

## Flickr 中的复合事件检测

周春姐 孟小峰 文洁

(中国人民大学信息学院 北京 100872)

(lucyzcj@ruc.edu.cn)

## Complex Event Detection on Flickr

Zhou Chunjie, Meng Xiaofeng, and Wen Jie

(School of Information, Renmin University of China, Beijing 100872)

**Abstract** Detecting events from Web resources has attracted increasing research interests in recent years. Flickr is one of Web resources, which is used to share photos. Complex event detection on Flickr includes the detection of tourist features, user's interest, and so on. We can provide more personalized and more intelligent services with complex event detection. In this work, we propose a method to detect tourist features of every scene and its difference among different seasons as a probabilistic combination of tags. The use of topic models enables the automatic detection of such patterns, which can translate unstructured tag information into structured event form. We evaluate the performance of our approach on a real dataset collected from Flickr. The experimental results show the feasibility and effectiveness of our approach to detect tourist features without user annotation.

**Key words** Flickr; complex event; topic model; tag

**摘要** Web资源上的复合事件检测已成为人们关注和研究的热点。Flickr作为一个照片共享的Web资源,其中的复合事件检测包括景点特色检测、用户兴趣检测等,从而主动地为用户提供个性化和智能化的服务。提出了一种新颖的、基于主题模型的方法来检测景点的旅游特色,以及同一景点在不同时节旅游特色的差异,并将其表示成一系列标签的概率组合。这种方法可以在没有用户标注的情况下进行自动检测,将非结构化的标签信息抽取成结构化的事件模式。Flickr中真实数据的实验结果表明该方法的可行性和有效性。

**关键词** Flickr;复合事件;主题模型;标签

**中图分类号** TP391

随着科技的进步与社会的发展,通信和计算机构成的信息空间与人们工作和生活的物理空间正在逐渐融为一体。数字技术和Web2.0的广泛应用,让越来越多的用户可以将自己拍摄的照片上传到Web上与世界各地的朋友分享,例如Flickr, Picasa以及Webshots,这些照片资源每日以PB量级的速度增长。如何根据用户的需求或者根据某些规则,自

动地从这些海量的、原子性的照片中抽取一些可直接服务于用户的复合事件(如景点旅游特色)成为一个亟待解决的问题。然而从图片这种二进制信息载体中提取事件等有用的知识是一直困扰人们的难点问题,因此Flickr中的多数照片都添加了标签。标签的引入使二进制的图片具有了一定的语义信息,从而更便于复合事件的检测和知识的抽取。这些添加

收稿日期:2010-06-25

基金项目:国家自然科学基金项目(60833005);国家“八六三”高技术研究发展计划基金项目(2009AA011904);教育部博士点基金项目(200800020002)

标签的照片弥补了传统网页搜索的不足,并带来很多实际应用,如:对于旅行者来说,那些有旅行目的地地理标签的照片可以帮助他们在定机票之前了解目的地.然而人们的需求正在由最初的存储、展现,向自动化、个性化、智能化转变.在任何一个信息应用的领域如何为用户提供个性化和智能化的服务已经成为人们关注和努力的热点问题.因此 Flickr 中现有的传统功能已经不能很好地满足用户的高端需求.例如:某用户计划在“五一”期间出游,但是没有明确的目的地,这时用户希望系统能够自动地检测出各个景点的旅游特色,并根据不同时节进行相应的分类,从而能够有效地制定旅游计划.但是当前的系统是无法满足这一需求的.现有的系统或者没有对景点的旅游特色进行分析,或者没有考虑不同时节中景点旅游特色的差异.因此如何从这些原子性的、零散的照片资源中抽取用户所需要的复合事件(例如景点的旅游特色),为用户提供个性化的服务就成为一个亟待解决的问题.

本文所描述的复合事件检测就是根据 Flickr 中照片对应的标签集合分析出反映某景点旅游特色的标签的概率组合,并且分析同一景点在不同时节的旅游特色的差异.复合事件检测的特点是以人为本,主动为用户提供更高效精确的个性化服务,即系统可以根据用户的指定或者根据某些规则,自动地从这些原子性的事件中抽取一些复合事件,对服务进行自由的裁剪和定制,从而将反映物理世界的原始数据实时地转化成可直接服务于终端应用的信息.类似地,我们的方法也可以应用于其他具有相似元数据的照片集合中.在 Flickr 的标签中进行复合事件检测是一个极具挑战的工作,因为:1) Flickr 中的数据存在很多噪声,并不是每一项 Flickr 数据都与主题相关;2) 与其他 Web 页面相比, Flickr 中包含了很少的文本信息,而是利用标签来反映照片的内容;3) 从非结构化的标签信息中抽取结构化的事件模式本身就是一项挑战.

