, 然后组合这些特征学习一个相似度函数进行聚类。Petkos等人 [86]利用多模态的相似度训练一个分类器预测“同类”的关系, 然后通过这种预测的同类关系聚类测试样本。另一类工作则序列的处理各种不同的异质数据进行社会事件检测。比如Zeppelzauer等人 [128]首先利用可靠的信息包括时间和GPS信息聚类获得鲁棒的候选事件, 然后使用一些情景信息比如文本标签和视觉信息过滤和改进候选事件。Wang等人 [111]的社会事件检测包括两个聚类过程, 第一个是基于时间的聚类, 第二个则是基于位置和文本信息的聚类。不同于以上方法, 我们构建星结构的K分图同时融合多种异质信息, 基于K分图联合聚类社会媒体和关联的异质元数据集去解决传统融合方法带来的问题。



### 5.2.2 联合聚类

不同于传统的基于一维的聚类方法，联合聚类能够考虑多种异质元数据的相关性同时在不同维度上进行聚类。Dhillon等人 [33] 将联合聚类看做是一个二分图分割问题，利用提出的谱聚类方法进行联合聚类。该方法的局限性在于每一种源的数据都至少和另一个源的数据结合在一起，这个条件在现实中很难满足，为了克服这个问题，Dhillon等人 [34] 又提出一个基于信息论的联合聚类方法以最小互信息损失的形式求解对应的映射，包括基于行的聚类和基于列的聚类。Banerjee等人 [6] 通过Bregman信息原则提出了一个泛化的最大熵联合聚类方法。除了成对的联合聚类的方法，Gao等人 [41] 提出了基于一致的信息论的联合聚类方法将基于两维的聚类方法扩展到多种类型。为了处理真实世界中的噪音数据，Deodhar [32] 提出了一种鲁棒重叠的联合聚类方法能够同时有效的挖掘稠密的、任意位置的和可能重叠的所有维度的类。

由于不同元数据集合的聚类数量的不同，受工作 [41] 提出的一致信息论的联合聚类方法的启发，我们提出了一种改进的高阶联合聚类模型对社会事件进行检测。和他们工作的不同在于我们的模型还可虑了基于时间空间的时间之间的关系，将该关系作为正则化项引入到最终的优化目标函数中。

## 5.3 方法框架

本章将社会事件检测看做是一个联合聚类问题，考虑社交媒体和他关联的多种异质元数据之间的依附关系，联合聚类社交媒体和关联的多种异质元数据。基于该方法，我们提出了一个社会事件检测框架，该框架包括两部分：（1）粗社会事件检测和（2）聚类和样本改进。

- （1）**粗社会事件检测** 我们提出了一个基于星结构的高阶联合聚类方法。首先通过构建星结构的 $K$ 分图建模社交媒体和关联的4种类型的元数据包括位置、时间、视觉和文本信息之间的关系。其次我们虑时间空间上时间与时间之间的关系，并作为正则化项引入到总的目标函数中。利用这两类关系我们构建总的目标函数迭代的联合聚类涉及到的所有 $K$ 个集合，得到初步的社会事件检测结果。
- （2）**聚类和样本改进** 对于初步的聚类结果，通过一些启发式的原则过滤和融合聚类，同时根据设定的阈值去除每个事件聚类中不相关的样本，以进

一步改进社会事件检测的性能。

首先，我们对使用的符号进行简要的介绍：集合用大写的欧拉字体表示，例如 $\mathcal{X}$ 。矩阵用大写的粗体表示，例如 $\mathbf{G}$ 。向量用小写粗体表示，例如 $\mathbf{h} = [h_1, h_2, \dots, h_n]^T$ 。大写字母如 $X$ 表示随机变量。假如我们有一个包含 $n$ 张图片的集合， $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ 。每一张图片 $x$ 包括四种元数据信息，即位置、时间、视觉和文本信息。并且我们将时间戳按照预定的时间间隔进行划分，此外将位置按照所在的城市进行划分。按照这种划分，我们得到了相应的时间集合 $\mathcal{T} = \{t_1, t_2, \dots, t_{|t|}\}$ 和位置集合 $\mathcal{L} = \{l_1, l_2, \dots, l_{|l|}\}$ 。其中 $t_i$ 表示第 $i$ 时间间隔而 $l_i$ 表示第 $i$ 个城市。对于视觉和文本内容，视觉词集合 $\mathcal{V} = \{v_1, v_2, \dots, v_{|v|}\}$ ，关键词集合 $\mathcal{W} = \{w_1, w_2, \dots, w_{|w|}\}$ 。我们的目的是通过联合聚类 $\mathcal{T}$ 、 $\mathcal{L}$ 、 $\mathcal{V}$ 、 $\mathcal{W}$ 和 $\mathcal{X}$ ，将 $\mathcal{X}$ 同时分割成 $|\hat{t}|$ 、 $|\hat{l}|$ 、 $|\hat{v}|$ 、 $|\hat{w}|$ 和 $|\hat{n}|$ 。我们定义 $X$ 、 $T$ 、 $L$ 、 $V$ 和 $W$ 为离散的随机变量，并且取值集合分别为 $\mathcal{X}$ 、 $\mathcal{T}$ 、 $\mathcal{L}$ 、 $\mathcal{V}$ 和 $\mathcal{W}$ 。 $X = x$ 、 $T = t$ 、 $L = l$ 、 $V = v$ 和 $W = w$ 的概率分别表示为 $p(x)$ 、 $p(t)$ 、 $p(l)$ 、 $p(v)$ 和 $p(w)$ 。 $(X, T) = (x, t)$ 、 $(X, L) = (x, l)$ 、 $(X, V) = (x, v)$ 和 $(X, W) = (x, w)$ 的联合概率分布被分别定义为 $p(x, t)$ 、 $p(x, l)$ 、 $p(x, v)$ 和 $p(x, w)$ 。离散的随机变量 $\hat{X}$ 、 $\hat{T}$ 、 $\hat{L}$ 、 $\hat{V}$ 和 $\hat{W}$ 的定义与上述类似。同样我们有 $p(\hat{X})$ 、 $p(\hat{T})$ 、 $p(\hat{L})$ 、 $p(\hat{V})$ 和联合的概率分布 $p(\hat{X}, \hat{T})$ 、 $p(\hat{X}, \hat{L})$ 、 $p(\hat{X}, \hat{V})$ 、 $p(\hat{X}, \hat{W})$ 。聚类的社交媒体集合表示为 $\hat{\mathcal{X}} = \{\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \dots, \hat{\mathcal{X}}_{|\hat{n}|}\}$ ，时间聚类 $\hat{\mathcal{T}} = \{\hat{\mathcal{T}}_1, \hat{\mathcal{T}}_2, \dots, \hat{\mathcal{T}}_{|\hat{t}|}\}$ ，位置聚类 $\hat{\mathcal{L}} = \{\hat{\mathcal{L}}_1, \hat{\mathcal{L}}_2, \dots, \hat{\mathcal{L}}_{|\hat{l}|}\}$ ，视觉词聚类 $\hat{\mathcal{V}} = \{\hat{\mathcal{V}}_1, \hat{\mathcal{V}}_2, \dots, \hat{\mathcal{V}}_{|\hat{v}|}\}$ ，关键词聚类 $\hat{\mathcal{W}} = \{\hat{\mathcal{W}}_1, \hat{\mathcal{W}}_2, \dots, \hat{\mathcal{W}}_{|\hat{w}|}\}$ 。列向量 $\mathbf{h}^x = \{h_1^x, \dots, h_{|n|}^x\}$ ， $\mathbf{h}^t = \{h_1^t, \dots, h_{|t|}^t\}$ ， $\mathbf{h}^l = \{h_1^l, \dots, h_{|l|}^l\}$ ， $\mathbf{h}^v = \{h_1^v, \dots, h_{|v|}^v\}$ 和 $\mathbf{h}^w = \{h_1^w, \dots, h_{|w|}^w\}$ 分别表示元素 $\mathcal{X}$ 、 $\mathcal{T}$ 、 $\mathcal{L}$ 、 $\mathcal{V}$ 和 $\mathcal{W}$ 的聚类索引。

## 5.4 基于高阶联合聚类的粗社会事件检测

### 5.4.1 星结构的K分图

一个 $K$ 分图被分成 $K$ 个非连接的集合，且通过边连接任意两个集合 [70]。星结构的 $K$ 分图是 $K$ 分图的一种特殊情况：包括一个中心顶点集合以及连接这个中心顶点的其他集合 [100]。如图5.2所示，我们定义星结构的 $K$ 分图为 $(\mathcal{X}, \mathcal{T}, \mathcal{L}, \mathcal{V}, \mathcal{W}, \mathbf{E}_T, \mathbf{E}_L, \mathbf{E}_V, \mathbf{E}_W)$ 。其中 $K = 5$ ， $\mathbf{E}_T(i, j) = e(x_i, t_j)$ 表示第 $i$ 个社交媒体样本和第 $j$ 个时间间隔的共生次数。 $\mathbf{E}_L(i, j) = e(x_i, l_j)$ 表示第 $i$ 个社交媒体样本和

第 $j$ 个位置的共生次数,  $\mathbf{E}_V(i, j) = e(x, v)$ 表示第 $i$ 个社交媒体样本和第 $j$ 个视觉词的共生次数。  $\mathbf{E}_W(i, j) = e(x, w)$ 表示第 $i$ 个社交媒体样本和第 $j$ 个关键词的共生次数。联合的概率分布 $p(x, t)$ 、 $p(x, l)$ 、 $p(x, v)$  and  $p(x, w)$  通过分别归一化相应的共生次数得到。例如 $p(x, t) = \frac{e(x, t)}{\mathbf{1}^T \mathbf{E}_T \mathbf{1}}$ 。

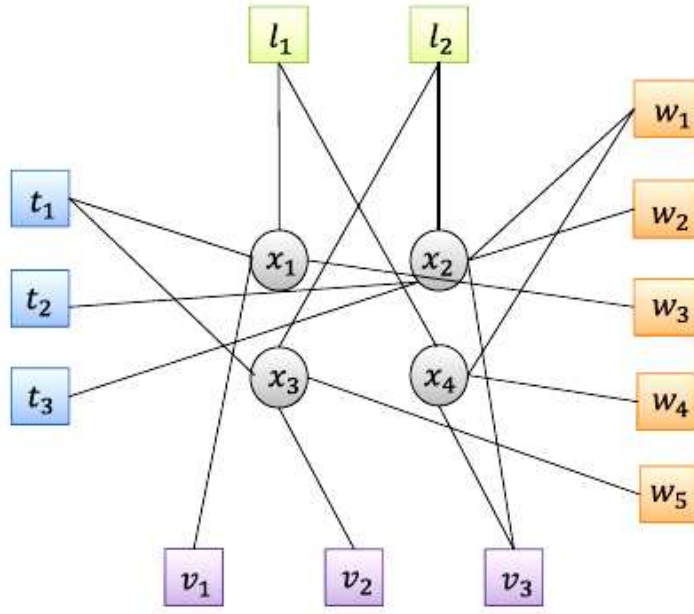


图 5.2: 星结构的K分图

#### 5.4.2 基于K分图的关系建模

在基于星结构 $K$ 分图的关系建模后, 我们采用一致信息论的联合聚类 (consistent information theoretic co-clustering, CIT) [41]方法将原始的高阶联合聚类问题分成 $\mathcal{X} - \mathcal{T}$ ,  $\mathcal{X} - \mathcal{L}$ ,  $\mathcal{X} - \mathcal{V}$ 和 $\mathcal{X} - \mathcal{W}$ 的聚类。聚类分割是在一个总的目标函数下进行的。相比于其他基于关系的联合聚类方法, 比如一致的二分图分割 [40], CIT 并没有要求每个社交媒体必须和某一类信息, 比如时间间隔、位置、视觉词和关键词关联。为了最小化基于 $K$ 分图的互信息的线性组合, 我们定义如下的目标函数:

$$J_{inter} = a[I(X; T) - I(\hat{X}; \hat{T})] + b[I(X; L) - I(\hat{X}; \hat{L})] \\ + c[I(X; V) - I(\hat{X}; \hat{V})] + d[I(X; W) - I(\hat{X}; \hat{W})] \quad (5.1)$$



其中 $a, b, c, d$  ( $a + b + c + d = 1$ )表示权重。 $I(X; T)$ 代表 $X$ 和 $T$ 之间的互信息, 定义为

$$I(X; T) = \sum_x \sum_t p(x)p(t|x) \log \frac{p(t|x)}{p(t)} \quad (5.2)$$

$I(\hat{X}; \hat{T})$ ,  $I(X; L)$ ,  $I(\hat{X}; \hat{L})$ ,  $I(X; V)$ ,  $I(\hat{X}; \hat{V})$ ,  $I(X; W)$  和  $I(\hat{X}; \hat{W})$  也是这样定义。公式(5.1)中第一项表示 $X - T$ 的联合聚类, 第二项代表 $X - L$ 的联合聚类, 第三项代表 $X - V$ 的联合聚类, 第四项表达 $X - W$ 的联合聚类。

经过推导, 互信息的损失表示为1) 行分布和行原型分布之间的相对熵的权重之和, 包括公式(5.3) (5.5) (5.7) (5.9) 或者2) 列分布和列原型分布之间的相对熵的权重之和, 包括公式(5.4) (5.6) (5.8) (5.10)。

$$I(X; T) - I(\hat{X}; \hat{T}) = \sum_{\hat{\mathcal{X}}} \sum_{x \in \hat{\mathcal{X}}} p(x) D(p(T|x) \| q(T|\hat{\mathcal{X}})) \quad (5.3)$$

$$I(X; T) - I(\hat{X}; \hat{T}) = \sum_{\hat{\mathcal{T}}} \sum_{t \in \hat{\mathcal{T}}} p(t) D(p(X|t) \| q(X|\hat{\mathcal{T}})) \quad (5.4)$$

$$I(X; L) - I(\hat{X}; \hat{L}) = \sum_{\hat{\mathcal{X}}} \sum_{x \in \hat{\mathcal{X}}} p(x) D(p(L|x) \| q(L|\hat{\mathcal{X}})) \quad (5.5)$$

$$I(X; L) - I(\hat{X}; \hat{L}) = \sum_{\hat{\mathcal{L}}} \sum_{l \in \hat{\mathcal{L}}} p(l) D(p(X|l) \| q(X|\hat{\mathcal{L}})) \quad (5.6)$$

$$I(X; T) - I(\hat{X}; \hat{V}) = \sum_{\hat{\mathcal{X}}} \sum_{x \in \hat{\mathcal{X}}} p(x) D(p(V|x) \| q(V|\hat{\mathcal{X}})) \quad (5.7)$$

$$I(X; T) - I(\hat{X}; \hat{V}) = \sum_{\hat{\mathcal{V}}} \sum_{v \in \hat{\mathcal{V}}} p(v) D(p(X|v) \| q(X|\hat{\mathcal{V}})) \quad (5.8)$$

$$I(X; T) - I(\hat{X}; \hat{W}) = \sum_{\hat{\mathcal{X}}} \sum_{x \in \hat{\mathcal{X}}} p(x) D(p(W|x) \| q(W|\hat{\mathcal{X}})) \quad (5.9)$$

$$I(X; T) - I(\hat{X}; \hat{W}) = \sum_{\hat{\mathcal{W}}} \sum_{w \in \hat{\mathcal{W}}} p(w) D(p(X|w) \| q(X|\hat{\mathcal{W}})) \quad (5.10)$$

其中 $q(t|\hat{\mathcal{X}}) = p(t|\hat{\mathcal{T}})p(\hat{\mathcal{T}}|\hat{\mathcal{X}})$ ,  $q(x|\hat{\mathcal{T}})$ 类似。 $q(l|\hat{\mathcal{X}})$ ,  $q(x|\hat{\mathcal{L}})$ ,  $q(v|\hat{\mathcal{X}})$ ,  $q(x|\hat{\mathcal{V}})$ ,  $q(w|\hat{\mathcal{X}})$  和  $q(w|\hat{\mathcal{W}})$ .  $D(\cdot \| \cdot)$  均表示相对熵, 也即Kullback-Leibler (KL) 散度。

为了解决总的联合聚类问题, 我们首先优化每一个子问题, 然后基于之前的 $\mathcal{T}$ 、 $\mathcal{L}$ 、 $\mathcal{V}$  和  $\mathcal{W}$ 的聚类结果通过最小化(5.1)决定中心集合的聚类结果 $\mathcal{X}$ 。

### 5.4.3 时间空间内的关系建模

除了考虑不同元数据类型和社会媒体之间的依附关系，我们同时要求基于 $\mathcal{T}$ 的聚类对齐应该是平滑的，因此我们在联合聚类方法中将 $\mathcal{T}$ 之间的关系作为全局正则化考虑进去。我们令 $\mathbf{S} \in \mathbb{R}^{|\mathcal{T}| \times |\mathcal{T}|}$ 为对称的权重矩阵。其中 $\mathbf{S}(i, j)$ 中的每个元素表示第 $i$ 个时间间隔和第 $j$ 个时间间隔的相似度，并且是非负的。在我们的问题中，我们将时间表示成Unix的时间，如果两个时间段超过一个周，相似度为0；否则相似度如下计算：

$$\mathbf{S}(i, j) = 1 - \frac{|t_i - t_j|}{\tau}, \quad (5.11)$$

其中 $\tau = 7$ 。考虑到列向量 $\mathbf{h}^t$ 是来自 $\mathcal{T}$ 中的元素的赋值聚类索引。和 [11]类似，我们有如下的目标函数：

$$J_{intra} = \sum_{i \neq j} (h_i^t - h_j^t)^2 \mathbf{S}(i, j) \quad (5.12)$$

直觉上看，如果 $t_i$ 和 $t_j$ 越相似，则 $\mathbf{S}(i, j)$ 越大， $|h_i^t - h_j^t|$ 越小。

我们定义拉普拉斯矩阵 $\mathbf{G}$ 和相应的对角阵 $\mathbf{D}$ 如下

$$\mathbf{G} = \mathbf{D} - \mathbf{S} \quad (5.13)$$

其中 $\mathbf{D}_{ii} = \sum_{i \neq j} \mathbf{S}(i, j)$  for  $\forall i$ 。则 $J_{intra}$ 最终推导为

$$J_{intra} = (\mathbf{h}^t)^T \mathbf{G} \mathbf{h}^t \quad (5.14)$$

### 5.4.4 统一的问题描述

通过以上两个方面，即基于 $K$ 分图不同维度之间的依附关系和基于时间空间 $\mathcal{T}$ 的内部关系，我们定义如下总的损失函数：

$$J = \alpha J_{inter} + (1 - \alpha) J_{intra}, \quad (5.15)$$

公式(5.15)的 $\alpha$ 表示 $J_{inter}$ 和 $J_{intra}$ 两者之间的平衡参数。我们的目标最终要

求解如下的优化问题：

$$\begin{aligned}
& \min_{\mathbf{h}^x, \mathbf{h}^t, \mathbf{h}^l, \mathbf{h}^v, \mathbf{h}^w} J \\
& = \min_{\mathbf{h}^x, \mathbf{h}^t, \mathbf{h}^l, \mathbf{h}^v, \mathbf{h}^w} \alpha \left\{ a[I(X; T) - I(\hat{X}; \hat{T})] + b[I(X; L) - I(\hat{X}; \hat{L})] \right. \\
& \quad \left. + c[I(X; V) - I(\hat{X}; \hat{V})] + d[I(X; W) - I(\hat{X}; \hat{W})] \right\} \\
& \quad + (1 - \alpha)(\mathbf{h}^t)^T \mathbf{G} \mathbf{h}^t
\end{aligned} \tag{5.16}$$

