

基于活动的社交网络中的群组推荐算法设计^①

朱成纯, 张 谥

(复旦大学 计算机科学技术学院, 上海 200203)

摘 要: 在基于活动的社交网络(EBSN)中, 群组中聚集了具有相似兴趣的用户, 并为用户组织并举办线下活动, 在社区的发展中起到了至关重要的作用, 因而理解用户加入群组的原因和群组形成的过程在社交网络的研究中是一个重要的议题. 本文通过基于活动的社交网络中的一些相关内容信息, 比如社交网络中的标签信息和地理位置信息, 来辅助推荐系统更好地为用户预测对于群组的偏好. 本文提出了 SEGELER(pair-wiSE Geo-social Event-based LatEnt factoR)模型, 并使用这些社交网络中的信息, 来为用户的兴趣进行预测. 通过在真实的 EBSN 数据集上进行实验与验证, 本文的模型不仅可以有效提升对于用户偏好的预测, 也可以缓解冷启动问题.

关键词: 社交网络; 推荐系统; 群组推荐; 贝叶斯个性化排序; 潜在因子

引用格式: 朱成纯, 张谥. 基于活动的社交网络中的群组推荐算法设计. 计算机系统应用, 2017, 26(9): 103-108. <http://www.c-s-a.org.cn/1003-3254/5940.html>

Predicting User Preferences for Groups in Event-Based Social Networks

ZHU Cheng-Chun, ZHANG Mi

(School of Computer Science and Technology, Fudan University, Shanghai 200203, China)

Abstract: In event-based social networks (EBSN), groups that aggregate users with similar interests for sharing events play important roles in community development. Understanding why people join a group and how groups are formed is particularly an interesting issue in social science. In this paper, we study predicting users' preferences on social groups by considering content information in EBSN, i.e., geographic-social event-based recommendation. Specifically, we consider two types of content information, i.e., the tags and geographical event locations about users/events. We propose the SEGELER (pair-wiSE Geo-social Event-based LatEnt factoR) to model the users behavior considering the information. Experiments on a real-world EBSN social network validate the effectiveness of our proposed approach for both normal users and cold start users.

Key words: social network; recommender system; group recommendation; Bayesian personalized ranking; latent factor

在社交网络中, 群组是一个基本的组成部分, 并在用户社交行为的参与中起到了非常重要的作用. Charles Horton Cooley 提出, 线下的社交群组可以被归类为一级组, 二级组和引用组, 而对于线上的社交网络而言, 由于其相对更快速的用户变化和活动举办, 对线上群组的分类是一个更加复杂的问题. 在线上社交网络中, 用户可以以不同的兴趣和理由加入不同的群组. 例如,

用户会加入一个产品交流讨论组来为想购买的产品寻求建议和讨论, 而同时用户也因自己的阅读兴趣而加入了阅读小组. 而关于群组形成的研究^[1,2], 以及预测用户对群组偏好的研究^[3]在社交网络中已经成为了非常重要的研究课题.

本文研究在特定社交网络——基于活动的社交网络(EBSN)中预测用户对群组偏好. 用户, 群组, 活动是

^① 收稿时间: 2016-12-14; 采用时间: 2017-01-16

EBSN 中三种最基本的实体. 用户可以加入不同的群组, 并且参加由不同群组举办的不同活动. 所以在 EBSN 中, 用户之间的关系和联系也通过用户加入的群组和参与的群组活动所建立. 因而在 EBSN 中, 预测用户对群组的偏好对于帮助用户发现自己的兴趣和想参加的活动是非常有帮助的.

EBSN 研究中很重要的一方面就是纳入并应用相关的信息. 很多相关的研究提出了一些使用相关信息的方法, 比如具有普适性的相关信息应用方法^[4,5], 和针对特定相关信息的应用方法, 例如地点推荐研究^[6]和活动推荐研究^[7,8]. 其他一些研究标签信息的相关工作的效果^[3]和活动位置信息^[9]的效果也被研究与提出, 但这些方法并没有考虑到活动的情况, 因而无法很好的使用这些相关信息.

