

利用最近邻域推荐且结合情境感知的个性化推荐算法

张宏丽¹ 白翔宇² 李改梅¹

(内蒙古师范大学教育信息技术学院 呼和浩特 010022)¹

(内蒙古大学计算机学院 呼和浩特 010021)²

摘 要 针对传统情境感知推荐算法推荐精确度低和适用环境受限等问题,提出了一种可行的解决方案。该方案可以根据检测到的情境信息找到相关的媒体内容,比仅依赖特征提取的方案更有效。首先,利用情境数据和搜索信息来识别所选项的情境与特定情境中用户的兴趣度之间的隐藏关系,并构建未知排名的推荐模型。然后,通过使用给定的情境列表来计算用户对项目的预期排名分数,从而进行情境感知评级。根据用户的情境参与选择新项目,从而使检测到的情境有助于促进对相关项目的搜索。进一步使用优化函数来最大化结果推荐的平均精度(MAP)。实验结果表明,与目前较为先进的两种算法相比,提出的方法表现出了比传统协同过滤算法更好的性能,且分别使平均绝对误差值降低了 1.8%和 1.2%,在推荐精确度和召回率方面也均优于两种对比方法。

关键词 个性化推荐算法,最近邻域推荐,隐式兴趣度,情境感知

中图法分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.04.037

Personalized Recommendation Algorithm Based on Recent Neighborhood Recommendation and Combined with Context Awareness

ZHANG Hong-li¹ BAI Xiang-yu² LI Gai-mei¹

(College of Educational Information Technology, Inner Mongolia Normal University, Hohehot 010022, China)¹

(College of Computer Science, Inner Mongolia University, Hohehot 010021, China)²

Abstract Aiming at the problem that the traditional context-aware recommendation algorithm is not accurate and the applicable environment is limited, this paper proposed a feasible solution. It can find relevant media content based on the detected context information, which is more effective than relying solely on feature extraction. First, the context data and the search information are used to identify a hidden relationship between the selected context and the user's interest in a particular context, and a recommendation model of the unknown ranking is constructed. The contextually perceived rating of the user's expected ranking score is then calculated by using the given contextual list. The user's situation is used to participate in the selection of new items so that the detected situation helps to facilitate the search for related items. The optimization function is further used to maximize the average accuracy (MAP) of the result recommendation. The experimental results show that compared with the more advanced algorithms, the proposed method shows better performance than the traditional collaborative filtering algorithm, and the absolute error value is reduced by 1.8% and 1.2% respectively in the recommendation accuracy and recall. The rate is also superior to the two comparison methods.

Keywords Personalized recommendation algorithm, Recent neighborhood recommendation, Implicit interest, Context awareness

情境感知推荐系统有助于针对特定项目预测特定用户的兴趣或偏好,建立用户兴趣模型的目的是从用户情境和用户行为中捕获用户的兴趣。情境感知推荐提供了挖掘社会内容的潜力,并利用相关标签和评级信息,根据给定的情境个性化地搜索内容。推荐系统解决了试图从大量在线选择中识别相关资源的问题。例如,用户可能喜欢特定的歌曲,并且可能给

该歌曲非常高的分数,但是他/她可能不想每天听它。在动态的普适计算环境中,用户任务的精确执行用户的个性化信息以及环境的情境紧密相关^[1-2]。

近年来出现了大量使用情境信息的算法来增强传统推荐方法的推荐性能,很多学者也对情境感知用户兴趣建模进行了多种方式的探索,以期提供更好的推荐结果。例如,文献

到稿日期:2018-09-27 返修日期:2018-12-11 本文受内蒙古自治区自然科学基金项目(2015MS0634),内蒙古自治区高等学校科学技术项目(NJZY033)资助。
张宏丽(1974—),女,博士,副教授,主要研究领域为个性化推荐、人工智能等;白翔宇(1974—),男,博士,教授,主要研究领域为数据挖掘、人工智能等;李改梅(1968—),女,硕士,副教授,主要研究领域为计算机应用,E-mail:zhouzihui1974@163.com(通信作者)。

