

引用格式:段宗涛,蔡丹丹,唐蕾,等.基于预测的 LBSN 兴趣点推荐算法[J].微电子学与计算机,2019,36(1):66-69. [DUAN Z T, CAI D D, TANG L, et al. Prediction-Based Point of Interest Recommendation Algorithm in LBSN[J]. Microelectronics & Computer, 2019, 36(1): 66-69.]

基于预测的 LBSN 兴趣点推荐算法

段宗涛,蔡丹丹,唐蕾,李菲菲

(长安大学 信息工程学院,陕西 西安 710064)

摘 要: 兴趣点(Point-Of-Interest,POD)推荐是基于位置的社交网络中(Location-Based Social Networks,LBSN)一种重要的个性化推荐功能.本文提出基于预测的兴趣点推荐算法.该算法根据 LBSN 中用户历史 POI 数据分布学习用户出行行为,利用变阶的马尔科夫算法根据当前位置预测用户未来到达 POI 的语义信息,最终推荐时考虑用户签到次数的差异为用户推荐 N 个具有高兴趣度的 POI.实验结果表明:本文提出的算法在准确率和召回率上均高于两个对比算法,说明该算法提高了兴趣点推荐效果,并可以有有效的推荐给用户下一个访问的兴趣点.

关键词: 基于位置的社交网络;个性化推荐;位置预测;POI

中图分类号: TP391.3

文献标识码: A

文章编号: 1000-7180(2019)01-0066-04

Prediction Based on Point-of-Interest Recommendation Algorithm in LBSN

DUAN Zong-tao,CAI Dan-dan,TANG Lei,LI Fei-fei

(School of Information Technology Changan University, Xi'an 710064,China)

Abstract: Point-Of-Interest (POI) recommendation is an important personalized recommendation function in Location-Based Social Networks (LBSN). This paper proposes a prediction point based recommendation algorithm. The algorithm learns the user's travel behavior according to the user history POI data distribution in the LBSN, and uses the variable Markov algorithm to predict the semantic information of the user's future arrival POI according to the current location. The final recommendation takes into account the difference in the number of user sign-ups and recommends N for the user. Highly interesting POI. The experimental results show that the proposed algorithm is higher than the two comparison algorithms in accuracy and recall rate, which indicates that the algorithm improves the recommendation effect of interest points and can effectively recommend the next interest points for users.

Key words: LBSN; personalized recommendation; location prediction; POI

1 引言

个性化推荐是 LBSN 中的重要任务.兴趣点一般推荐的是用户从未去过但有可能感兴趣的点.然而,以前的方法主要考虑在整体签到点中为用户推荐用户最感兴趣的 POI,并没有考虑到用户的出行习惯^[1-2].而用

户每天的出行具有一定的规律性,他们的 POI 签到更是具有一定的兴趣偏好^[3].基于以上考虑,本文所针对的个性化推荐是为用户推荐用户即将签到的兴趣点.本文提出了一种基于预测的兴趣点推荐算法.在 LBSN 中利用用户历史 POI 数据分布学习用户出行行为,利用变阶的马尔科夫算法根据当前位置预测用户未来到

收稿日期:2018-05-20;修回日期:2018-06-15

基金项目:陕西省重点科技创新团队项目(2017KCT-29);陕西省重点研发计划项目(2017GY-072,2018GY-022,2018GY-032,2018GY-136);陕西省国际科技合作计划项目(2017KW-015);陕西省工业科技攻关项目(2015GY002)