在 EBSN 社交网络中拥有很多相关信息, 例如标签信息、地理位置信息、时间信息等等, 其中地理位置信息作为 EBSN 中的重要信息, 用于帮助判断用户活动范围; 而标签信息作为对用户和群组的补充描述也有助于了解用户的兴趣. 本文同时纳入并使用了两种社交网络中的相关信息——用户和群组的标签信息和地理位置的活动信息, 来帮助更好的预测用户对 EBSN 社交群组的偏好. 本文提出了基于 EBSN 社交网络中地理位置信息和标签信息的 SEGELER (pairwise Geo-social Event-based LatEnt factoR)模型, 通过利用标签信息和地理位置信息来判断用户的行为. 标签的语义相似度信息和近邻影响在模型中被考虑, 而 pairwise 的目标函数也被提出用于最大化用户对群组的偏好. 对于 EBSN 经常受缺少负例的影响, 例如只有用户参与群组的信息被记录下来, 而用户讨厌或不喜欢群组的信息则会相对而言较少甚至缺失, 本文也提出了有效的算法来解决这一问题.

本文通过在真实的 EBSN 大数据集 meetup 上进行实验验证得出 SEGELER 模型比其他的算法有明显的提升, 而推荐系统中冷启动问题也可以通过使用这些社交网络中的信息来得到缓解.

1 符号定义

基于活动的社交网络(EBSN)通常拥有标签信息与地理位置的活动信息, EBSN 的数据可以被定义为集合 $\{U, G, E, T, L\}$, 其中 $U = \{u_i\}_{i=1}^{|U|}$, $G = \{g_i\}_{i=1}^{|G|}$, $E = \{e_i\}_{i=1}^{|E|}$, $T = \{t_i\}_{i=1}^{|T|}$, $L = \{l_i\}_{i=1}^{|L|}$ 分别代表了群组集合, 用户集合, 活动

集合, 标签集合与地点集合. 活动 E 可以被群组 G 举办, 而用户 U 可以参与不同的群组 G , 并根据自己的兴趣参加群组举办的活动. 对于用户 u , 他/她所加入的群组的集合记为 $GP_u \in G$, 而他/她不感兴趣的群组的集合记为 $GN_u \in G \setminus G_u$, 其中 $GP_u \cap GN_u = \emptyset$. 用户 u 和群组 g 所拥有的标签分别记为 T_u 和 T_g . 用户 u 所居住的地点记为 L_u , E_g 表示群组 g 举办的活动集合, E_u 表示用户 u 参与的活动集合, 活动 e 的地点记为 L_e . 用户, 群组和活动通常由其用户 ID, 群组 ID 和活动 ID 来表示, 而标签由有意义的描述用户和群组的词语组成, 地点则由其经纬度来表示. 图 1 展示了一个 EBSN 的数据结构.

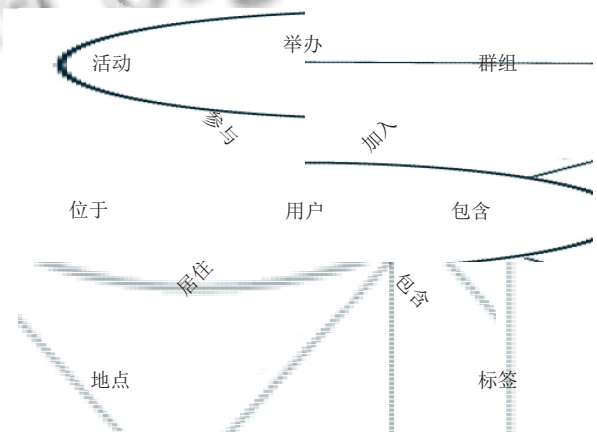


图 1 基于活动的社交网络基本结构

本文定义每一个用户和群组的活动范围为用户和群组去过的地点的集合:

$$ER_u = L_u \cup_{e \in E_u} L_e; ER_g = \cup_{e \in E_g} L_e$$

2 SEGELER 模型

对于预测用户感兴趣的群组的问题而言, 准确预测用户最感兴趣的前几位群组是非常重要的, 因而本文使用排序的目标函数来优化此问题, 即需要把用户感兴趣的群组排序高于用户不感兴趣的群组.

