



位置社交网络的潜在好友推荐模型研究^{*}

孙晓晨¹, 徐雅斌^{1,2}

(1. 北京信息科技大学计算机学院 北京 100101;

2. 网络文化与数字传播北京市重点实验室 北京 100101)

摘要: 为了提高位置社交网络的服务便捷性和用户感受度, 与位置相关的推荐服务越来越具有重要意义和应用需求。提出的潜在好友推荐模型主要是根据签到位置的相似度及好友相似度进行潜在用户推荐。通过用户的好友关系、签到特性及签到历史记录, 计算用户在各个位置兴趣点的位置权重, 再分别利用位置权重及好友关系计算用户的位置相似度和好友相似度, 最后根据用户位置和好友关系的综合相似度进行潜在用户推荐。实验结果表明, 提出的潜在好友推荐模型是切实有效的。

关键词: 位置服务; 位置社交网络; 潜在好友; 推荐模型; 位置相似度

doi: 10.3969/j.issn.1000-0801.2014.10.012

A Potential Friends Recommendation Model for Location-Based Social Network

Sun Xiaochen¹, Xu Yabin^{1,2}

(1. School of Computer, Beijing Information Science & Technology University, Beijing 100101, China;

2. Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing 100101, China)

Abstract: To make service convenient and improve user experience degrees, recommendation service has become more and more important to users in the location-based social network (LBSN). An improved potential friends recommendation in LBSN, which is based on the relationship among friends and the similarity of locations, was proposed. First of all, the user interest points in each position weight with friends' relationship, check-in feature and historical record were calculated. Then position weight and the friend relationship were used to calculate the friends' similarity and location similarity. Finally, potential friends to user based on the friends' similarity and location similarity were recommended. The experimental results show that the proposed model is effective.

Key words: location-based service, location-based social network, potential friend, recommendation model, location similarity

1 引言

位置服务(location-based service, LBS)又称定位服务, 是通过GPS(global positioning system, 全球定位系统)、

WLAN(wireless local area network, 无线局域网)等技术获得移动终端的位置信息(如经纬度坐标数据), 并将这些位置信息提供给移动通信用户及通信系统, 实现各种与位置相关的业务。位置服务与传统在线社交网络逐渐融合, 从

^{*} 国家自然科学基金资助项目(No.61370139), 网络文化与数字传播北京市重点实验室基金资助项目(No.ICDD201309), 北京市属高等学校创新团队建设与教师职业发展计划基金资助项目(No.IDHT20130519)



而产生了位置社交网站(location-based online social network, LBSN)。国外典型的 LBSN 主要有 Foursquare、Geolife 等,国内发展比较好的 LBSN 主要有街旁、陌陌等。随着位置服务的广泛应用,传统在线社交网络也开始引入位置服务,如微博的“微领地”应用。

位置社交网络不仅可以像传统在线社交网络那样分享博客、照片、视频进行信息交互,而且可以随时随地地定位和分享位置信息^[1]。用户可以通过电脑、手机及其他移动终端进行签到,并将所处的位置信息及时告诉他的朋友。当签到信息发生变化时,用户能通过社交网站进行同步更新,这样就可以方便快捷地与好友分享自己的最新动态。

位置社交网络的意义在于分享与位置相关的内容,并借此结识朋友。因此,通过研究和分析位置社交网络中与位置相关的历史数据及用户个人兴趣信息,并建立位置社交网络的潜在好友推荐模型,就可以为用户推荐一些与他们行为、兴趣相似的好友,从而帮助用户结识更多在实际生活中未曾见面但却与自己兴趣相投的其他用户,从而更好地发展自己的社交圈。同时,为用户推荐潜在好友,使用户体会到社交网络应用的方便性,从而活跃于社交网络中,有助于位置社交网络的健康发展。