达 POI 的语义信息,其次考虑用户签到次数的差异,基于用户好友的 POI 计算特定用户未来到达各候选兴趣点的签到概率,从而生成推荐列表。

2 语义 POI 预测

本文利用带有逃逸码的变阶马尔科夫模型进行下一个出行 POI 预测,解决预测算法的零频率问题。POI 的语义信息则作为推荐的关键约束,更准确的为用户推荐未知 POI^[4]。假设用户 u 的历史到访兴趣点集 $\text{POI} = \{\text{POI}_1, \text{POI}_2, \dots, \text{POI}_w\}$, 其中 POI_i 表示兴趣点名称, w 表示到访兴趣点总数。在一条序列 Seq_Loc 中,若给定位置 $\text{loc}_{n+1} \in \text{POI}$, 则 loc_{n+1} 的 N 阶上下文模型 loc_n^N 指的是 Seq_Loc 中以 loc_n 为下一个出行兴趣点,且长度为 N 的序列,即 $\text{loc}_n^N = \langle \text{loc}_{n-N+1}, \dots, \text{loc}_{n-1}, \text{loc}_n \rangle$ 。变阶的马尔科夫模型在每个上下文模型中都添加一个初始频度为 $N-1$ 阶的逃逸码,通过在用户所有历史轨迹集合 Seq 中搜索与 loc_n^N 具有相同长度的轨迹,根据满足条件的轨迹次数来预测位置点 loc_{n+1} 在紧邻时间的兴趣点中出现概率。对于轨迹 Seq_Loc , 本文用 P_i 表示用户到达 POI_i 的概率,则计算 P_i 的流程如下:

步骤 1: 当 Seq 中包含位置 POI_i 的上下文模型 loc_n^N , 预测 POI_i 的出现概率 P_i 为 $P(\text{loc}_{n+1} | \text{loc}_n^N)$, 并停止计算; P_i 由公式(1)表示:

$$P_i = P(\text{loc}_{n+1} | \text{loc}_n^N) = \frac{f(\text{loc}_{n+1} | \text{loc}_n^N)}{S(\text{loc}_n^N)} \quad (1)$$

式中, $f(\text{loc}_{n+1} | \text{loc}_n^N)$ 表示 Seq 中以 POI_i 为下一个出行兴趣点的上下文模型 loc_n^N 出现个数。 $S(\text{loc}_n^N)$ 表示 Seq 中以不同兴趣点为下一个出行兴趣点的上下文模型 loc_n^N 出现的总数。 $S(\text{loc}_n^N) = \sum_{i=1}^{|A(\text{loc}_n^N)|} f(x_i | \text{loc}_n^N)$, $x_i \in A(\text{loc}_n^N)$, 其中 $A(\text{loc}_n^N)$ 表示 Seq 中匹配上下文模型 loc_n^N 的不同的兴趣点集合, $|A(\text{loc}_n^N)|$ 为 Seq 中匹配上下文模型 loc_n^N 的不同的兴趣点的类型个数;

步骤 2: 当 Seq 中没有 loc_n^N 时, 逃逸码频度减 1, 同时上下文模型更改为 loc_{n-1}^{N-1} , 并计算逃逸码预测概率 $p(\text{esc} | \text{loc}_n^N)$, 由公式(2)表示:

$$p(\text{esc} | \text{loc}_n^N) = \frac{f(\text{esc} | \text{loc}_n^N)}{S(\text{loc}_n^N) + f(\text{esc} | \text{loc}_n^N)} \quad (2)$$

式中, $p(\text{esc} | \text{loc}_n^N)$ 表示上下文模型 loc_n^N 的逃逸码预测概率, $f(\text{esc} | \text{loc}_n^N)$ 表示上下文模型 loc_n^N 的逃逸码频度, 由 $|A(\text{loc}_n^N)|$ 计算, $f(\text{esc} | \text{loc}_n^N) = |A(\text{loc}_n^N)|$;

步骤 3: 再次统计 loc_{n-1}^{N-1} 的出现次数, 直到 Seq

中能匹配到 loc_{n-1}^{N-1} 时, 这时上下文模型为 loc_n^k , 输出预测 POI_i 的出现概率 P_i 由公式(3)表示:

$$P_i = p(\text{esc} | \text{loc}_n^{k+1}) * P(\text{loc}_{n+1} | \text{loc}_n^k) \quad (3)$$

