

DOI:10.3969/j.issn.1673-4785.201406017
网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20150302.1106.007.html>

融合上下文信息的社会网络推荐系统

李慧^{1,2}, 马小平¹, 胡云², 施珺²

(1. 中国矿业大学 信电学院, 江苏 徐州 221008; 2. 淮海工学院 计算机工程学院, 江苏 连云港 222005)

摘 要: 上下文环境和社会网络信息已经成为推荐系统所需的重要信息来源, 在推荐系统中融入这些信息将进一步改进推荐系统的精度和用户满意度。为了提高用户对推荐系统的满意度, 提出一种融入上下文信息与社交网络信息的个性化推荐系统 CS。该算法应用随机决策树划分原始的用户-商品评分矩阵来进行上下文信息的处理, 使得具有相似上下文信息的评分被分为一组。随后应用矩阵因式分解来预测用户对未评分项的预测。为了整合社交网络信息, 在考虑上下文信息的环境下提出了一种融入社会网络关系的增强推荐模型, 使用一种基于信任度的皮尔逊相关系数来衡量用户的相似度。在真实的实验数据集上进行验证, 表明 CS 系统推荐较传统的基于上下文的和基于社交网络的推荐算法在性能上和推荐性能上有了很大的改善。

关键词: 上下文; 信息; 社会网络; 矩阵因式分解; 推荐; 协同过滤

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2015)02-0293-08

中文引用格式: 李慧, 马小平, 胡云, 等. 融合上下文信息的社会网络推荐系统[J]. 智能系统学报, 2015, 10(2): 293-300.

英文引用格式: LI Hui, MA Xiaoping, HU Yun, et al. Social network recommendaton system mixing contex information[J]. CAAI Transactions on Intelligent Systems, 2015, 10(2): 293-300.

Social network recommendaton system mixing contex information

LI Hui^{1,2}, MA Xiaoping¹, HU Yun², SHI Jun²

(1. School of Information & Electrical Engineering, China University of Mining & Technology, Xuzhou 221008, China; 2. Department of Computer Science, Huaihai Institute of Technology, Lianyungang 222005, China)

Abstract: Contexts and social network information is valuable information for building an accurate recommender system. The merging of such information could further improve accuracy of the system and user satisfaction. This paper proposes the context and social (CS) network, which is novel context-aware recommender system incorporating elaborately processed social network information, in order to increase the user satisfaction on the recommendation system. The contextual information happens by applying random decision trees to partition the original user-item-rating matrix such that the ratings with similar contexts are together. The matrix factorization functionality is to predict missing preference of a user for an item using the partitioned matrix. An enhanced recommendation model aided by social relationships considering the context information is proposed. A trust-based Pearson Correlation Coefficient is proposed to measure user similarity. Real datasets based experiments showed that CS enhances its performance compared with traditional recommendation algorithms based on context and social networks.

Keywords: context; information; social network; matrix factorization; recommendation; collaborative filtering

收稿日期: 2014-06-11. 网络出版日期: 2015-03-02.

基金项目: 国家自然科学基金资助项目 (61403156, 61403155); 江苏省
高校自然科学基金资助项目 (13KJB520002, 14KJB520005).

通信作者: 李慧. E-mail: shufanzs@126.com.

社交网络服务是 Web 2.0 应用中的重要一类。这种系统帮助用户在互联网上与现实中的好友保持联络、分享内容等。然而随着社交网络的使用人数增加, 用户分享的数据也不断上升, 信息过载的问题

也随之出现。另外,社交网络第 1 次提供了丰富的真实世界中人们的社交信息,利用这些信息的研究在各个学科中也不断出现。对于推荐系统来说,如何利用社交网络所提供的这些信息,进行更精确地推荐,成为一个重要研究方向^[1-2]。大部分推荐系统都采用协同过滤技术,它通过分析用户对相似项目以往的评分记录或相似用户对项目的评分模式来预测用户对某未知项目的评分。虽然协同过滤技术已成为解决推荐问题的标准方法,但传统的推荐系统只利用相关用户或相关商品的评分数据进行推荐,而并未考虑任何其他的信息。但当分析的信息量不断增大时,传统的推荐算法就会遇到诸多挑战,如数据稀疏问题、低质量的推荐以及信息源的不均衡性。