2 相关工作

迄今为止,国内外学者在用户推荐服务方面做了大量工作。参考文献[2]中根据好友的好友进行潜在用户推荐,因为用户的两个朋友成为好友的概率比随机两个人成为好友的概率高。参考文献[3]和参考文献[4]都是通过对用户的轨迹进行建模,以此来分析用户的轨迹相似度,进而度量用户之间的相似性。但是它们的建模方法有所不同,参考文献[3]中的建模方法采用 HGSM(hierarchical-graph-based similarity measurement, 基于等级结构图的相似度测量)算法,而参考文献[4]中的建模方法采用 PST(probabilistic suffix tree, 概率后缀树)。参考文献[5]利用在线社交关系,并利用矩阵分解的方法进行潜在用户推荐,取得了很好的效果。参考文献[6]中通过用户的历史位置进行建模,可以分析出用户的行为模式,然后根据用户行为模式之间的相似性,向用户推荐潜在好友或进行用户行为的异常检测。参考文献[7]利用用户的标签信息和时间信息建立推荐模型,较好地分析了用户的兴趣爱好,具有很好的性能。参考文献[8]

采用改进协同过滤算法对交友网站中的用户进行网上交友推荐,将交友双方的兴趣和吸引力等因素都考虑到推荐模型中,从而提高推荐效果。

综合分析发现:参考文献[1,2,5,7,8]主要是根据传统的社交关系进行潜在用户推荐;参考文献[3,4]利用社交网站中的位置轨迹信息进行潜在用户推荐;参考文献[6]则利用社交网站中的历史位置信息分析用户的行为模式,然后利用用户的行为模式相似性进行潜在用户推荐及行为异常检测。

考虑用户签到历史信息、用户影响力等因素,本文提出了一种利用用户对位置的隐性评价计算用户之间位置相似度的方法,进行潜在用户推荐。该方法首先对用户签到的位置兴趣点进行聚类,得到位置兴趣区域;然后利用用户的好友关系、用户影响力、签到特性来计算用户在各个位置兴趣点的位置权重,再利用向量空间模型计算用户位置相似性及好友相似度;最后根据用户综合相似度进行潜在好友推荐。

3 用户影响力分析与计算

本文对 Gowalla、Brightkite 等 LBSN 的签到数据进行挖掘分析,发现 53%左右的 LBSN 用户完成了 85%的签到,如图 1 所示,而且统计得到 1 年内 35%的用户签到次数少于 10 次,由此可以看出 LBSN 中存在核心用户。这些核心用户在 LBSN 中处于领导地位,他们在 LBSN 中很活跃,拥有很多 LBSN 好友,并且分享位置信息的意愿很强。

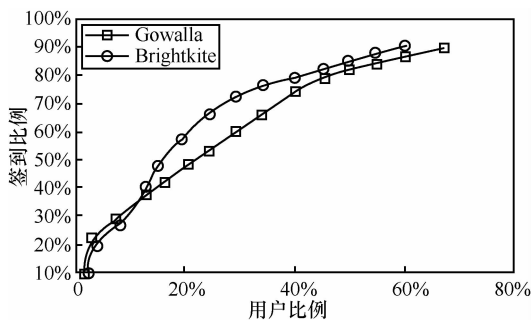


图 1 Gowalla 和 Brightkite 签到数据分析

位置社交网络的用户影响力是指用户在位置社交网络中对其他用户的影响和带动能力。它是用户在社交网络中所处重要程度的判断标准,反映了用户之间的交互关系和亲密程度,同时对社交网络的发展产生一定的推动力。

用户的影响力越大,别人对他的关注度就越高,对网络信息的传播推动作用也就越强。也就是说,在位置社交网络中如果一个用户的影响力很高,那么他所推荐的事物(无论是位置信息还是各类网络广告、网络应用)都很容易被传播,很容易被别人接受。因此,位置社交网络的用户影响力是影响推荐服务的重要因素之一。