[3]提出结合用户显式和隐式兴趣度的计算,精确执行用户的个性化信息,并根据给定情境的个性化搜索内容,建立一种M-C-W用户兴趣模型。该模型从多角度综合计算用户的兴趣度,挖掘出用户的显式兴趣和隐式兴趣的关联性,实时且准确地表达用户的兴趣爱好,实现多维动态情境兴趣的结合。文献[4]在基于语义关联的实例相似度算法的基础上,提出了一种基于语义关联和情景感知的推荐方法,其通过情景感知推荐提升了系统的情景敏感性。文献[5]融合时间情境因素,提出了一种基于情境信息降低维度的关联规则推荐模型;为了找到内容的情境或用户的情境,以视频网站的Web日志为数据源,实现了基于时间情境划分的关联规则推荐算法,并将其与传统推荐算法进行了对比分析。文献[6]针对当前各种情境下推荐系统存在的问题,从实际应用的角度考虑,将传统推荐算法和情境感知作为研究前提,提出了一种推荐精度和用户满意度均较高的改进推荐算法,并将其应用于实际。这几种算法在处理多维数据时都表现出了较为出色的性能,但没有考虑用户所处的实际场景,用户兴趣变化通常会导致传统信息推送系统预测用户兴趣的准确性下降,因此需要新的方式来解决情境信息引入而导致的多维数据处理问题。用户兴趣会因情境、时间、行为等因素而发生变化,具有多样性和动态性。

与所提到的其他情境感知推荐算法不同,本文提出的算法更加重视个人用户的实际情境,包括实际场景(不需要极度耗费资源的可用在线内容的特征提取过程来分析情境);构建了3个模型来搜索其他类似用户、项目和情境日志中的隐藏偏好,依赖于社交网络上可用的在线用户项目对媒体信息进行注释和分类。本文的主要创新点如下:

- 1)提出了一个模型,用于识别用户对情境的隐式兴趣度以及用户对新项目和选定项目的隐式兴趣度。该模型使用社交标签来探索隐式兴趣度,以反映收集的情境信息。
- 2)提出了基于项目的最近邻域推荐,它弥合了媒体资源、用户个人、联合兴趣度以及识别的情境信息之间的差距,并通过最大化平均精度均值(Mean Average Precision, MAP)以及识别用户查询的相关项来评估所提出的算法。

1 构建模型

本文所提出的推荐方法利用情境数据和搜索来识别所选项的情境与特定情境中的用户的兴趣度之间的隐藏关系。这两种类型的目标关联用于构建不同可能情境中的项目的未知排名的推荐模型。然后,本文使用优化函数来最大化所推荐结果的MAP度量。

所提模型中考虑的情境参数是位置、事物、人、时间、用户配置文件以及天气等。给定与用户(u)和项目(i)交互且关联的情境标签,识别给定用户感兴趣的项目列表 I_x ,其中 $|I_x| \leq k$ 且评级 R_{u,i_x} 未知。因此,提出的算法将隐藏的情境 c_2 暴露给项目 i_1 用于 u_1 的推荐,并且可以为用户 u_2 推荐 i_1 。

将可能的情境列表表示为 $C=\{c_1, c_2, \dots, c_{|C|}\}$,可能的项目(资源)表示为 $I=\{i_1, i_2, \dots, i_{|I|}\}$,并且将用户的集合表示为 $U=\{u_1, u_2, \dots, u_{|U|}\}$ 。本文的推荐问题具有3个维度,可

以使用一个具有 $|U| \times |I| \times |C|$ 个实体的频率模型来进行求解。项目 I 指的是应用程序可以推荐的任何资源,例如电影、音乐或新闻等。本文将张量模型的已识别特征分解为3个不同的矩阵:用户项、情境项和情境用户矩阵;还构建了其他支持的矩阵,以发现每个单独维度之间的相关相似性。表1列出了本文中使用的符号。

表 1 本文所使用的符号

Table 1 Symbols used in this article

符号	意义
U	用户的集合
I	项目的集合
C	情境的集合
$A_{ C \times I }$	情境-项目矩阵
$B_{ U \times C }$	用户-项目矩阵
$R_{ U \times I }$	用户-情境矩阵
$S_{ U \times U }$	用户-用户相似性矩阵
$E_{ I \times I }$	项目-项目相似性矩阵
$Q_{ C \times C }$	情境-情境相似性矩阵
$U\mathbf{r}C_{ U \times C }$	用户对情境的隐式兴趣度
$C\mathbf{r}I_{ C \times I }$	情境对项目的隐式兴趣度
$U\mathbf{r}I_{ U \times I }$	用户对项目的隐式兴趣度