针对上述问题,本文提出了一种新颖的、基于主题模型的、结合照片-标签-景点信息的复合事件检测方法.本文的检测方法分为3步:首先利用数学建模的方法得到每个标签关键词的先验概率,并根据概率模型为每张照片选取相应的标签关键词;其次,通过主题模型,结合 Flickr 中照片、景点和标签信息,分析标签关键词和景点的逻辑关系,找出能够很好地反映景点旅游特色的标签的概率组合;最后,从时间的角度出发,根据不同时节拍摄的照片集合发

现各景点在不同时节的旅游特色的差异.

本文主要的贡献如下:

1) 提出了 Flickr 中照片-标签-景点的数据表示模型.给出了数据的完整定义和描述,并引入邻接矩阵来表示其关联关系;

2) 提出了一种基于主题模型的新方法用来检测景点旅游特色,并将其表示成一系列标签的概率组合;

3) 考虑了同一景点在不同时节的旅游特色的差异,并能够在没有用户标注的情况下自动完成检测.

## 1 相关工作

复合事件检测的目标就是发现新的或者已有的、但未被识别的事件.这里的事件就是在特定的时间、特定的地点发生的特定的事情<sup>[1]</sup>.复合事件检测可大概分为两种类型:一种是回顾性检测;另一种是实时性检测<sup>[2]</sup>.回顾性检测就是从历史数据中发现已有的、但未被识别的事件;实时性检测就是实时、实地发现新的事件.本文我们综合了回顾性检测和实时性检测的优点.在检测景点旅游特色时我们用的是回顾性检测方法,但是考虑不同时节中同一景点旅游特色的变化时,又用的是实时性检测的思想.

作为事件检测的前期工作,文献[2]提出了一种简单的聚类算法(group average clustering),用来从文集中发现事件.文献[3]给出了一种概率方法,对文档中的内容和时间信息建模.这些方法都是基于文本进行的复合事件检测,后来 Kleinberg 又提出了基于特征进行检测的方法.基于特征进行检测的方法很多,比如无限自动机模型<sup>[4]</sup>,满足二项分布的特征聚类方法<sup>[5]</sup>,利用离散傅里叶变换的特征分类方法<sup>[6]</sup>.这些工作只考虑利用特征关键字进行分类,文献[7]同时借助用户指定打分来分类.由于 Flickr 中包含很少的文本信息,并且不是每一项 Flickr 数据都与主题相关,所以本文采用的是基于特征的检测方法.

近来,很多 Web 页面允许用户自主地添加标签,例如 Del.icio.us, Flickr 和 Last.fm 等.这一服务不仅吸引了很多用户来贡献标签,而且有利于系统的动态性和协同性. Halpin 等人利用 Del.icio.us 中书的标签测试出这些标签的分布随着时间的变化逐渐趋于稳定,并满足幂律分布<sup>[8]</sup>.标签元数据已经被广泛应用于很多领域,如企业搜索<sup>[9]</sup>、Web 检索<sup>[10]</sup>等.文献[11]尝试从 Flickr 的标签中抽取语义,找到与事件相关的标签.他们认为一个标签或者是与事件相关

的或者不是. 然而标签数据的模糊性和多义性使某些与事件相关的标签却看似无关. 为了解决这一问题, 文献[12]同时考虑了标签的时间和位置分布特性, 并将每个标签映射到经度-纬度-时间的三维空间中. 其他一些 Flickr 数据上的研究工作致力于寻找风景或界标<sup>[13-14]</sup>, 他们既考虑用户提供的标签, 又关注图片本身的内容.

## 2 预备知识

本节中我们给出一些基本概念, 数据的表示模型, 问题定义和主题模型的基本思想.