在 LBSN 中用户影响力受到很多因素的影响,其中包括发挥自身影响力的意愿、自身活跃程度、好友数量、好友质量、好友活跃程度等。综合考虑各方面因素,本文的用户影响力考虑了用户自身发挥影响力的意愿和用户影响度,计算方法如式(1)所示:

$$I_i = \frac{N_i}{T} \times LIR(i) \quad (1)$$

其中, I_i 表示用户 i 的影响力, N_i 表示用户 i 在时间 T 内的签到次数, N_i/T 表示用户发挥自身影响力的意愿,可以理解为:用户在单位时间内签到的次数越多,用户发挥其影响力的意愿就越高。 $LIR(i)$ 表示用户的影响度,关于它的计算,借鉴 PageRank 的思想,其值应介于 0 和 1 之间,在包括自身影响度的同时,也要包括其好友对他的影响度,计算方法如式(2)所示:

$$LIR(i) = (1-d) + d \times \sum_{v \in F_i} LIR(v) \times C(v) \quad (2)$$

其中, $LIR(i)$ 、 $LIR(v)$ 分别表示用户 i 和用户 v 的影响力; d 是权重系数,表示用户好友对其影响所占的比重(本文选取 0.85); $C(v) = P_i / (P_i + P_v)$, 表示用户 v 的影响能力分配给用户 u 的比例因子, $P_i = \sum_{v \in F_i} N_v$, F_i 是用户 u 的好友集合, N_v 是用户 v 在 LBSN 中的等级。用户等级可以通过用户的签到次数来衡量,因为在位置社交网站上,如果用户在某一位置签到次数最多,那么他会成为这一位置的地主,具有很强的号召力和说服力。将所有 LIR 的初始值设为 0.1, 通过迭代到收敛为止,可以得到所有用户的 LIR 。该方法涉及的数据随着用户数量的不断增加可以进行增量更新。

4 潜在好友推荐模型

位置社交网络中的签到数据包含用户签到位置的经度和纬度这一新维度,记录了用户访问的关键地点,形成了用户的离散化行为轨迹。位置社交网络的用户行为轨迹由一系列离散的时空点组成,虽然它不像 GPS 那样记录用户行为的连续轨迹,但是其记录的离散化轨迹更能体现用户强烈的目的性。每一个签到位置对于用户来说都有一

定的意义,能够表现出用户的兴趣爱好等,同时,每个位置对用户的重要程度是不一样的。因此,本文根据用户之间的位置相似度和好友相似度,计算用户的综合相似度,从而进行潜在好友推荐。

根据实际情况可知,用户的两个朋友成为好友的概率比随机两个人成为好友的概率要高,同时,相近的位置一般属于同一个区域,因而当用户访问相近或相同位置时,他们可能有相似或相同的行为目的。例如,当两个用户都经常在雍和宫这类古代寺庙签到时,说明他们具有一定的相似性或相同性,进而可以进行好友推荐。此外,每个位置对于用户来说重要程度是不一样的,如果两个用户对于很多位置持有相同的位置权重,那么可以认为他们具有一定的相似度。因此,本文判断用户相似性基于几条经验,包括:用户之间拥有共同的好友、用户具有相同或相似的位置签到历史记录、用户持有相同位置权重的位置等的数量,这些数量越多,用户之间的相似度越高。

4.1 位置聚类

在位置社交网络中,用户会在自己感兴趣的地方进行签到,每一个签到位置称为位置兴趣点^[9](location point of interest, LPOI)。LPOI 中包含位置的唯一标识号、经纬度和名称,即 $LPOI = (locationID, latitude, longitude, name)$ 。用户在位置兴趣点签到时,位置社交网络会留下用户的签到记录(checkin),其中包括用户 ID、用户访问的 LPOI、签到时间、对位置兴趣点的评价或此时的心情,表示为 $checkin = (userID, LPOI, time, text)$ 。其中, $text$ 是可选内容。

因为在某一个实际位置附近可能有很多不同的 LPOI,所以在用户访问相近位置时可能不同的 LPOI 上签到。如果只是简单地通过判断用户是否访问同一个 LPOI 来得到用户之间的位置相似度,那么这种数据是很稀疏的。因此,本文采用聚类的方法将相近的 LPOI 聚到同一区域,然后再根据用户访问的各个聚类区域,计算用户之间的位置相似性。利用聚类的方法,可以将用户经常访问的 LPOI 聚集到一块,而用户偶尔访问或很少有人访问的位置将被视为噪音而过滤掉。