推荐系统通过识别和预测用户偏好,提供个性化服务,是缓解“信息过载”问题的重要手段之一。上下文感知计算是指系统能发现并有效利用上下文信息进行计算的一种计算模式,已经广泛应用于许多领域^[3-4]。上下文感知推荐系统通过将上下文信息引入推荐系统,成为一个刚刚兴起的领域,还有许多问题等待解决,如上下文用户偏好提取、高维数据稀疏性问题等^[5]。

本文提出了一个新颖的融入上下文信息与社会网络信息的推荐系统——CS 推荐算法,该系统首先抽取与用户兴趣相关的上下文信息,然后应用随机决策树算法分解用户—项目评分矩阵,最终生成的用户—项目评分子矩阵(即决策树的叶子节点)包含了相似的评分。在产生的子矩阵中应用了矩阵因式分解技术进行用户对某一缺失项的评分预测。基于矩阵因式分解模型,本文还引进社会规范术语来改善推荐系统质量。为了衡量用户之间的相似程序,本文使用了上下文相关的皮尔逊相关系数来计算用户相似度。最后通过在 2 个真实的数据集上验证 CS 推荐系统的有效性。

1 相关工作

目前,社会化推荐算法大都集中在如何利用用户间的社会关系来提高推荐的准确率。Yang 等在文献[6]中认为用户间的信任关系强度并不是唯一的,而是受到其所处的朋友圈的影响,在不同的朋友圈中,用户间的信任关系也是不同的。他们通过给用户的的朋友赋予不同的权重来对用户间的信任关系进行建模,并提出了一种基于朋友圈的社会化推荐算法。Noel 等^[7]针对社会化推荐算法在计算用户相似性时没有考虑到用户本身的特征、不能直接对

用户间信息传播建模,不能反映用户间只有部分兴趣相同的问题,给出了一种将多种目标函数融合在一起的方法,并在矩阵分解的框架下进行求解。

然而,目前大部分社会网络推荐模型在度量 2 个用户的相似性时,基本都忽视了上下文信息。比如,即使一个用户和其朋友的品味极其相似,她对一部电影的评价可能还受其他因素影响(比如,她在看电影时候的情绪和陪她看电影的人)。因此近期的研究开始关注社交网络中的上下文信息。文献[8]提出了将用户和项目进行群组的方法,在协同过滤算法中利用了这些子群信息(一种上下文信息)来提高用推荐系统的质量。Liu 等^[9]利用推荐对象的属性上下文信息来对它们之间的关联关系进行度量,并通过估计出的关联关系信息来改善推荐的效果。文献[10]提出了将社会网络上下文信息(个人表现和交际影响)整合到一个矩阵分解模型中。但是,这样的上下文信息仅仅与社交关系有关,大量的非社交的上下文信息却被忽视了。相反,本文提出的 CS 算法运用机器学习技术和矩阵分解技术,不仅包含了大量的上下文信息,而且对上下文信息没有限定信息类型:上下文信息被显式地应用到矩阵划分中。基于信任度的皮尔逊相关系数提高了计算用户相似度的准确性。

2 CS 推荐系统

2.1 预备知识

2.1.1 相关概念

传统的推荐系统通常只考虑用户—项目评分矩阵来进行推荐。然而,在许多系统中,可以通过丰富的上下文信息来为推荐系统提供新的信息维度。本文把上下文信息分为 2 类:1) 静态上下文,它描述用户的特性,例如年龄、性别、会员身份、角色等,或者是一种商品、种类、价钱、物理特性等;2) 动态上下文,是一种与等级相关的即时信息(例如当一个用户评价一个产品时,他的心情和位置信息)。另一方面,在线社交网络也带来一些其他资源,通过分析这些资源一个用户的喜好可以由与他有相同品味的的朋友推断出。因此,本文试图系统地融合上下文信息和社交网络信息来改善推荐性能。

用 $u = \{U_1, U_2, \dots, U_m\}$ 表示用户集合, $v = \{V_1, V_2, \dots, V_n\}$ 表示项目集合。所有用户可以根据自己的喜好为项目评分。假设分值为离散变量,范围为 $L = \{L_1, L_2, \dots, L_m\}$ 。比如,许多推荐系统(如 MovieLens)使用五分制进行评分(例如[1, 2, 3, 4, 5])。用户 U_u 对项目 V_v 的评分表示为 $R_{u,v}$,所有的