### 2.1 数据表示模型

表 1 列出了本文所使用的主要符号及其描述. 我们采集了多个不同景点的照片. 每张照片  $r_i$  对应一个位置  $l(r_i)$  和一个时间  $t(r_i)$ . 每张照片  $r_i \in R$  都对应标签的一个子集合  $Q(r_i) = \{q_1, q_2, \dots, q_m\} \subseteq Q$ . 照片  $r_i$  的一个标签  $q_j \in Q(r_i)$  与  $r_i$  的位置和时间也是有关联的. 标签  $q_j \in Q$  可以用来标注  $R$  中的多张照片, 我们用  $R(q_j)$  来表示被  $q_j$  标注的所有照片的集合, 则  $R(q_j) = \{r_1, r_2, \dots, r_n\} \subseteq R$ . 因此, 标签  $q_j$  与一系列的位置和时间相关联, 即  $L(q_j) = \{l(r_1), l(r_2), \dots, l(r_n)\}$  和  $T(q_j) = \{t(r_1), t(r_2), \dots, t(r_n)\}$ .

表 1 符号表

符号	定义和描述
$R$	Flickr 照片的集合: $R = \{r_1, r_2, \dots, r_n\}$
$Q$	Flickr 标签的集合: $Q = \{q_1, q_2, \dots, q_m\}$
$\mathcal{S}$	所有景点的集合: $\mathcal{S} = \{s_1, s_2, \dots, s_k\}$
$W^{x,y}$	对象 $x$ 和对象 $y$ 组成的邻接矩阵
$W = [W^{x,y}]$	整个邻接矩阵 ( $n \times n$ )
$N$	所有照片的数目
$M$	所有标签关键词的数目
$K$	景点的数目

图 1 中我们列举了 8 张照片 ( $r_1, \dots, r_8$ ), 这些照片来源于 5 个景点 ( $s_1, \dots, s_5$ ), 每张照片包含 2 个标签. 为了描述方便, 我们把“景点”、“照片”和“标签”看作不同类型的对象, 它们之间的关系用邻接矩阵来表示 ( $W^{x,y}$ ). 例如, 用  $W^{s,r}$  来表示景点和照片之间的关系. 如果  $W^{s,r}(i, j) = 1$ , 当且仅当第  $j$  张照片拍摄于第  $i$  个景点, 否则  $W^{s,r}(i, j) = 0$ . 类似地, 我们可以用  $W^{r,q+x}$  表示照片和第  $x$  类标签之间的关系. 如果  $W^{r,q+x}(i, j) = 1$ , 当且仅当第  $i$  张照片包含  $x$  类标签的第  $j$  个实例, 否则  $W^{r,q+x}(i, j) = 0$ .

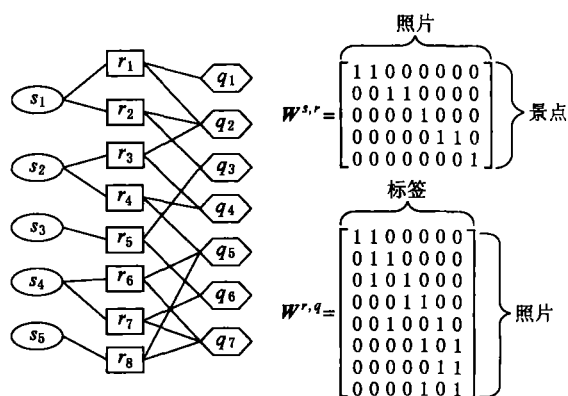


图 1 数据表示模型

### 2.2 问题定义

如我们在文献[15]中定义的, 事件就是在特定的时间、特定的地点发生的特定的事情. 因此, 在 Flickr 中抽取复合事件(如景点旅游特色的检测)必须满足下列 3 个约束条件: 1) 这些照片的标签内容要相应一致; 2) 这些照片要在一个特定的时间段内拍摄; 3) 这些照片要在一个特定的区域内拍摄.

我们将复合事件定义为“检测反映某个时节中某个景点的旅游特色的标签概率组合”. 因此, 给定一系列 Flickr 的照片  $R$  和其对应的标签集合  $Q$ , 本文要做的事情就是找出  $Q$  的一个子集  $Q_k \subset Q$ , 使其能够更好地反映  $R$  对应的景点  $\mathcal{S}$  的旅游特色, 并将该景点表示成一系列标签的概率组合. 由于每张照片都有时间属性, 因此结合景点-标签的概率分布我们可以进而推导出同一景点在不同时节的旅游特色的差异.