本文采用 DBSCAN 算法^[10,11]对用户访问的位置兴趣点进行聚类。该算法是一种基于密度的空间聚类算法,与层次聚类方法不同,它将簇定义为密度相连的点的最大集合,能够把达到一定高密度的区域划分为簇,并可以在有噪声的空间数据库中发现任何形状的聚类。也就是说,该



算法可以将一些高密度的 LPOI 划分为簇,并且将 LPOI 数据库中的噪声 LPOI 排除在外。本文将聚集的 LPOI 集合定义为位置兴趣点区域 (regional point of interest position, RPIP)。

对 LPOI 的 DBSCAN 聚类需要给出邻域 ε 和 $\minPts$ 两个对象参数,它们可以根据区域含 LPOI 的密度来决定。

LPOI 的 DBSCAN 聚类算法描述如下。

- (1) 输入 LPOI 集合 $L=\{L_1, L_2, \dots, L_n\}$;
- (2) 通过计算每平方千米含有 LPOI 的个数,选取适合的邻域参数 ε 和聚类簇内最少包含 LPOI 的个数 $\minPts$;
- (3) 遍历 LPOI 整个集合,计算从 L_i 到 L_j 在平面坐标上的距离,是否满足密度可达的条件;
- (4) 若密度可达点的数量大于 $\minPts$,则构成一个 RPIP; 否则,继续遍历 LPOI 集合;
- (5) 遍历完 LPOI 集合,输出 RPIP 集合 $R=\{R_1, R_2, \dots, R_n\}$ 。

4.2 位置权重表示

定义 1 位置权重 (position weight) 就是用户对聚类后的位置兴趣点区域 (RPIP) 的重视程度,也可以理解为用户对 RPIP 的评分。

在位置社交网络中,用户对某一地理位置的重视程度可以用在某位置的签到次数来衡量,在一定程度上反映了用户对该位置的兴趣,是一种隐式评价。但是不能准确地表示用户对位置的兴趣度,仅表明用户自身对位置兴趣点的评价,忽略了用户好友对用户的影响。在位置社交网络中,用户在某一位置兴趣点签到不仅要考虑自身的意愿,还要考虑用户好友对他的影响,因为在现实生活中用户去一个新位置进行活动时经常会考虑自己好友所给的意见或者好友的经历。基于此思想,本文给出了计算用户位置权重的方法。

位置权重用 W_{ij} 表示,即用户 i 在位置 j 的位置权重, W_{ij} 的数值越大则表示位置 j 对于用户 i 越重要,其计算式为:

$$W_{ij} = \alpha W_{\text{user}} + (1-\alpha) W_{\text{friend}} \quad (3)$$

$$W_{\text{user}} = N_{ij}/N_i, W_{\text{friend}} = I \times (N_{jf}/N_j), I = \frac{\sum_{k \in M_k} I_k}{M} \quad (4)$$

其中, N_{ij} 表示用户 i 在位置兴趣点区域 j 的签到次数; N_i 表示用户 i 在各个位置兴趣点区域的总签到次数; N_{jf} 表示用户 i 的朋友 f 在位置兴趣点区域 j 的总签到次数; N_j 表示所有用户在位置兴趣点区域 j 的签到总数; W_{user} 表示用户在某一位置兴趣点区域的自身位置权重; W_{friend} 表示

用户的好友在某一位置兴趣点区域的位置权重; α 是权重因子 (本文选择 0.85); I 为影响因子,表示用户的好友对用户在某位置兴趣点区域的平均影响力; I_k 表示用户的朋友 k 对用户的影响因子,是用户好友的影响力,其值通过对式 (1) 迭代到收敛取得; M 表示用户的好友个数。