对于排序问题而言之前的研究提出了许多排序的方法, 例如 pairwise 排序^[10]和 listwise^[11]排序. 由于在本问题中只需要将感兴趣的群组排序靠前, 本文使用 pairwise 排序的方法使用户感兴趣的群组排序高于用户不感兴趣的群组. 即对于每一个用户 u , 以及 u 感兴趣的群组集合 GP_u 和不感兴趣的群组集合 GN_u , 需要使 $gp \in GP_u$ 的排序高于 $gn \in GN_u$. 本文基于贝叶斯个性化排序(BPR)^[12]方法提出本文的 pairwise 排序目标函数

来提取用户对群组的偏好特征,通过最大化如下的后验概率(MAP):

$$\begin{aligned} & \operatorname{argmax}_{\theta} \prod_{u \in U} \prod_{gp \in GP_u} \prod_{gn \in GN_u} \Pr(r_{u,gp} > r_{u,gn} | \theta) \Pr(\theta) \\ & \propto \operatorname{argmax}_{\theta} \prod_{u \in U} \prod_{gp \in GP_u} \prod_{gn \in GN_u} \sigma(\hat{r}_{u,gp} - \hat{r}_{u,gn}) \Pr(\theta) \quad (1) \\ & \propto \operatorname{argmax}_{\theta} \sum_{u \in U} \sum_{gp \in GP_u} \sum_{gn \in GN_u} \log \sigma(\hat{r}_{u,gp} - \hat{r}_{u,gn}) - \lambda \|\theta\|^2 \end{aligned}$$

其中 θ 表示相关的模型参数,这些参数会在后面的章节进行详细解释。 $\hat{r}_{u,g} \in R$ 表示用户 u 对群组 g 的估计偏好值,此值越大表示用户 u 对群组 g 的偏好越大。 $\sigma(x)$ 是sigmoid函数 $\sigma(x) = 1/[1 + \exp(-x)]$ 。 $\Pr(\theta) \sim N(0, \sigma^2)$ 表示SEGELER模型中参数的高斯先验概率分布,这部分会在后面进行详述。用户感兴趣的群组 GP_u 与不感兴趣的群组 GN_u 之间的预测值之间的差距应当变大。下一章将会详细地阐述 $r_{u,g}$ 的具体模型。

2.1 算法设计

潜在因子模型(LFM)假设用户和群组对每一个隐式因子都有偏好。对于每一个用户 u 和群组 g , $M_u \in R^k$ 和 $V_g \in R^k$ 代表 u 与 g 的隐式向量表示,即 u 与 g 在这个隐式因子下的偏好。 M_u 与 V_g 的乘积 $M_u^T V_g$ 表示用户 u 对群组 g 的偏好 $r_{u,g}$,即 u 对于 g 的偏好可以表示为:

$$r_{u,g} = M_u^T V_g \quad (2)$$

考虑标签信息和活动范围信息之后,本文提出公式(3)以更好的预测用户对群组的偏好:

$$M_u = \omega_1 v_{t_u} + \omega_2 v_{l_u} + \omega_3 P_u \quad (3)$$

$$V_g = \omega_1 v_{t_g} + \omega_2 v_{l_g} + \omega_3 Q_g$$

P_u 和 Q_g 代表用户 u 和群组 g 的通用潜在因子向量, $v_{t_u}, v_{t_g} \in R^k$ 代表与标签语义文字相关的潜在因子向量, $v_{l_u}, v_{l_g} \in R^k$ 代表与活动范围相关的潜在因子向量。参数 $\omega_i \in [0, 1]$ 调节每一部分的贡献,其中 $\sum \omega_i = 1$ 。 $v_{t_u}, v_{t_g}, v_{l_u}, v_{l_g}$ 的具体形式为:

$$v_{t_u} = \frac{1}{|T_u|} \sum_{t \in T_u} \pi_t U T_t; v_{l_u} = \frac{1}{|ER_u|} \sum_{l \in ER_u} \pi_l U L_l \quad (4)$$

$$v_{t_g} = \frac{1}{|T_g|} \sum_{t \in T_g} \pi_t G T_t; v_{l_g} = \frac{1}{|ER_g|} \sum_{l \in ER_g} \pi_l G L_l \quad (5)$$