## 5.5 聚类 and 样本改进

经过高阶联合聚类，我们得到初步的社会事件结果。本节我们采用一个启发式的方法去除和改进获得的聚类 and 每个聚类的社会媒体样本。这个过程包括两个步骤：1) 过滤和融合聚类；2) 去除和填充社会媒体样本。

对于聚类，由于我们并不知道最后聚类的数量，因此经验设定一个阈值。如果一些带有错误样本值的类别阈值超过一个预定义的阈值，我们将该聚类去除，同样如果两个聚类的之间的距离很小，我们将他们融合到一个新的聚类。

而对于社会媒体样本，我们需要去除有噪音的样本。我们假定属于同一个社会事件的社会媒体样本在给定的四种类型的元数据样本中都应该相似的，定义样本  $x$  和他的聚类中心  $\hat{\mathcal{X}}$  的距离为

$$dis = a|t - t_c| + b\sigma(l - l_c) + c\|\mathbf{v} - \mathbf{v}_c\|_2^2 + d\|\mathbf{w} - \mathbf{w}_c\|_2^2, \tag{5.17}$$

其中  $\sigma(l - l_c) = 0$ , if  $l = l_c$ ; 否则 1.  $\|\cdot\|_2$  是  $\ell_2$  范数的向量。  $t_c = \sum_{j: x_j \in \hat{\mathcal{X}}} t_j$ ,  $l_c$  是  $\hat{\mathcal{X}}$  中最关联的城市,  $\mathbf{v}_c = \sum_{j: x_j \in \hat{\mathcal{X}}} \mathbf{v}_j$ , and  $\mathbf{w}_c = \sum_{j: x_j \in \hat{\mathcal{X}}} \mathbf{w}_j$ . 我们去除距离超过设定阈值的样本。此外，和 [69] 的处理类似，我们假定一个人不可能同时参与两个事件，因此当一个事件发生的时候，被同一个用户拍摄的所有的图片属于同一个聚类。

## 5.6 社会事件检测算法

目标函数(5.15) 可以通过一个迭代的过程进行优化。如果我们固定聚类赋

值  $\mathbf{h}^t, \mathbf{h}^l, \mathbf{h}^v$  and  $\mathbf{h}^w$ , 最好的赋值  $(h_i^x)^*$  of  $x_i$  能够通过如下公式(5.18)求解。

$$\begin{aligned} (h_i^x)^* = \arg \min_{h_i^x} & aD(p(T|x_i)||q(T|\hat{\mathcal{T}})) + bD(p(L|x_i)||q(L|\hat{\mathcal{L}})) \\ & + cD(p(V|x_i)||q(V|\hat{\mathcal{V}})) + dD(p(W|x_i)||q(W|\hat{\mathcal{W}})) \end{aligned} \quad (5.18)$$

同样  $(h_i^t)^*$ 、 $(h_i^l)^*$ 、 $(h_i^v)^*$  和  $(h_i^w)^*$  分别通过公式(5.19)、(5.20)、(5.21) 和(5.22)进行优化

$$(h_i^t)^* = \arg \min_{h_i^t} \alpha aD(p(X|t_i)||q(X|\hat{\mathcal{T}})) + (1 - \alpha)(\mathbf{h}^t)^T \mathbf{G} \mathbf{h}^t \quad (5.19)$$

$$(h_i^l)^* = \arg \min_{h_i^l} bD(p(X|l_i)||q(X|\hat{\mathcal{L}})) \quad (5.20)$$

$$(h_i^v)^* = \arg \min_{h_i^v} bD(p(X|v_i)||q(X|\hat{\mathcal{V}})) \quad (5.21)$$

$$(h_i^w)^* = \arg \min_{h_i^w} bD(p(X|w_i)||q(X|\hat{\mathcal{W}})) \quad (5.22)$$

基于鲁棒的高阶联合聚类的伪代码算法如1所示, 算法开始于一个初始的联合聚类  $(\mathbf{h}^x)^{(0)}$ ,  $(\mathbf{h}^t)^{(0)}$ ,  $(\mathbf{h}^l)^{(0)}$ ,  $(\mathbf{h}^v)^{(0)}$  和  $(\mathbf{h}^w)^{(0)}$  (见步骤一)。然后我们通过步骤2进行粗社会事件检测。在步骤2.1, 我们固定  $\mathbf{h}^t$ ,  $\mathbf{h}^l$ ,  $\mathbf{h}^v$  和  $\mathbf{h}^w$  最小化公式5.18, 同时计算  $q(\cdot)$  所需要的边界。接下来的  $\mathcal{T}$ ,  $\mathcal{L}$ ,  $\mathcal{V}$  和  $\mathcal{W}$  聚类等价于最小化公式(5.19) (5.20) (5.21) 和(5.22) (见步骤2.2-2.5)。当目标函数不在下降时 (见步骤2.6), 迭代过程结束。在步骤3, 我们进行后处理过滤和融合聚类和本地样本。

## 5.7 实验结果与分析

我们用MediaEval SED的数据库<sup>2</sup>评估提出的方法的性能。

### 5.7.1 数据库

#### 5.7.1.1 MediaEval SED 数据库

MediaEval SED 数据库在2012年发布, 共包括从Flickr收集的167,332图片集。拍摄时间范围为2009-01-01到2011-12-31, 共有4,222用户贡献这个数据库。每张图片包括Flickr ID、拍摄时间、用户、标题、描述和关键词。其中所有的图片集合中只有20%的图片有经纬度信息。MediaEval 社会事件检测2012年包

<sup>2</sup><http://mklab.it/iti.gr/project/sed2012>

含以下3个任务。第一个挑战检测发生在德国的技术事件；第二个挑战是检测发生在德国汉堡和西班牙马德里的足球事件；第三个挑战是检测发生在马德里的抗议运动事件。对于这些挑战，该数据库提供了标准数据用以评估方法的检测性能，对于第一个挑战将2,234样本分成18个事件；第二个挑战将1,684样本分到79个事件；第三个挑战共52个事件3,992样本。图5.3给出了不同足球事件的一些样本。



图 5.3: 不同足球事件的例子图片

在我们的实验中，我们主要考虑第二个任务足球事件检测。我们构建两个数据库，第一个数据库是我们选择所有的足球事件的标准数据建立一个足球事件子集，该数据并不包括任何非事件样本，用于验证给予高阶联合聚类的粗社会事件检测的有效性。第二个数据库则是选择整个的MediaEval SED数据库去评估我们提出的框架（Social Event Detection with Robust High-Order Co-Clustering, SED-RHOCC）的有效性。

### 5.7.1.2 预处理

由于数据库中提供的文本信息（标题、描述和标签）中含有一些非英语的单词，我们利用Google 翻译API将这些翻译成英语。然后我们对文本数据进行预处理，包括去除停用词、html标签和与摄像机参数相关的词，比如“Canon”和“35mm”。接下来我们借助WordNet去掉一些拼写错误的词 [76]，同时按照拍摄时间分组图片，将关联的经纬度信息解析成对应的城市。

### 5.7.2 评估测度和特征

在该评估中，既然我们采用的是联合聚类方法，我们选择归一化的互信息（Normalized Mutual Information, NMI）[18]测量聚类的性能。NMI被广泛应用到聚类质量的评测上。对于两个随机变量 $X$ 和 $Y$ ，NMI定义为[104]：

$$NMI = \frac{I(X, Y)}{H(X)H(Y)} \quad (5.23)$$

其中 $I(X, Y)$ 是 $X$ 和 $Y$ 之间的互信息， $H(X)$ 和 $H(Y)$ 分别表示 $X$ 和 $Y$ 的熵。给定一个聚类结果，公式(5.23)的NMI可以如下进行估计[104]：

$$NMI = \frac{\sum_{i=1}^{\hat{m}} \sum_{j=1}^{\hat{m}} \hat{m}_{ij} \log \left( \frac{m \hat{m}_{ij}}{\hat{m}_i (\hat{m}^*)_j} \right)}{\sqrt{\left( \sum_{i=1}^{\hat{m}} \hat{m}_i \log \frac{\hat{m}_i}{m} \right) \left( \sum_{j=1}^{\hat{m}} (\hat{m}^*)_j \log \frac{(\hat{m}^*)_j}{m} \right)}} \quad (5.24)$$

其中 $\hat{m}_i$ 是第 $i$ 个聚类的数据的量， $(\hat{m}^*)_j$ 是第 $j$ 个标准类的数据的数量； $\hat{m}_{ij}$ 表示第 $i$ 个的聚类和第 $j$ 标准类数据的交叉部分。从公式我们可以看出NMI值越大，聚类的性能越好。我们使用的时间戳、位置、可视化内容和文本内容的特征具体如下：

- 时间戳：我们表示时间值是从Unix epoch开始所走过的天数。
- 位置：我们根据样本的地理标签转换成对应的城市。对于非地理标注的样本，我们设置为0。
- 视觉内容：每一张图片由词袋模型（Bag-of-Word, BOW）1024维的特征向量表示。
- 文本内容：我们用term frequency - inverse document frequency (TF-IDF)权重表示，具体从标题、描述和标签合并的文本信息中计算。

### 5.7.3 实验结果

#### 5.7.3.1 足球事件子集

该子集是由所有的足球事件样本组成，并不包含任何的非事件样本，也就是标准事件集合。我们使用该子集验证粗社会事件检测的有效性。考虑到一些事件的样本太少，我们只考虑超过10个样本的事件，最终我们的数据子集包括30个事件和1,456个样本。在该实验中选择两个基准方法作为比较。

- 归一化的割 (Normalized Cut, N-cut) [101]: 在该方法中, 样本对之间总的相似度是通过不同类别元数据相似度的线性组合计算的。
- 基于一致信息论的联合聚类 (CIT): 该方法融合多种元数据但是并没有考虑时间空间中时间之间的内在关系。

在N-cut, 两个时间的相似度通过公式(5.11)计算。如果两个位置关联同一个城市被设为1, 而否为0。我们用余弦相似度计算文本之间的相似度, 而视觉之间的相似度是通过欧式距离来计算的。融合参数经验设置为 $a = 0.1, b = 0.1, c = 0.1$ 和 $d = 0.7$  在CIT方法中, 权重参数被经验设置为 $a = 0.5, b = 0.1, c = 0.3$ 和 $d = 0.1$ 。在SED-RHOCC, 权重参数的设置和CIT一致,  $\alpha = 0.9$ 。

表5.1显示了我们的方法和两种基准方法的NMI值。我们能够看到联合聚类方法CIT 和SED-RHOCC的性能超过了单聚类方法N-cut。而且由于我们的方法由于考虑了时间空间内的时间关系因而性能相比于CIT 又提高了大概2%。

图5.4 展示了在不同的 $\alpha$ 下不同的NMI值。当 $\alpha = 0.9$ 时, 结果达到了最佳值。

表 5.1: 不同方法的NMI

N-cut	CIT	SED-RHOCC
0.7457	0.9363	<b>0.9575</b>

### 5.7.3.2 MediaEval SED 数据库

在这次实验中, 我们考虑数据库中所有的79个足球事件。由于MediaEval SED 数据库混有大量的非事件和非足球事件样本, 这些样本和足球事件没有关系。我们需要首先将该数据样本分成足球事件和非足球事件。为此, 我们从每一个城市鉴别出多个足球俱乐部, 提取所有在2009年到2011年的足球比赛的日期和时间`playerhistory.com`。更多的细节可以参考文献 [17]。利用这些信息我们可以从数据库中选择出一些正样本和负样本, 利用线性的SVM [37]进行分类。最终我们将样本的数量减少到2152。表5.2展示了不同方法的NMI的结果, 在没有进行聚类 and 样本改进前, 第一步我们方法的NMI值为0.6834, 而经过第二步, NMI增加到了0.7076。



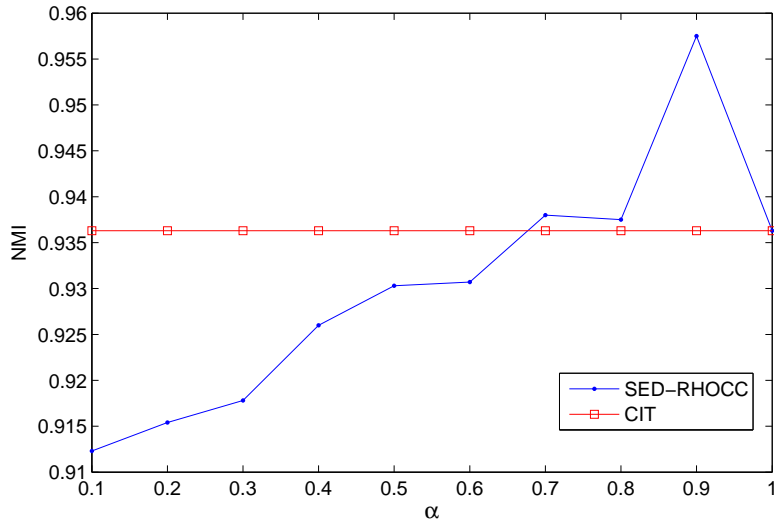
图 5.4: 不同 $\alpha$ 值下的性能比较

表 5.2: 不同方法的NMI比较

N-cut	Coarse Event Detection	SED-RHOCC
0.3930	0.6834	<b>0.7076</b>

## 5.8 本章小结

本章提出了一个社会事件检测框架用于社会事件的检测。我们提出的方法的优点包括以下几点：1) 通过构建星结构的 $K$ 分图融合了多种异质元数据信息，避免了因早融合或者晚融合所带来的信息损失。2) 引入了时间空间内的关系作为全局正则化改进了聚类性能；3) 我们的方法由于引入了信息论的联合聚类框架因而不需要对每个元数据集合的聚类数量有所限制。最终实验结果验证了我们方法的有效性。



---

**Algorithm 1** 基于鲁棒高阶的联合聚类的社会事件检测 (Social Event Detection with Robust High-order Co-clustering, SED-RHOCC) 算法

---

输入:

共生计数矩阵  $\mathbf{E}_T, \mathbf{E}_L, \mathbf{E}_V, \mathbf{E}_W$ , 聚类数量:  $|\hat{n}|, |\hat{t}|, |\hat{l}|, |\hat{v}|, |\hat{w}|$   
 权重因子:  $a, b, c, d$ , 目标函数参数:  $\alpha$

输出:

聚类赋值  $\mathbf{h}^x, \mathbf{h}^t, \mathbf{h}^l, \mathbf{h}^v$  and  $\mathbf{h}^w$

1: 初始化:

$$p(x, t) = \frac{\mathbf{E}_T}{\mathbf{1}^T \mathbf{E}_T \mathbf{1}}, p(x, l) = \frac{\mathbf{E}_L}{\mathbf{1}^T \mathbf{E}_L \mathbf{1}},$$

$$p(x, v) = \frac{\mathbf{E}_V}{\mathbf{1}^T \mathbf{E}_V \mathbf{1}}, p(x, w) = \frac{\mathbf{E}_W}{\mathbf{1}^T \mathbf{E}_W \mathbf{1}};$$

初始化聚类集合  $\mathbf{h}^x(0), \mathbf{h}^t(0), \mathbf{h}^l(0), \mathbf{h}^v(0)$  and  $\mathbf{h}^w(0)$ ;

计算  $q^{(0)}(\hat{\mathcal{X}}, \hat{\mathcal{T}}), q^{(0)}(X|\hat{\mathcal{X}}), q^{(0)}(T|\hat{\mathcal{T}}), q^{(0)}(T|\hat{\mathcal{X}}), q^{(0)}(\hat{\mathcal{X}}, \hat{\mathcal{L}}), q^{(0)}(L|\hat{\mathcal{L}}),$   
 $q^{(0)}(L|\hat{\mathcal{X}}), q^{(0)}(\hat{\mathcal{X}}, \hat{\mathcal{V}}), q^{(0)}(V|\hat{\mathcal{V}}), q^{(0)}(V|\hat{\mathcal{X}}), q^{(0)}(\hat{\mathcal{X}}, \hat{\mathcal{W}}), q^{(0)}(W|\hat{\mathcal{W}}), q^{(0)}(W|\hat{\mathcal{X}})$

2: 粗社会事件检测

- 2.1: 使用公式(5.18) 在  $\mathcal{X}$  进行聚类赋值和计算分布
- 2.2: 使用公式(5.19) 在  $\mathcal{T}$  进行聚类赋值和计算分布
- 2.3: 使用公式(5.20) 在  $\mathcal{L}$  进行聚类赋值和计算分布
- 2.4: 使用公式(5.21) 在  $\mathcal{V}$  进行聚类赋值和计算分布
- 2.5: 使用公式(5.22) 在  $\mathcal{W}$  进行聚类赋值和计算分布
- 2.6: 使用公式(5.15) 计算目标函数值和目标函数的变化  $\Delta J$  如果  $\Delta J \geq 10^{-3}$ , 继续步骤2.1.

