

基于应答关系网络的QQ群意见领袖发现*

冉义军, 毋亚男, 许爽, 许小可[†]

(大连民族大学信息与通信工程学院, 辽宁大连116600)

摘要: 意见领袖是社交网络和社交媒体中的重要节点,是信息传播的关键性因素。在QQ群聊天中由于参与用户较多,各种话题比较繁杂,所以识别其中的意见领袖比较困难。为此提出了一种基于应答关系网络来挖掘QQ群中意见领袖的方法。该方法首先构建回应词词库;然后基于Aho-Corasick算法来匹配聊天文本中的回应词数据,构建出用户应答关系的网络结构;最后使用社交网络中重要节点识别的方法来发现意见领袖。该方法对QQ群中的意见领袖发现具有较高的准确率,在融合QQ群用户交互社交网络的节点重要性特征后,能够达到更好的意见领袖发现效果。

关键词: 意见领袖; 应答关系; 文本挖掘; 社交网络

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2017)10-2993-04

doi:10.3969/j.issn.1001-3695.2017.10.025

Recognition and analysis of opinion leaders in QQ group based on response relationship network

Ran Yijun, Wu Yanan, Xu Shuang, Xu Xiaoke[†]

(College of Information & Communication Engineering, Dalian Minzu University, Dalian Liaoning 116600, China)

Abstract: Opinion leaders are the important nodes in social networks and social media, and they also are the key factors of information spreading. It is difficult to identify the opinion leaders from a large number of members who participate in different topics in QQ group. This paper proposed a novel way to mine the opinion leaders based on the response relationship in the QQ group. Firstly, it built a response word bank, and then constructed the user-response relationship network using Aho-Corasick algorithm. Lastly, it identified the opinion leaders by analyzing the user's node statistic information of the QQ social network. The experimental results show that there is a high accuracy by using the proposed method to mine the opinion leaders of the QQ group data. After merging the importance of the user-interaction network nodes feature, the method can achieve a better performance for mining opinion leaders.

Key words: opinion leaders; response relationship; text mining; social network

0 引言

近年来,随着信息技术和移动互联网的迅速发展,QQ、微博、微信等在线社交网络吸引着越来越多的用户,大量用户每天产生海量的结构化或非结构化的数据并传播高达上亿的信息。QQ群作为一种QQ用户中拥有共性的小世界,建立了一个多人即时通信平台,吸引了大量互联网用户发布并传播各种各样的信息。QQ群社交网络中的意见领袖挖掘具有非常重要的研究价值。

目前关于社交网络意见领袖挖掘的研究,主要分为网络结构分析法和用户属性分析法两类。基于网络结构的相关工作中,Cho等人^[1]利用用户间的关系建立了新的社交网络模型,通过分析用户在网络中的度中心性来挖掘意见领袖。Bai等人^[2]也研究了免疫在复杂网络中的局部效率,把网络中度最大的节点视为意见领袖,这一方法被其他研究人员应用于多种社交网络的意见领袖挖掘中。Kitsak等人^[3]认为网络中节点

的传播能力与节点所处的位置有重要关系,处于网络中心位置的节点具有较高的传播能力。

基于用户属性的意见领袖挖掘相关工作中,王珏等人^[4]从各个论坛中提取用户特征,采用期望最大化算法来挖掘意见领袖。丁雪峰等人^[5]通过构造网络话题参与者的属性矩阵,采用用户属性矩阵权重排序的方法挖掘意见领袖。Bodendorf等人^[6]通过分析用户交流间的关系挖掘意见领袖。

上述文献大多是从社交网络结构或用户属性单方面进行意见领袖的挖掘,相关文献中均没有综合利用文本和用户交互等多种特性来挖掘意见领袖。此外,相关文献中对社交网络的分析还较为粗糙,均没有对实际的网络进行细粒度分析。在本文中,针对QQ群聊天记录,首先利用文本挖掘技术构建回应词词库来建立用户之间的应答关系网络;然后研究QQ群好友之间的网络属性并利用网络中心性指标来挖掘意见领袖;最后通过基于应答关系网络 and 用户交互社交网络特征相融合的方法来联合挖掘QQ群意见领袖,取得了较为准确的效果。

收稿日期: 2016-07-24; **修回日期:** 2016-09-19 **基金项目:** 国家自然科学基金面上资助项目(61374170);辽宁省自然科学基金资助项目(2015020084);辽宁省教育厅科学研究资助项目(L2015127, L2014548)