其中 $U T_t, G T_t \in R^k$ 分别代表标签 t 用于描述用户和群组时的隐式因子表示, $U L_l, G L_l \in R^k$ 分别代表地点 l 用于描述用户和群组时的隐式因子表示。参数 $\pi_t, \pi_l \in [0, 1]$ 调节了每一部分的贡献,其中 $\sum \pi_t = 1$,

$$\sum \pi_l = 1.$$

此外,David^[13]提出拥有相似标签/活动范围的用户更有可能出现在相同的群组中,因此本文定义另两个隐式向量表示 $v_{nt_u}, v_{nl_u} \in R^k$ 以考虑用户的近邻对用户的影响:

$$v_{nt_u} = \frac{1}{|N_u|} \sum_{v \in N_u} v_{t_v}; v_{nl_u} = \frac{1}{|N_u|} \sum_{v \in N_u} v_{l_v} \quad (6)$$

在这里 v_{nt_u}, v_{nl_u} 代表用户 u 对于他/她近邻的标签信息和活动范围信息的隐式向量。 N_u 代表 k 个在 G_u 中与用户 u 最相似的近邻用户,用户之间的相似度由用户标签和活动范围的Jaccard系数计算得到:

$$\begin{aligned} \text{Jaccard}(u, v) &= \rho \text{Jac}_{T_{ag}}(u, v) + (1 - \rho) \text{Jac}_{ER}(u, v) \\ &= \rho \frac{|T_u \cap T_v|}{|T_u \cup T_v|} + (1 - \rho) \frac{|ER_u \cap ER_v|}{|ER_u \cup ER_v|} \end{aligned}$$

其中 T_u, T_v 和 ER_u, ER_v 分别代表用户 u, v 的标签集合和活动范围集合, $\rho \in [0, 1]$ 调节两部分的比重。

将公式(6)合并入公式(2),用户对群组偏好的式子可以延伸为:

$$\hat{r}_{u,g} = M_u^T V_g + v_{nt_u}^T v_{nt_g} + v_{nl_u}^T v_{nl_g} \quad (7)$$

2.2 参数学习

为求得公式(1)中的参数,本文采用了SGD(随机梯度下降)算法来最大化公式(1)。在每一轮SGD迭代中,通过随机选择一个用户 u ,用户感兴趣的正例群组 gp ,与用户不感兴趣的负例群组 gn ,组成一个三元组 $\{u, gp, gn\}$,并使用梯度下降法来求解参数 θ :

$$\theta = \theta + \alpha \left(\frac{\partial}{\partial \theta} \log \sigma(e_{u,gp,gn}) - \lambda \theta^2 \right) \quad (8)$$

其中 $e_{u,gp,gn} = \hat{r}_{u,gp} - \hat{r}_{u,gn}$, α 是SGD的学习步长。

在SGD学习的过程中,对于每一个用户,本文只随机选取一小部分的负例(u, gn)来减少计算的复杂度,这样算法的计算复杂度为 $O(|D|(|L|+|T|))$,其中 $|D|$ 是训练集的大小。在实验中,对于每一个用户 u 和 u 感兴趣的正例群组 gp ,本文选取 c 个 u 不感兴趣的负例群组 gn ,加入训练集。 c 值越大,模型训练与收敛时间越长,效果越好。通过平衡训练时间与最后的模型效果,最终本文设置参数 $c=5$ 作为选取负例的数量。

在EBSN中,负例即用户不喜欢的群组很少被记录下来,这对SEGELER模型的学习和参数优化提出了很大的挑战。本文提出了一个有效的策略来选择潜在负例。此外,考虑到标签 w 和标签 t 之间拥有语义相

似度, 本文为 $\{UT_t\}_t$ 引入了一个相应的先验概率, 标签 u 和标签 t 之间的语义相似度 $s_{w,t}$ 由 WordNet 计算得到, N_t 表示和标签 t 相似度高的标签集合.

$$\Pr(UT_t|N_t) = N\left(\frac{\sum_{w \in N_t} s_{w,t} UT_w}{\sum_{w \in N_t} UT_w}, \sigma^2\right) \quad (9)$$

2.3 负例选择

由于用户加入不感兴趣的群组的概率较低, 可以把一些偏好值较低的群组 $\{(u, g)\}$ 作为负例, 本文提出了如下的相似度计算方法来得到用户对群组的偏好值:

$$S(u, g) = \max(Jac_{Tag}(u, g), Jac_{ER}(u, g), S_{dis}(u, g)) \quad (10)$$

其中 $Jac_{Tag}(u, g)$, $Jac_{ER}(u, g)$ 分别代表用户 u 和群组 g 的 Jaccard 相似度, 由 u 与 g 的标签信息和活动范围信息算得. $S_{dis}(u, g) = 1/(1+d(u, g))$ 代表用户 u 和群组 g 的地理位置距离, 其中群组的地理位置距离由这个群组举办的活动的地理位置的平均值代表, 而设定 $S(u, g)$ 小于一定阈值 ζ 的群组为负例.