评分集合 $R = \{R_{u,v} | U_u \in u, V_v \in v\}$ 构成一个用户—项目评分矩阵(如图 1(a))。正如上面提到的,假设对用户的每一个评分级 R_i 都存在与其相关的上下文信息集合,用 $C_i = \{c_1, c_2, \dots\}$ 来表示。对每种类型的上下文信息的值域没有限制,也就是说,离散值和连续值都是合法的。

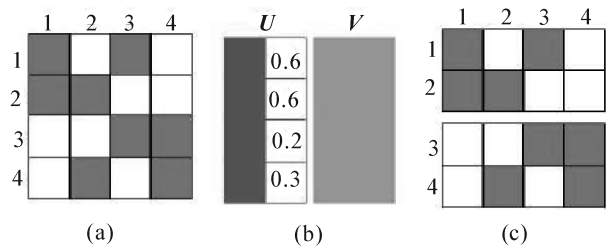


图 1 用户—项目评分矩阵分割过程

Fig.1 Illustration for division

在社会网络中将用户信息及用户之间的关系可以抽象表示为有向带权值的社会网络图的形式: $G=(V,E,C)$ 。其中, V 表示节点集合, 每个节点代表网络中的用户个体; E 表示边的集合, 表示 2 个个体之间存在的关系; $C = \{c_{uv}\}$ 表示边的权重值, 此值越大表示信任程度越大, 本文将其定义为用户间的信任度。由于信任关系不是对称的, 所以图中的边是有向的, 网络图为有向图。

2.1.2 矩阵因式分解

矩阵因式分解的目的是将一个矩阵分解成 2 个以上的矩阵, 使得将矩阵因子相乘后可以重构或者近似原始矩阵。在推荐问题中, 一个矩阵因式分解模型是将用户—项目评分矩阵 R , $R \in \mathbf{R}^{m \times n}$ (m 是用户数量, n 是项目数量) 分解成一个用户特征矩阵 U , $U \in \mathbf{R}^{m \times l}$ 和一个项目特征矩阵 V , $V \in \mathbf{R}^{l \times n}$ 。

$$R \approx U^T V \tag{1}$$

式中: l 是一个潜在特征向量的维度, 它标志着一个用户或者一个项目的特征。对于一个用户 a 来说, U 的元素(即 U_a) 衡量了用户 a 对项目的兴趣度; 对于项目 b , V 的元素(即 V_b) 衡量了 b 和相应的潜在特征的相关程度。因此, $U_a^T V_b$ 表示用户 a 和项目 b 之间的关联度, 即考虑了所有潜在特征后用户 a 对项目 b 的偏好度。

为了计算 R , 考虑到用户—项目评分矩阵的稀疏性, 定义了以下的目标函数, 即使预测评分与用户实际评分的误差最小化:

$$\arg \min_{U,V} \sum_{j=1}^m \sum_{k=1}^n I_{i,j} (R_{i,j} - U_i^T V_j)^2 \tag{2}$$

式中: $I_{i,j}$ 为一个指示变量, 即如果用户 i 对商品 j 进行了打分, 则 $I_{i,j}$ 为 1, 否则为 0。另外, 为了避免过度拟合, 在公式中加入了规范化系数, 即

$$\arg \min_{U,V} \sum_{j=1}^m \sum_{k=1}^n I_{i,j} (R_{i,j} - U_i^T V_j)^2 + \lambda (\|U\|_F^2 + \|V\|_F^2) \tag{3}$$

式中: $\|A\|_F^2$ (A 是 $X \times Y$ 的矩阵) 是 Frobenius 范数, 是通过 $\sqrt{\sum_x \sum_y |A_{xy}|^2}$ 计算得到。参数 λ 控制规范化的范围。

式 3 可以通过 2 种方式解得: 1) 随机梯度算法 (SGD), 通过迭代更新潜在用户特征因子和潜在项目特征因子; 2) 交替最小二乘算法 (ALS), 通过修正矩阵 U (或者 V) 以优化 V (或者 U), 并且轮转迭代。