3: 聚类 and 样本改进

- 3.1: 过滤带有大的错误的聚类 and 融合相似的聚类
- 3.2: 过滤聚类中离聚类中心较远的社交媒体样本 and 填充错误丢失的社交媒体样本

4: 返回: 鉴别的聚类

---



## 第六章 基于位置情景的跨平台个性化推荐

### 6.1 引言

Web2.0技术的发展促进了各种各样的社会平台的发展。不同的社会平台提供不同的特色服务。例如Flickr是一个图片分享网站，能够使用户上传和分享个人图片，而Foursquare提供一个个性化的位置服务用于推荐周边的商户信息。GPS移动设备的广泛应用使得这些分享的媒体信息通常都和位置信息关联在一起。如何打破平台的边界充分利用这些平台的社会媒体数据解决相关领域的应用（比如基于Foursquare和Twitter的社会链接预测 [131]和基于新浪微博和淘宝的市场营销 [130]等）已经成为一个重要的研究主题。本工作中，我们研究跨平台多模态数据的关联和分析。具体来说，我们提出一个模型关联来自两个社会平台Flickr和Foursquare的数据，并将其应用到个性化的平台间推荐中。在该应用中，在给定位置情景信息的条件下，Flickr用户去该地方旅游时可以快速找到他们感兴趣的来自Foursquare平台且和该位置相关的商户信息。同样，Foursquare用户也可以享受他们感兴趣的和该位置相关的Flickr图片。

现有的基于跨平台的工作 [95], [58], [131]集中利用辅助平台的信息以便于目标平台的多媒体应用。例如Zhong等人 [131]基于两个平台Twitter和Foursquare共有的用户作为桥梁借助用户在Twitter上的用户信息和关系实现基于Foursquare平台的朋友推荐。但在许多实际情况下，我们并不容易获取同一用户在不同平台的账户信息。为了克服这个限制，Roy等人 [95]利用来自Twitter平台的事件信息关联Twitter和YouTube实现基于Youtube平台的社会化的视频搜索和推荐。和以上工作不同，我们的工作考虑更一般的主题信息去关联带有不同模态的不同平台实现平台间的应用，使得不同平台的用户同时受益。

每个平台都有大量用户分享的内容信息，比如来自Flickr的带有标签的图片和来自Foursquare的带有标签的商户信息。我们从不同平台的内容信息挖掘的知识通常分成两部分：所有平台共享的知识和每个平台特有的知识。如图6.1的例子展示，我们能够看到红色标注的“音乐会”和紫色标注的“动物”是两个平台共享的知识。而在Flickr上关于“日落”的图片（蓝色标注）表示Flickr平台特有的知识而来自Foursquare绿色标注的“书店”则对

知识类型	图片 <b>flickr</b> 标签	商户 <b>foursquare</b> 标签
共享的知识	 <b>concert</b> jake paleschic joey gorman <b>live</b> oak spune longshots pit <b>music band</b> concert <b>performance</b>	<ul style="list-style-type: none"> <li>• irving plaza music venue</li> <li>• pianos rock club</li> <li>• music hall of williamsburg</li> <li>• fontana's bar rock club</li> </ul> <b>dancing</b> irish pub <b>live music</b> pool tables pub grub funk soul <b>rock</b> classical performing arts
	 <b>phoenix zoo</b> mountain lion phoenix andean bear luka edmonton valley red <b>wild</b> panda african dog	<ul style="list-style-type: none"> <li>• central park zoo</li> <li>• bronx zoo</li> <li>• prospect park zoo</li> <li>• children's zoo</li> <li>• jungleworld zoo</li> </ul> street <b>animals</b> borough bronx great outdoors nature park photo booth <b>zoo</b> animal house beer college pasta pizza
平台特有的知识	 <b>sunset</b> <b>clouds</b> outdoor <b>ocean</b> usa newyork bay boat fireisland sky sun aerial beemo	<ul style="list-style-type: none"> <li>• bent pages bookstore</li> <li>• mercer street bookstore</li> <li>• namaste bookshop</li> <li>• posman book bookstore</li> </ul> books chainl <b>bookstore</b> brooklyn starbucks park slope public <b>readings</b> bathroom yuppies

图 6.1: 来自Flickr和Foursquare不同类型的知识：红色和紫色标注的代表共享的知识，绿色标注的代表Flickr特有的知识，蓝色标注的代表Foursquare特有的知识

应Foursquare平台特有的知识。由于共享的知识属于所有的平台，我们可以充分利用这类知识作为桥梁连接不同的社会平台。因此我们的工作目的也是要通过区分共享的知识和平台特有的知识以提取共享的知识连接不同的平台。

提取来自不同平台共享的知识需要解决以下两个问题：（1）每个平台共享的知识和平台特有的知识是混合在一起的。如果我们直接简单的融合所有平台的内容去提取共享的知识，提取的知识必定带有大量的噪音，因此如何有效的区分共享的知识和平台特有的知识是我们解决的第一个问题；（2）对于每个平台，除了文本内容，每个平台都有自己特有的模态，比如来自Flickr的图片和来自Foursquare的商户信息。为了更好的实现平台间的多媒体应用，我们需要对每一类主题对齐不同的模态。因而我们还需要解决同一主题来自不同平台的不同模态的对齐问题。

为了解决以上问题，本章提出了一个概率主题模型，即跨平台多模态主题模型（Cross-platforM Multi-Modal Topic Model, CM<sup>3</sup>TM）用于挖掘来自不同平台的主题。我们定义两种不同类型的主题，即平台共享的主题和每个平台特有的主题。共享的主题描述来自所有平台共享的知识，由共享的主题子空间表示。相比之下，平台特有的主题描述特定平台的知识，且有相应平台特有的主题子空间表示。CM<sup>3</sup>TM 通过分别学习这两类主题子空间区分这两类主题，此外CM<sup>3</sup>TM能够通过学习主题子空间关联不同模态的信息，从而对齐每一个主

题来自不同平台不同模态。最终我们将提出的模型应用到基于位置情景的平台间推荐问题中。本章的具体贡献总结如下：

- 我们通过挖掘来自不同平台的多模态内容信息解决平台间的关联问题。
- 我们提出了一个跨平台多模态主题模型，该模型能够同时发现共享的主题和平台特有的主题，同时能够对齐同一主题来自不同平台的多种模态。
- 我们提出了一个基于位置情景的平台间的推荐应用：在给定某一位置情景和两个不同的社交媒体平台Flickr和Foursquare的条件下，1）推荐来自Foursquare的当前位置的商户信息服务给Flickr用户和2）推荐来自Flickr的和当前位置相关的图片给Foursquare用户。

## 6.2 相关工作

我们的工作和以下两个研究领域相关：（1）跨平台个性化推荐和（2）跨平台主题模型。

### 6.2.1 跨平台个性化推荐

现有的大部分推荐工作是基于单平台的，例如兴趣点（Point-of-interest, POI）推荐 [118, 137] 和图像推荐 [36, 98]。但随着各种各样社会媒体的发展，许多研究者开始逐渐转移到基于多平台的协同应用以解决单平台无法克服的问题。具体主要包括以下两类研究方向。

（1）融合来自不同平台的信息进行用户的建模和推荐 Abel等人 [1] 通过聚合来自Flickr、Twitter和Delicious的用户信息解决冷启动的推荐问题。Zahalka等人 [126] 从Foursquare、Picasa 和Flickr的网站中收集图片集和关联的文本标签，然后通过一个交互式的接口学习用户的兴趣实现个性化的推荐。Zhong等人 [140, 141] 提出了一个基于主题模型的方法自适应的融合来自不同平台的信息转移知识分别实现商品和朋友的推荐。Liu 等人 [68] 同时分析来自Flickr的带有地理标注的图片和来自Foursquare的签到数据从中挖掘旅游者的兴趣和当地居民活动的偏好，进而发现和推荐一个城市感兴趣的地理位置。这些工作主要是收集来自不同平台的数据用于推荐，而我们则是通过不同平台之间的知识转移进行平台间的推荐，使得来自不同平台的用户同时受益。

（2）充分利用不同平台各自的特色实现协同应用 Yan等人 [116] 以不同平台共有的用户作为桥梁提出了基于随机游走的方法解决朋友推荐中的冷启动问

题。Deng [31] 则是通过融入来自Twitter用户的信息实现个性化的Youtube 视频推荐。Yan等人 [117] 则是在给定Youtube视频的情况下, 在Twitter上找到能够最大化进行该视频推广的Twitter追随者。Zhang等人 [131, 132] 基于两个平台共有的用户作为桥梁借助两个平台的位置、时间和文本信息转移来自源平台的链接信息用于预测新用户的社会链接信息。以上方法均是以不同平台间共有用户作为桥梁, 而Qi等人 [88, 89] 则以用户的内容属性作为桥梁充分使用有偏的跨平台采样方法进行跨平台的链接预测。Roy等人 [95]通过挖掘博文的主题将Twitter平台的知识转移到Youtube, 从而进行视频的个性化推荐。尽管我们也是以不同平台共享的内容信息关联不同平台, 但和以上应用不同的是, 这些工作借助其他平台的信息最终仍服务于目标平台或者目标域, 而我们则进行平台间的推荐应用使得不同的平台同时受益。此外, 在方法上考虑到来自每个平台的模态种类和数量的不同, 我们提出了一个多模态情景下的跨平台推荐框架。

### 6.2.2 概率主题模型

概率主题模型 [14]旨在从大量的文档集合中挖掘主题, 这里一个主题是由词的分布表示而一篇文档则是由主题的分布表示。和本章相关的两类主题模型分别是多模态主题模型和跨域主题模型。基于多模态的主题模型旨在建立不同模态比如图片和文本等之间的关联。例如Blei [13] 提出了两类方法: 多模态潜在狄利克雷分布和对应性潜在狄利克雷分布。这些方法用一组共享的潜在变量去直接建模图像和文本, 捕获不同模态数据的相关性。Putthividhy等人 [87] 提出了一个基于回归的主题模型, 该模型能够捕获更一般的关联结构, 允许不同模态的主题数量可以有不同的设置。以上工作均假设文本都和相应的图片成对关联的, 但是现实很多情况下, 文本和图片的关系并不是完全对应的, 比如一些文本并没有与之关联的图片。为了能够更一般的建模多模态主题, Jia 等人 [57]提出了一个基于马尔科夫场的主题模型, 该模型既能建模不同模态之间的关系, 还可以建模单个模态特有的主题。Virtanen 等人 [108] 将主题空间分成三类: 文本特有的主题子空间、文本和图片共享的主题子空间和图片特有的主题子空间, 提出了一种可分解的主题模型学习这三类主题子空间。

此外随着社交媒体各种平台的发展, 国内外一些学者提出了基于跨域的主题模型, 比如, Tan等人 [105] 提出了一个主题模型融合多种文本信息通过挖掘不同领域共享的用户兴趣主题进行跨域推荐。Li 等人 [65]将不同域的主题类型划分成共享的主题和域特有的主题, 通过主题模型发现这两类主题并应用到文





图 6.2: 基于位置情景的跨平台个性化推荐框架

本分类任务中。Tang等人 [106]采用相似的思想提出了一个主题模型用于跨域的科研合作者推荐。和以上方法的不同在于，我们基于社交媒体情景下研究一个新奇的平台间的推荐问题。除了文本信息外，我们提出了一个多模态跨平台主题模型框架同时建模来自不同平台的视觉信息和商户信息。

### 6.3 方法框架

如图6.2,我们提出了一个基于位置情景的跨平台推荐框架，该框架包括以下两个部分：

- 主题建模。我们提出了CM<sup>3</sup>TM主题模型，该模型能够学习两类不同的主题空间和用户的主题分布。我们以Flickr和Foursquare平台作为研究平台，输入为（a）来自Flickr的图片和标注的文本；（b）来自Foursquare的商户信息和标注的标签。输出包括1）学习的共享的主题空间和平台特有的主题空间和2）用户共享的主题分布和平台特有的主题分布。
- 基于位置情景的跨平台个性化推荐。在特定城市和用户主题分布的条件下，该推荐主要包括1）个性化的Foursquare商户推荐给Flickr用户和2）个性化的Flickr图片推荐给Foursquare用户。

表6.1给出了本章中常用的符号。

表 6.1: 关键符号列表

符号	描述
$c_m$	第 $m$ 个社交平台
$U, U^{c_1}, U^{c_2}$	用户集, Flickr 用户集, Foursquare 用户集
$I_u, T_u$	条目集合(例如, 来自Flickr的图片或者来自Foursquare的商户), 用户 $u$ 关联的文本
$W, V, L$	文本词典, 视觉词典, 商户词典
$K, K^{c_1}, K^{c_2}$	共享主题、Flickr特有主题和Foursquare特有主题的数目
$n_u^w, n_u^v, n_u^l$	用户文档 $u$ 的文本词数量, 视觉词数量和商户词数量
$\psi, \phi, \varphi$	共享主题的文本词, 视觉词和商户词的多项式分布
$\psi^{c_1}, \phi^{c_1}$	Flickr特有主题的文本词和视觉词的多项式分布
$\psi^{c_2}, \varphi^{c_2}$	Foursquare特有主题的文本词和商户词的多项式分布
$\theta'^{c_m}, \theta$	平台 $c_m$ 特有的用户主题和用户共享主题的多项式分布
$s^w, s^v, s^l$	二值标签指示文本词、视觉词、商户是由 $\theta$ 或 $\theta'^{c_m}$ 产生
$z^w, z^v, z^l$	文本词、视觉词和商户词的主题赋值
$\rho^{c_m}$	来自平台 $c_m$ 的采样二值变量的参数 $s^w, s^v$ 或 $s^l$
$\lambda_1^{c_m}, \lambda_2^{c_m}$	用于产生 $\rho^{c_m}$ 的Beta先验
$\alpha_1^{c_m}, \alpha_2^{c_m}$	用于产生多项式分布 $\theta^{c_m}$ , $\theta^{c_m}$ 和 $\theta'^{c_m}$ 的狄利克雷先验
$\beta, \beta^{c_m}$	用于产生多项式分布 $\psi, \psi^{c_m}$ 的狄利克雷先验
$\gamma, \gamma^{c_1}$	用于产生多项式分布 $\phi, \phi^{c_1}$ 的狄利克雷先验
$\eta, \eta^{c_2}$	用于产生多项式分布 $\varphi, \varphi^{c_2}$ 的狄利克雷先验

## 6.4 跨平台多模态主题建模

本章主要引入了跨平台多模态主题模型（Cross-platform Multi-Modal Topic Model, CM<sup>3</sup>TM）用于从用户文档<sup>1</sup>中发现来自Flickr和Foursquare的共享主题

<sup>1</sup>这里, 我们聚合每一用户的内容, 这样有利于基于面向用户的应用, 比如个性化的搜索和推荐。一个Flickr用户文档包括共享的图片和关联的文本信息而一个Foursquare用户文档包括该用户签到的所有商户以及和这些商户关联的标签信息



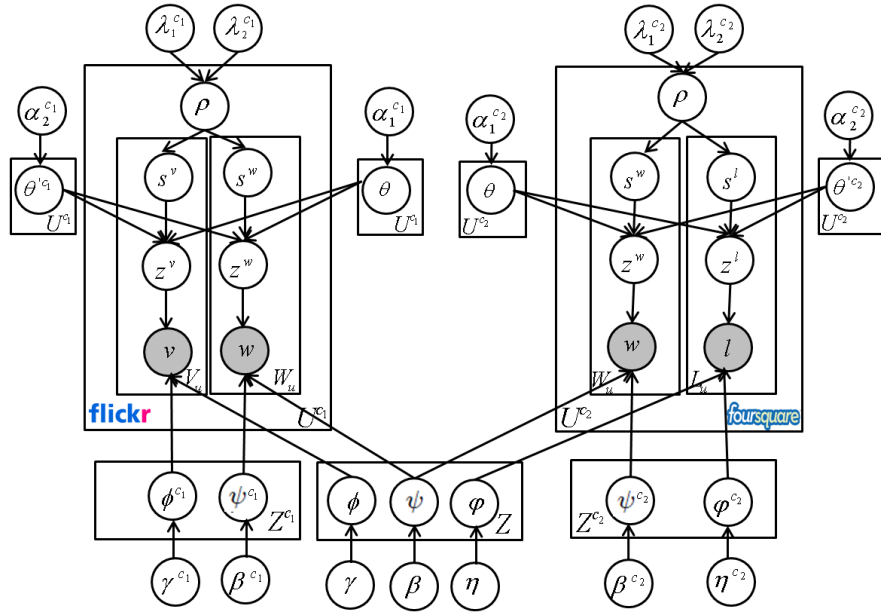


图 6.3: CM³TM的生成式模型

和每个平台特有的主题

#### 6.4.1 问题定义

我们认为来自Flickr和Foursquare的用户文档包括三类主题：1) 共享的主题, 该主题由所有社会平台共同享有；2) **Flickr**平台特有的主题, 该主题只和Flickr平台有关和3) **Foursquare**平台特有的主题, 该类主题只Foursquare平台有关。此外，每一类主题都是由多种模态共同表示。基于以上概念，CM³TM可以如下定义：给定Flickr平台 $c_1$ 和Foursquare平台 $c_2$ 的用户集合， $U = \{u_1^{c_1}, u_2^{c_1}, \dots, u_{|U^{c_1}|}^{c_1}, u_1^{c_2}, u_2^{c_2}, \dots, u_{|U^{c_2}|}^{c_2}\}$ ，其中 $u_i^{c_m}$ 表示社会平台 $c_m$ 的一个二维的元组 $[I_{u_i^{c_m}}, T_{u_i^{c_m}}]$ 。CM³TM的目标是学习：1) 三类主题空间，也即共享的主题空间 $\psi, \phi, \varphi$ ；Flickr平台特有的主题空间 $\psi^{c_1}, \phi^{c_1}$ 和Foursquare平台特有的主题空间 $\psi^{c_2}, \varphi^{c_2}$ ；2) 用户分别基于以上三类主题的分布 $\theta, \theta'^{c_1}$ 和 $\theta'^{c_2}$ 。

#### 6.4.2 CM³TM

我们的CM³TM主要基于以下思想。（1）我们的模型应该能够区分共享的主题和每个平台特有的主题以便关联不同的平台。为了实现它，我们将主题集合分成三组，也即 $K$ 个共享主题 $\psi = \{\psi_1, \dots, \psi_K\}$ ， $K^{c_1}$ 个Flickr特有的主