作者简介: 冉义军(1989-),男(土家族),贵州铜仁人,硕士,主要研究方向为复杂网络、数据挖掘;毋亚男(1990-),女,硕士,主要研究方向为复杂网络、数据挖掘;许爽(1978-),女,副教授,博士,主要研究方向为图像处理、模式识别;许小可(1979-),男(通信作者),教授,博士,主要研究方向为非线性时间序列分析、复杂网络(xuxiaoke@foxmail.com)。

1 基于应答关系网络的意见领袖挖掘方法

1.1 意见领袖挖掘的整体框架

1.1.1 数据说明

本文所使用的数据是大连民族大学本科生班级 QQ 群聊天记录,主要为信息与通信工程学院 2011、2012、2013 级中 23 个班级的日常群聊数据,每个班级平均为 30 人。该数据均是从 2013 年 10 月~2014 年 12 月结束,完整地记录了 23 个班级为期一年两个月的聊天内容,总共得到了 47 785 条信息,首先人工分好话题,合计共有 5 056 个话题,然后使用自然语言处理的方法对话题间的内容进行分析。

1.1.2 整体系统框架

本文意见领袖挖掘系统分为文本数据处理、匹配与网络构建和社交网络分析三个模块,如图 1 所示。

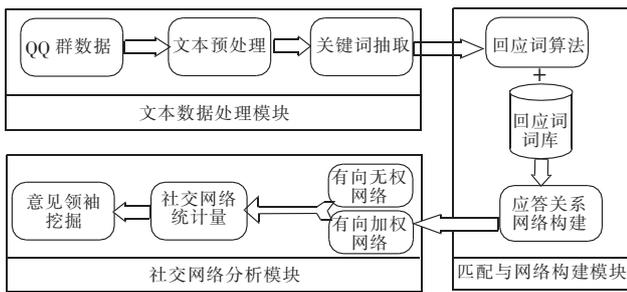


图 1 基于应答关系网络的 QQ 群意见领袖发现框架

其中:文本数据处理模块首先对 QQ 群聊天记录的文本信息使用 Python 第三方开源软件 jieba 分词进行文本预处理,包括分词、词性标注和过滤掉无用信息等;然后使用 jieba 中基于 TF-IDF 算法抽取关键词。匹配与网络构建模块是系统的核心模块,它决定了最终挖掘意见领袖核心率的高低,此模块包含回应词词库的构建和基于 AC 的回应词匹配算法设计,其中回应词词库的构建是关键环节,然后使用基于 AC 的回应词匹配算法与文本数据处理模块抽取的关键词相匹配,将匹配成功的应答关系数据使用基于 Python 语言的 NetworkX 函数库构建社交网络。NetworkX 是图论与复杂网络的建模与仿真工具,内置了常用的图与复杂网络分析算法,NetworkX 中内置了无向网络、有向网络、加权网络等图数据结构,因此本文首先根据由回应词生成 Python 语言中常用的列表结构,然后使用 NetworkX 里的有向图 DiGraph 方法来生成社交网络(图结构)。社交网络分析模块通过分析所构建有向无权网络和有向加权网络中每个节点的度中心性、接近度中心性、中介中心性和 PageRank 统计量,使之挖掘出 QQ 群中的意见领袖。

1.2 评价方法

为了验证意见领袖挖掘算法的有效性,需要采用相应的评价指标来对算法进行衡量。在现有的研究中多数研究人员还是采用人工评价的方法,此外,Song 等人^[7]采用覆盖度(coverage)作为衡量指标。覆盖度是从用户交互构成的拓扑结构角度,计算直接或间接影响用户数来衡量意见领袖的能力。Miao 等人^[8]采用核心率(core ratio)作为评价指标。核心率是指其他用户对该用户的交互信息量占总的交互信息量的比率。本文研究对象 QQ 群作为一个交互性很强的社交网络,因此本文采用核心率作为评价指标。

2 基于回应词构建应答关系网络

2.1 回应词词库的构建

快速发展的移动网络为大众提供了即时通信、聊天室、微信、微博等新型的交互方式,与此同时专门用于这类场合的一些回应词也相应派生出来。例如,在 QQ 群聊天的交际活动中,受话者常使用一些回应词语,如“好的”“嗯嗯”“知道了”等等,来表达对话题发起人提出内容的回复,认可话题提出者的观点,或者表明理解此话题的含义。随着回应词的广泛使用,聊天活动中由回应词作为指示的应答关系也很容易建立起来。构建出优质的、完整健全的回应词词库能够帮助有效创建用户之间的应答关系,提高 QQ 群中意见领袖挖掘的效率和准确性,因此,回应词词库的构建是本项目研究的一个重要环节。