2 多重情境推荐

在特定情境中选择项目的用户可能在类似的情境中选择类似的项目。本文的主要技术问题是确定用户和项目的情境的潜在特征。为了解决这个问题,分析了与数据集中的交互相关的情境信息,对情境选择的模式进行了跟踪,以填补用户和新项目之间的差距以及项目和新情境之间的差距。为了更好地反映用户的个性化偏好,计算用户的隐式兴趣偏好,受文献[7]中三维推荐模型的启发,在情境的集合 C 中, U 中的用户有一些项目,其中用户的偏好未知。但是,可以建立一个有效的隐式用户兴趣模型。用户的隐式模型如图1所示。构建3个潜在模型来表示用户对情境的隐式兴趣度 $U\mathbf{r}C_{|U| \times |C|}$ 、情境对项目的隐式兴趣度 $C\mathbf{r}I_{|C| \times |I|}$,以及用户对项目的兴趣度 $U\mathbf{r}I_{|U| \times |I|}$ 。

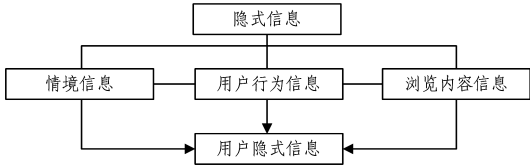


图 1 用户的隐式兴趣模型

Fig. 1 User implicit interest model

2.1 相似性度量

为了在本文所提出的模型中找到隐式兴趣度,首先为每个单独的维度计算3种类型的相似性:用户-用户、项目-项目和情境-情境。可以从原始3-D张量的一组分解矩阵中计算相似性度量。首先构造用户项矩阵 $R_{|U| \times |I|}$,计算用户 $S_{|U| \times |U|}$ 之间的相似性。矩阵 R 可以由布尔函数创建,其中矩阵中的每个真值表示用户 u 选择项目 i 的事实,而不管其情境。它也可以根据用户分配给项目的评级值进行构建,这也是本文采用的方法。类似地,可以根据相同的矩阵 $R_{|U| \times |I|}$ 来计算项目的相似性 $R_{|I| \times |I|}$ 。为了找到情境的相似性,本文创

建了解析矩阵 $\mathbf{A}_{|C| \times |I|}$, 表示在特定情境中选择项目的频率。情境的相似性通过确定项目在每个情境中出现的频率来显示两个情境的语义关系。因此, 经常与项目一起使用的情境描述方法可以在文献[8-10]中找到。注意, 情境的相似性是根据项目而不是用户来计算的, 因为用户对情境的相关行为是不可靠的。

为了确定属于维度的条目在任何两个向量 \mathbf{v}_x 和 \mathbf{v}_y 上的相似性, 可以使用各种相似性度量指标进行度量, 例如余弦相似性^[11]、Jaccard、Overlap^[12], 以及 Pearson 相关系数^[13]。本文使用余弦相似性度量(见式(1))和其他相似性度量(如文献[21]中提出的相似性度量)来研究算法的性能。

$$\text{sim}(\mathbf{v}_x, \mathbf{v}_y) = \cos(\mathbf{v}_x, \mathbf{v}_y) = \frac{\mathbf{v}_x \cdot \mathbf{v}_y}{\|\mathbf{v}_x\|_2 \cdot \|\mathbf{v}_y\|_2} \quad (1)$$

2.2 基于项目的最近邻域推荐

当前, 百万级甚至更高用户数量的电商网站等平台一般都会采用基于项目的最近邻推荐, 该方法是一种能够适应相当数量级的评分矩阵, 并且依然能够做到实时计算和推荐, 在线下可以方便地实现对数据的预处理。

基于用户的最近邻域推荐在推荐时需要扫描当前用户全部的潜在邻。假设两个项目 a 和 b 对应的评分向量为 \vec{a} 和 \vec{b} , 则其余弦相似度的定义如下:

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|} \quad (2)$$

其数值介于 0 到 1 之间, 数值越大表示相关性越高, 当评分矩阵量级逐渐变大时, 计算效率也会逐渐下降^[14-15]。改进的余弦相似度计算方法如下^[16-17]:

$$\text{sim}(a, b) = \frac{\sum_{u \in U} (r_{a,u} - \bar{r}_u)(r_{b,u} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{a,u} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{b,u} - \bar{r}_u)^2}} \quad (3)$$

其中, u 为当前用户, p 为待预测项目, U 为所有同时给物品 a 和 b 评分的用户集。这种余弦相似度计算考虑到了用户评分的平均值之间的差异。

确定项目间的相似度后, 就需要预测当前用户对待预测用户的评分, 这需要计算 u 对所有 p 的加权评分总和, 如式(4)所示:

$$\text{pred}(u, p) = \frac{\sum_{i \in \text{ratedItems}(u)} \text{sim}(i, p) * r_{u,i}}{\sum_{i \in \text{ratedItems}(a)} \text{sim}(i, p)} \quad (4)$$