题  $\psi^{c_1} = \{\psi_1^{c_1}, \dots, \psi_{K^{c_1}}^{c_1}\}$  和  $K^{c_2}$  个 Foursquare 特有的主题  $\psi^{c_2} = \{\psi_1^{c_2}, \dots, \psi_{K^{c_2}}^{c_2}\}$ ; (2) 为了便于平台间的应用, 我们的模型应该能够对齐同一个主题的不同模态, 尤其是在跨平台的情况下。为了关联同一主题的不同模态, 我们从相同的用户主题分布中采样主题。由于该模型能够学习来自所有平台的共享主题-文本词的分布  $\psi$ , 来自共享主题的文本模态、视觉模态和商户模态也同时通过该共享主题-文本词分布实现对齐。具体来说, 在 Flickr, 除了共享主题-文本词分布  $\psi$  和 Flickr 特定的主题-文本词分布  $\psi^{c_1}$  外, 同时对应着共享主题-视觉词分布  $\phi$  和 Flickr 特定的主题-视觉词分布  $\phi^{c_1}$ 。在 Foursquare 上, 共享主题-文本词分布  $\varphi$  和 Foursquare 特有的主题-文本词分布  $\varphi^{c_2}$ , 则分别对应  $\psi$  和  $\psi^{c_2}$ 。对于每一个平台  $m$ , 我们引入一个潜在的开关变量  $s^x$ ,  $x = \{w, v, l\}$  去控制文本词、视觉词或者商户是由  $\theta$  还是由  $\theta'^{c_m}$  生成。图 6.3 表示了该模型的概率生成式的图表示。具体生成过程如下:

1. 对于共享的主题  $z \in \{1, \dots, K\}$ , 采样  $\psi_z \sim \text{Dir}(\beta)$ ,  $\phi_z \sim \text{Dir}(\gamma)$ ,  $\varphi_z \sim \text{Dir}(\eta)$
2. 对于 Flickr 平台  $c_1$  特有的主题  $z \in \{1, \dots, K^{c_1}\}$ , 采样  $\psi_z^{c_1} \sim \text{Dir}(\beta^{c_1})$ ,  $\phi_z^{c_1} \sim \text{Dir}(\gamma^{c_1})$
3. 对于 Foursquare 平台  $c_2$  特有的主题  $z \in \{1, \dots, K^{c_2}\}$ ,  $\psi_z^{c_2} \sim \text{Dir}(\beta^{c_2})$ ,  $\varphi_z^{c_2} \sim \text{Dir}(\eta^{c_2})$
4. 对于每一个平台  $m \in \{c_1, c_2\}$ 
  - (a) 如果  $m = c_1$ , 对于每一个用户  $u \in U^{c_1}$ 
    - i. 采样  $\theta_u^{c_1} \sim \text{Dir}(\alpha_1^{c_1})$
    - ii. 采样  $\theta'_u{}^{c_1} \sim \text{Dir}(\alpha_2^{c_1})$
    - iii. 采样  $\rho_u^{c_1} \sim \text{Beta}(\lambda_1^{c_1}, \lambda_2^{c_1})$
    - iv. 对于每一个文本词  $w_{u,j} \in \mathbf{w}_u$ 
      - A. 采样一个开关变量  $s_{u,j}^w \sim \text{Binomial}(\rho_u^{c_1})$
      - B. 如果  $s_{u,j}^w = 0$ , 采样一个主题  $z_{u,j}^w \sim \text{Multi}(\theta_u^{c_1})$ , 然后采样一个词  $w_{u,j} \sim \text{Multi}(\psi_{z_{u,j}^w}^{c_1})$
      - C. 如果  $s_{u,j}^w = 1$ , 采样一个主题  $z_{u,j}^w \sim \text{Multi}(\theta'_u{}^{c_1})$  然后采样一个词  $w_{u,j} \sim \text{Multi}(\psi_{z_{u,j}^w}^{c_1})$

- v. 对于每一个视觉词  $v_{u,j} \in \mathbf{v}_u$ 
  - A. 采样一个开关变量  $s_{u,j}^v \sim \text{Binomial}(\rho_u^{c_1})$
  - B. 如果  $s_{u,j}^v = 0$ , 采样一个主题  $z_{u,j}^v \sim \text{Multi}(\theta_u^{c_1})$ , 然后采样一个词  $v_{u,j} \sim \text{Multi}(\phi_{z_{u,j}^v}^v)$
  - C. 如果  $s_{u,j}^v = 1$ , 采样一个主题  $z_{u,j}^v \sim \text{Multi}(\theta_u^{c_1'})$ , 然后采样一个词  $v_{u,j} \sim \text{Multi}(\phi_{z_{u,j}^v}^{c_1'})$
- (b) 如果  $m = c_2$ , 对于每一个用户  $u \in U^{c_2}$ 
  - i. 采样  $\theta_u^{c_2} \sim \text{Dir}(\alpha_1^{c_2})$ ,
  - ii. 采样  $\theta_u^{c_2'} \sim \text{Dir}(\alpha_2^{c_2})$ ,
  - iii. 采样  $\rho_u^{c_2} \sim \text{Beta}(\lambda_1^{c_2}, \lambda_2^{c_2})$ ,
  - iv. 对于每一个文本词  $w_{u,j} \in \mathbf{w}_u$ 
    - A. 采样一个开关变量  $s_{u,j}^w \sim \text{Binomial}(\rho_u^{c_2})$
    - B. 如果  $s_{u,j}^w = 0$ , 采样一个主题  $z_{u,j}^w \sim \text{Multi}(\theta_u^{c_2})$ , 然后采样一个词  $w_{u,j} \sim \text{Multi}(\psi_{z_{u,j}^w}^w)$
    - C. 如果  $s_{u,j}^w = 1$ , 采样一个主题  $z_{u,j}^w \sim \text{Multi}(\theta_u^{c_2'})$ , 然后采样一个词  $w_{u,j} \sim \text{Multi}(\psi_{z_{u,j}^w}^{c_2'})$
  - v. 对于每个商户  $l_{u,j} \in \mathbf{l}_u$ 
    - A. 采样一个开关变量  $s_{u,j}^l \sim \text{Binomial}(\rho_u^{c_2})$
    - B. 如果  $s_{u,j}^l = 0$ , 采样一个主题  $z_{u,j}^l \sim \text{Multi}(\theta_u^{c_2})$ , 然后采样一个词  $l_{u,j} \sim \text{Multi}(\varphi_{z_{u,j}^l}^l)$
    - C. 如果  $s_{u,j}^l = 1$ , 采样一个主题  $z_{u,j}^l \sim \text{Multi}(\theta_u^{c_2'})$ , 然后采样一个词  $l_{u,j} \sim \text{Multi}(\varphi_{z_{u,j}^l}^{c_2'})$

### 6.4.3 模型推断

CM<sup>3</sup>TM 模型包括四类变量集合, 即来自Flickr的二值开关标签 $\mathbf{s}^w$ ,  $\mathbf{s}^v$ , 主题赋值 $\mathbf{z}^w$ ,  $\mathbf{z}^v$ ; 来自Foursquare的二值开关标签 $\mathbf{s}^w$ ,  $\mathbf{s}^l$ 和主题赋值 $\mathbf{z}^w$ ,  $\mathbf{z}^l$ 。对于来自Flickr<sub>c1</sub>的潜变量 $\mathbf{z}^w$ 和 $\mathbf{s}^w$ , 我们使用一个两步的吉布斯采样过程: (i) 给定

当前的 $\mathbf{z}^w$ 估计, 采样 $\mathbf{s}^w$

$$\begin{aligned} p(s_i^w = 0 | \mathbf{s}_{-i}^w, \mathbf{z}^w, \mathbf{w}) \\ \propto \frac{n_{u,z_i,-i}^{w,s_i^w=0,c_1} + \alpha_1^{c_1}}{\sum_z n_{u,z,-i}^{w,s_i^w=0,c_1} + K\alpha_1^{c_1}} (n_{u,s_i^w=0,-i}^{w,c_1} + \lambda_1^{c_1}) \end{aligned} \quad (6.1)$$

$$\begin{aligned} p(s_i^w = 1 | \mathbf{s}_{-i}^w, \mathbf{z}^w, \mathbf{w}) \\ \propto \frac{n_{u,z_i,-i}^{w,s_i^w=1,c_1} + \alpha_2^{c_1}}{\sum_z n_{u,z,-i}^{w,s_i^w=1,c_1} + K\alpha_2^{c_1}} (n_{u,s_i^w=1,-i}^{w,c_1} + \lambda_2^{c_1}) \end{aligned} \quad (6.2)$$

(ii) 给定当前的 $\mathbf{s}^w$ , 采样 $\mathbf{z}^w$

$$\begin{aligned} p(z_i^w | s_i^w = 0, \mathbf{z}_{-i}^w, \mathbf{s}_{-i}^w, \mathbf{w}) \\ \propto \frac{n_{z_i,w_i,-i} + \beta}{\sum_{w'} n_{z_i,w',-i} + W\beta} (n_{u,z_i,-i}^{w,s_i^w=0,c_1} + \alpha_1^{c_1}) \end{aligned} \quad (6.3)$$

$$\begin{aligned} p(z_i^w | s_i^w = 1, \mathbf{z}_{-i}^w, \mathbf{s}_{-i}^w, \mathbf{w}) \\ \propto \frac{n_{z_i,w_i,-i}^{c_1} + \beta^{c_1}}{\sum_{w'} n_{z_i,w',-i}^{c_1} + W\beta^{c_1}} (n_{u,z_i,-i}^{w,s_i^w=1,c_1} + \alpha_2^{c_1}) \end{aligned} \quad (6.4)$$

其中 $i = (u, j)$ 代表当前的索引。下标 $-i$ 表示排除文档中第 $i$ 个词索引的计数变量;  $n_{u,z_i,-i}^{w,s_i^w=0,c_1}$ 表示由词 $w$ 表示的共享主题 $z_i$  ( $s_i^w = 0$ )赋予Flickr的用户 $u$ 的次数;  $n_{u,z_i,-i}^{w,s_i^w=1,c_1}$ 表示由词 $w$ 表示的Flickr特有的主题 $z_i$  ( $s_i^w = 1$ )赋予Flickr的用户 $u$ 的次数;  $n_{z_i,w_i,-i}$ 表示词 $w_i$ 赋予共享的主题 $z_i$  ( $s_i^w = 0$ )的次数;  $n_{z_i,w_i,-i}^{c_1}$ 表示词 $w_i$ 赋予Flickr特有的主题 $z_i$  ( $s_i^w = 1$ )的次数;  $n_{u,s_i^w=0,-i}^{w,c_1}$ 表示 $s_i^w = 0$ 被赋予Flickr用户 $u$ 的次数;  $n_{u,s_i^w=1,-i}^{w,c_1}$ 表示 $s_i^w = 1$ 被赋予Flickr用户 $u$ 的次数; 对于来自Flickr的视觉词 $\mathbf{z}^v$ 和 $\mathbf{s}^v$ 更新规则可以类比的进行推导出来。

对于来自Foursquare $c_2$ 的潜变量 $\mathbf{z}^w$ 和 $\mathbf{s}^w$ , 更新规则如下:

(i) 给定当前的 $\mathbf{z}^w$ 估计, 采样 $\mathbf{s}^w$

$$\begin{aligned} p(s_i^w = 0 | \mathbf{s}_{-i}^w, \mathbf{z}^w, \mathbf{w}) \\ \propto \frac{n_{u,z_i,-i}^{w,s_i^w=0,c_2} + \alpha_1^{c_2}}{\sum_z n_{u,z,-i}^{w,s_i^w=0,c_2} + K\alpha_1^{c_2}} (n_{u,s_i^w=0,-i}^{w,c_2} + \lambda_1^{c_2}) \end{aligned} \quad (6.5)$$

$$\begin{aligned} p(s_i^w = 1 | \mathbf{s}_{-i}^w, \mathbf{z}^w, \mathbf{w}) \\ \propto \frac{n_{u,z_i,-i}^{w,s_i^w=1,c_2} + \alpha_2^{c_2}}{\sum_z n_{u,z,-i}^{w,s_i^w=1,c_2} + K\alpha_2^{c_2}} (n_{u,s_i^w=1,-i}^{w,c_2} + \lambda_2^{c_2}) \end{aligned} \quad (6.6)$$

(ii) 给定当前的 $s^w$  估计, 采样 $\mathbf{z}^w$

$$\begin{aligned} p(z_i^w | s_i^w = 0, \mathbf{z}_{\neg i}^w, \mathbf{s}_{\neg i}^w, \mathbf{w}) \\ \propto \frac{n_{z_i, w_i, \neg i} + \beta}{\sum_{w'} n_{z_i, w', \neg i} + W\beta} (n_{u, z_i, \neg i}^{w, s_i^w=0, c_2} + \alpha_1^{c_2}) \end{aligned} \quad (6.7)$$

$$\begin{aligned} p(z_i^w | s_i^w = 1, \mathbf{z}_{\neg i}^w, \mathbf{s}_{\neg i}^w, \mathbf{w}) \\ \propto \frac{n_{z_i, w_i, \neg i}^{c_2} + \beta^{c_2}}{\sum_{w'} n_{z_i, w', \neg i}^{c_2} + W\beta^{c_2}} (n_{u, z_i, \neg i}^{w, s_i^w=1, c_2} + \alpha_2^{c_2}) \end{aligned} \quad (6.8)$$

这些参数的意义可以类比来自Flickr平台的参数的意义。同样来自Foursquare的 $\mathbf{z}^l$ 和 $\mathbf{s}^l$ 的更新规则也可以类比的推导出来。

#### 6.4.4 参数估计

经过充分的采样迭代次数, 我们进行如下的参数估计:

$$\begin{aligned} \hat{\rho}_{u, s=0}^{c_1} &= \frac{n_{u, s=0}^{w, c_1} + n_{u, s=0}^{v, c_1} + \lambda_1^{c_1}}{\sum_{s'} n_{u, s'}^{w, c_1} + \sum_{s'} n_{u, s'}^{v, c_1} + \lambda_1^{c_1} + \lambda_2^{c_1}} \\ \hat{\rho}_{u, s=1}^{c_1} &= \frac{n_{u, s=1}^{w, c_1} + n_{u, s=1}^{v, c_1} + \lambda_2^{c_1}}{\sum_{s'} n_{u, s'}^{w, c_1} + \sum_{s'} n_{u, s'}^{v, c_1} + \lambda_1^{c_1} + \lambda_2^{c_1}} \\ \hat{\rho}_{u, s=0}^{c_2} &= \frac{n_{u, s=0}^{w, c_2} + n_{u, s=0}^{l, c_2} + \lambda_1^{c_2}}{\sum_{s'} n_{u, s'}^{w, c_2} + \sum_{s'} n_{u, s'}^{l, c_2} + \lambda_1^{c_2} + \lambda_2^{c_2}} \\ \hat{\rho}_{u, s=1}^{c_2} &= \frac{n_{u, s=1}^{w, c_2} + n_{u, s=1}^{l, c_2} + \lambda_2^{c_2}}{\sum_{s'} n_{u, s'}^{w, c_2} + \sum_{s'} n_{u, s'}^{l, c_2} + \lambda_1^{c_2} + \lambda_2^{c_2}} \\ \hat{\theta}_{u, z}^{c_1} &= \frac{n_{u, z}^{w, s^w=0, c_1} + n_{u, z}^{v, s^v=0, c_1} + \alpha_1^{c_1}}{\sum_{z'} n_{u, z'}^{w, s^w=0, c_1} + \sum_{z'} n_{u, z'}^{v, s^v=0, c_1} + K\alpha_1^{c_1}} \\ \hat{\theta}_{u, z}^{c_1'} &= \frac{n_{u, z}^{w, s^w=1, c_1} + n_{u, z}^{v, s^v=1, c_1} + \alpha_2^{c_1}}{\sum_{z'} n_{u, z'}^{w, s^w=1, c_1} + \sum_{z'} n_{u, z'}^{v, s^v=1, c_1} + K\alpha_2^{c_1}} \\ \hat{\theta}_{u, z}^{c_2} &= \frac{n_{u, z}^{w, s^w=0, c_2} + n_{u, z}^{l, s^l=0, c_2} + \alpha_1^{c_2}}{\sum_{z'} n_{u, z'}^{w, s^w=0, c_2} + \sum_{z'} n_{u, z'}^{l, s^l=0, c_2} + K\alpha_1^{c_2}} \\ \hat{\theta}_{u, z}^{c_2'} &= \frac{n_{u, z}^{w, s^w=1, c_2} + n_{u, z}^{l, s^l=1, c_2} + \alpha_2^{c_2}}{\sum_{z'} n_{u, z'}^{w, s^w=1, c_2} + \sum_{z'} n_{u, z'}^{l, s^l=1, c_2} + K\alpha_2^{c_2}} \end{aligned} \quad (6.9)$$

$$\begin{aligned}
\hat{\psi}_{z,w} &= \frac{n_{z,w} + \beta}{\sum_{w'} n_{z,w'} + W\beta} \\
\hat{\psi}'_{z,w} &= \frac{n_{z,w}^{c_m} + \beta^{c_m}}{\sum_{w'} n_{z,w'}^{c_m} + W\beta^{c_m}} \\
\hat{\phi}_{z,v} &= \frac{n_{z,v} + \gamma}{\sum_{v'} n_{z,v'} + V\gamma} \\
\hat{\phi}'_{z,v} &= \frac{n_{z,v}^{c_1} + \gamma^{c_1}}{\sum_{v'} n_{z,v'}^{c_1} + V\gamma^{c_1}} \\
\hat{\varphi}_{z,l} &= \frac{n_{z,l} + \eta}{\sum_{l'} n_{z,l'} + L\eta} \\
\hat{\varphi}'_{z,l} &= \frac{n_{z,l}^{c_2} + \eta^{c_2}}{\sum_{l'} n_{z,l'}^{c_2} + L\eta^{c_2}}
\end{aligned} \tag{6.10}$$

其中  $m \in \{1, 2\}$ 。  $s \in \{0, 1\}$ ,  $mod \in \{w, v, l\}$ 。  $n_{u, s_i^w=s}^{w, c_m}$  表示  $s_i^w = s$  被赋予平台  $m$  的用户  $u$  的次数。  $n_{u, s_i^v=s}^{v, c_1}$  表示  $s_i^v = s$  被赋予 Flickr 用户  $u$  的次数。  $n_{u, s_i^l=s}^{l, c_2}$  表示  $s_i^l = s$  被赋予来自 Foursquare 用户  $u$  的次数。  $n_{u, z}^{x, s_i^x=s, c_m}$  表示由  $x$  表示的主题  $z$  被赋予来自平台  $c_m$  的用户  $u$  的次数。  $n_{z,w}$  表示词  $w$  被赋予共享的主题  $z$  的次数。  $n_{z,w}^{c_m}$  表示词  $w$  被赋予平台  $c_m$  特定的主题  $z$  的次数。  $n_{z,v}$  表示视觉词  $v$  被赋予共享主题  $z$  的次数。  $n_{z,v}^{c_1}$  表示视觉词  $v$  被赋予平台  $c_1$  特定的主题  $z$  的次数。  $n_{z,l}$  表示商户词  $l$  被赋予共享主题  $z$  的次数。  $n_{z,l}^{c_2}$  表示商户词  $l$  被赋予平台  $c_2$  特有的主题  $z$  的次数。