### 2.3 主题模型

主题模型的核心思想认为, 文档是若干主题的混合分布, 而每个主题又是一个关于单词的概率分布. 主题模型强调文档是由文档-主题-关键词 3 层关系组成, 而不仅仅是文档-关键词, 即文档不是仅由单个主题组成, 而是由多个主题组成. 主题模型是一种生成概率模型, 可以应用于文本数据、图像、生物图像以及其他多维数据的识别、分类和数据挖掘.

本文将采用目前被广泛使用的主题模型 LDA. LDA 将任意一个文档表示为一系列主题的组合, 同时每个主题是由单词表的一系列关键词的分布组成. LDA 模型可以用图 2 所示的概率图模型来表示, LDA 模型有 3 层表述, 参数  $\alpha$  和  $\beta$  是整个文档库层面的表述, 用以在概率生成过程中生成整个文档集合. 变量  $\theta$  是单个文档层面的表述, 在生成某个文档时用到. 最后变量  $z_n$  和  $w_n$  是词层面的表述, 用以生成每个文档中的每个词语. 对于文档集合  $D$  中

的任意一篇文档  $W$ , LDA 使用如下方法生成:

1)  $N$  符合泊松分布  $N \sim \text{Poisson}(\epsilon)$ ;

2)  $\theta \sim \text{Dir}(\alpha)$ ;

3) 对于任一关键词  $w_n (1 \leq n \leq N)$ , 选择一个主题  $z_n$  符合多项式分布, 即  $z_n \sim \text{Multinomial}(\theta)$ ; 在主题  $z_n$  条件下的多项式概率公式  $p(w_n | z_n, \beta)$  下选择一个关键词  $w_n$ .

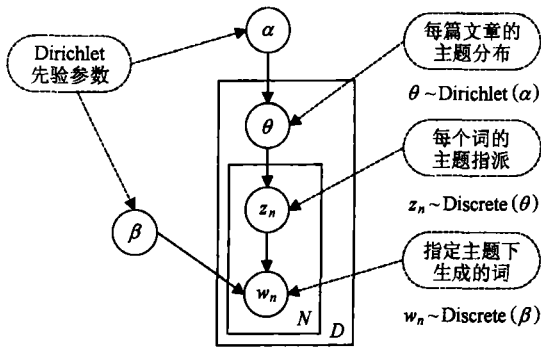


图2 LDA 主题模型

给定参数  $\alpha$  和  $\beta$  以及主题的混合参数  $\theta$ , 文档中的主题  $z$  和关键词  $w$  的分布可用式(1)来表示:

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta). \quad (1)$$

对于一篇文档可用式(2)来表示:

$$p(w | \alpha, \beta) = \int p(\theta | \alpha) \left( \prod_{n=1}^N \sum_{z_n} p(z_n | \theta) p(w_n | z_n, \beta) \right) d\theta. \quad (2)$$

### 3 基于主题模型的复合事件检测

我们的检测方法分为 3 步: 首先利用数学建模的方法得到每个标签关键词的先验概率, 并根据概率模型为每张照片选取相应的标签关键词(见 4.2 节).

其次, 通过主题模型, 结合 Flickr 中照片、景点和标签信息, 分析标签关键词和景点的逻辑关系, 找出能够很好地反映景点旅游特色的标签的概率组合. 第 1 步中我们得到了每个标签关键词的先验概率, 接下来的主要目标就是利用 2.3 节所描述的主题模型 LDA, 从这些概率分布中分析出各个景点的旅游特色. 给定一个标签关键词的集合, 要想进一步得到各景点的旅游特色, 就需要将这些标签关键词分类, 例如: 历史古迹、自然风光、小吃等. 然而这些景点特色并不能简单地从其对应的标签关键词获得, 因为: 1) 同一个标签关键词可能对应多个景点;

2) 同一个景点也会包含多个标签; 3) 同一个景点在不同的时间, 其旅游特色可能是不同的.

最后, 从时间的角度出发, 根据不同时节拍摄的照片集合发现各景点在不同时节的旅游特色的差异. 例如, 假设某用户想制定一次旅游计划, 其中包含“历史古迹”、“自然风光”、“小吃”3 个旅游特色. 那么制定这次旅游计划的本质就是从所选择的景点特色列表表中选取  $N$  次, 并利用概率分布为每个景点特色选取相应的标签关键词. 这一过程保证了所选取的景点中包含  $N$  个标签关键词.