近邻集合也会受限于固定的阈值, 这就如同基于用户的推荐方法, 以免产生数据量过大而引发的计算效率过低的问题。

2.3 搜索用户隐式兴趣度

通过查看情境的不同属性的用户和项目之间的关系, 可能会发现在特定情境中选择项目的隐藏原因, 以及用户更喜欢在给定情境中选择某些项目的隐藏原因。寻找这些隐藏原因可能会涉及对人类行为和反应的分析。但是, 通过使用创建的相似性矩阵, 本文方法可以构建一个模型, 该模型反映给定用户对情境下隐式兴趣度, 以及给定情境下给定项目的隐式兴趣度。

前面提到可以通过了解用户在项目选择方面的行为来预测用户在当前情境中的潜在偏好。矩阵 $\mathbf{UtC}_{|U| \times |C|}$ 根据式

(5), 捕获用户 u 对另一个类似情境的情境偏好。潜在的用户情境偏好是 $\bar{\mathbf{B}}$ 的归一化频率矩阵和相似性矩阵 $(\mathbf{Q})^T$ 的转置的乘积。 \mathbf{B} 的矩阵归一化, 使得列矢量获得了归一化, 从而使得频率值归一化到 0 和 1 之间, 如式(6)所示。设 $n_{u,c}$ 表示用户 u_x 所选项目列表中情境 c_y 出现的次数, N_{u,c_y} 表示所有用户列表中情境 c_y 选择的项目数, 如式(7)所示。

$$\mathbf{UtC}(u, c) = \sum_{m=1}^{|C|} (\bar{\mathbf{B}})_{u,m} \times (\mathbf{Q})_{m,c}^T \quad (5)$$

$$b(u_x, c_y) = \frac{n_{u,c}(u_x, c_y)}{N_{u,c_y}} \quad (6)$$

$$N_{u,c_y} = \sqrt{\sum_{x=1}^{|U|} (\delta_{x,y} f_{x,y})^2} \quad (7)$$

$$\delta_{x,y} = \begin{cases} 1, & c_y \text{ occurred in } u_x \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

如果情境类似于给定的特定情境, 则使用式(5)捕获用户 u 对类似于另一个情境的情境偏好。本文分析了特定情境是如何影响用户的项目选择的。情境 c_y 对用户的情境越相似, 它对用户的兴趣度值的影响就越大。规范化步骤减小了许多用户选择的情境对较少用户选择的情境的影响。

同样, 还可以预测项目对其检测到的情境的隐式兴趣度。这是一种捕获特定情境的行为方式, 用于优化给定情境中用户的项目选择。矩阵 $\mathbf{CtI}_{|C| \times |I|}$ 是为此目的而建的。 \mathbf{CtI} 表示 $\bar{\mathbf{A}}$ 的归一化频率矩阵与相似矩阵 $(\mathbf{E})^T$ 的转置的乘积, 如式(9)所示。本文使用相同的规范化方法来构造 $\bar{\mathbf{B}}$, 构造过程与构造 $\bar{\mathbf{A}}$ 时类似。

$$\mathbf{CtI}(c, i) = \sum_{m=1}^{|I|} (\bar{\mathbf{A}})_{c,m} \times (\mathbf{E})_{m,i}^T \quad (9)$$

最后在用户和由矩阵 $\mathbf{UtI}_{|U| \times |I|}$ 表示的项目中找到隐式兴趣度。为了获得用户对项目的潜在偏好, 本文方法搜索与其偏好或类似用户的选择相似的项目。可以根据式(10)构建 \mathbf{UtI} 矩阵。

$$\mathbf{UtI}(u, i) = \sum_{m=1}^{|U|} (\bar{\mathbf{R}})_{i,m}^T \times (\mathbf{S})_{m,u}^T \quad (10)$$

矩阵 $(\bar{\mathbf{R}})_{i,m}^T$ 是原始归一化等级矩阵 \mathbf{R} 的转置, \mathbf{S} 包含前 k 个用户-用户相似度矩阵。两个矩阵的乘积表示用户的偏好以及他们最近邻居对给定项目的兴趣度。