2.2 上下文感知的推荐模型

本节首先介绍一下如何结合上下文信息来提高推荐系统的推荐准确度, 在此暂不考虑社会关系。为了有效结合不同的上下文信息, 使用一种具有较高学习精度的随机决策树算法。该算法的目标是对原始即用户—项目评分矩阵 R 使用随机划分策略将相似用户或相似项目的评分划分到树的同一结点中, 即将具有相似上下文的评分划分在一个组内。由于是在相似的上下文中产生, 因此在相同组里的评分将会比在原始评分矩阵中的评分具有更高的相关性, 有助于提高推测缺失值的准确性。

对每个决策树中的每一个结点, 利用式 (2) 对评分矩阵 R 进行基本的矩阵因式分解。经过分解之后, 分别得到用户潜在特征向量与项目潜在特征向量(如图 1(b))。用户特征因子表明了用户在一些潜在主题上的兴趣分布, 而项目特征因子代表了与这些主题相关的项目成员。为了划分评分矩阵 R , 选择了一个潜在特征(如图 1(b) U 的第 2 列) 和随机选取了一个分割值(本例中假设选择的分割值为 0.4)。设定之后, 则当前的评分矩阵被划分为 2 部分, 如图 1(c) 所示。在本例中, 根据 U 中第 2 个潜在特征向量和随机选定的分割值, 评分矩阵被从第 2 行和第 3 行之间分割成了 2 部分。由于第 1 个和第 2 个用户的潜在特征值比较相似, 因此他们给出的评分被决策树划分到同一个结点中。

在为每个上下文信息构建决策树时, 在树的每一层, 算法都会从上下文信息集合 C 中随机选择一个上下文信息 c_r 来划分评分矩阵(见图 2)。具体来说, 评分矩阵是根据 c_r 的值进行划分的。例如, 如果假设上下文信息 c_r 是一周时间, 则评分矩阵可以根据每一天(即从周一到周六, 工作日或者周末)来进行有意义的划分。另一方面, 如果 c_r 的值没有任何语义信息, 则首先要对每一个评分 c_r 进行标准化到某一特定区间(如 $[0, 1]$), 然后选择一个随机的阈值(如 $\in [0, 1]$) 来划分评分。一旦在树中的某一层上完成了评分划

分,则随机选取的上下文信息 c_r 就会从上下文信息集合中被删除: $C = C/c_r$, 从而保证了一个上下文信息在一条路径上只被操作一次。

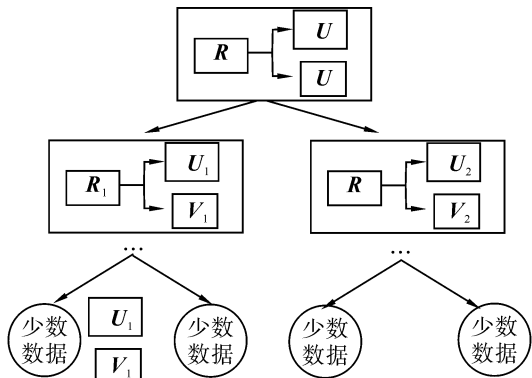


图2 随机决策树划分过程

Fig.2 Main partition flow of random decision tree

树的划分过程将一直继续,直到满足下列要求之一:1)所有的上下文信息都已经处理完毕;2)决策树的深度达到了预定阈值。划分结束之后,评分矩阵 R 被不同的上下文信息分类到不同类别中(即图2中树的叶子节点)。当预测一个缺失评分 R_m , 在每个决策树中,根据 R_m 的上下文信息对决策树进行划分,将其分类到用户—项目评分子矩阵 $R_i^s \subset R$ 中。对于每一个 R_i^s , 将其分解后分别得到用户特征因子 U_i^s 和项目特征因子 V_i^s , 它们能够被用来预测缺失评分 R_m (见式(4)、(5))。

$$L_1 = \argmin_{U_i^s, V_i^s} \sum_{j=1}^{|U_i^s|} \sum_{k=1}^{|V_i^s|} I_{j,k} (R_{i,j,k}^s - (U_{i,j}^s)^T V_{i,k}^s)^2 + \lambda (\|U_i^s\|_F^2 + \|V_i^s\|_F^2) \quad (4)$$