4.3 用户相似度计算

位置社交网络上包含 3 种关系:人与人的关系、人与位置的关系及位置之间的关系,这些关系会影响用户相似度的计算。LBSN 用户之间拥有共同好友,在一定程度上说明用户之间具有相似度,用户之间共同好友的个数越多,说明他们越相似;用户的签到位置一般会选在自己喜欢的位置或者对自己有意义的位置,位置记录能够表现用户的爱好及行为习惯,因此,签到位置的相似性也能体现用户之间的相似度,用户之间所选择的签到位置越相似,他们越相似。综合以上两点,本文计算相似度的方法采用两种相似度的加权之和,如式 (5) 所示:

$$\text{Sim} = b \text{Sim}_L + (1-b) \text{Sim}_F \quad (5)$$

其中, Sim_L 表示位置相似度, Sim_F 表示好友相似度, b 为加权系数。

4.3.1 位置相似度计算

本文利用空间向量模型计算用户之间的位置权重相似度,即用户的位置相似度。向量 L_i 表示用户在各个位置兴趣点区域的位置权重向量,其中 $L_i = [W_{i1}, W_{i2}, \dots, W_{in}]$, W_{ij} 表示用户 i 在位置兴趣点区域 j 的位置权重。用户之间位置权重相似的位置越多,可以理解为他们相似度越高,也可以说他们成为好友的可能性越大。所有用户访问的位置兴趣区域的位置权重构成用户位置权重矩阵,然后用余弦相似性方法计算用户之间的相似度。

用户位置权重矩阵表示为:

$$V_{(m \times n)} = \begin{bmatrix} W_{1,1} & \dots & W_{1,j} & \dots & W_{1,n} \\ W_{2,1} & \dots & W_{2,j} & \dots & W_{2,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ W_{i,1} & \dots & W_{i,j} & \dots & W_{i,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ W_{m,1} & \dots & W_{m,j} & \dots & W_{m,n} \end{bmatrix} \quad (6)$$

其中, m 为位置社交网络的用户数, n 为位置兴趣区域的总数, W_{ij} 为用户 i 在位置兴趣点 j 的位置权重。

把用户在各个位置兴趣区域的位置权重看成 n 维向量空间上的向量,用向量余弦夹角度量用户之间的相似

度。设用户 X 与用户 Y 在 n 维向量空间中表示为 W_X 和 W_Y , 通过式(7)计算用户之间相似度:

$$\begin{aligned} Sim_L(X, Y) &= \cos(W_X, W_Y) = \frac{W_X \times W_Y}{\|W_X\| \|W_Y\|} \\ &= \frac{\sum_{i \in L_{XY}} W_{X_i} W_{Y_i}}{\sqrt{\sum_{i \in L_{XY}} W_{X_i}^2} \sqrt{\sum_{i \in L_{XY}} W_{Y_i}^2}} \quad (7) \end{aligned}$$

其中, $L_{XY} = \{i \in I | L_{X_i} \neq \emptyset \text{ 或 } L_{Y_i} \neq \emptyset\}$ (I 表示全部位置兴趣区域的集合), 表示用户 X, Y 的签到位置兴趣区域并集, W_{X_i} 表示用户 X 在位置兴趣区域 i 处的位置权重, W_{Y_i} 表示用户 Y 在位置兴趣区域 i 处的位置权重。

4.3.2 好友相似度计算

根据六度分割原理可以知道, 通过一定次数的传递, 人能通过朋友的朋友认识社交网络中的任何人。因此, 两个用户具有共同好友, 他们成为朋友的概率比两个没有任何关系的陌生人成为朋友的概率大。本文利用共同好友比重度量好友相似度, 如式(8)所示:

$$Sim_F(X, Y) = \frac{|U_X \cap U_Y|}{\sqrt{|U_X| |U_Y|}} \quad (8)$$

其中, U_X 为用户 X 的好友集合, U_Y 为用户 Y 的好友集合。

4.4 产生 TopK 推荐

根据前文所述, 本文分别计算出任意用户之间位置相似度和好友相似度, 再计算用户的综合相似度, 并根据综合相似度进行 TopK 推荐, 得到最终的推荐列表。通过位置相似度和好友相似度进行好友推荐可以看成两个属性的 TopK 推荐问题, 位置相似度是一个属性, 好友相似性是另一个属性, 通过前文的计算得到各自对应的值, 最后综合计算最优值, 从而得到最适合推荐的 K 个好友。