现有词库构建的研究中,在构建结构的设计上主要分为树状层次结构、矩阵模型、网状结构和褒贬二性等几类。知网(HowNet)主要根据自身的网状形式进行构建^[9],WordNet 利用词语语义的矩阵模型构建词语^[10],《同义词词林扩展版》根据不同词语在语义上的不同构建成树状层次结构^[11],《台湾中文情感词典》根据褒贬义将词语构建成两种类别^[12]。本文采用褒贬义的思想,根据回应词的肯定、否定和中立意思,从聊天语料和网络用语中收集回应词,按照肯定、否定和中立三种构建回应词词库,如表 1 所示。

表 1 回应词词库

褒贬义	词库							
肯定	嗯嗯	恩恩	好的	好好	ok	是	是的	没有问题
	嘻嘻	啊啊	对	对的	佩服	同乐	同祝	喜欢
	支持	收到	点赞	额	顶	么么哒	强	强悍
	我爱你	Byebye	THS	了不起	厉害	3Q	En	这么牛
中立	好帅啊	去吧	真好	不错	再见	牛逼	不介意	我愿意
	小事情	小 case	愿意	See you	不错	Bye	哇塞	真的吗
	呃呃	汗	买醉	哇哇	呵呵	路过	飘过	没什么
	咋	鬼啊	打酱油	对不起	哈哈	讨厌	臭美	倒床
否定	什么	睡着了	晕	滚蛋	醉了	哎哎	怎么样	逗比
	不要	听不懂	屁	闭嘴	我操	不用	鄙视	我去
	恶心	切	没有	不是	木有	靠	气死我了	你去死

2.2 回应词的匹配与应答关系网络的构建

模式匹配算法作为信息检索和内容过滤的核心,引起了国内外研究人员的广泛关注。模式匹配算法分为单模式和多模式匹配算法,在单模式匹配算法中,KMP 算法能够消除回溯的匹配模式,对文本从左到右匹配,但是匹配效率较低^[13]。BM 算法利用好后缀规则和坏字符规则实现字符串的跳跃式匹配,效率比 KMP 算法提高很多,但是时间复杂度较高^[14]。在多模式匹配算法中,Aho-Corasick (AC) 算法结合了 KMP 算法的思想和有限状态自动机的理论,只需扫描一次文本串就可以识别出全部模式串,时间复杂度低^[15]。与单模式匹配算法相比,AC 算法可对所有模式串并行处理,提高了模式串的匹配速度。本文研究所使用的回应词词典较小,能够很好地利用 AC 算法的特点,使之得到较高的匹配率,因此本文采用基于 AC 的回应词匹配算法进行数据的匹配。

基于 AC 的回应词匹配算法的基本原理是在预处理过程中将模式串中各个字符所在节点当做状态,root 节点当做初始状态,如图 2 所示的‘0’节点,标签模式的节点作为终结状态,如图 2 所示的‘2’‘4’节点。该算法匹配的整个过程中分为三

个功能函数,即转向函数、失效函数和输出函数,由这三个函数构成一个树型有限自动机,如图 2 中的(a)~(c)三个函数。使用该算法进行匹配时,先根据文本数据处理模块提取的关键词的词性标注来判断用户使用回应词的褒贬性;然后与对应的回应词词库进行扫描读取,读取后的保存状态如图 2(a)所示。如当用户 A 在 QQ 群中说“今天晚上 21 点之前交信号与系统作业”,用户 B 使用“收到”来应答主用户 A 提出的话题,这时算法会先找到图 2 中的 root 节点,然后进行匹配,算法就会跳到‘1’节点,继续查找,算法到达‘2’节点,整个过程匹配成功,输出用户 B,此过程要是用户 B 使用的回应词没有‘到’字,算法就会跳到 root 节点,这就是图 2(b)中失效函数的功能。又当用户 C 使用“ok”来应答主用户 A 提出的话题,与上面一样,算法会跳到‘3’节点,然后继续查找,到达‘4’节点,匹配成功,则输出用户 C。然后将所有成功输出的匿名列表处理成构建网络需要的格式,得出(B,A),(C,A)各自有一条有向连边,并且方向都是指向 A 用户。如果已经存在连边的话,那么在加权网络中就将连边的权重加 1。