利用 2.3 节所描述的主题模型, 我们可以在没有用户干预的情况下, 自动地检测各景点的旅游特色, 以及不同时节中同一景点旅游特色的差异. 下面仍以制定旅游计划为例来说明其实现过程. 将景点的集合记为  $\mathcal{S}$ , 假设不同照片对同一景点的影响因子不同, 照片  $r$  的概率分布记为  $p(r | \mathcal{S})$ , 而照片  $r$  中的每个标签关键词  $w$  的概率分布记为  $p(w | r)$ . 在给定这两个概率分布的前提下, 景点  $\mathcal{S}$  中每个标签关键词  $w$  的概率为:

$$p(w | \mathcal{S}) = \sum_{r=1}^N r(w | r) p(r | \mathcal{S}), \quad (3)$$

$N$  为景点  $\mathcal{S}$  中照片的个数, 这时概率分布  $p(w | \mathcal{S})$  已经不包含照片的概念. 当有很多景点时,  $p(w | \mathcal{S})$  就可以表示成如图 3 左边所示的邻接矩阵. 根据式(3), 该邻接矩阵可以通过标签关键词-照片矩阵和照片-景点矩阵计算得到. 也就是说, 左边的邻接矩阵可以拆分成右边的两个矩阵, 反之亦然.

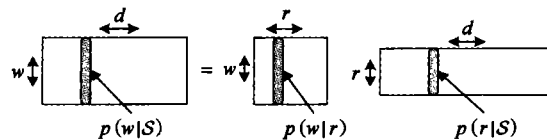


图3 景点特色分解模式

在下面的实验中, 利用 2.3 节介绍的 LDA 主题模型. 其中  $\alpha$  是景点-照片的概率分布控制参数,  $\beta$  是照片-标签关键词的控制参数. 本文要解决的问题就是找到符合狄利克雷分布的参数  $\alpha$  和  $\beta$ , 使式(4)中的概率值最大, 即找出景点中概率值最大的那些标签关键词. 式(4)是由 2.3 节中式(2)扩展来的, 其中,  $N$  表示照片的个数,  $M$  表示标签关键词的个数,  $K$  表示景点的个数.

$$p(w | \alpha, \beta) = \prod_{s=1}^K \int p(\theta_s | \alpha) \left( \prod_{n=1}^M \sum_{r=1}^N p(w_n | r, \beta) p(r | \theta_s) \right) d\theta_s. \quad (4)$$

### 4 实验结果及分析

在本节中,我们利用 Flickr 上的真实数据测试了基于主题模型的复合事件检测方法的可行性和有效性.首先我们对实验中用到的数据集进行了描述,然后对实验结果进行了分析.

#### 4.1 数据集

我们通过 Flickr 的 API 接口从 Flickr 站点上提取了大量的照片信息.搜集了从 2008 年 1 月 1 日到 2010 年 1 月 1 日期间北京的 30 个主要景点的所有照片,共计 700 000 张,其中约 302 000 张拍摄于 2008 年,398 000 张拍摄于 2009 年.平均每天拍摄的照片有 960 张,最少 236 张,最多 1 630 张.对于每张照片,我们都抽取了用户提供的标签关键词.与这些照片相关的标签关键词有 4 100 000 个,平均每张照片有 5.86 个标签,每个标签平均与 42.36 张照片相关联.每张照片都有一个时间属性,表示该照片的拍摄时间.

#### 4.2 实验结果

在实验过程中,我们设定狄利克雷先验参数  $\alpha = 0.01$ .表 2 显示的是照片-标签概率分布情况(在此只列出概率值  $p(w|r) \geq 0.02$  的标签关键词),由表 2 可知同一张照片包含多个标签,这些标签存在不同的概率.比如照片“天安门”可能包含“国徽”、“国旗”、“喷泉”、“花草”等多个标签,显然“国徽”所占的比重比“花草”大,因此我们只取最能反映照片内容的那些标签作概率组合.同时,我们可以看出不同的照片是可以共享同一个标签的,如  $q_1, q_3$ .

表 2 照片-标签概率分布

序号	照片	标签	概率
1	$r_1$	$q_1$	0.4402
2	$r_1$	$q_2$	0.2396
3	$r_1$	$q_3$	0.1088
4	$r_1$	$q_4$	0.0224
5	$r_2$	$q_1$	0.3427
6	$r_2$	$q_3$	0.2456
7	$r_2$	$q_5$	0.1176

图 4 表示的是拍摄于景点  $i$  的不同照片的比重关系.由图 4 可知,不同照片对景点特色的反映因子不同,即对于某一景点而言,只有少数照片最能体现该景点的旅游特色.