5 实验数据与实验结果

5.1 实验数据集

本文采用 LBSN 网站 Gowalla 的 2009–2010 年的签到数据集^[12], 其中包含 640 万条签到记录、70 万个签到位置、19 万个用户的 200 万条好友关系。实际上对位置数据中的无效用户(即注册后很少签到的用户)和很少有人签到的位置兴趣点(即到访人数很少的点)进行挖掘是没有意义的, 因此需要对数据进行预处理, 移除无意义点, 减少数据量。本文将一些很少访问的位置和具有很少好友的用户

进行清理, 清洗过程中设定用户在规定时间内 ΔT 内的最小签到频率 $F_{\min}$ (本文 $F_{\min}=1$) 和最大签到频率 $F_{\max}$ (本文 $F_{\max}=8$), 如果用户签到频率小于 $F_{\min}$, 则该用户的相应数据视为无效; 如果用户签到频率大于 $F_{\max}$, 则视为虚假签到, 将其相应的数据进行清除。经过数据清洗后, 最终得到 2 689 个用户的 5 万多条好友关系、135 万多条签到记录、23 万多个签到位置。从所清洗的数据集中随机选取 500 个用户进行实验研究, 这些用户的平均好友数为 20 个, 平均签到次数为 689 次。将所选取的数据集存储在 MySQL 数据库中, 建立新的数据库表, 分别为用户好友表(user_friend)、用户签到表(user_checkin)、RPIP 表(location_PRIP)、位置用户签到表(location_user), 其中用户好友表记录用户的好友, 用户签到表记录用户访问位置兴趣区域, RPIP 表记录每个 RPIP 具体包含的 LPOI, 位置用户表记录在每个 RPIP 签到的用户。

5.2 方法评测标准

本文的推荐效果由召回率和精确度进行度量^[13]。将数据集上的好友关系划分为训练集(training set)与测试集(test set), 训练集用于建立潜在好友推荐模型, 测试集用于对潜在好友推荐模型进行验证。 $L(u)$ 是对训练集上的用户做出的推荐列表, 即推荐结果, $R(u)$ 是测试集上已存在的真实关系。推荐结果可能包括相关结果(relevant result, RR)和不相关结果(irrelevant result, IR), 即 $L(u) = RR + IR$, 其中相关结果就是实际存在的好友关系, 不相关结果则是在测试集中不存在的好友关系。测试集中应该包含所有存在的结果, 即包括推荐结果中相关结果和未推荐的相关结果(no relevant result, NRR), 则 $R(u) = RR + NRR$ 。

定义 2 召回率(recall)就是推荐结果中相关结果占好友总数的比重, 如式(9)所示:

$$recall = \frac{RR}{R(u)} = \frac{RR}{RR + NRR} \quad (9)$$

定义 3 精确度(accuracy)就是推荐结果中相关结果占所有推荐结果的比重, 如式(10)所示:

$$accuracy = \frac{RR}{L(u)} = \frac{RR}{RR + IR} \quad (10)$$

定义 4 F 指标是一种综合考虑召回率和精确度的指标, 如式(11)所示:

$$F = \frac{2 \times recall \times accuracy}{recall + accuracy} \quad (11)$$



在本文中,如果算法 A 的召回率和精确度总体高于算法 B,认为算法 A 比算法 B 具有更高的推荐性能^[14]。

5.3 实验结果及分析

本文根据 LPOI 的密度,将 DBSCAN 聚类的邻域 ε 取 300, $\minPts$ 取 10,图 2 为聚类结果中部分 RPIP 显示结果,横纵坐标分别为 LPOI 的纬度和经度,由图 2 可知聚类效果还可以。