式中, $0 < k < N$,

$$P(\text{loc}_{n+1} | \text{loc}_n^k) = \frac{f(\text{loc}_{n+1} | \text{loc}_n^k)}{S(\text{loc}_n^k) + f(\text{esc} | \text{loc}_n^{k+1})}$$

步骤 4: 如此重复统计, 直至逃逸码频度为 1, 如果最终没有搜索到与上下文模型匹配的轨迹, 则确定预测 POI_i 的出现概率为 P_i 由公式(4)表示:

$$P_i = \frac{1}{w} \quad (4)$$

系统中的每个兴趣点都对应分别的 POI 类别, 下一个出行兴趣点预测的目的是, 推荐给用户与预测概率最大的兴趣点具有相同 POI 类别的 POI。

3 Top-N 兴趣点推荐算法

在第 3 节中利用变阶马尔科夫算法结合 POI 序列, 预测当前位置的下一个出行 POI。而本文提出的 POI 推荐算法通过预测的 POI 语义产生候选 POI 集合, 通过计算兴趣度发现 TOP-N 个用户最可能去的 POI。针对这样的 POI, 用户具有很大概率会选择访问, 并添加到接下来的出行任务中^[5]。

一般来说, 用户对候选地点兴趣度可以用签到次数来衡量, 如果用户对某一个地点的签到次数越多, 说明这个用户对该地点的兴趣度越高, 仅仅通过单纯的次数并不能判断用户是否对该地点感兴趣^[6]。比如, 用户 u_i 在地点 l_j 签到 5 次, 相比其他用户, 用户 u_i 的签到次数高于其他用户, 但是, 用户在其他地点签到的次数都大于 5 次, 这说明用户 u_i 对地点 l_j 并不感兴趣。所以本文利用签到地点 l_j 在相同类型地点集合中所有签到地点的签到数总和的比重表示用户 u_i 对地点 l_j 的兴趣度。用户 u_i 对地点 l_j 的签到次数表示为 $\text{check_in}(u_i, l_j)$, 本文利用公式(5)表明 u_i 对地点 l_j 的兴趣度 $\text{Interest}(u_i, l_j)$:

$$\text{Interest}(u_i, l_j) = \frac{\text{check_in}(u_i, l_j) - \overline{\text{check_POI}_{l_j}}}{(\sigma^2)_{u_i}^{l_j}} \quad (5)$$

式中, $\overline{\text{check_POI}_{l_j}}$ 为用户 u_i 对与地点 l_j 具有相同 POI 类型的地点包括地点 l_j 在内的签到地点平均签到次数, $(\sigma^2)_{u_i}^{l_j}$ 为用户 u_i 对于与地点 l_j 具有相同 POI 类型的地点包括地点 l_j 在内的签到地点的签到次数的方差。

对于每个用户 $u \in U$, 都有其对应的朋友用户集合 F 。在第 3 节中预测到用户下一个出行最有可能的 POI 类型, 在 F 中每个用户对应一个位置集合, 其中每个位置都有一个 POI 类型标签。每个用户所

签到的 POI 类别所对应的次数由向量 $\overrightarrow{\text{POI}}$ 表示: $\overrightarrow{\text{POI}}_{u_i} = \{\text{POI}f_1, \text{POI}f_2, \dots, \text{POI}f_x\}$, 其中 $\text{POI}f$ 为用户出行中兴趣点类型的访问次数. 搜索与用户下一个 L_i 具有相同 POI 类型标签的未知地点放入候选位置集合中 L^* 中.

在用户 u_i 的候选位置集合 $L_{u_i}^*$ 中取一位置点 l_k , l_k 对应的用户 u_i 的好用用户集合为 F_i , 用户 u_i 有个 Y 好友用户 $F_i = \{v_1, v_2, \dots, v_Y\}$. 用户 u_i 对 l_k 的兴趣度 $\text{Interest}(u_i, l_k)$ 表示为公式(6):

$$\text{Interest}(u_i, l_k) = \frac{\sum_{v_j \in F_i} \text{Sim}(u_i, v_j) \cdot \text{Interest}(v_j, l_k)}{\sum_{v_j \in F_i} \text{Sim}(u_i, v_j)} \quad (6)$$

式中, $\text{Sim}(u_i, v_j) = \frac{\overrightarrow{\text{POI}}_{u_i} \cdot \overrightarrow{\text{POI}}_{v_j}}{\|\overrightarrow{\text{POI}}_{u_i}\| \|\overrightarrow{\text{POI}}_{v_j}\|}$. 通过计算用户与候选位置集中每个位置的兴趣度, 取 TOP- N 作为兴趣点推荐集合 $R(u)$.