2.4 情境感知评级

3 个模型(\mathbf{UtC} , \mathbf{CtI} 和 \mathbf{UtI})通过使用给定的情境列表 m 作为查询来计算用户 u 对项目 i 的预期排名分数。因此, 本文建议让用户的情境参与选择新项目, 从而使检测到的情境有助于促进对相关项目的搜索。但是, 仅依赖于情境和项目之间的关系(如在 \mathbf{CtI} 中)可能不会显著改善基于用户需求的相关项目的搜索。因此, 对于给定的用户查询(m), 本文提出了一种平衡算法, 其根据式(11)和式(12)考虑用户对情境的隐式兴趣度、情境对项目的隐式兴趣度以及用户对项目的评级。

$$\text{Score}_{u,c}(i) = \alpha \mathbf{CtI}_{c,i} \times \beta \mathbf{UtI}_{u,i} \quad (11)$$

$$\text{Rank}_{u,c}(i) = \sum_{c \in m} \text{Score}_{u,c}(i) \times \mathbf{UtC}_{u,c} \quad (12)$$

其中, m 包含一些使用给定查询提交的情境参数, $m = \{c_1, c_2, \dots, c_m\}$, $m \leq |C|$ 。 $\mathbf{CtI}_{c,i}$ 是一个表示 \mathbf{CtI} 矩阵的第 c 行和第 i 列的矩阵。 $\mathbf{UtI}_{u,i}$ 是一个表示 \mathbf{UtI} 矩阵中的第 u 行和第 i 列的矩阵。参数 α 和 β 是两个衰减因子, 其中 $\alpha, \beta \in (0, \dots, 1)$ 用

于减小不太敏感的情境的权重。在实验后设定 α 和 β 值的衰减,建议用户使用排名得分最高的项目($\mathbf{Rank}_{u,m}$)。

3 优化平均精度均值

利用与推荐的用户-项目关系模型相关联的情境信息,本文希望优化给定情境中用户的项目选择。根据为用户计算的排名值,可以为用户生成前 k 个推荐项目,并计算精度和召回率,以衡量项目与用户任务的相关程度。MAP 函数根据式(13)测量所提算法的性能,其中 t_u 表示用户 u 的测试用例的数量, P_n 是在顶部 n 的精度。

$$MAP=\frac{1}{|U|}\sum_{u=1}^{|U|}\sum_{c=1}^{|C|}\frac{1}{|t_u|}\sum_{n=1}^{|t_u|}P_n\times R_n$$
 (13)

其中, $R_n=\begin{cases} 1, & \text{item is relevant at rank } n \\ 0, & \text{otherwise} \end{cases}$ 。

由式(11)和式(12)可知,排名值根据查询中的情境和用户而改变,因此精度和 MAP 以非平滑的方式变化。在使用任何优化函数之前,首先需要对 U 、 I 和 C 应用平滑的精度函数。式(14)可以近似表示顶部 n 的精度,其中两个变量 r_i 和 r_j 为式(13)中提到的等级 n 的召回。然后就可以将计算 MAP 的早期函数修改为平滑的函数,如式(15)所示:

$$R_n(r_i\leqslant r_j)\approx s(R_n)=0.5+0.5\tanh(\frac{R_n-1}{1})$$
 (14)

$$MAP=\frac{1}{|U|}\sum_{u=1}^{|U|}\sum_{c=1}^{|C|}\frac{1}{|t_u|}\sum_{n=1}^{|t_u|}P_n\times s(R_n)$$
 (15)

在平滑所提出的 MAP 函数之后,就可以根据式(16)的标准梯度上升方法来对其进行优化。

$$L(U,I,C)=\sum_{u=1}^{|U|}\sum_{c=1}^{|C|}\frac{1}{|t_u|}\sum_{n=1}^{|t_u|}P_n\times s(R_n)-\frac{\lambda}{2}(|U|^2+|I|^2+|C|^2)$$
 (16)

4 实验评估

本节将研究所提出的推荐算法的性能,并评估其项目预测的准确性,同时将其与传统协同过滤算法进行比较,还将调查在推荐过程中使用的用户的情境是否有助于提高推荐性能。硬件环境是 MacBook Pro (13", 2017),CPU_Intel Core i5/显卡_Intel Iris Plus Graphics 640,2.3GHz,硬盘 256 G,内存 8GB。

4.1 实验配置和数据集

利用在线社交音乐数据库 last.fm 创建了一个数据集,这是一个在线社交音乐无线电资源,用户可以订阅、收听和标记自己喜欢的专辑或曲目。用户对不同专辑和曲目的注释有助于构建我们的情境标记。最初得到的数据集非常稀疏,因此必须在开始实验之前对其进行预处理,包括删除由极少数用户以及选择少量项目的用户选择的项目。