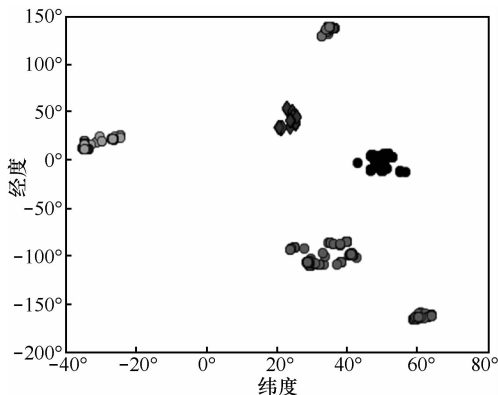


图 2 部分 RPIP 结果

在本实验中将数据集的 90%作为训练集,10%作为测试集。式(5)中的系数 $b \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ 。当 $b=0$ 时,推荐方法为只考虑用户好友相似性,而不考虑用户的位置签到信息,所得推荐列表仅与社交网络好友相似度有关。当 $b=1$ 时,推荐方法为只考虑位置相似性,所得推荐列表仅与用户的签到位置信息有关。其余几组数据表示综合考虑用户好友相似度和用户位置相似度,这就是本文提出的结合好友相似度和签到位置相似度的推荐算法。通过上述参数设置,可以对比得出本文算法中好友相似度和位置相似性所占比重对推荐结果产生的影响。推荐列表大小 N 取值为 25,系数 b 作为横轴,计算此推荐模型中所有用户的召回率和精确度,进而计算其平均 F 指标,如图 3 所示。

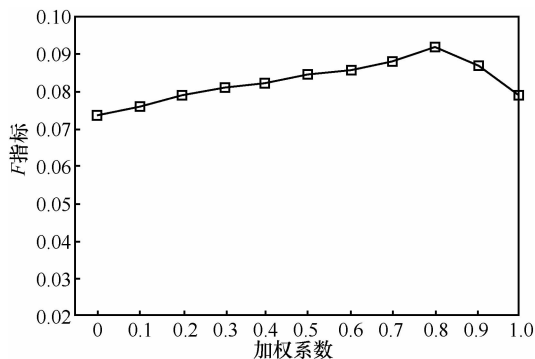


图 3 不同系数下的 F 指标

F 指标综合考虑了召回率和精确度,其大小可以衡量推荐算法的性能。由图 3 可知,完全基于位置相似度的推荐算法(即 $b=1$ 时)性能要略高于完全基于好友相似度的推荐算法(即 $b=0$ 时),同时,本文推荐算法的效果随着系数 b 的增加而先增加后下降,当 $b=0.8$ 时本算法的推荐质量最高。

由上述实验可知; $b=0.8$ 时,本推荐方法效果最好,因此以下实验中取 $b=0.8$ 。以下为本方法分别与基于 Jaccard 系数的推荐方法(利用 Jaccard 系数^[15]求位置相似度,从而得到用户间相似度,再根据用户间相似度的高低进行潜在好友推荐)和基于好友相似度的推荐方法(简称 UserRec,其方法是利用第 3.3 节计算用户间的相似度,再根据用户间相似度的高低进行潜在好友推荐)的对比试验。其中推荐列表个数 N 分别为 3、5、10、15、20、25,其精确度和召回率结果如图 4、图 5 所示。通过实验结果可以看出,本文提出的推荐算法的精确度和召回率比其他两种方法都高。虽然召回率在 $N=3$ 和 $N=5$ 时没有表现出很好的优越性,但是 $N>10$ 时开始明显优于上述其他两种方法。总体来说,本文推荐方法具有较好的性能。