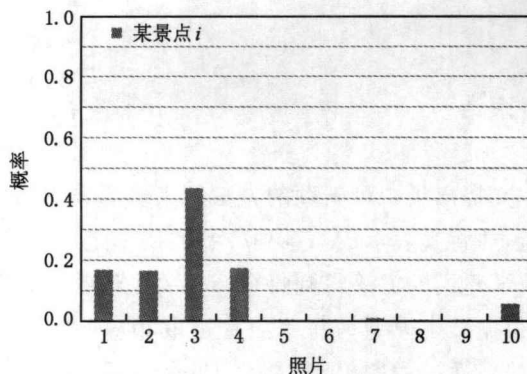


图 4 景点-照片概率分布

图 5(a) 给出了景点-标签的概率分布情况.由图 5 可知,表示不同景点旅游特色的标签的概率组合是不同的,每个景点的概率组合中都有 2~3 个标签最能体现其旅游特色.并且同一标签在不同景点中所占的比重也存在很大差异,最能体现景点 1 旅

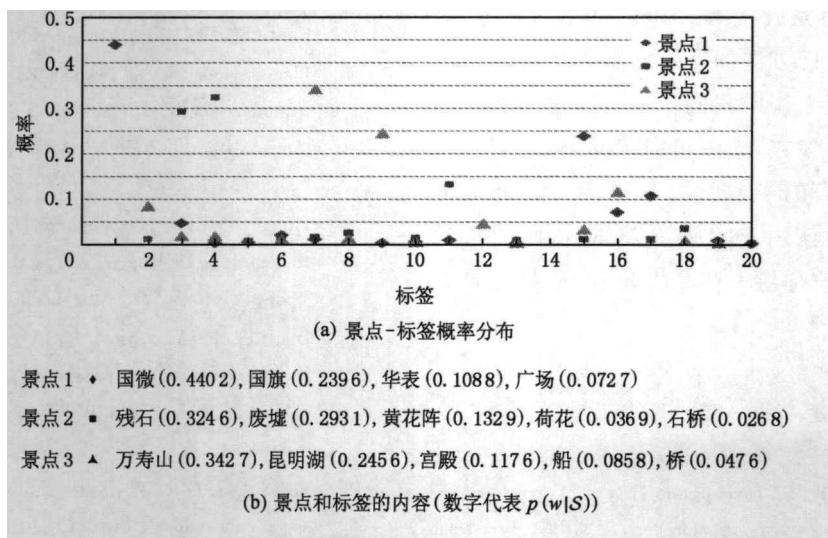


图 5 景点-标签概率分布

游特色的照片可能在景点 2 的概率组合中却占有很少的比例. 例如标签 1 更能体现景点 1 的旅游特色, 而标签 5 更能体现景点 3 的特色. 图 5(b) 显示的是景点和标签的内容, 其中括号里面的数字代表  $p(w|\mathcal{S})$ , 即此标签在该景点所占的概率值(其中  $p(w|\mathcal{S}) < 0.002$  的值没有显示).

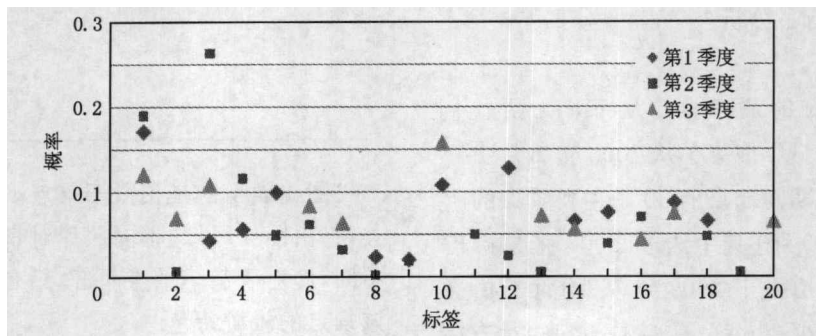


图 6 景点-时节概率分布

## 5 结束语

本文提出了一种新颖的方法用来检测景点的旅游特色, 并将其表示成一系列标签的概率组合. 首先利用数学建模的方法得到每个标签关键词的先验概率, 并根据概率模型为每张照片选取相应的标签关键词; 然后通过主题模型, 结合 Flickr 中照片、景点和标签信息, 将照片表示成含有一系列标签关键词的文档, 将不同景点定义成不同的主题, 从而分析标签关键词和景点的逻辑关系, 找出能够很好地反映景点旅游特色的标签的概率组合; 最后从时间的角度出发, 根据不同时节拍摄的照片集合发现各景点在不同时节的旅游特色的差异, 并可以在没有用户标注的情况下自动完成检测. Flickr 上真实数据的实验分析表明了我们方法的有效性和可行性.