$$R_{m,i} = (U_i^s)^T V_i^s \quad (5)$$

最终,将所有决策树得到的预测组合起来就可以得到对 R_m 的最终预测得分。

$$R_m = \frac{\sum_{i=1}^T R_{m,i}}{T} \quad (6)$$

式中: T 为决策树的数量。通过组合不同决策树获得的多个预测得分,所有上下文信息被完全整合,产生了个性化和准确的上下文感知推荐系统。此外通过删除较少联系的(考虑到上下文信息)评分,生成的子矩阵比原始的评分矩阵要小很多,这将在很大程度上减少算法的计算复杂度。影响算法复杂度的一个重要因素是决策树的数量。在后续的实验部分验证了其实只要少部分树就可以获得高质量的推荐,从而说明了本文所提出方法的实用性。

2.3 融合社会网络关系的增强模型

在2.2节提出的上下文感知推荐模型的基础上,本节将社交网络信息加入到推荐中以改善推荐

质量。这种模型的假设前提与现实非常相符:在现实生活中,当用户在决定是否购买一个产品时往往会参考与自己有相似品味与兴趣的朋友所给出的建议。通过结合多个朋友的意见,用户最终会做出明智的选择。

尽管朋友能够提供有用的信息来帮助推荐系统为用户做出高质量的推荐,但现有的研究大部分都是在利用社会网络中所有的可用信息进行推荐,没有对这些信息进行细致地过滤;或者并没有深入地调查怎样精确计算用户之间的品味相似性。为了解决这些问题,本文引进一个新的社会规范化系数来对用户和他朋友之间的品味差异进行约束。在真实生活中,一个用户可能会有成百上千个朋友,因此同等对待朋友(或者朋友所给出的推荐信息)是没有意义的,因为其中的一些朋友可能与用户具有非常相似的品味,而与另一些朋友可能拥有完全不同的品味。为了解决这种社会网络中品味差异性,本文引入了社会规范化系数,在推荐系统中考虑了用户与其朋友之间的品味相似性:

$$\alpha \sum_{j=1} \sum_{f \in F_j} s(j,f) \|U_{i,j} - U_{i,f}\|_F^2 \quad (7)$$

式中: α 是控制社会规范化范围的常量。 $S(j,f)$ 表示基于以往评分模式计算出的用户 u_j 与朋友 u_f 之间的品味相似性。相似性分值较高表明 u_j 和 u_f 有着非常相似的品味,分值越低表明 u_j 和 u_f 有不同品味。

从式(7)中可以看出,融合社交网络信息的有效方法是通过研究用户与朋友之间的相似性来精确衡量每个朋友的观点,这种相似性可以通过他们以往的评分模式来推测,比如可以从他们共同评价过的商品的特性进行推测。现有许多相似性计算方法,其中皮尔逊相关系数已被证明比其他方法(如向量空间相似性)具有更高的精确性。因此本文也使用皮尔逊相关系数来衡量 u_j 与其朋友 u_f 的相似性。

$$S(j,f) = \frac{\sum_{v \in V(j) \cap V(f)} (R_{i,v} - \bar{R}_j) (R_{j,v} - \bar{R}_f)}{\sqrt{\sum_{v \in V(j) \cap V(f)} (R_{i,v} - \bar{R}_j)^2} \cdot \sqrt{\sum_{v \in V(j) \cap V(f)} (R_{j,v} - \bar{R}_f)^2}} \quad (8)$$

式中: $V(j) \cap V(f)$ 表示 u_j 和 u_f 共同评级过的项目集合, \bar{R}_j 和 \bar{R}_f 分别是 u_j 和 u_f 的平均评分。

在社会网络中,每一个用户 u 都会有邻居集合 N_u , 用 t_{uv} 表示节点 u 对节点 v 的社会信任度,其取值范围在 $[0,1]$ 。值为0表示完全不信任,值为1表示完全信任。在社会网络中, t_{uv} 的值可以解释为用

户 u 对用户 v 的了解与信任程度。但由于该权值包含一些噪音数据,不能体现社交网络中的整体结构信息,这就类似于在网页分析中的忽略了网页的链接结构信息。但实际在一个信任网络中,如果某个用户 u 信任大部分的用户,则其信任度应当被降低;反之,如果某个用户 v 被大部分用户所信任,则其信任度应该被增强。因此,综合考虑了社交网络的整体结构信息,利用式(9)对信任度进行了修正,修正后的信任度记为 t_{uv}^* :

$$t_{uv}^* = \sqrt{\frac{d_{in}(v)}{d_{out}(u) + d_{in}(v)}} \times t_{uv} \tag{9}$$