此外,本文清理了少于 5 个情境的所有用户标记的项目,并删除了少于 5 个用户注释的标签。生成的数据集包含 397 个用户、1298 个轨道、746 个情境标记和 17 036 个项目标记注释。

本文通过离线实验并按照文献[18]中描述的程序测量基于情境的建议的准确性。首先将数据集随机分为两个池,将原始数据集的 20%用作测试集,剩下的 80%用作训练集。为了构建测试集,随机隐藏了一个项目,包括用户配置文件中分配的情境标记,随后将它们用作每个用户的测试查询。当想要测试算法的性能时,会根据测试集运行查询。根据式(12)的结果,可以确定一个项目是否是前 k 个结果项目之一。将本文的数据集划分为训练集和测试集可能会对随机选择的项目/情境敏感。因此,为了确保本文所提出的评估不易受到划分步骤的随机性的影响,使用了不同的分区分别运行算法 5 次。

使用精确度和召回率来测量算法的检索精度。精确度被确定为推荐项目与已识别为相关项目的比率,如式(17)所示。召回率可以被确定为相关内容与推荐内容列表的比率,如式(18)所示。为了将提出的算法与其他基线方法进行比较,使用式(19)报告 F1 度量。F1 将精确度和召回率整合为一个可比较的值。

$$P_{top-k}=\frac{1}{|U|}\times\sum_{u=1}^{|U|}\frac{I_u^k}{k}$$
 (17)

$$R_{top-k}=\frac{1}{|U|}\times\sum_{u=1}^{|U|}I_u^k$$
$$I_u^k=\begin{cases} 1, & \text{item appears in the returned results at top-}k \\ 0, & \text{otherwise} \end{cases}$$
 (18)

$$F1=\frac{2\times Pn\times Rn}{Pn+Rn}$$
 (19)

4.2 3 种推荐模型的推荐性能对比

为验证本文方法的有效性,同样选取两种最先进的算法即文献[19]中的方法和文献[20]中的方法进行对比分析。

表 2 对比了 3 种推荐模型的推荐性能。由表可知,在不同的近邻数情况下,本文方法的性能表现均优于文献[19]中的方法及文献[20]中的方法,即不论是在近邻搜索还是融入情境因素的推荐过程中,本文方法分别使绝对误差值平均降低了 1.8%和 1.2%,表现出了比传统协同过滤算法更好的性能。就客观现实来看,评分标准以及对项目资源的客观需求存在差异,因此推荐系统最终给出的预测评分往往与用户的实际评分不一致。同时,对于文献[19]中的方法及文献[20]中的方法这两种协同过滤算法,融入一定的情境因素确实能够在一定程度上优化推荐系统的算法性能。

表 2 3 种推荐模型的推荐性能对比

Table 2 Performance comparison of three recommended models

方法	$n=5$	$n=15$	$n=25$	$n=35$	$n=45$	$n=55$	$n=65$	$n=75$	$n=85$	$n=95$
本文方法	0.782	0.0070	0.759	0.745	0.736	0.741	0.744	0.742	0.743	0.746
文献[20]中的方法	0.785	0.766	0.757	0.747	0.740	0.744	0.745	0.747	0.747	0.749
文献[19]中的方法	0.786	0.785	0.766	0.760	0.749	0.750	0.748	0.750	0.752	0.753

根据实验结果不难看出,目标用户的近邻数的确能够影响推荐算法的预测精确度,两者之间呈现正相关的关系。若想得到较为精确的预测精确度,则必须依据合适数量的近邻评分来预测用户的兴趣度和行为偏好,否则容易出现较大偏差。而近邻数过大也不能一味追求数量,近邻数过大往往会导致最终的预测评分结果的精度降低,这是由于随着近邻数的增多,用户之间的差异性也逐渐增大,进而影响了预测精度。当所取近邻规模过小时,由于能提取出的共同特征偏少,难以反映当前用户的兴趣度属性,因此造成预测精度偏低,这一点与实际应用的经验是相符合的。

4.3 检索准确性的分析及比较

由于本文的情境维度来自于社交标记,因此考虑将文献[19]中的方法与本文所提出的模型进行比较。文献[20]中的方法类似于本文方法,但是其提出了具有优化的矩阵因子分解功能,用于挖掘针对情绪特定标签的电影。图 2 给出了每种推荐方法在检索相关项目时的表现,以及它们在推荐列表中的排名位置。