这个策略在试验中被证明十分有效, 并比其他的策略效果, 比如 Kabbur^[14] 和 Cheng^[6] 更好. 本文在实验中设定 $\zeta=0.015$ 作为选取负例的阈值.

3 实验与分析

3.1 数据集介绍

本文在真实的 EBSN 大数据集 Meetup^[15] 中验证了模型的效果, Meetup 是一个在线社交活动网站, 帮助用户发布和参与到线下的活动中. 由于只有 1.7% 的活动是跨城市的活动, 本文从 Meetup 中选取了三个大城市 LA(洛杉矶), SF(旧金山) 和 NYC(纽约) 作为三个数据集来验证 SEGELER 模型.

3.2 对比算法介绍

本文采用并实现了六个推荐系统领域的算法作为对比算法.

基于用户的协同过滤算法(UBCF)通过将用户的邻居信息考虑其中来预测用户对群组的偏好. 由于在 EBSN 中, 用户行为容易受到邻居影响, UBCF 很适合作为这个问题的对比算法.

SVD++把用户的隐式反馈加入了考虑中, 在矩阵分解中提供了额外的考虑信息. 在 EBSN 中, 用户的隐式反馈对于预测群组偏好也是很有帮助.

概率矩阵分解(PMF)是一个基于概率的因子算法,

在 EBSN 中效果很好并被广泛应用.

NSVD 是一个基于物品-物品因子的协同过滤算法, NSVD 使用了用户的过往记录和相关信息.

BPR-MF 是一个潜在因子协同过滤算法, 通过使用 BPR 优化来针对个性化的排序问题.

PTARMIGAN(PTA)是一个 pairwise 的基于标签信息和特征的矩阵分解算法, PTA 使用了诸如标签信息和地理位置信息的相关信息来预测用户对群组的偏好, PTA 也是为 EBSN 进行特定设计的预测算法.

3.3 评价指标介绍

本文使用了三个评价指标——准确率(precision), 召回率(recall)以及 NDCG, 来为所有方法的 top-k 预测进行衡量与评价.

准确率 $Precision = |A \cap B|/|A|$ 测量了一份预测列表 A 中正确的群组数 B 占的比重, 召回率 $Recall = |A \cap B|/|B|$ 则测量了所有正确的群组中被预测到的群组所占的比重. NDCG 则考虑到了排序的效果, 比如更相关的群组应当在推荐的列表中排位更高. NDCG 值的计算方式如下:

$$NDCG_K = \frac{DCG_K}{IDCG_K} \quad (11)$$

$$DCG_K = \sum_{i=1}^k \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

rel_i 在排序第 i 的群组相关时为 1, 否则为 0, $IDCG_k$ 是对于一个推荐群组的序列中最大的可能的 DCG 值. 注意到准确率, 召回率, 以及 NDCG 值越大, 则说明算法的效果越好.

对于每一个数据集, 本文随机选择了 80% 的用户和与之相关的群组/活动/标签作为训练集, 而将剩下的作为测试集. 公式(1)中的 λ 在 $[0, 0.02]$ 中取值, 所有模型的实验都经过了交叉验证.

3.4 SEGELER 实验结果分析

本文在三个数据集 LA, SF 和 NYC 上将 SEGELER 模型与六个对照模型进行了比对. 图 2 与图 3 列出了所有模型的准确率, 召回率以及 NDCG 值.

图 3 的结果可以得出, SEGELER 模型在三个数据集下的 NDCG 值的评价指标下明显超过了其他的对照算法. 例如, SEGELER 模型在 LA 数据集上的 NDCG 值相较其他对照算法至少 17%. 而图 2 中准确率和召回率的结果也验证了 SEGELER 模型相较其他模型拥有非常大的提升. 这说明通过利用好标签信息

和地理位置信息. 并且最大化用户喜欢的群组和不喜欢的群组间的不同, SEGELER 模型可以将用户喜欢的群组排序更高, 从而得到更好的预测结果.