4 实验结果与分析

4.1 实验数据集

本文利用两个典型 LBSN 服务平台 Four-square 和 Gowalla 的纽约真实签到数据. 每个数据集都包括签到记录和社交内容. 在获得数据集后, 对短时间内进行了大量签到的虚假行为进行了过滤预处理之后, 筛选用户在某一天中只有一个签到点的数据, 得到以下数据 Foursquare 的用户数为 3 357 个用户, 兴趣点为 3 543 个兴趣点, 签到数 168 297. 而 Gowalla 的用户数为 5 419 个用户, 兴趣点为 6 742 个兴趣点, 签到数 330 724 个.

4.2 推荐效果的评测方法

为了评价兴趣点推荐算法的效果. 本文选用推荐问题通用的指标: 准确率和召回率. 准确率指推荐结果中用户将来真正去的数量占推荐总数的比例, 反映了推荐的准确性. 召回率指推荐结果中用户将来真正去的数量占用户将来访问兴趣点总量的比例, 反映了推荐的全面性. 对用户 u 进行兴趣点推荐的准确率定义如公式(7)所示:

$$\text{Precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (7)$$

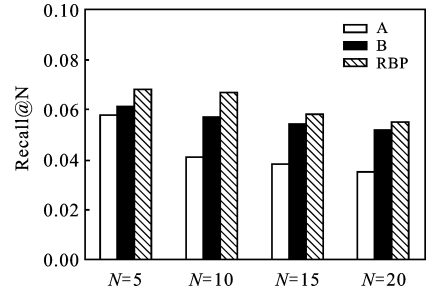
$$\text{Recall} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (8)$$

式中, $R(u)$ 为对用户 u 进行推荐的兴趣点集合; $T(u)$ 为用户 u 在测试集上实际的签到集合. 准确率和召回率相互制约, 综合利用二者可以对预测结果

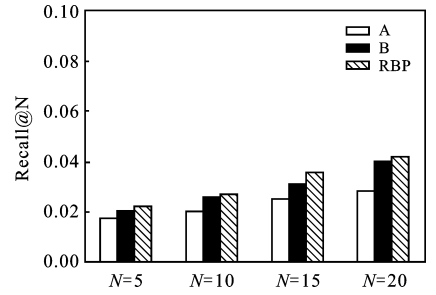
做出客观的评价.

4.3 实验结果与分析

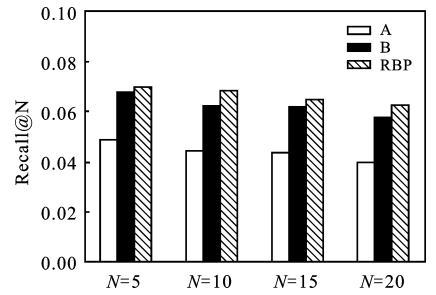
为验证本文提出的基于预测的兴趣点推荐算法的推荐效果以及引入相似用户的合理性, 本实验将提出的算法与两个典型兴趣点推荐算法进行比较和分析, 这两个算法一个是经典的协同过滤推荐算法, 一个是经典的基于用户评级的兴趣点推荐算法. 实验中将前一个算法作为算法 A, 后一个算法为算法 B. 本文实现并比较了这 3 种推荐算法, 推荐结果的准确率和召回率如图 1 所示.