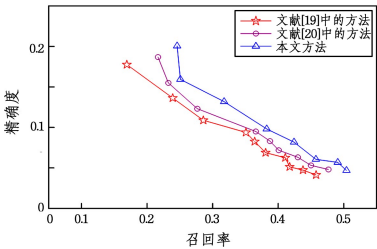


图 2 3 种算法在 Last.fm 数据集上的召回率与精确度
Fig.2 Recall rate and accuracy of three algorithms on Last.fm dataset

图 2 中所示的精确度和召回率曲线表示当返回项目的数量(top- k)发生改变时每种算法的性能。每条曲线上最左边的点代表前 1 位的精确度和召回率,而最右边的点代表前 10 位的结果。所提出的方法使用潜在偏好模型以及优化函数,所得结果优于其他算法。当 k 值增加时,文献[19]中的方法、文献[20]中的方法和本文方法的精确度和召回率会降低。平均而言,在前 1 位,本文方法比文献[20]中的方法的精确度提高了 2%,比文献[19]中的方法的精确度提高了 3%;此外,本文方法比文献[20]中的方法的精确度高 4%,比文献[19]中的方法的精确度高 7%。

使用两个配对 t 检验,本文算法($M=0.099,SD=0.002$)和文献[19]中的方法($M=0.086,SD=0.002$)的精确度存在显著差异, $t(9)=2.26,p<0.01$,所提出的算法与文献[19]中的方法在精确度和召回率上的比较均达到统计显著性,即 $p<0.01$ 。本文方法在精确度方面的表现与文献[20]中的方法没有统计学差异。使用两个配对 t 检验,本文方法($M=0.395,SD=0.009$)和文献[20]中的方法($M=0.356,SD=0.008$)的召回率存在显著差异, $t(9)=2.26,p<0.01$ 。然而,本文方法确实在召回时获得了更好的性能,具有统计意义。结果表明,文献[20]中的方法和本文方法仍然可能为用户查询并推荐不相关的项目,但本文方法带来了比文献[20]中的方法更多的相关项目。

使用 F1 度量各种算法的性能,如表 3 所列。当 $k=1$ 时,

本文所提出的算法的结果比文献[20]中的方法和文献[19]中的方法的结果分别高 3.24%和 4.48%。当 $k=10$ 时,本文所提出的算法的结果比文献[20]中的方法和文献[19]中的方法的结果分别高 0.57%和 0.53%。

表 3 每种方法 top-1 和 top-10 的 F1 度量
Table 3 F1 metrics for each method at top-1 and top-10
(单位:%)

方法	$k=1$	$k=10$
文献[19]中的方法	18.32	8.44
文献[20]中的方法	19.56	8.40
本文方法	22.80	8.97

结束语 随着社交媒体资源和服务的广泛使用,互联网用户可以使用推荐系统从在线获得的丰富信息中受益。本文分析了情境标记的使用以及使用用户情境的可能性,以提高推荐结果的准确性;还构建了一种在张量模型中搜索与用户、项目和情境的多个维度相关联的潜在偏好的新模型。本文所提出的算法能够识别分配给项目的潜在情境,并从类似用户中找到潜在情境偏好。此外,实验证明该算法在提高符合用户情境的相关项目的搜索方面提供了有利的优势。

最后对实验结果做了对比分析,结果表明本文方法优于传统的协同过滤算法,前者表现出的推荐精确度和召回率都优于后者,取得了预期的结果。未来希望融入更多情境因素,以进一步完善本文方法。

参 考 文 献

[1] JUNG H,CHUNG K. P2P context awareness based sensibility design recommendation using color and bio-signal analysis[J]. Peer-to-Peer Networking and Applications, 2016, 9 (3): 546-557.

[2] WANNENBURG J,MALEKIAN R. Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing[J]. IEEE Transactions on Systems Man & Cybernetics Systems, 2017, 47(12): 3142-3149.

[3] GE G L, YUAN L Y, WANG X C. Personalized user interest modeling based on context aware[J]. Application Research of Computers, 2017, 34(4): 995-999. (in Chinese)

葛桂丽,袁凌云,王兴超. 基于情境感知的用户个性化兴趣建模[J]. 计算机应用研究, 2017, 34(4): 995-999.

[4] LI F L, CHEN D X, LIANG S X. Research on Personalized Recommendation Method Based on Semantic Association and Context Awareness[J]. Journal of Intelligence, 2015, 37(10): 189-195. (in Chinese)

李枫林,陈德鑫,梁少星. 基于语义关联和情景感知的个性化推荐方法研究[J]. 情报杂志, 2015, 37(10): 189-195.