得知, 本文的负例选择算法相较其他算法对最后结果有更大的提升. 这也说明了在 EBSN 中, 用户通常更喜欢更相似的群组, 这个相似度可以通过不同方面来计算, 诸如相似的标签或相似的地点.

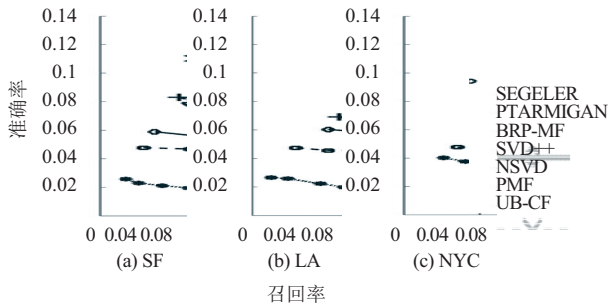


图2 各个模型的 Precision 和 Recall 值

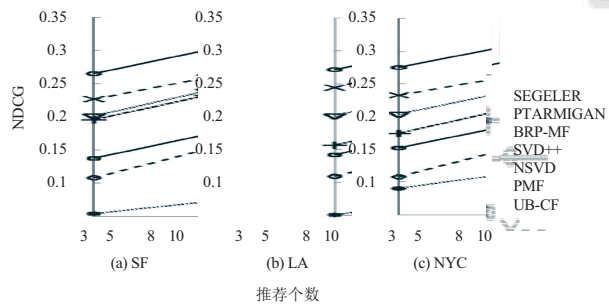


图3 各个模型的 NDCG 值

3.5 冷启动问题分析

本文也同样针对冷启动用户的预测效果进行了实验研究. 在此实验中, 本文将参与群组数较少的用户定义为“冷启动用户”. 由于在数据集中, 每一个用户平均参与了 40 个群组, 本文中参与少于 10 个群组的用户和少于 15 个群组的用户定义为两种情况下的“冷启动用户”. 图 4 展示了在 LA 数据集上 top-3 准确率上不同算法的结果. 从图 4 中可以看出, 在两种情况下, SEGELER 模型都比其他对照算法效果更好, 并取得了至少 48.21% 的提升. 这说明通过对于相关标签信息和地理位置信息的挖掘, SEGELER 算法可以更好的帮助新用户来选择感兴趣的群组.

3.6 负例选取方法实验分析

最后本文比较不同的负例选择方法对实验结果产生的影响. 本文将本文提出的负例选择算法与随机负例选择法, 只通过标签信息计算得到的标签策略, 只通过活动信息计算得到的活动策略, 以及只通过地理位置信息计算得到的地点策略进行比较. 图 5 展示了在 LA 数据集上不同负例选择方法的效果, 从图 5 中可以