## 6.5 基于位置情景的跨平台推荐

我们利用主题模型推导的共享的主题和相应的用户主体分布可以将 CM<sup>3</sup>TM 应用到许多新的平台间的应用中。在本节中，我们设计一个基于位置情景的平台间的应用：在给定某一个平台的用户  $u$  和一个查询城市  $r$ ，我们的目标是根据该用户的兴趣推荐来自其他平台的商品列表。具体包括以下两个方面：1) 个性化的 Foursquare 商户推荐给 Flickr 用户和 2) 个性化的 Flickr 图片推荐给 Foursquare 用户。

### 6.5.1 个性化的 Foursquare 商户推荐给 Flickr 用户

对于一个 Flickr 用户  $u_{c_1}$  和一个查询城市  $r$ ，来自 Foursquare 推荐的商户的排

序通过如下公式计算:  $p(l_{c_2}|u_{c_1}, r) = \sum_{k \in K} p(l_{c_2}|z_k)p(z_k|u_{c_1}) =$

$$\begin{cases} \sum_{k \in K} \hat{\varphi}_{z_k, l_{c_2}} \hat{\theta}_{u_{c_1}, z_k}^{c_1}, l_{c_2} \in L_r \\ 0, l_{c_2} \notin L_r \end{cases} \quad (6.11)$$

其中  $L_r$  代表某个城市  $r$  对应的商户集合。  $\hat{\theta}_{u_{c_1}, z_k}^{c_1}$  表示 Flickr 用户的共享的主题分布,  $\hat{\varphi}_{z_k, l_{c_2}}$  表示共享的主题-商户分布。通过共享的主题空间我们将 Flickr 用户和 Foursquare 商户信息建立关联, 这样得到的排序的商户列表中很好的考虑了 Flickr 用户的兴趣。同时为了推荐与该城市相关的商户信息, 我们对推荐的商户是否在该城市进行了判断, 如果  $l_{c_2} \notin L_r$ , 我们将对应的商户得分设置为 0。

### 6.5.2 个性化的 Flickr 图片推荐给 Foursquare 用户

对于一个 Foursquare 用户  $u_{c_2}$  和一个查询城市  $r$ , 来自 Flickr 推荐的图片的排序通过如下公式计算:  $\text{sim}(u_{c_2}, I_{c_1}) =$

$$\begin{cases} \omega \text{sim}(\hat{\theta}_{u_{c_2}}^{c_2}, \mathbf{d}_{I_{c_1}}^w) + (1 - \omega) \text{sim}(\hat{\theta}_{u_{c_2}}^{c_2}, \mathbf{d}_{I_{c_1}}^v), \text{geo}_{I_{c_1}} \in \text{Geo}_r \\ 0, \text{geo}_{I_{c_1}} \notin \text{Geo}_r \end{cases} \quad (6.12)$$

其中  $\omega$  是一个权重参数用于控制图片  $I_{c_1}$  视觉内容和文本内容的强度。  $\text{geo}_{I_{c_1}}$  表示图片  $I_{c_1}$  的地理标签信息。  $\text{Geo}_r$  表示城市  $r$  的经纬度范围。  $\hat{\theta}_{u_{c_2}}^{c_2}$  表示  $u_{c_2}$  的主题分布。  $\mathbf{d}_{I_{c_1}}^w$  表示图片文本内容的主题分布。  $\mathbf{d}_{I_{c_1}}^w = \langle d_{I_{c_1}, z_1}^w, \dots, d_{I_{c_1}, z_k}^w, \dots, d_{I_{c_1}, z_K}^w \rangle$ 。  $\mathbf{d}_{I_{c_1}}^v$  表示图片视觉内容的主题分布。  $\mathbf{d}_{I_{c_1}}^v = \langle d_{I_{c_1}, z_1}^v, \dots, d_{I_{c_1}, z_k}^v, \dots, d_{I_{c_1}, z_K}^v \rangle$ 。由于引入了共享的主题空间, 通过 Foursquare 用户共享的主题分布和 Flickr 图片共享的主题分布的相似度计算, 我们可以得到每一个 Foursquare 用户感兴趣的 Flickr 图片列表。同样我们根据每张图片的 GPS 信息判断该图片是否在相应的查询城市, 这样可以保证我们推荐的 Flickr 图片不仅和 Foursquare 的用户兴趣相关, 而且图片的视觉内容是和查询城市相关的。

$$\begin{aligned} d_{I_{c_1}, z_k}^w &= \frac{1}{n_{I_{c_1}}^w} \sum_{i=1}^{n_{I_{c_1}}^w} p(z_k | w_{I_{c_1}, i}) = \frac{p(z_k)}{n_{I_{c_1}}^w} \sum_{i=1}^{n_{I_{c_1}}^w} \frac{p(w_{I_{c_1}, i} | z_k)}{p(w_{I_{c_1}, i})} \\ &= \frac{p(z_k)}{n_{I_{c_1}}^w} \sum_{i=1}^{n_{I_{c_1}}^w} \frac{\hat{\psi}_{z_k, w_{I_{c_1}, i}}}{p(w_{I_{c_1}, i})} \end{aligned} \quad (6.13)$$

同样

$$d_{I_{c_1}, z_k}^v = \frac{p(z_k)}{n_{I_{c_1}}^v} \sum_{i=1}^{n_{I_{c_1}}^v} \frac{\hat{\phi}_{z_k, v_{I_{c_1}, i}}}{p(v_{I_{c_1}, i})} \quad (6.14)$$

其中 $n_{I_{c_1}}^w$ 和 $n_{I_{c_1}}^v$ 图片 $I_{c_1}$ 的文本词和视觉词的数量。 $p(w_{I_{c_1},i})$ 和 $p(v_{I_{c_1},i})$ 则表示文本词和视觉词的先验分布。我们假定视觉词先验和标签词先验服从均匀分布,也就是 $p(w_{I_{c_1},i}) = \frac{1}{W}$ 和 $p(v_{I_{c_1},i}) = \frac{1}{V}$ 。 $p(z_k)$ 表示主题的先验分布,计算如下:

$$p(z_k) = \sum_u p(z_k|u)p(u) = \frac{\sum_{u \in U^{c_1}} \hat{\theta}_{u,z_k}^{c_1} (n_u^w + n_u^v) + \sum_{u \in U^{c_2}} \hat{\theta}_{u,z_k}^{c_2} (n_u^w + n_u^l)}{\sum_{u \in U^{c_1}} (n_u^w + n_u^v) + \sum_{u \in U^{c_2}} (n_u^w + n_u^l)} \quad (6.15)$$

其中 $n_u^w$ ,  $n_u^v$ 和 $n_u^l$ 表示用户 $u$ 文本词、视觉词和商户词的数量。

## 6.6 实验结果与分析

本节首先介绍相关的实验设置,包括数据库和一些实现细节,然后定性和定量的评估CM<sup>3</sup>TM模型的性能。最后我们验证提出的个性化平台间的推荐的性能。

### 6.6.1 实验设置

#### 6.6.1.1 数据库

我们选择Foursquare和Flickr两个平台作为测试平台。其中Foursquare是一个著名的基于位置的社会媒体平台而Flickr则是一个比较流行的图片分享平台。为了进行基于位置情景的平台间的推荐评估,我们选择纽约城是作为测试城市,收集两个平台的数据。数据爬取的过程如下:

- **Flickr** 为了下载足够的图片,我们使用两种查询:城市的名称和相应的GPS信息检索种子图片。然后我们基于种子图片的用户ID作为查询,利用Flickr API<sup>2</sup>下载所有的地理标签的图片。对于每张图片,我们也下载与之关联的元数据,包括标签、标题、描述、地理标签和最受欢迎的数量(favourite-count)信息。
- **Foursquare** 由于我们无法直接从Foursquare中直接获得个性化的签到信息,因而我们间接通过Twitter流<sup>3</sup>爬取包含签到信息的推文。和[135]类似,我们通过爬取带有关键词“4sq”的推文得到签到数据。每一条推文包含签到信

<sup>2</sup><https://www.flickr.com/services/api/>

<sup>3</sup><https://dev.twitter.com/docs/streaming-api>



息和链接到Foursquare的URL网址。然后我们可以直接从该URL页中得到更多的信息。最终我们的数据库包括1,224,899个用户，7,383,865的签到数据和2,509,641商户信息。时间范围从2014-04-01到2014-07-01。对于初步获得的数据，我们首先基于经纬度信息提取来自纽约城市的签到数据。对于这些签到数据的用户，我们爬取他们在其他城市的签到信息。除此之外，既然发送一条评论可以等价于一条签到信息 [7]，因此我们用Foursquare API<sup>4</sup>爬取所有的评论。对于每一个商户，我们也同时爬取关联的其他信息包括标签、经纬度、签到数量和评论数量。

既然我们关注以用户为中心的应用，我们去掉每个平台用户记录数小于8的用户。对于来自Foursquare和Flickr的文本信息，我们通过去除停用词、html和标签等进行预处理。文本的词典大小为22,532。最终的数据库统计如下表6.2所示。

表 6.2: 数据统计

# 平台	# 用户	# 图片	# 商户	# 签到	# 推文
Flickr	11,996	856,795	-	-	-
Foursquare	13,143	-	40,166	144,805	516,655

### 6.6.1.2 实现细节

相比基于关键点的描述子，MSER区域能够指示对象的局部同质的部分，展示比较高的判别性 [75]。在我们的实现中，我们在来自MSER的每个检测的椭圆区域中计算SIFT描述子，我们从采样的20,000张图片中提取大概447,030个MSER描述子，这些描述子进一步量化成大小为1024的视觉词词典。对于来自Foursquare的商户信息，和 [59]类似，每个用户的所有的商户信息被聚合成一个商户文档，这里每个商户可以看做一个词。此外在没有先验知识的情况下，这些模型的超参数被经验的如下设置： $\alpha_1^{c_1} = \alpha_2^{c_1} = \alpha_1^{c_2} = \alpha_2^{c_2} = 1.0$ ， $\lambda_1^{c_1} = \lambda_2^{c_1} = \lambda_1^{c_2} = \lambda_2^{c_2} = 1.0$ ， $\beta = \beta^{c_1} = \beta^{c_2} = 0.01$ ， $\gamma = \gamma^{c_1} = 0.01$ ， $\eta = \eta^{c_2} = 0.01$ 。

<sup>4</sup><https://developer.foursquare.com/docs/users/tips>

## 6.6.2 跨平台多模态主题模型评估

### 6.6.2.1 定性评估

我们首先通过检查发现的主题展示我们模型的有效性。具体来说，我们通过提供排序最靠前的5个文本词、图片和商户信息可视化这些主题。每一个共享的主题是由相关的文本词、视觉词和商户信息表示。Flickr特有的主题则是由相关的词和图片表示。Foursquare特有的主题则是由相关的词和商户信息表示。图6.4展示了一些挖掘的主题，包括三个共享的主题，两个Flickr特有的主题和两个Foursquare特有的主题。文本词是根据相应的概率分布 $p(w_i|z_k)$ 排序而商户信息根据 $p(l_i|z_k)$ 排序。图片是通过如下的相似度计算排序：

$$\text{sim}(Z_i, I_j) = \nu \frac{\mathbf{w}_{z_i} \mathbf{w}_{I_j}}{\|\mathbf{w}_{z_i}\| \|\mathbf{w}_{I_j}\|} + (1 - \nu) \frac{\mathbf{v}_{z_i} \mathbf{v}_{I_j}}{\|\mathbf{v}_{z_i}\| \|\mathbf{v}_{I_j}\|} \quad (6.16)$$

其中 $\nu$ 是权重参数（本实验中 $\nu = 0.6$ ）。 $\mathbf{w}_{z_i}$ 和 $\mathbf{v}_{z_i}$ 分别表示学习的主题文本词分布和主题视觉词分布； $\mathbf{w}_{I_j}$ 和 $\mathbf{v}_{I_j}$ 分别表示归一化的文本词的Term Frequency(TF)向量和图片 $I_j$ 视觉词的向量。

正如图6.4所示，我们可以看到一些共享的主题包括食物（#5）、艺术（#29）和运动（#87），这些主题由所有平台共享。相比之一，Flickr平台特有的主题包括一些政治事件（#54）和自然景观（#64），这些主题一般都不会在Foursquare平台出现。同样Foursquare平台特有的主题包括阅读（#44）和电子发烧友（#62）。通过提供代表性的词、图片和商户的组合，我们很容易解释和每个主题关联的领域知识。从代表性的主题，我们能够看到1）这两个平台确实存在三类主题，包括共享的主题、Flickr特有的主题和Foursquare特有的主题；2）通过考虑文本词、视觉词和商户信息，发现的主题展示了语义概念、视觉主题和商户主题高度的一致性。

### 6.6.2.2 定量评估

我们使用perplexity值进行模型的评估，该测度为主题模型的标准定量评估测度。值越低，模型的泛化性能越好。在我们的评估中，我们采用caption perplexity [13]作为性能测量，这个评估适合用在带有多种模态的主题模型中。具体定义如下

$$\text{perplexity}(D_{\text{test}}) = \exp\left(-\frac{\sum_{d \in D_{\text{test}}} \ln p(\mathbf{w}_d | \mathbf{v}_d)}{\sum_{d \in D_{\text{test}}} n_d^w}\right) \quad (6.17)$$

共享的主题 #5					
词	ramen 0.08639	pork 0.0470	japanese 0.0444	restaurant 0.0427	noodle 0.0221
商户	ippudo ramen / noodle house 0.05327	momofuku noodle bar 0.0374	totto ramen ramen / noodle house 0.0300	chuko ramen / noodle house 0.0179	rai rai ken ramen / noodle house 0.0151
图片	 0.3955	 0.3719	 0.3697	 0.3577	 0.3189
共享的主题 #29					
词	art 0.1250	meum 0.08652	painting 0.01369	sculpture 0.01362	modern 0.01270
商户	whitney museum of american art museum 0.05327	the cloisters museum 0.04493	moma ps1 contemporary art center art museum 0.0412	new art museum 0.02497	neue galerie art museum 0.0150
图片	 0.6887	 0.6656	 0.6488	 0.6426	 0.6319
共享的主题 #87					
词	baseball 0.07335	stadium 0.06574	game 0.03480	field 0.03339	sports 0.03214
商户	yankee stadium 0.2088	citi field baseball stadium 0.0933	basketball stadium 0.0793	mcu park baseball stadium 0.0148	stan's sports bar 0.0104
图片	 0.6770	 0.6178	 0.5768	 0.5567	 0.5397
Flickr特有的主题 #54					
词	march 0.0817	day 0.0717	wall 0.0582	occupy 0.0351	protest 0.0338
图片	 0.5805	 0.5424	 0.5183	 0.4991	 0.4398
Flickr特有的主题 #64					
词	sunset 0.0989	clouds 0.0700	sky 0.04417	island 0.0363	ocean 0.0224
图片	 0.6241	 0.5565	 0.5373	 0.4984	 0.4792
Foursquare特有的主题 #44					
词	book 0.04859	books 0.0481	bookstore 0.03829	store 0.0379	library 0.0213
商户	strand book store 0.0301	forbidden planet bookstore 0.0268	jhu comic books comic shop 0.0167	barnes & noble bookstore 0.0100	brooklyn public library 0.008
Foursquare特有的主题 #62					
词	store 0.0918	apple 0.0824	electronics 0.0545	iphone 0.0317	mac 0.0219
商户	apple store 0.1475	b&h photo video camera store 0.0369	apple store, roosevelt field electronics store 0.0167	best buy electronics store 0.0148	apple store electronics store 0.0111

图 6.4: CM<sup>3</sup> TM挖掘的三类主题例子

其中 $D_{test}$ 为测试集， $\mathbf{w}_d$ 是文本文档的文本词而 $\mathbf{v}_d$ 则表示视觉词的向量。

我们定义CM<sup>3</sup>TM的caption perplexity如下：

$$perplexity(D_{test}) = \exp\left(-\frac{\sum_{u \in U_{test}^{c1}} \ln p(\mathbf{w}_u | \mathbf{v}_u) + \sum_{u \in U_{test}^{c2}} \ln p(\mathbf{w}_u | \mathbf{l}_u)}{\sum_{u \in U_{test}} n_u^w}\right) \quad (6.18)$$

$$\begin{aligned} p(\mathbf{w}_u | \mathbf{v}_u) &= \prod_{w_u \in \mathbf{w}_u} \sum_{k \in K_1} p(w_u | z_k) p(z_k | \mathbf{v}_u) \\ p(\mathbf{w}_u | \mathbf{l}_u) &= \prod_{w_u \in \mathbf{w}_u} \sum_{k \in K_2} p(w_u | z_k) p(z_k | \mathbf{l}_u) \end{aligned} \quad (6.19)$$

其中 $\mathbf{w}_u$ 、 $\mathbf{v}_u$ 和 $\mathbf{l}_u$ 分别表示用户 $u$ 的文本词向量、视觉词向量和商户词向量。经过模型训练，我们得到主题词分布 $p(w_u | z_k)$ 、主题视觉词分布 $p(v_u | z_k)$ 和主题商户词分布 $p(l_u | z_k)$ 。和方法 [25]类似， $p(z_k | \mathbf{v}_u)$ 的估计是通过使用 $p(v_u | z_k)$ 对来自测试用户的所有图片的视觉词进行吉布斯采样得到（不使用文本词）。同样， $p(z_k | \mathbf{l}_u)$ 的估计是通过使用 $p(l_u | z_k)$ 对来自测试用户的所有商户文档的商户词上进行吉布斯采样得到（不使用文本词）。 $K_1$ 和 $K_2$ 分别是来自Foursquare和Flickr的主题数量和。 $K_1 = K + K^{c1}$ ,  $K_2 = K + K^{c2}$ 。

我们设计如下基准方法进行比较：**Basic Version: CM<sup>3</sup>TM\_BV**。该模型并不区分共享的主题和平台特有的主题，而只考虑两个平台共享的主题空间。

我们随机将数据库分成两部分：训练集和测试集。我们随机选择每个平台80%的用户和他们的数据作为训练集，剩余的数据为测试集。在没有任何先验的情况下，我们设置每个平台共享主题的比例为0.5，也就是 $K = K^{c1} = K^{c2}$ 。图6.5展示了对不同主题数量 $K^{all} = K + K^{c1} + K^{c2}$ 的测试集的perplexity值。我们能够看到我们提出的方法由于能够区分共享的主题和平台特有的主题而达到了较好的perplexity值。区分这两类主题能够比较好的分离出不相关或者不一致的主题知识。此外，perplexity值在 $K^{all} = 240$  ( $K = K^{c1} = K^{c2} = 80$ )达到稳定，因此在此后的实验中，我们选择主题的数量 $K = K^{c1} = K^{c2} = 80$ 进行如下的实验。