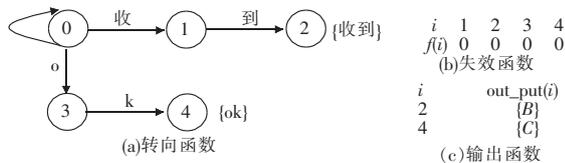


图 2 基于 AC 的回应词匹配算法示意图

在构建应答关系网络中,将 QQ 群中的成员视为节点,成员与成员之间的回应视为一条连边,在分话题的情况下构建有向无权网络和有向加权网络。定义 $G_1 = (V_1, E_1)$ 为有向无权网络, $G'_1 = (V_1, E_1, W_1)$ 为有向加权网络,其中 V_1 表示话题间回应词匹配的用户节点集合, E_1 表示有向连边集合,方向由话题回应者指向话题提出者, W_1 表示边的权重集合,其权重等于话题间某一用户用回应词回复话题提出者的次数。本文建立有向无权网络和有向加权网络两种网络的目的一方面是为了比较权重信息(用户交互的次数)对意见领袖识别的影响,另一方面是想对这两种网络特征和结果进行综合。

3 应答网络中的意见领袖发现

3.1 社交网络中意见领袖的挖掘方法

通过上面的回应词匹配算法后,每个 QQ 群都可以构建出一个应答关系的有向无权网络和有向加权网络。该网络中每个节点的重要程度是不同的,因此可以使用网络中重要节点的度量算法来挖掘意见领袖。目前网络中有关节点重要性的指标中,基于局域属性的统计量主要有度中心性、局部中心性等;基于全局属性的统计量有 Katz 指标、接近度中心性和中介中心性等;基于随机游走的有 PageRank^[16] 和 LeaderRank^[17] 等。本文主要使用度中心性、接近度中心性、中介中心性和 PageRank 四种经典统计量来挖掘 QQ 群的意见领袖。

节点的度中心性体现了该节点和周围节点之间建立直接联系的能力,中心性较高的节点具有较多的连接关系。设网络中有 n 个节点,节点 x 的度中心性指标计算公式为

$$C_d(x) = d(x)$$

其中: $d(x)$ 表示节点 x 直接相连的节点数,也就是该节点的度值。接近度中心性^[13]指标刻画了网络中节点到达网络中其他节点的难易程度,其值为该节点到达所有其他节点的最短距离

之和的倒数。设网络中有 n 个节点,则节点 x 的接近度中心性计算公式为

$$C_c(x) = 1 / \sum_{y=1}^n d_{xy}$$

其中: d_{xy} 是节点 x 和 y 之间的最短路径。接近度中心性指标反映了节点通过网络对其他节点施加影响的能力。中介中心性指标刻画了网络中节点对于信息流动的影响力,因此利用中介中心性指标可以确定信息负载繁重的网络节点。网络中节点 x 的中介中心性计算公式为

$$C_b(x) = \sum_{j < k} g_{jk}(x) / g_{jk}$$

其中: $g_{jk}(x)$ 表示节点 j 和 k 之间经过节点 x 的最短路径数, g_{jk} 表示节点 j 和 k 之间的最短路径数。中介中心性除了能衡量节点的重要性外,它的另一个重要作用是能够分辨出跨界者,即那些在两个或多个社团结构中扮演桥梁作用的个体。PageRank 算法的基本思想是:当节点 a 有一个连接指向节点 b , b 就获得了 a 对它贡献的分值,该值的多少取决于节点 a 本身的重要程度,即节点 a 的重要性越大,节点 b 获得的贡献值就越高。网络中节点 x 的 PageRank 计算公式为

$$R(x) = c \sum_{y \in B_x} B_y / N_y$$

其中: B_x 表示指向节点 x 的节点集合,设 B_y 表示指向节点 y 的节点集合, $N_y = |B_y|$ 表示连接 y 的数目, c 表示标准化因子。

3.2 结果说明与分析

根据构建的应答关系社交网络,计算了网络中每个节点的度中心性、接近度中心性、中介中心性和 PageRank 这四种统计量。基于四种统计量,计算出其核心率如表 2 所示。四种意见领袖统计量的核心率都在 60% ~ 70% 左右,由表 2 中的结果可知,有向无权网络与有向加权网络的核心率相差很小,从这一点上可以认为网络的权重信息(用户交互的次数)对本文的结果没有太大影响,即构建有向无权网络来识别意见领袖即可。同时本文也尝试了对有向无权网络与有向加权网络的结果进行了融合,发现融合后核心率的指标几乎没有提高,因此在文中没有显示这一融合后的结果。由于考虑到单从回应词构建的应答网络不能体现出用户之间应答的时间属性和用户间强的交互性,为此本文又从用户交互社交网络进行进一步研究。