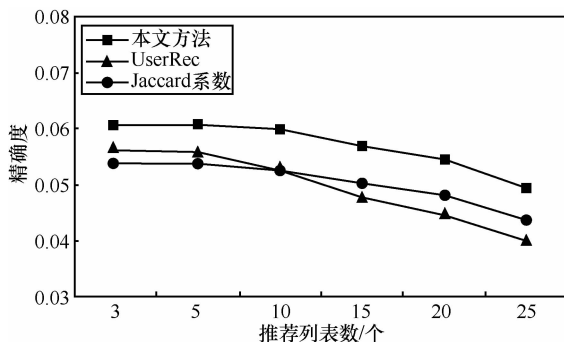


图 4 精确度结果

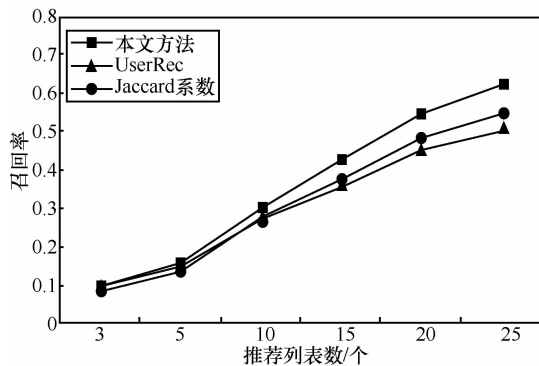


图 5 召回率结果

6 结束语

本文对潜在好友推荐模型进行了研究,旨在主动为用户有效信息,解决信息过载等问题,有效地为用户推荐合适的潜在好友,更好地发展用户的社交圈,同时为位置社交网络的健康发展贡献一份力量。首先,本文利用基于密度聚类的 DBSCAN 算法,将各个位置兴趣点聚类为不同的位置兴趣区域,这样比只用位置兴趣点进行相似度计算更加快捷准确;然后,再计算所有用户在各个位置兴趣区域的位置权重,如果用户在很多位置兴趣区域的位置权重相似,那么用户之间的相似度就很高;最后,得到用户间的位置相似度及好友相似度,再对它们进行权重叠加,进而得到潜在好友推荐列表,实现 TopK 推荐。利用真实数据对潜在好友推荐模型进行实验验证,结果表明,本文方法的推荐性能相比基于 Jaccard 系数的推荐方法和基于社交网络关系的推荐方法有明显提高,由此证明了该方法的有效性。

参考文献

- 翟红生, 于海鹏. 在线社交网络中的位置服务研究进展与趋势. 计算机应用研究, 2013, 30(11): 3221~3227
- Hruschka D J, Henrich J. Friendship, cliquishness, and the emergence of cooperation. Journal of Theoretical Biology, 2006, 239(1): 1~15
- Li Q, Zheng Y, Xie X, *et al.* Mining user similarity based on location history. Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, USA, 2008:31~34
- Hung C C, Chang C W, Peng W C. Mining trajectory profiles for discovering user communities. Proceedings of the International Workshop on Location Based Social Networks, New York, USA, 2009: 1~8
- Ma H, Zhou D, Liu C. Recommender systems with social regularization. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, New York, USA, 2011: 287~296
- Zhang D, Li N, Zhou Z H, *et al.* IBAT: detecting anomalous taxi trajectories from GPS traces. Proceedings of the 13th International Conference on Ubiquitous Computing, New York, USA, 2011: 99~108
- Zheng N, Li Q. A recommender system based on tag and time information for social tagging systems. Expert Systems with Applications, 2011, 38(4): 4575~4587
- Zha o K, Wang X, Yu M, *et al.* User recommendation in reciprocal and bipartite social networks—a case study of online dating. IEEE Intelligent Systems, 2013, 17(3): 29~30
- Liu B, Xiong H, Liu B, *et al.* Point-of-interest recommendation in location based social networks with topic and location awareness. Proceedings of SIAM International Conference on Data Mining, Austin, Texas, USA, 2013: 396~404
- 邢冬丽, 赵美红, 陈文成. 基于密度的 DBSCAN 算法. 计算机工程与应用, 2007, 43(20): 216~221
- Dunham H M. 数据挖掘教程. 郭崇慧, 田凤占, 靳晓明译. 北京:清华大学出版社, 2005
- Cho E, Myers S A, Leskovec J. Friendship and mobility: user movement in location-based social networks. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 2011: 1082~1090
- Yang T, Cui Y, Jin Y. BPR-UserRec: a personalized user recommendation method in social tagging systems. The Journal of China Universities of Posts and Telecommunications, 2013, 20(1): 122~128
- Spertus E, Sahami M, Buyukkokten O. Evaluating similarity measures: a large-scale study in the orkut social network. Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, New York, USA, 2005: 678~684
- Tan P N, Steinbach M, Kumar V. 数据挖掘导论. 范明, 范宏建译. 北京:人民邮电出版社, 2010

[作者简介]



孙晓晨,女,北京信息科技大学教授,主要研究方向为社交网络、数据挖掘。



徐雅斌,男,北京信息科技大学教授,主要研究方向为社交网络、云计算。

(收稿日期:2014-08-20)