在未来的研究工作中我们将进一步增加实验中的衡量检测指标和参数, 进一步完善其测试标准. 并将考虑标签的时间和地理属性, 根据用户之前上传的照片集合分析出该用户的旅游兴趣, 从而结合景点旅游特色, 为用户提供一些个性化的、满足需求的景点推荐.

## 参 考 文 献

- [1] Allan J, Carbonell J G, Doddington G, et al. Topic detection and tracking pilot study: Final report //DARPA Broadcast News Trans and Understanding Workshop. 1998: 194-218
- [2] Yang Y, Pierce T, Carbonell J G. A study of retrospective and on-line event detection //Proc of the 21th Annual Int ACM SIGIR Conf (SIGIR). New York: ACM, 1998: 28-36
- [3] Li Z, Wang B, Li M, et al. A probabilistic model for retrospective news event detection //Proc of the 28th Annual Int ACM SIGIR Conf (SIGIR). New York: ACM, 2005: 106-113
- [4] Kleinberg J M. Bursty and hierarchical structure in streams //Proc of the 9th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining (KDD). New York: ACM, 2003: 373-397
- [5] Fung G P C, Yu J X, Yu P S, et al. Parameter free bursty events detection in text streams //Proc of the 31th Int Conf on Very Large Data Bases (VLDB). New York: ACM, 2005: 181-192
- [6] He Q, Chang K, Lim E P. Analyzing feature trajectories for event detection //Proc of the 30th Annual Int ACM SIGIR Conf (SIGIR). New York: ACM, 2007: 207-214
- [7] Plangprasopchok A, Lerman K. Constructing folksonomies from user-specified relations on flickr //Proc of the 18th Int Conf on World Wide Web (WWW). New York: ACM, 2009: 781-790
- [8] Halpin H, Robu V, Shepherd H. The complex dynamics fo collaborative tagging //Proc of the 16th Int Conf on World Wide Web (WWW). New York: ACM, 2007: 211-220
- [9] Dmitriev P A, Eiron N, Fontoura M, et al. Using annotations in enterprise search //Proc of the 15th Int Conf on World Wide Web (WWW). New York: ACM, 2006: 811-817
- [10] Bao S, Xue G R, Wu X, et al. Optimizing Web search using social annotations //Proc of the 16th Int Conf on World Wide Web (WWW). New York: ACM, 2007: 501-510

- [11] Rattenbury T, Good N, Naaman M. Towards automatic extraction of event and place semantics from flickr tags //Proc of the 30th Annual Int ACM SIGIR Conf (SIGIR). New York: ACM, 2007: 103-110
- [12] Chen L, Roy A. Event detection from flickr data through wavelet-based spatial analysis //Proc of the 18th ACM Conf on Information and Knowledge Management (CIKM). New York: ACM, 2009: 523-532
- [13] Popescu A, Grefenstette G, Moellic P A. Mining tourist information from user-supplied collections //Proc of the 18th ACM Conf on Information and Knowledge Management (CIKM). New York: ACM, 2009: 1713-1716
- [14] Kennedy L S, Naaman M. Generating diverse and representative image search results for landmarks //Proc of the 17th Int Conf on World Wide Web (WWW). New York: ACM, 2008: 297-306
- [15] Zhou C, Meng X. IO3: Interval-based out-of-order event proc in pervasive computing //Proc of the 15th Int Conf on Database Systems for Advanced Applications (DASFAA). Berlin: Springer, 2010: 261-268

**周春姐** 女,1981年生,博士研究生,主要研究方向为普适计算、移动数据管理和模式挖掘。

**孟小峰** 男,1964年生,教授,博士生导师,中国计算机学会理事、高级会员,主要研究方向为 Web 数据管理、XML 数据库、移动数据管理。

**文洁** 女,1989年生,硕士研究生,主要研究方向为移动数据管理。