表 2 基于应答关系网络意见领袖的核心率

网络结构	核心率			
	度中心性	接近度中心性	中介中心性	PageRank
有向无权网络	0.624	0.632	0.643	0.664
有向加权网络	0.618	0.626	0.649	0.667

3.3 融合用户交互社交网络的意见领袖挖掘

除了采用上述的应答关系来构建社交网络,还可以根据信息交互的时间顺序构建网络。即将 QQ 群中成员视为节点,话题内前一条消息发布的节点与紧接着后一条消息发布的节点之间产生连边。如果已经存在连边的话,那么在加权网络中就将连边的权重加 1。群中所有话题则形成用户交互社交网络。本文最终构建了两种类型网络,即有向无权网络 $G_2 = (V_2, E_2)$ 和有向加权网络 $G'_2 = (V_2, E_2, W_2)$ 。其中: V_2 表示用户节点集合; E_2 表示用户间的有向边集合,方向是由话题间下一条消息指向上一条消息; W_2 表示边权重集合,其权重等于话题间某一用户在另一用户后面发言的次数。构建出用户交互社交网络之后,同样也可以使用度中心性、接近度中心性、中介中心性和 PageRank 四种统计量分别算出用户交互网络意见领袖的核心率,其核心率如表 3 所示。

表 3 基于用户交互社交网络意见领袖的核心率

网络结构	核心率			
	度中心性	接近度中心性	中介中心性	PageRank
有向无权网络	0.617	0.621	0.613	0.631
有向加权网络	0.634	0.627	0.636	0.648

最后本文采取特征融合的方法对两种社交网络中节点的四种统计量值进行融合,其使用的带权平均公式为

$$y_i = \frac{1}{L} \sum_{j=1}^L \delta_j d_{ij}$$

其中: d_{ij} 是每种网络使用统计量计算节点的值; L 为构建网络方法的个数; δ_j 表示每种网络对节点统计量值的权重。采取投票表决的思想,当两种网络对 QQ 群意见领袖的影响都相等即 $\delta_1 = \delta_2 = 0.5$ 时,融合后意见领袖的核心率提高很小。由表 2 与 3 对比可知,基于应答关系网络对 QQ 群意见领袖的影响较大,所以本文采用 $\delta_1 = 0.75, \delta_2 = 0.25$ 进行特征融合。

使用上述特征融合的方法,得出融合后意见领袖的核心率,如表 4 所示。两种方法融合后将准确率提升了 11.5% 左右。研究发现,在融合 QQ 群用户交互社交网络的节点重要性特征后,能够达到更好的意见领袖发现效果。

表 4 融合用户交互网络后意见领袖的核心率

网络结构	核心率			
	度中心性	接近度中心性	中介中心性	PageRank
有向无权网络	0.751	0.747	0.736	0.766
有向加权网络	0.740	0.753	0.745	0.772

4 结束语

本文综合使用了文本挖掘技术和社交网络分析方法,对一种典型移动社交网络 QQ 群中的意见领袖进行了研究。本文首先按照褒贬义构建了回应词词库,基于回应词词库对 QQ 群数据进行话题分割,构建了用户间的应答关系网络;其次,对该网络采用常用的四种节点重要性统计量进行分析,得出了意见领袖的核心率。鉴于单从回应词构建的社交网络不能涵盖用户交互这一有用的信息,提出使用用户交互社交网络的中心性特征进行分析;最后,使用带权平均法融合两种方法提供的不同维度特征,从而能够达到更好的意见领袖发现效果。

参考文献:

[1] Cho Y, Hwang J, Lee D. Identification of effective opinion leaders in

the diffusion of technological innovation: a social network approach [J]. *Technological Forecasting and Social Change*, 2012, 79 (1): 97-106.

[2] Bai Wenjie, Zhou Tao, Wang Binghong. Immunization of susceptible-infected model on scale-free networks [J]. *Physica A: Statistical Mechanics and Its Applications*, 2007, 384(2): 656-662.

[3] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. *Nature Physics*, 2010, 6(11): 888-893.

[4] 王珏,曾剑平,周葆华,等.基于聚类分析的网络论坛意见领袖发现方法[J].*计算机工程*,2011,37(5):44-46,49.

[5] 丁雪峰,胡勇,赵文,等.网络舆论意见领袖特征研究[J].*四川大学学报:工程科学版*,2010,42(2):145-149.