### 6.6.2.3 计算复杂度分析

不失为一般性，令 $K = K^{c1} = K^{c2}$ ，吉布斯采样的迭代次数为 $I$ 。在每一次迭代中，它需要遍历来自两个平台所有的用户 $U$ 。对于每个用户 $u \in U^{c1}$ ，需

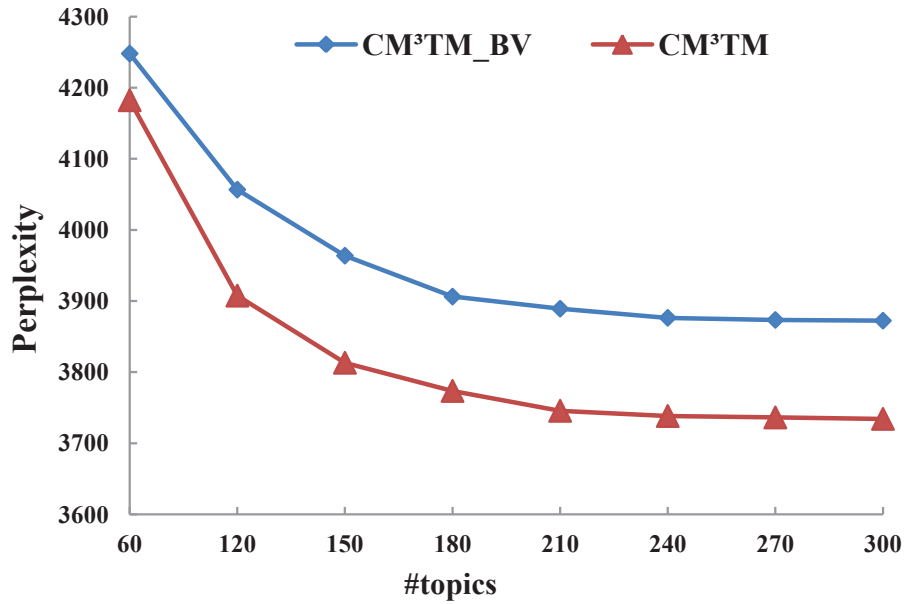


图 6.5: Perplexity 比较

要 $O((W_u + V_u)K)$ 操作进行模型推断。同样对于每个用户 $u \in U^{c2}$ ，需要 $O((W_u + L_u)K)$ 操作进行模型推断。其中 $W_u$ 是用户文档 $u$ 中词的总数， $V_u$ 是用户文档 $u$ 总的视觉词， $L_u$ 是用户文档 $u$ 总的商户数量。因此整个的CM<sup>3</sup>TM的时间复杂度为 $O(U^{c1}(W_u + V_u)KI + U^{c2}(W_u + L_u)KI) = O((UW_u + U^{c1}V_u + U^{c2}L_u)KI)$ 。在我们的实验中，我们在Inter Core 3.40GHz处理器的PC上进行训练，训练过程包括1000次的吉布斯采样，大概持续462,000秒。平均每次迭代的时间大概为462秒。

### 6.6.3 个性化平台间推荐评估

#### 6.6.3.1 评估策略

跨平台个性化推荐的目的是基于一个平台用户的兴趣将另一平台的商品推荐给该用户。但是我们很难获得Foursquare和Flickr两个平台同一个用户的不同账号信息。因此我们采用如下的评估策略。对于个性化的Foursquare商户推荐给Flickr用户，我们让20个用户标注返回的商户列表。他们可以查看该用户上传的图片集帮助判断。对于每个商户，如果超过10个用户认为它和该用户有关，那么标注1，否则标注0。同样在个性化的Flickr图片推荐给Foursquare用

户，我们也同样让这20个用户去标注返回的图片列表，对于每张图片，如果超过10个用户标注他和该Foursquare用户相关，则标为1，否则为0。

### 6.6.3.2 评估测度

我们的目的是提供给用户一个排序的商品列表，因此和传统的信息检索方法一致，我们使用Precision@K和MAP@K作为评估测度。

对于一个给定的查询 $u \in U$ ，Precision@K如下定义：

$$Precision@K = \frac{\sum_{k=1}^K r(k)}{K} \quad (6.20)$$

MAP@K是所有用户 $U$ 的平均准确度得分之和的平均，定义如下

$$MAP@K = \frac{1}{U} \sum_{u=1}^U \frac{\sum_{k=1}^K Precision@uk * r(uk)}{\sum_{k=1}^K r(uk)} \quad (6.21)$$

其中 $r_k$ 是在位置 $k$ 的相关等级，0表示“Not Relevant”，而1表示“Relevant”。 $r_{uk}$ 是用户 $u$ 在位置 $k$ 的相关度。Precision@uk是用户 $u$ 在位置 $k$ 的准确度。并且 $K$ 是截断级。在我们的研究中， $K \in \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ 。经过模型学习，我们随机选择1000个用户作为每个平台的测试查询，其中每个查询包括查询用户的主题分布和查询城市。我们分别根据(6.11)和(6.12)排序商户和图片。在个性化的Flickr图片推荐给Foursquare用户中，我们设置权重 $\omega = 0.6$ 。

### 6.6.3.3 基准方法

既然我们很难获得来自不同平台的共有用户，我们不能比较我们的方法和现有的基于共有用户连接平台的方法，具体来说，我们考虑如下方法进行比较。

- **流行度 (Popularity, POP)**：该方法根据商品的流行度对所有用户提供相同的推荐列表。在Foursquare，商户 $j$ 的流行度得分 $score\_l_j$ 为

$$score\_l_j = \tau \#check\_in_j + (1 - \tau) \#user\_tip_j \quad (6.22)$$

其中 $\#check\_in_j$ 为商户 $j$ 总的签到次数。 $\#user\_tip_j$ 为商户 $j$ 留下的评论条数。 $\tau = 0.5$ 。在Flickr，该图片受喜爱的数量直接表明该图片有多少人喜欢，因此他直接反映出一张图片的流行度，因而图片 $p_j$ 的流行度得分 $score\_p_j$ 为

$$score\_p_j = \#favorite_j \quad (6.23)$$



其中 $\#favorite_j$ 为图片 $j$ 总的喜欢的人的数量。

- **基于向量跨空间模型的K近邻算法 (Vector Space Model-based K-Nearest Neighbors Algorithm, VSM\_KNN)**: 该方法充分利用用户活动历史和商品的文本信息创建一个由TF-IDF表示的文本向量。对于每个Flickr用户, 我们将来自每张图片的文本聚合成一个用户文档。同样, 每一个Foursquare用户签到的所有商户的标签和用户评论信息作为一个文档。每张图片 and 每个商户也通过他们相关的文本的TF-IDF向量表示。对于Flickr用户, VSM\_KNN检索所有商户信息基于用户也文本向量和商户文本向量的余弦相似度找到 $k$ 个最近邻。

$$sim(u_{c_1}, l_{c_2}) = \frac{\mathbf{w}_{u_{c_1}} \mathbf{w}_{l_{c_2}}}{\|\mathbf{w}_{u_{c_1}}\| \|\mathbf{w}_{l_{c_2}}\|} \quad (6.24)$$

同样对于Foursquare用户, VSM\_KNN基于用户的文本向量和图片的文本向量的相似度找到 $k$ 个最近邻。

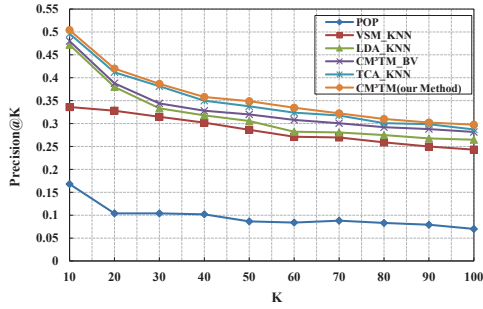
$$sim(u_{c_2}, I_{c_1}) = \frac{\mathbf{w}_{u_{c_2}} \mathbf{w}_{I_{c_1}}}{\|\mathbf{w}_{u_{c_2}}\| \|\mathbf{w}_{I_{c_1}}\|} \quad (6.25)$$

其中 $\mathbf{w}_{u_{c_1}}$ ,  $\mathbf{w}_{I_{c_1}}$ ,  $\mathbf{w}_{u_{c_2}}$ ,  $\mathbf{w}_{l_{c_2}}$ 分别表示Flickr用户、Flickr图片、Foursquare用户和Foursquare商户的特征向量。

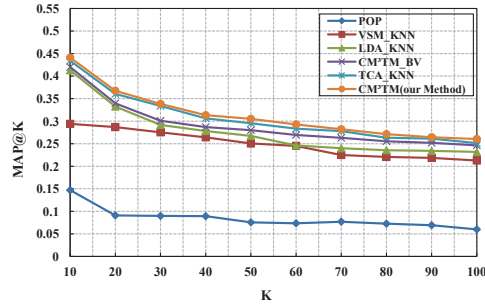
- **基于潜在狄利克雷分布的K近邻算法 (LDA-based K-Nearest Neighbors Algorithm, LDA\_KNN)**: LDA 用于发现每个用户和商品的主题分布, 但这个模型只考虑文本信息, 并不区分共享的主题和平台特有的主题。相似度的计算和公式(6.24) (6.25)相似。
- **CM<sup>3</sup>TM\_BV**: 该模型融入多模态信息。推荐策略和CM<sup>3</sup>TM相似, 相似度通过(6.11) and (6.12)计算。
- **基于TCA的K近邻算法 (TCA-based K-Nearest Neighbors Algorithm, TCA\_KNN)** [65]: 相比于LDA\_KNN, 该模型区分共享的主题和平台特有的主题。每一用户和商品都通过共享的主题分布表示。相似度通过(6.11) 和(6.12)分别计算。

#### 6.6.3.4 结果分析

我们给出两种推荐情况下的性能比较。图6.6 和6.7展示了在两种设置下的Precision@K 和MAP@K。对于第一种设置, 我们可以看到CM<sup>3</sup>TM超出了基准方法。首先, POP 和VSM\_KNN 比其他的方法性能要差, 这表

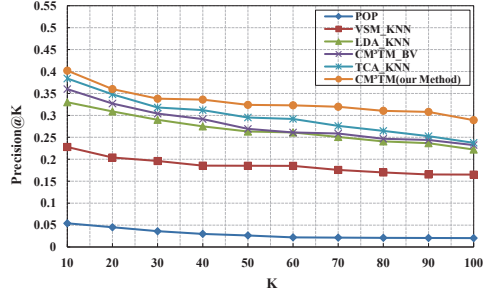


(a) Precision@K

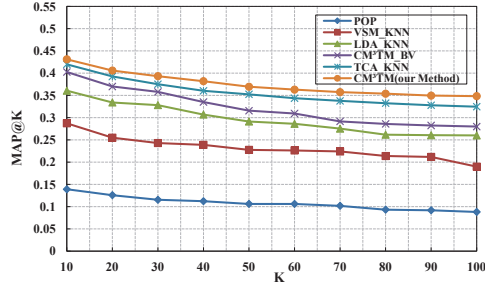


(b) MAP@K

图 6.6: Precision 和MAP 在个性化的Foursquare商户推荐给Flickr用户的性能比较



(a) Precision@K



(b) MAP@K

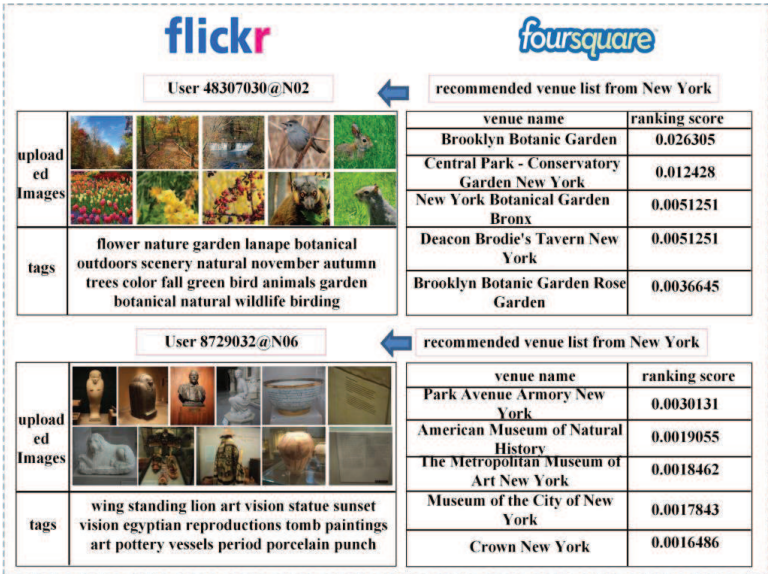
图 6.7: Precision 和MAP 在个性化的Flickr图片推荐给Foursquare用户的性能比较

明使用潜主题模型能够更加有效的建模用户兴趣。其次， $CM^3TM.BV$  超过了 $LDA\_KNN$ ，这验证了文本主题和商户主题一致性带来的优势。相比于 $CM^3TM.BV$  和 $LDA\_KNN$ ， $TCA\_KNN$ 进一步改进了性能。这是因为 $TCA\_KNN$ 能够区分共享的主题和平台特定的主题，因而能够将平台特定的主题作为不相关的信息给予过滤掉。这保证了平台间知识转移的一致性。最终，我们提出的方法超过了 $TCA\_KNN$ 达到了最佳推荐性能。这主要是因为我们的方法对每一个共享主题能够对齐文本模态和商户模态。这保证了文本主题和商户主题的异质性。同样我们可以在第二种设置下得出相似的实验分析结论。

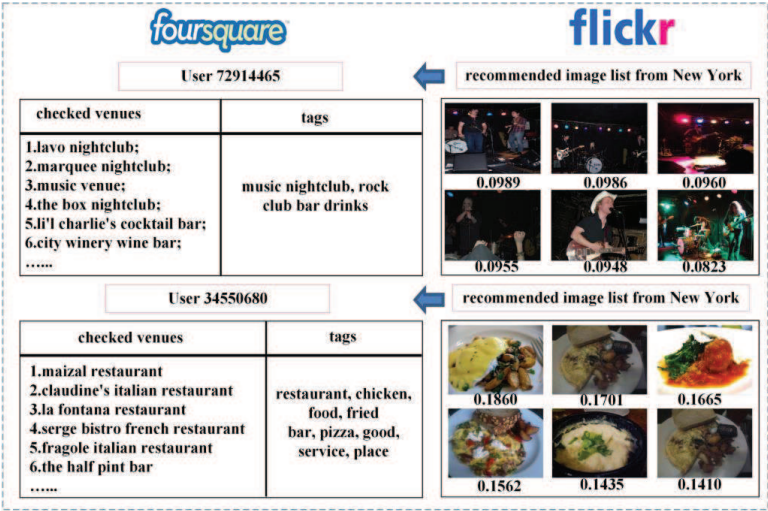
### 6.6.3.5 定性的事例分析

我们进一步通过提供用户推荐的结果展示平台间的推荐性能，包括如下两





(a) 个性化的Foursquare商户推荐给Flickr用户的例子



(b) 个性化的Flickr图片推荐给Foursquare用户的例子

图 6.8: 个性化的跨平台推荐的事例研究

种设置：1) 个性化的来自Foursquare的商户推荐给Flickr用户(图6.8(a))和2) 个性化的来自Flickr的图片推荐给Foursquare用户(图6.8(b))。在(1)中，图6.8(a)展示了两个Flickr用户，他们分享的图片以及对应的来自Foursquare推荐的商户信息。既然用户能够通过上传他们的图片表达他们的兴趣，我们展示相应的上传图片的图片和标签。以Flickr用户“48307030@N02”作为一个例子，我们可以

看到该用户喜欢场景和动物。来自Foursquare的相应的推荐的商户列表包括一些公园，这比较好的满足了Flickr用户的兴趣。同样在(2)，图6.8(b)也展示了两个Foursquare用户和相应的推荐的来自Flickr的图片列表。以Foursquare用户“72914465”作为例子，Foursquare用户经常去一些俱乐部和音乐商户，我们断定他可能会比较喜欢音乐。我们的方法相应的推荐该Foursquare用户一些关于音乐会的图片。

## 6.7 本章小结

本章提出了一个跨平台多模态概率模型 $CM^3TM$ 解决平台间的推荐问题。 $CM^3TM$ 能够区分共享的主题和平台特有的主题，同时可以对齐来自不同平台同一主题的不同模态。而且 $CM^3TM$ 能够比较容易被泛化到其他社会多媒体研究的跨平台的推荐问题中。实验评估展示了我们方法在连接不同平台的有效性。我们未来的工作主要包括两个方面1) 基于我们提出的主题模型，我们设计更多的应用，比如平台间的朋友推荐等。2) 我们通过引入社会关系 [124] [119]来扩展我们的模型，来自不同平台的内容信息和社会关系是互补的因而可以增强模型的性能。

## 第七章 结束语

Web2.0技术的迅速发展和手持移动设备的广泛应用使得用户在使用各类社会媒体的同时以各种各样的方式分享自己的地理位置。大量丰富的带有地理位置的社会媒体为我们基于地理位置的社会媒体研究提供了许多机会。本文从社会媒体下地理数据处理和应用两个方面进行了研究和探讨，具体取得的工作进展如下：

- 提出了基于场景角度主题模型的地标总结方法。（1）该模型同时考虑图片的拍摄角度和场景进行地标总结；（2）基于拍摄角度的图片聚类集合，该模型能够区分和学习来自不同角度聚类共享的场景主题子空间以及同一场景下不同角度所特有的场景角度主题子空间，通过学习的两类子空间获得主题的场景角度主题的图片表示。
- 提出了基于多模态时空主题模型的地标分析方法。（1）该模型通过学习地标的三类主题，即全局主题、位置主题和时间主题的子空间发现地标的三类主题；（2）基于地标和位置主题的关联，地标-时间和时间主题的关联引入了基于互信息的正则化的优化目标函数；（3）利用贝叶斯定理从地标的时间和空间两个方面对挖掘的主题进行了分析。
- 提出了基于鲁棒的高阶联合聚类的社会事件检测方法。（1）构建星结构的K分图表示该社会媒体本身和位置信息、时间信息、文本信息和视觉信息之间的依附关系，通过基于一致信息(CIT)论的理论建模这些关系；（2）该方法提出了基于时间空间内时间之间的关系的正则化的优化目标函数；（3）采用启发式的方法去除或者融合聚类中的样本。
- 提出了基于多模态跨平台主题模型的个性化跨平台推荐框架。（1）提出的模型能够区分平台共享的主题和每个平台特有的主题（2）该模型能够对齐每一个主题来自不同平台所有的模态；（3）基于共享的主题以及对应的多模态表示，设计了一个基于位置情景的跨平台推荐应用。