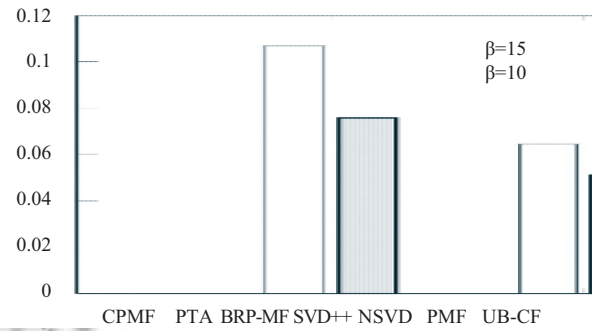


图4 SEGELER 模型在冷启动用户上的 Precision

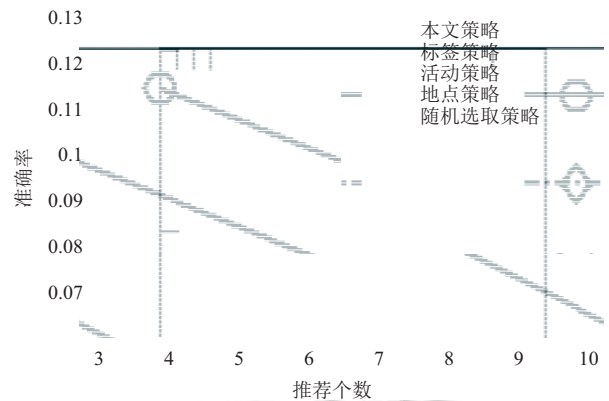


图5 不同负例选取策略的 Precision 值

4 结语

本文研究在社交网络中为用户预测对群组的偏好的问题, 并利用了 EBSN 中的标签信息以及地理位置信息来提升预测的效果. 本文提出了 SEGELER(pair-wiSE Geo-social Event-based LatEnt factoR)模型将这些相关信息进行使用来更好地分析用户偏好, 并提出了基于贝叶斯个性化排序的目标函数来优化预测用户在社交网络中喜欢的群组对于社交网络中负例缺失的问题, 本文提出了有效的算法来解决这一问题. 通过 EBSN 数据集 Meetup 上的实验, 验证了提出的算法在效果上的提升, 同时可以缓解冷启动问题. 在未来工作中, 作者会继续分析并纳入更多社交网络中相关信息, 诸如时间信息和情感信息, 来提升预测的效果. 同时除了预测准确度之外, 其他的评价标准诸如预测区分度也将进行研究.

参考文献

- 1 Sun YZ, Tang J, Han JW, *et al.* Co-evolution of multi-typed objects in dynamic star networks. *IEEE Trans. on Knowledge and Data Engineering*, 2014, 26(12): 2942–2955. [doi: [10.1109/TKDE.2013.103](https://doi.org/10.1109/TKDE.2013.103)]
- 2 Dong YX, Zhang J, Tang J, *et al.* CoupledLP: Link prediction in coupled networks. *Proc. of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Sydney, NSW, Australia. 2015. 199–208.
- 3 Zheng N, Li QD, Liao SC, *et al.* Flickr group recommendation based on tensor decomposition. *Proc. of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Geneva, Switzerland. 2010. 737–738.
- 4 Rendle S, Gantner Z, Freudenthaler C, *et al.* Fast context-aware recommendations with factorization machines. *Proc. of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Beijing, China. 2011. 635–644.
- 5 Pham TAN, Li XT, Cong G, *et al.* A general graph-based model for recommendation in event-based social networks. *Proc. of 2015 IEEE the 31st International Conference on Data Engineering*. Seoul, South Korea. 2015. 567–578.
- 6 Cheng C, Yang HQ, Lyu M R, *et al.* Where you like to go next: Successive point-of-interest recommendation. *Proc. of the 23rd International Joint Conference on Artificial Intelligence*. Beijing, China. 2013. 2605–2611
- 7 Ji XC, Qiao Z, Xu MZ, *et al.* Online event recommendation for event-based social networks. *Proc. of the 24th International Conference on World Wide Web*. Florence, Italy. 2015. 45–46.
- 8 Qiao Z, Zhang P, Cao YN, *et al.* Combining heterogenous social and geographical information for event recommendation. *Proc. of the 28th AAAI Conference on Artificial Intelligence*. Québec City, Québec, Canada. 2014. 145–151.
- 9 Zhang W, Wang JY. A collective bayesian poisson factorization model for cold-start local event recommendation. *Proc. of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Sydney, NSW, Australia. 2015. 1455–1464.
- 10 Sellamanickam S, Garg P, Selvaraj SK. A pairwise ranking based approach to learning with positive and unlabeled examples. *Proc. of the 20th ACM International Conference on Information and Knowledge Management*. Glasgow, Scotland, UK. 2011. 663–672.
- 11 Cao Z, Qin T, Liu TY, *et al.* Learning to rank: From pairwise approach to listwise approach. *Proc. of the 24th International Conference on Machine Learning*. Corvallis, Oregon, USA. 2007. 129–136.
- 12 Rendle S, Freudenthaler C, Gantner Z, *et al.* BPR: Bayesian personalized ranking from implicit feedback. *Proc. of the 25th Conference on Uncertainty in Artificial Intelligence*. Montreal, Quebec, Canada. 2012. 452–461.
- 13 Crandall DJ, Backstrom L, Cosley D, *et al.* Inferring social ties from geographic coincidences. *Proc. of the National Academy of Sciences of the United States of America*, 2010, 107(52): 22436–22441. [doi: [10.1073/pnas.1006155107](https://doi.org/10.1073/pnas.1006155107)]
- 14 Kabbur S, Ning X, Karypis G. Fism: Factored item similarity models for top-n recommender systems. *Proc. of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Chicago, Illinois, USA. 2013. 659–667.
- 15 Liu XJ, He Q, Tian YY, *et al.* Event-based social networks: Linking the online and offline social worlds. *Proc. of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Beijing, China. 2012. 1032–1040.