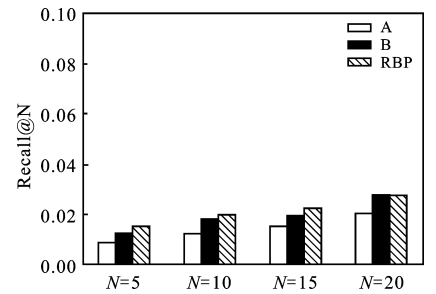
(a) Gowalla 纽约数据集推荐结果的准确率



(b) Gowalla 纽约数据集推荐结果的召回率



(c) Foursquare 纽约数据集推荐结果的准确率



(d) Foursquare 纽约数据集推荐结果的召回率

图 1 各算法 TOP- N ($N=5, 10, 15, 20$) 的推荐性能

图1展示了各算法TOP- N ($N=5, 10, 15, 20$)的推荐性能。

由图可以看出,无论 N 取何值,本文所提出的基于预测的兴趣点推荐算法在2个数据集上准确率和召回率普遍优于其他推荐算法。本文所提出的基于预测的兴趣点推荐算法有效地利用了用户的出行行为偏好,在签到数的计算方法上也较传统方法更为合理化。在算法应用研究中实现了个性化推荐,丰富了推荐系统的推荐结果,提高了推荐系统的综合质量。

5 结束语

LBSN的兴趣点推荐对基于位置的社交网络吸引用户和合作商家有着关键的作用^[7-8]。本文针对推荐内容的及时性提出了基于当前位置的POI推荐算法,该算法与前人研究不同的有两点,一是通过预测确定下一个兴趣点的服务类型,二是通过计算位置签到次数在所有签到数总和中的比重表示用户对地点的兴趣度。从而提高了推荐算法的准确性和召回率。实验表明,本文提出的算法在LBSN真实数据集上相比于其他推荐算法具有更高的准确率。

今后还可以从两方面展开进一步的研究:1、如何将时间因素应用到挖掘用户行为的模型中以提高推荐效果,以及好友推荐和活动推荐也是基于位置的社交网络重要的功能,值得进行更深入的讨论。

参考文献:

- [1] Sen Wang. A location recommendation algorithm based on location-based social networks [J]. Computer Engineering & Science, 2016, 38(4): 1007-130.
- [2] Deshpande M, Karypis G. Item-based top- N , recommendation algorithms [J]. Acm. Transactions on Information Systems, 2004, 22(1): 143-177.
- [3] Yoshii K, Goto M, Komatani K, et al. An efficient hybrid music recommender system using an incremen-

tally trainable probabilistic generative model[J]. IEEE Transactionson Audio Speech & Language Processing, 2008, 16(2): 435-447.

- [4] Berjani B, Strufe T. A recommendation system for spots in location-based online social networks [C]// Proc of the 4th Workshop on Social Network Systems. New York: ACM Press, 2011: 4-20.
- [5] Shi Yue, Serdyukov P, Hanjalic A, et al. Personalized landmark recommendation based on geotags from photo sharing sites [C]// Proc of ICWSM. Barcelona, Spain, 2011: 622-625.
- [6] Mao Ye, Yin Peifeng, Wang C L. Location recommendation for location-based social networks [C]// Proc of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2010: 458-461.
- [7] Ference G, Ye M, Lee W C. Location recommendation for out-of-town users in location-based social network [C]// Proc of the 22nd ACM International Conference on Information & Knowledge Management. San Francisco, 2013: 721-726.
- [8] Ye M, Yin P, Lee W C, et al. Exploiting geographical influence for collaborative point-of-interest recommendation [C]// Proc of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan, 2011: 325-334.

作者简介:

段宗涛 男, (1977-), 博士, 教授. 研究方向为交通信息综合处理及应用。

蔡丹丹 女, (1994-), 硕士研究生. 研究方向为数据挖掘、出行行为建模与分析。

唐 蕾(通信作者) 女, (1983-), 博士, 副教授. 研究方向为智能交通系统、交通信息服务计算。

E-mail: 824212713@qq.com.

李菲菲 女, (1993-), 硕士研究生. 研究方向为数据挖掘。