虽然我们对社会媒体下地理数据处理和应用两个方面的研究取得一定的成果，但是在当前社会媒体数据爆炸的时代下需要我们亟待解决和深入研究不

断出现的新问题。展望未来的工作，我认为可以从以下两个方面继续探索和研究：

- 社交媒体下地理大数据分析。社交媒体下的地理大数据分析是当前和未来比较重要的课题，地理大数据分析需要有相应的新的学习方法来处理。现有的新的深度学习技术具有强大的表示学习能力。借鉴深度学习的思想如何将我们现有的方法扩展成相应的深度学习模型是一个值得深入研究的课题。
- 大规模社交媒体地理位置估计。地理位置估计有利于社交媒体数据的总结、组织和检索。尽管现有的社交媒体鼓励用户用地理位置标注他们分享的媒体，但是现实中带有地理标签的媒体数据相比于总的分享的社会媒体数据占有的比例极小（3%<sup>1</sup>）。因此我们需要通过利用现有和地理位置关联的媒体数据训练模型以便能够标注大量的没有地理标签的媒体。但是社交媒体数据模态多样、数量巨大，并且不同位置的数据分布的有偏性，含有大量噪音 [27]，这为我们训练模型提出了挑战。因此如何设计方法应对这些挑战实现大规模社交媒体地理位置预测是一个值得深入研究的课题。
- 基于地理-社会媒体的用户建模。基于GPS的移动设备的广泛应用使得地理位置获取的技术和服务得到迅速发展，这些位置服务和社会网络结合产生了地理-社交媒体。在地理-社交媒体中，我们的工作主要讨论了基于地理位置的用户分享的内容分析技术，除此之外，社会关系也是地理社交媒体分析一个重要的元素。但是社会关系异质复杂，尤其是在跨平台条件下，不同平台的社会关系如何有效的融合是解决用户建模的重要问题之一。因此如何综合考虑用户关系、用户分享的内容和情景信息进行用户建模，比如用户行为分析，地理社会媒体的链接预测等也是未来研究的重要方向。

---

<sup>1</sup><https://www.flickr.com/map>

## 参考文献

- [1] ABEL, F., ARAÚJO, S., GAO, Q., AND HOUBEN, G.-J. Analyzing cross-system user modeling on the social web. In *Web Engineering*. Springer, 2011, pp. 28–43.
- [2] ALLAN, J. *Topic detection and tracking: event-based information organization*, vol. 12. Springer, 2002.
- [3] ALSUMAIT, L., BARBARÁ, D., AND DOMENICONI, C. On-line lda: Adaptive topic models for mining text streams with applications to topic detection and tracking. In *Data Mining. IEEE International Conference on* (2008), IEEE, pp. 3–12.
- [4] AVRITHIS, Y., KALANTIDIS, Y., TOLIAS, G., AND SPYROU, E. Retrieving landmark and non-landmark images from community photo collections. In *Proceedings of the international conference on Multimedia* (2010), ACM, pp. 153–162.
- [5] AVRITHIS, Y., KALANTIDIS, Y., TOLIAS, G., AND SPYROU, E. Retrieving landmark and non-landmark images from community photo collections. In *Proceedings of the international conference on Multimedia* (2010), ACM, pp. 153–162.
- [6] BANERJEE, A., DHILLON, I., GHOSH, J., MERUGU, S., AND MODHA, D. A generalized maximum entropy approach to bregman co-clustering and matrix approximation. *Journal of Machine Learning Research* 8 (2007), 1919–1986.
- [7] BAO, J., ZHENG, Y., AND MOKBEL, M. F. Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems* (2012), ACM, pp. 199–208.

- [8] BAO, Y., COLLIER, N., AND DATTA, A. A partially supervised cross-collection topic model for cross-domain text classification. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management* (2013), ACM, pp. 239–248.
- [9] BECKER, H., NAAMAN, M., AND GRAVANO, L. Event identification in social media. In *Proceedings of the ACM SIGMOD Workshop on the Web and Databases* (2009).
- [10] BECKER, H., NAAMAN, M., AND GRAVANO, L. Learning similarity metrics for event identification in social media. In *Proceedings of the third ACM international conference on Web search and data mining* (2010), ACM, pp. 291–300.
- [11] BELKIN, M., AND NIYOGI, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15, 6 (2003), 1373–1396.
- [12] BLEI, D., CARIN, L., AND DUNSON, D. Probabilistic topic models. *Signal Processing Magazine, IEEE* 27, 6 (2010), 55–65.
- [13] BLEI, D. M., AND JORDAN, M. I. Modeling annotated data. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (2003), ACM, pp. 127–134.
- [14] BLEI, D. M., NG, A. Y., AND JORDAN, M. I. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.
- [15] BRANTS, T. H., TSOCHANTARIDIS, I., HOFMANN, T., AND CHEN, F. R. Methods, apparatus, and program products for performing incremental probabilistic latent semantic analysis, May 5 2009. US Patent 7,529,765.
- [16] BRENNER, M., AND IZQUIERDO, E. Mediaeval benchmark: Social event detection in collaborative photo collections. In *MediaEval* (2011).
- [17] BRENNER, M., AND IZQUIERDO, E. QMUL@ MediaEval 2012: Social event detection in collaborative photo collections. In *Proceedings of the MediaEval 2012 Workshop* (2012).

- [18] CAI, D., MEI, Q., HAN, J., AND ZHAI, C. Modeling hidden topics on document manifold. In *Proceedings of the 17th ACM conference on Information and knowledge management* (2008), ACM, pp. 911–920.
- [19] CANINI, K. R., SHI, L., AND GRIFFITHS, T. L. Online inference of topics with latent dirichlet allocation. In *International conference on artificial intelligence and statistics* (2009), pp. 65–72.
- [20] CHARRAD, M., AND AHMED, M. B. Simultaneous clustering: A survey. In *Pattern Recognition and Machine Intelligence*. Springer, 2011, pp. 370–375.
- [21] CHEN, D. M., BAATZ, G., KOSER, K., TSAI, S. S., VEDANTHAM, R., PYLVANAINEN, T., ROIMELA, K., CHEN, X., BACH, J., POLLEFEYS, M., ET AL. City-scale landmark identification on mobile devices. In *Computer Vision and Pattern Recognition, 2011. IEEE Conference on* (2011), IEEE, pp. 737–744.
- [22] CHEN, L., AND ROY, A. Event detection from flickr data through wavelet-based spatial analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management* (2009), ACM, pp. 523–532.
- [23] CHEN, L., AND ROY, A. Event detection from flickr data through wavelet-based spatial analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management* (2009), ACM, pp. 523–532.
- [24] CHEN, W.-C., BATTESTINI, A., GELFAND, N., AND SETLUR, V. Visual summaries of popular landmarks from community photo collections. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on* (2009), IEEE, pp. 1248–1255.
- [25] CHEN, X., LU, C., AN, Y., AND ACHANANUPARP, P. Probabilistic models for topic learning from images and captions in online biomedical literatures. In *Proceedings of the 18th ACM conference on Information and knowledge management* (2009), ACM, pp. 495–504.

- [26] CHIEN, J.-T., AND WU, M.-S. Adaptive bayesian latent semantic analysis. *Audio, Speech, and Language Processing, IEEE Transactions on* 16, 1 (2008), 198–207.
- [27] CHOI, J., THOME, B., FRIEDLAND, G., CAO, L., NI, K., BORTH, D., ELIZALDE, B., GOTTLIEB, L., CARRANO, C., PEARCE, R., ET AL. The placing task: A large-scale geo-estimation challenge for social-media videos and images. In *Proceedings of the 3rd ACM Multimedia Workshop on Geotagging and Its Applications in Multimedia* (2014), ACM, pp. 27–31.
- [28] CHOU, T.-C., AND CHEN, M. C. Using incremental plsi for threshold-resilient online event analysis. *Knowledge and Data Engineering, IEEE Transactions on* 20, 3 (2008), 289–299.
- [29] CRANDALL, D., BACKSTROM, L., HUTTENLOCHER, D., AND KLEINBERG, J. Mapping the world’s photos. In *Proceedings of the 18th international conference on World wide web* (2009), ACM, pp. 761–770.
- [30] DALAL, N., AND TRIGGS, B. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on* (2005), vol. 1, IEEE, pp. 886–893.
- [31] DENG, Z., SANG, J., AND XU, C. Personalized video recommendation based on cross-platform user modeling. In *Multimedia and Expo, International Conference on* (2013), IEEE, pp. 1–6.
- [32] DEODHAR, M., GUPTA, G., GHOSH, J., CHO, H., AND DHILLON, I. A scalable framework for discovering coherent co-clusters in noisy data. In *Proceedings of the 26th International Conference on Machine Learning* (2009), pp. 241–248.
- [33] DHILLON, I. S. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining* (2001), pp. 269–274.



- [34] DHILLON, I. S., MALLELA, S., AND MODHA, D. S. Information-theoretic co-clustering. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining* (2003).
- [35] E GOHR, A., HINNEBURG, A., E SCHULT, R., AND SPILIOPOULOU, M. Topic evolution in a stream of documents, 2009.
- [36] FAN, J., KEIM, D. A., GAO, Y., LUO, H., AND LI, Z. Justclick: Personalized image recommendation via exploratory search from large-scale flickr images. *Circuits and Systems for Video Technology, IEEE Transactions on* 19, 2 (2009), 273–288.
- [37] FAN, R.-E., CHANG, K.-W., HSIEH, C.-J., WANG, X.-R., AND LIN, C.-J. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research* 9 (2008), 1871–1874.
- [38] FERNÁNDEZ-TOBÍAS, I., CANTADOR, I., KAMINSKAS, M., AND RICCI, F. Cross-domain recommender systems: A survey of the state of the art. In *Spanish Conference on Information Retrieval* (2012).
- [39] FIRAN, C., GEORGESCU, M., NEJDL, W., AND PAIU, R. Bringing order to your photos: event-driven classification of flickr images based on social knowledge. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (2010), pp. 189–198.
- [40] GAO, B., LIU, T., ZHENG, X., CHENG, Q., AND MA, W. Consistent bipartite graph co-partitioning for star-structured high-order heterogeneous data co-clustering. In *Proceedings of the 8th International Conference on Knowledge Discovery in Data Mining* (2005).
- [41] GAO, B., LIU, T. Y., AND MA, W. Y. Star-structured high-order heterogeneous data co-clustering based on consistent information theory. In *Proceedings of the International Conference on Data Mining* (2006), pp. 880–884.

- [42] GAO, H., TANG, S., ZHANG, Y., JIANG, D., WU, F., AND ZHUANG, Y. Supervised cross-collection topic modeling. In *Proceedings of the 20th ACM international conference on Multimedia* (2012), ACM, pp. 957–960.
- [43] GAO, Y., TANG, J., HONG, R., DAI, Q., CHUA, T., AND JAIN, R. W2go: a travel guidance system by automatic landmark ranking. In *Proceedings of the international conference on Multimedia* (2010), ACM, pp. 123–132.
- [44] GAVVES, E., AND SNOEK, C. G. Landmark image retrieval using visual synonyms. In *Proceedings of the international conference on Multimedia* (2010), ACM, pp. 1123–1126.
- [45] GAVVES, E., SNOEK, C. G., AND SMEULDERS, A. W. Visual synonyms for landmark image retrieval. *Computer Vision and Image Understanding* 116, 2 (2012), 238–249.
- [46] GRAHAM, A., GARCIA-MOLINA, H., PAEPCKE, A., AND WINOGRAD, T. Time as essence for photo browsing through personal digital libraries. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries* (2002), ACM, pp. 326–335.
- [47] HAO, Q., CAI, R., LI, Z., ZHANG, L., PANG, Y., AND WU, F. 3d visual phrases for landmark recognition. In *Computer Vision and Pattern Recognition, 2012. IEEE Conference on* (2012), IEEE, pp. 3594–3601.
- [48] HEINRICH, G. Parameter estimation for text analysis. Tech. rep., Technical report, 2005.
- [49] HOFFMAN, M., BACH, F. R., AND BLEI, D. M. Online learning for latent dirichlet allocation. In *advances in neural information processing systems* (2010), pp. 856–864.
- [50] HOFFMAN, M. D., BLEI, D. M., AND COOK, P. R. Content-based musical similarity computation using the hierarchical dirichlet process. In *ISMIR* (2008), pp. 349–354.

- [51] HOFMANN, T. Unsupervised learning by probabilistic latent semantic analysis. *Machine learning* 42, 1-2 (2001), 177–196.
- [52] HONG, L., AHMED, A., GURUMURTHY, S., SMOLA, A. J., AND TSIOUTSIOLIKLIS, K. Discovering geographical topics in the twitter stream. In *Proceedings of the 21st international conference on World Wide Web* (2012), ACM, pp. 769–778.
- [53] HU, B., JAMALI, M., AND ESTER, M. Spatio-temporal topic modeling in mobile social media for location recommendation. In *Data Mining, 13th International Conference on* (2013), IEEE, pp. 1073–1078.
- [54] JI, R., DUAN, L.-Y., CHEN, J., YAO, H., YUAN, J., RUI, Y., AND GAO, W. Location discriminative vocabulary coding for mobile landmark search. *International Journal of Computer Vision* 96, 3 (2012), 290–314.
- [55] JI, R., GAO, Y., ZHONG, B., YAO, H., AND TIAN, Q. Mining flickr landmarks by modeling reconstruction sparsity. *ACM Transactions on Multimedia Computing, Communications, and Applications* 7, 1 (2011), 31.
- [56] JI, R., XIE, X., YAO, H., AND MA, W. Mining city landmarks from blogs by graph modeling. In *Proceedings of the 17th ACM international conference on Multimedia* (2009), ACM, pp. 105–114.
- [57] JIA, Y., SALZMANN, M., AND DARRELL, T. Learning cross-modality similarity for multinomial data. In *Computer Vision, IEEE International Conference on* (2011), IEEE, pp. 2407–2414.
- [58] JIANG, M., CUI, P., LIU, R., YANG, Q., WANG, F., ZHU, W., AND YANG, S. Social contextual recommendation. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (2012), ACM, pp. 45–54.
- [59] JOSEPH, K., TAN, C. H., AND CARLEY, K. M. Beyond local, categories and friends: clustering foursquare users with latent topics. In *Proceedings of ACM Conference on Ubiquitous Computing* (2012), ACM, pp. 919–926.

- [60] KENNEDY, L., AND NAAMAN, M. Generating diverse and representative image search results for landmarks. In *Proceeding of the 17th international conference on World Wide Web* (2008), ACM, pp. 297–306.
- [61] KENNEDY, L., NAAMAN, M., AHERN, S., NAIR, R., AND RATTENBURY, T. How flickr helps us make sense of the world: context and content in community-contributed media collections. In *Proceedings of the 15th international conference on Multimedia* (2007), ACM, pp. 631–640.
- [62] KLING, C. C., KUNEGIS, J., SIZOV, S., AND STAAB, S. Detecting non-gaussian geographical topics in tagged photo collections. In *Proceedings of the 7th ACM international conference on Web search and data mining* (2014), ACM, pp. 603–612.
- [63] LEE, Y., AND K. GRAUMAN. Object-graphs for context-aware category discovery. In *IEEE Conference on Computer Vision and Pattern Recognition* (2010), pp. 1–8.
- [64] LI, B. Cross-domain collaborative filtering: A brief survey. In *Tools with Artificial Intelligence, 2011 IEEE International Conference on* (2011), IEEE, pp. 1085–1086.
- [65] LI, L., JIN, X., AND LONG, M. Topic correlation analysis for cross-domain text classification. In *AAAI* (2012).
- [66] LI, X., WU, C., ZACH, C., LAZEBNIK, S., AND FRAHM, J.-M. Modeling and recognition of landmark image collections using iconic scene graphs. In *Computer Vision–ECCV 2008*. Springer, 2008, pp. 427–440.
- [67] LI, Y., CRANDALL, D., AND HUTTENLOCHER, D. Landmark classification in large-scale image collections. In *Computer Vision, 2009. 12th International Conference on* (2009), IEEE, pp. 1957–1964.
- [68] LIU, J., HUANG, Z., CHEN, L., SHEN, H. T., AND YAN, Z. Discovering areas of interest with geo-tagged images and check-ins. In *Proceedings of the 20th ACM international conference on Multimedia* (2012), ACM, pp. 589–598.

- [69] LIU, X., HUET, B., AND TRONCY, R. Eurecom@ mediaeval 2011 social event detection task. In *MediaEval* (2011).
- [70] LONG, B., WU, X., ZHANG, Z., AND YU, P. Unsupervised learning on k-partite graphs. In *Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining* (2006), pp. 317–326.
- [71] LU, X., WANG, C., YANG, J.-M., PANG, Y., AND ZHANG, L. Photo2trip: generating travel routes from geo-tagged photos for trip planning. In *Proceedings of the international conference on Multimedia* (2010), ACM, pp. 143–152.
- [72] LUO, J., JOSHI, D., YU, J., AND GALLAGHER, A. Geotagging in multimedia and computer vision — a survey. *Multimedia Tools and Applications* 51, 1 (2011), 187–211.
- [73] MAKKONEN, J., AHONEN-MYKA, H., AND SALMENKIVI, M. Simple semantics in topic detection and tracking. *Information Retrieval* 7, 3 (2004), 347–368.
- [74] MEI, Q., LIU, C., SU, H., AND ZHAI, C. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In *Proceedings of the 15th international conference on World Wide Web* (2006), ACM, pp. 533–542.
- [75] MIKOLAJCZYK, K., AND SCHMID, C. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27, 10 (2005), 1615–1630.
- [76] MILLER, G., AND FELLBAUM, C. Wordnet: An electronic lexical database, 1998.
- [77] NEAL, R. M., AND HINTON, G. E. A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*. Springer, 1998, pp. 355–368.
- [78] OLIVA, A., AND TORRALBA, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision* 42, 3 (2001), 145–175.