[5] SHEN N Y, LI Q Y. Personalized Recommendation Algorithm Based on Context-Aware Technology[J]. Computer Systems & Applications, 2017, 26(9): 135-139. (in Chinese)

时念云,李秋月. 基于情境感知的个性化推荐算法[J]. 计算机系统应用, 2017, 26(9): 135-139.

[6] LIANG Z Y. Research and Application of Personalized Recommendation Algorithm Based on Context Awareness [D]. Taiyuan: North University of China, 2017. (in Chinese)

梁卓越. 基于情境感知的个性化推荐算法研究与应用[D]. 太原:中北大学,2017.

[7] ALHAMID M F,RAWASHDEH M,OSMAN H A,et al. Towards context-sensitive collaborative media recommender system[J]. Multimedia Tools & Applications,2015,74(24):11399-11428.

[8] CHEN J M,CHEN M C,SUN Y S. A tag based learning approach to knowledge acquisition for constructing prior knowledge and enhancing student reading comprehension[M]. Elsevier Science Ltd,2014.

[9] LI J W,LIU Y L,QIN X L. Context-dependent Double-layered Data Model for Indoor Space[J]. Computer Science, 2017, 44 (8):187-192. (in Chinese)
李敬雯,刘宇雷,秦小麟. 一个情境相关的双层室内空间数据模型[J]. 计算机科学,2017,44(8):187-192.

[10] LEE H S,KWON S Y,LIM J H. A Development of a Lighting Control System Based on Context-Awareness for the Improvement of Learning Efficiency in Classroom[J]. Wireless Personal Communications,2016,86(1):165-181.

[11] LIU B,LI W S. Indoor Positioning Method Based on Cosine Similarity of Fingerprint Matching Algorithm[J]. Bulletin of Science and Technology,2017,33(3):198-202. (in Chinese)
刘冰,李文书. 基于余弦相似度的指纹匹配算法的室内定位方法[J]. 科技通报,2017,33(3):198-202.

[12] CHEN D,JIN D,GOH T T, et al. Context-Awareness Based Personalized Recommendation of Anti-Hypertension Drugs[J]. Journal of Medical Systems,2016,40(9):1-10.

[13] HAO D H,GUAN W G,ZOU L J,et al. Fast virtual grid matching localization algorithm based on Pearson correlation coefficient[J]. Journal of Computer Applications, 2018, 38 (3): 763-768. (in Chinese)
郝德华,关维国,邹林杰,等. 基于 Pearson 相关系数的快速虚拟网格匹配定位算法[J]. 计算机应用,2018,38(3):763-768.

[14] ZHANG J,ZHANG C,ZHU Q S. Adaptive Nearest Neighbor Algorithm with Dynamic Neighborhood[J]. Computer Science, 2017,44(12):194-201. (in Chinese)
冯骥,张程,朱庆生. 一种具有动态邻域特点的自适应最近邻居算法[J]. 计算机科学,2017,44(12):194-201.

[15] SUN B,LI L,WU X,et al. Combining feature-level and decision-level fusion in a hierarchical classifier for emotion recognition in the wild[J]. Journal on Multimodal User Interfaces, 2016, 10(2):125-137.

[16] A H K,B M P. Template Extraction from Heterogeneous Web Pages with Cosine Similarity[J]. International Journal of Computer Applications,2014,87(3):4-8.

[17] WANG W,JIANG Q,LV T,et al. An improved text similarity algorithm research for clinical decision support system[C]//International Conference on Cloud Computing and Intelligence Systems. IEEE,2016:155-159.

[18] CHEN X,ZHENG Z,LYU M R. QoS-Aware Web Service Recommendation via Collaborative Filtering[M] // Web Services Foundations. New York:Springer,2014.

[19] KURAHASHI A M,STINSON J N,WYK M V,et al. The Perceived Ease of Use and Usefulness of Loop:Evaluation and Content Analysis of a Web-Based Clinical Collaboration System[J]. Jmir Hum Factors,2018,5(1):217-230.

[20] SHI Y,LARSON M,HANJALIC A. Mining contextual movie similarity with matrix factorization for context-aware recommendation[J]. Acm Transactions on Intelligent Systems & Technology,2013,4(1):1-19.

[21] ZHONG X Y,LIU Y B,XIAO Y P. A user recommendation scheme based on similar community and node role division in social network[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2016, 28 (4): 525-532. (in Chinese)
钟晓宇,刘宴兵,肖云鹏. 一种基于相似社团和节点角色划分的社交网络用户推荐方案[J]. 重庆邮电大学学报(自然科学版), 2016,28(4):525-532.