[6] Bodendorf F, Kaiser C. Detecting opinion leaders and trends in online communities [C]//Proc of the 4th International Conference on Digital Society. Washington DC: IEEE Computer Society, 2010: 124-129.

[7] Song Xiaodan, Chi Yun, Hino K, et al. Identifying opinion leaders in the blogosphere [C]//Proc of the 16th ACM Conference on Information and Knowledge Management. New York: ACM Press, 2007: 971-974.

[8] Miao Qingliang, Zhang Shu, Meng Yao, et al. Domain-sensitive opinion leader mining from online review communities [C]//Proc of the 22nd International Conference on World Wide Web. New York: ACM Press, 2013: 187-188.

[9] WordNet [EB/OL]. [2013-04-11]. <http://wordnet.princeton.edu/>.

[10] HowNet [EB/OL]. [2013-08-11]. <http://www.keenage.com>.

[11] 哈工大社会计算与信息检索研究中心同义词词林扩展版 [EB/OL]. [2012-07-02]. <http://www.datatang.com/data/42306/>.

[12] NT-USD [EB/OL]. [2013-08-02]. <http://www.datatang.com/data/44317/>.

[13] Knuth D E, Morris J H, Pratt V R. Fast pattern matching in strings [J]. *SIAM Journal on Computer*, 1977, 6(2): 323-350.

[14] Boyer R S, Moore J S. A Fast string searching algorithm [J]. *Communications of the ACM*, 1977, 20(10): 762-772.

[15] Kilpelainen P. Set matching and Aho-Corasick algorithm [J]. *Biosequence Algorithms*, 2005(4): 1-24.

[16] Fushimi T, Saito K, Kimura M, et al. Finding relation between PageRank and Voter model [C]//Proc of the 11th International Conference on Knowledge Management and Acquisition for Smart Systems and Services. 2010: 208-222.

[17] Lyu Linyuan, Zhang Yicheng, Yeung C H, et al. Leaders in social networks, the delicious case [J]. *PLoS ONE*, 2011, 6(6): e21202.

puter Society, 2012: 1038-1043.

[45] Xue A Y, Zhang Rui, Zheng Yu, et al. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction [C]//Proc of the 29th IEEE International Conference on Data Engineering. 2013: 254-265.

[46] Yuan Quan, Cong Gao, Ma Zongyang, et al. Time-aware point-of-interest recommendation [C]//Proc of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2013: 363-372.

[47] Yuan Quan, Cong Gao, Zhao Kaiqi, et al. Who, where, when, and what: a nonparametric bayesian approach to context-aware recommendation and search for Twitter users [J]. *ACM Trans on Information Systems*, 2015, 33(1): 1-33.

[48] Liu Bin, Fu Yanjie, Yao Zijun, et al. Learning geographical preferences for point-of-interest recommendation [C]//Proc of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2013: 1043-1051.

[49] Ying J J C, Kuo Wenning, Tseng V S, et al. Mining user check-in behavior with a random walk for urban point-of-interest recommendations [J]. *ACM Trans on Intelligent Systems and Technology*, 2014, 5(3): 1-26.

(上接第 2983 页)

[39] Zhao Yiliang, Nie Liqiang, Wang Xiangyu, et al. Personalized recommendations of locally interesting venues to tourists via cross-region community matching [J]. *ACM Trans on Intelligent Systems & Technology*, 2014, 5(3): 1-26.

[40] 高榕,李晶,杜博,等.一种融合情景和评论信息的位置社交网络兴趣点推荐模型[J].*计算机研究与发展*,2016,53(4):752-763.

[41] Yang Dingqi, Zhang Daqing, Yu Zhiyong, et al. A sentiment-enhanced personalized location recommendation system [C]//Proc of the 24th ACM Conference on Hypertext and Social Media. New York: ACM Press, 2013: 119-128.

[42] Cheng Chen, Yang Haiqin, Lyu M R, et al. Where you like to go next: successive point-of-interest recommendation [C]//Proc of the 23rd International Joint Conference on Artificial Intelligence. 2013: 2605-2611.

[43] Monreale A, Pinelli F, Trasarti R, et al. WhereNext: a location predictor on trajectory pattern mining [C]//Proc of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2009: 637-646.

[44] Noulas A, Scellato S, Lathia N, et al. Mining user mobility features for next place prediction in location-based services [C]//Proc of the 12th International Conference on Data Mining. Washington DC: IEEE Com-