- [79] PAN, W., LIU, N. N., XIANG, E. W., AND YANG, Q. Transfer learning to predict missing ratings via heterogeneous user feedbacks. In *Proceedings-International Joint Conference on Artificial Intelligence* (2011), vol. 22, p. 2318.
- [80] PAN, W., XIANG, E. W., LIU, N. N., AND YANG, Q. Transfer learning in collaborative filtering for sparsity reduction. In *AAAI* (2010), vol. 10, pp. 230–235.
- [81] PANG, Y., HAO, Q., YUAN, Y., HU, T., CAI, R., AND ZHANG, L. Summarizing tourist destinations by mining user-generated travelogues and photos. *Computer Vision and Image Understanding* 115, 3 (2011), 352–363.
- [82] PAPADOPOULOS, S., SCHINAS, E., MEZARIS, V., TRONCY, R., AND KOMPATSIARIS, I. Social event detection at MediaEval 2012: Challenges, dataset and evaluation. In *Proceedings of the MediaEval 2012 Workshop* (Pisa, ITALY, 2012).
- [83] PAPADOPOULOS, S., TRONCY, R., MEZARIS, V., HUET, B., AND KOMPATSIARIS, I. Social event detection at MediaEval 2011: Challenges, dataset and evaluation. In *Proceedings of the MediaEval 2012 Workshop* (Pisa, ITALY, 2011).
- [84] PAPADOPOULOS, S., ZIGKOLIS, C., KOMPATSIARIS, Y., AND VAKALI, A. Cluster-based landmark and event detection on tagged photo collections. *Multimedia, IEEE*, 99 (2010), 1–1.
- [85] PAUL, M., AND GIRJU, R. Cross-cultural analysis of blogs and forums with mixed-collection topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3* (2009), Association for Computational Linguistics, pp. 1408–1417.
- [86] PETKOS, G., PAPADOPOULOS, S., AND KOMPATSIARIS, Y. Social event detection using multimodal clustering and integrating supervisory signals. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval* (2012), ACM, pp. 23–30.

- [87] PUTTHIVIDHY, D., ATTIAS, H. T., AND NAGARAJAN, S. S. Topic regression multi-modal latent dirichlet allocation for image annotation. In *Computer Vision and Pattern Recognition, IEEE International Conference on* (2010), IEEE, pp. 3408–3415.
- [88] QI, G., AGGARWAL, C., AND HUANG, T. Breaking the barrier to transferring link information across networks.
- [89] QI, G.-J., AGGARWAL, C. C., AND HUANG, T. Link prediction across networks by biased cross-network sampling. In *Data Engineering, IEEE 29th International Conference on* (2013), IEEE, pp. 793–804.
- [90] QUACK, T., LEIBE, B., AND VAN GOOL, L. World-scale mining of objects and events from community photo collections. In *Proceedings of the 2008 international conference on Content-based image and video retrieval* (2008), ACM, pp. 47–56.
- [91] RAGURAM, R., WU, C., FRAHM, J.-M., AND LAZEBNIK, S. Modeling and recognition of landmark image collections using iconic scene graphs. *International journal of computer vision* 95, 3 (2011), 213–239.
- [92] RATTENBURY, T., GOOD, N., AND NAAMAN, M. Towards automatic extraction of event and place semantics from flickr tags. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (2007), ACM, pp. 103–110.
- [93] RATTENBURY, T., GOOD, N., AND NAAMAN, M. Towards extracting flickr tag semantics. In *Proceedings of the 16th international conference on World Wide Web* (2007), ACM, pp. 1287–1288.
- [94] REUTER, T., AND CIMIANO, P. Event-based classification of social media streams. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval* (2012), ACM, p. 22.
- [95] ROY, S. D., MEI, T., ZENG, W., AND LI, S. Socialtransfer: cross-domain transfer learning from social streams for media applications. In *Proceedings of*

- the 20th ACM international conference on Multimedia* (2012), ACM, pp. 649–658.
- [96] RUDINAC, S., HANJALIC, A., AND LARSON, M. Generating visual summaries of geographic areas using community-contributed images. *IEEE Transactions on Multimedia* 15, 4 (2013), 921–932.
- [97] SANG, J., AND XU, C. Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications. In *Proceedings of the 20th ACM international conference on Multimedia* (2012), ACM, pp. 19–28.
- [98] SANG, J., AND XU, C. Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications. In *Proceedings of the 20th ACM international conference on Multimedia* (2012), ACM, pp. 19–28.
- [99] SCHINAS, E., PETKOS, G., PAPADOPOULOS, S., AND KOMPATSIARIS, Y. CERTH@ MediaEval 2012 social event detection task. In *Proceedings of the MediaEval 2012 Workshop* (2012).
- [100] SHAO, J., YIN, W., MA, S., AND ZHUANG, Y. Topic discovery of web video using star-structured k-partite graph. In *Proceedings of the International Conference on Multimedia* (2010), pp. 915–918.
- [101] SHI, J., AND MALIK, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 8 (2000), 888–905.
- [102] SIMON, I., SNAVELY, N., AND SEITZ, S. M. Scene summarization for online image collections. In *Computer Vision, 2007. 11th International Conference on* (2007), IEEE, pp. 1–8.
- [103] SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics* (2006), vol. 25, ACM, pp. 835–846.



- [104] STREHL, A., AND GHOSH, J. Cluster ensembles – a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research* 3 (2003), 583–617.
- [105] TAN, S., BU, J., QIN, X., CHEN, C., AND CAI, D. Cross domain recommendation based on multi-type media fusion. *Neurocomputing* 127 (2014), 124–134.
- [106] TANG, J., WU, S., SUN, J., AND SU, H. Cross-domain collaboration recommendation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (2012), ACM, pp. 1285–1293.
- [107] VERMA, D., AND MEILA, M. A comparison of spectral clustering algorithms.
- [108] VIRTANEN, S., JIA, Y., KLAMI, A., AND DARRELL, T. Factorized multi-modal topic model. *arXiv preprint arXiv:1210.4920* (2012).
- [109] WANG, C., AND BLEI, D. M. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (2011), ACM, pp. 448–456.
- [110] WANG, C., WANG, J., XIE, X., AND MA, W.-Y. Mining geographic knowledge using location aware topic model. In *Proceedings of the 4th ACM workshop on Geographical information retrieval* (2007), ACM, pp. 65–70.
- [111] WANG, Y., XIE, L., AND SUNDARAM, H. Social event detection with clustering and filtering. In *MediaEval* (2011), Citeseer.
- [112] XIAO, X., XU, C., AND WANG, J. Landmark image classification using 3d point clouds. In *Proceedings of the international conference on Multimedia* (2010), ACM, pp. 719–722.
- [113] XIAO, X., XU, C., WANG, J., AND XU, M. Landmark recognition and retrieval: from 2d to 3d. In *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding* (2011), ACM, pp. 77–78.

- [114] XIAO, X., XU, C., WANG, J., AND XU, M. Enhanced 3-d modeling for landmark image classification. *Multimedia, IEEE Transactions on* 14, 4 (2012), 1246–1258.
- [115] XUE, Y., AND QIAN, X. Visual summarization of landmarks via viewpoint modeling. In *Image Processing, 2012. 19th IEEE International Conference on* (2012), IEEE, pp. 2873–2876.
- [116] YAN, M., SANG, J., MEI, T., AND XU, C. Friend transfer: cold-start friend recommendation with cross-platform transfer learning of social knowledge. In *Multimedia and Expo, IEEE International Conference on* (2013), IEEE, pp. 1–6.
- [117] YAN, M., SANG, J., AND XU, C. Mining cross-network association for youtube video promotion. In *Proceedings of the ACM International Conference on Multimedia* (2014), ACM, pp. 557–566.
- [118] YE, M., YIN, P., LEE, W.-C., AND LEE, D.-L. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (2011), ACM, pp. 325–334.
- [119] YE, M., YIN, P., LEE, W.-C., AND LEE, D.-L. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (2011), ACM, pp. 325–334.
- [120] YIN, H., SUN, Y., CUI, B., HU, Z., AND CHEN, L. Lcars: A location-content-aware recommender system. In *Proceedings of the 19th ACM international conference on Knowledge discovery and data mining* (2013), ACM, pp. 221–229.
- [121] YIN, Z., CAO, L., HAN, J., ZHAI, C., AND HUANG, T. Geographical topic discovery and comparison. In *Proceedings of the 20th international conference on World wide web* (2011), ACM, pp. 247–256.

- [122] YUAN, J., LUO, J., KAUTZ, H., AND WU, Y. Mining gps traces and visual words for event classification. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval* (2008), ACM, pp. 2–9.
- [123] YUAN, J., LUO, J., AND WU, Y. Mining compositional features from gps and visual cues for event recognition in photo collections. *Multimedia, IEEE Transactions on* 12, 7 (2010), 705–716.
- [124] YUAN, Q., CHEN, L., AND ZHAO, S. Factorization vs. regularization: fusing heterogeneous social relationships in top-n recommendation. In *Proceedings of the fifth ACM conference on Recommender systems* (2011), ACM, pp. 245–252.
- [125] YUAN, Q., CONG, G., AND SUN, A. Graph-based point-of-interest recommendation with geographical and temporal influences. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (2014), ACM, pp. 659–668.
- [126] ZAHÁLKA, J., RUDINAC, S., AND WORRING, M. New yorker melange: Interactive brew of personalized venue recommendations. In *Proceedings of the ACM International Conference on Multimedia* (2014), ACM, pp. 205–208.
- [127] ZAHARIEVA, M., ZEPPELZAUER, M., AND BREITENEDER, C. Automated social event detection in large photo collections. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval* (2013), ACM, pp. 167–174.
- [128] ZEPPELZAUER, M., ZAHARIEVA, M., AND BREITENEDER, C. A generic approach for social event detection in large photo collections. In *Proceedings of the MediaEval 2012 Workshop* (2012).
- [129] ZHAI, C., VELIVELLI, A., AND YU, B. A cross-collection mixture model for comparative text mining. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (2004), ACM, pp. 743–748.

- [130] ZHANG, F., YUAN, N. J., LIAN, D., AND XIE, X. Mining novelty-seeking trait across heterogeneous domains. In *Proceedings of the 23rd international conference on World wide web* (2014), International World Wide Web Conferences Steering Committee, pp. 373–384.
- [131] ZHANG, J., KONG, X., AND YU, P. S. Predicting social links for new users across aligned heterogeneous social networks. In *Data Mining, IEEE 13th International Conference on* (2013), IEEE, pp. 1289–1294.
- [132] ZHANG, J., KONG, X., AND YU, P. S. Transferring heterogeneous links across location-based social networks. In *Proceedings of the 7th ACM international conference on Web search and data mining* (2014), ACM, pp. 303–312.
- [133] ZHANG, Y., CAO, B., AND YEUNG, D.-Y. Multi-domain collaborative filtering. *arXiv preprint arXiv:1203.3535* (2012).
- [134] ZHAO, Y.-L., CHEN, Q., YAN, S., CHUA, T.-S., AND ZHANG, D. Detecting profilable and overlapping communities with user-generated multimedia contents in lbsns. *ACM Transactions on Multimedia Computing, Communications, and Applications* 10, 1 (2013), 3.
- [135] ZHAO, Y.-L., NIE, L., WANG, X., AND CHUA, T.-S. Personalized recommendations of locally interesting venues to tourists via cross-region community matching. *ACM Transactions on Intelligent Systems and Technology* 5, 3 (2014), 50.
- [136] ZHAO, Y.-L., ZHENG, Y.-T., ZHOU, X., AND CHUA, T.-S. Generating representative views of landmarks via scenic theme detection. In *Advances in Multimedia Modeling*. Springer, 2011, pp. 392–402.
- [137] ZHENG, V. W., ZHENG, Y., XIE, X., AND YANG, Q. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th international conference on World wide web* (2010), ACM, pp. 1029–1038.
- [138] ZHENG, Y., ZHAO, M., SONG, Y., ADAM, H., BUDDEMEIER, U., BISSACCO, A., BRUCHER, F., CHUA, T., AND NEVEN, H. Tour the world: build-

- ing a web-scale landmark recognition engine. In *Computer Vision and Pattern Recognition, 2009. IEEE Conference on* (2009), IEEE, pp. 1085–1092.
- [139] ZHENG, Y.-T., ZHA, Z.-J., AND CHUA, T.-S. Research and applications on georeferenced multimedia: a survey. *Multimedia Tools and Applications* 51, 1 (2011), 77–98.
- [140] ZHONG, E., FAN, W., WANG, J., XIAO, L., AND LI, Y. Comsoc: adaptive transfer of user behaviors over composite social network. In *Proceedings of the 18th ACM international conference on Knowledge discovery and data mining* (2012), ACM, pp. 696–704.
- [141] ZHONG, E., XIANG, E. W., FAN, W., LIU, N. N., AND YANG, Q. Friendship prediction in composite social networks. *arXiv preprint arXiv:1402.4033* (2014).
- [142] ZHUANG, J., MEI, T., HOI, S. C., HUA, X.-S., AND LI, S. Modeling social strength in social media community via kernel-based learning. In *Proceedings of the ACM international conference on Multimedia* (2011), pp. 113–122.



## 发表文章目录

- [1] **Weiqing Min**, Changsheng Xu, Min Xu, Xian Xiao, Bing-kun Bao. Mobile Landmark Search with 3D Models. IEEE Trans. on Multimedia (TMM) 16(3):623-636(2014)
- [2] **Weiqing Min**, Bing-kun Bao, Changsheng Xu. Multi-modal Spatio-Temporal Theme Modeling for Landmark Analysis. IEEE Multimedia (IEEE MM). 21(3): 20-29 (2014)
- [3] **Weiqing Min**, Bing-kun Bao, Changsheng Xu. An Incremental Probabilistic Model for Temporal Theme Analysis of Landmarks. Multimedia Systems Journal . Accepted.
- [4] **Weiqing Min**, Bing-Kun Bao, Changsheng Xu :Scene and Viewpoint Based Visual Summarization for Landmarks. ICIIP 2014. : 3112 - 3116
- [5] **Weiqing Min**, Bing-Kun Bao, Changsheng Xu :Landmark History Visualization. MMM 2013: 121-132.
- [6] **Weiqing Min**, Bing-Kun Bao, Changsheng Xu :What Happened Near Big Ben: Event-Driven Landmark Mining from Flickr. PCM 2012: 769-778
- [7] Bing-Kun Bao, **Weiqing Min**, Jitao Sang, Changsheng Xu :Multimedia news digger on emerging topics from social streams. ACM Multimedia (ACM MM) 2012: 1357-1358
- [8] Bing-Kun Bao, **Weiqing Min**, Ke Lu, Changsheng Xu :Social event detection with robust high-order co-clustering. ICMR 2013: 135-142
- [9] Bing-kun Bao, **Weiqing Min**, Teng Li, Changsheng Xu. Joint Local and Global Consistency on Inter-document and Inter-word Relationships for Co-clustering. IEEE Trans. Cybernetics. 45(1): 15-28 (2015).



- 
- [10] Bing-Kun Bao, Changsheng Xu, **Weiqing Min** , M. Shamim Hossain: Cross-Platform Emerging Topic Detection and Elaboration from Multimedia Stream. ACM Transactions on Multimedia Computing Communications and Applications. (TOMM) 2015. Accepted.
  - [11] Chao Liang, Changsheng Xu, Jian Cheng, **Weiqing Min**, Hanqing Lu: Script-to-Movie: A Computational Framework for Story Movie Composition. IEEE Trans. Multimedia (TMM), 15(2): 401-414 (2013)
  - [12] **Weiqing Min**, Bing-kun Bao, Changsheng Xu. Cross-Platform Multi-Modal Topic Modeling for Personalized Inter-Platform Recommendation. IEEE Trans. on Multimedia (TMM) 2015.**Under Review**

## 致 谢

时光荏苒，岁月如梭，转眼间五年的博士生涯即将画上句号。在这五年的博士生涯中，自己在科研上取得了进步，在生活中获得了成长。但这些点点滴滴的成长离不开老师、同学、朋友和家人的鼓励、帮助和指导。藉此论文完成之际，谨借此文向你们致以最诚挚的感谢。

首先要衷心感谢我的导师徐常胜研究员。徐老师带我们进入多媒体挖掘和检索的研究领域，并不遗余力的为我们组创造宽松的科研环境，对我们科研要求严格，同时鼓励我们发挥自己在科研方面的主观能动性，努力让我们取得科研的自由。徐老师严谨的治学态度和谦和的处事风范，是我终身学习的榜样。

感谢鲍秉坤老师给我科研方面的指导和帮助。鲍老师在科研中对我的悉心指导和严格要求让我能够突破科研的瓶颈，比较顺利进入科研。同时鲍老师学术上的造诣和治学精神经过和她不断的讨论和交流，逐渐感染了我，让我继续沿着正确的科研道路前进。

感谢桑基韬老师和张天柱老师在科研方面的指导和帮助，和各位老师在对研究内容和研究方法的讨论对我影响很大。

感谢卢汉清研究员、程健研究员、刘静副研究员、王金桥副研究员和张一帆副研究员对我工作上的帮助和学术上的启发，令我受益匪浅。

感谢龚怡宏教授在我实习期间给我的帮助，他的宝贵意见和建议给予了我莫大的帮助。

感谢模式识别国家重点实验室综合办的连国臻、赵薇、马丽霞和姜文利老师多年来在科研和生活中给我们提供的无微不至的关心和帮助。感谢研究生部和科技处的邸凌、李磊、胡蓉、曹娟和鞠召艳老师在我学习和生活中提供的支持和帮助。

感谢在MMC组一起工作学习的同学和同事：肖宪、郭文、应龙、俞雷、孙超、张歆明、邓拯宇、杨小汕、方全、袁召全、严明、钱胜胜、邱泽宇和项连城等。感谢我们MMC组的秘书孙金艳在科研和工作中给我的帮助。感谢在IVA组一起工作学习的同学和同事：梁超、张淳杰、王博、刘晶晶、李泽超、李鹏、蒋瑜、钮彪和付伟等，和他们学术上的讨论和知识的分享让我体验到学

习和进步的快乐。

感谢10级博士班的所有同学给予我的帮助，感谢我的舍友戴龙泉给予我科研上的帮助，和他学术上的讨论让我的科研有了更进一步的提高。

感谢我的女朋友韩亚冬在科研上给我的鼎力支持，能够让我全身心的投入到科研当中。

感谢我的弟弟闵超庆和弟妹赵字蕾，他们无时无刻不在鼓励和支持我。

最后感谢我的父母，他们含辛茹苦的养育我，默默无闻的鼓励我，他们是我前进的不竭动力。

闵巍庆

2015年5月于北京