式中: $d_{out}(u)$ 表示以 u 为起点, u 所指向的所有节点集合,即 u 的出度集合, $d_{in}(v)$ 表示以 v 为终点,所有指向 v 的节点集合,即 v 的入度集合。

经典的相似性计算方法仅仅利用了评分的值,并未考虑用户之间的信任关系。为了进一步改善相似性计算的精确度,本文使用一种基于信任度的皮尔逊相关系数来计算用户之间的相似度:

$$S(j,f) = \frac{\sum_{v \in V(j) \cap v(f)} t_{uv}^* (R_{i,v} - \overline{R_j})(R_{j,v} - \overline{R_f})}{\sqrt{\sum_{v \in V(j) \cap v(f)} t_{uv}^* (R_{i,v} - \overline{R_j})^2} \cdot \sqrt{\sum_{v \in V(j) \cap v(f)} t_{uv}^* (R_{f,v} - \overline{R_f})^2}} \tag{10}$$

很明显,权重 t_{uv}^* 较大意味着用户之间的信任度越大,因此对整体的相似度计算时将会产生非常大的影响。应用式(7)和式(10),将式(4)重新定义为

$$L_2 = \underset{U_{i,j}^s, V_{i,j}^s}{\operatorname{argmin}} \sum_{j=1}^{|U_i^s| \cup |V_i^s|} \sum_{k=1}^{|U_{i,j}^s|} I_{i,j}(R_{i,j,k}^s - (U_{i,j}^s)^T V_{i,k}^s)^2 + \alpha \sum_{j=1} \sum_{f \in F_j} S(j,f) \|U_{i,j} - U_{i,f}\|_F^2 + \lambda (\|U_i^s\|_F^2 + \|V_i^s\|_F^2) \tag{11}$$

式(11)可以通过对 $U_{i,j}^s$ 和 $V_{i,k}^s$ 应用梯度下降算法进行求解,这是一个不断迭代更新的过程。

$$U_{i,j}^s \leftarrow U_{i,j}^s + \gamma \frac{\partial L_2}{\partial U_{i,j}^s} \tag{12}$$

$$V_{i,k}^s \leftarrow V_{i,k}^s + \gamma \frac{\partial L_2}{\partial V_{i,k}^s} \tag{13}$$

式中: γ 代表学习率。迭代的数量对推荐效率的影响将在实验部分进行验证。

3 实验评估

3.1 实验方法

3.1.1 数据集

豆瓣网(www.douban.com)是中国最大的社交平台之一,许多人在这里分享对书、电影、音乐的评

价。每个用户可以对书、电影、音乐进行评级(从一星到五星),表达他们对这些产品的喜好。在社交网络中如果某用户的评论被认为是有趣且有用的,则他就可能被其他用户所跟随。表 1 列出了数据集的统计数据。

表 1 豆瓣网的统计数据

Table 1 Statistics of the Douban dataset

类别	评分数	用户数	项目数
书	812 037	8 598	169 982
电影	1 336 484	5 227	48 381
音乐	1 387 216	23 822	185 574
所有	3 535 737	25 560	403 937

选择豆瓣的数据因为它不仅包含了相关的时间/数据和其他可推断的上下文信息,而且还包含了社交网络信息,因此非常适合用于评估应用了多种类型信息的 CS 算法的性能。从豆瓣数据集中,随机选择 80% 的评价来训练推荐模型,使用剩下的 20% 比较它们的性能。

3.1.2 比较对象

本文将 CS 推荐系统和目前主流的几种推荐方法进行了对比实验:传统的基于上下文感知推荐系统 RPMF^[14]、基于社交网络的推荐系统 SoReg^[11]、应用基本的矩阵分解模型构建的推荐系统 BMF^[12]。

与所有的上下文推荐系统相似,从数据集中可获得的上下文化信息中选取了 5 种类型的上下文信息:1) 小时信息,即用户给出评分的时刻;2) 日期信息,即用户给出评分的日期;3) 当一个评价被给出的时候,对目标商品产生“期待”的数量;4) 当目标用户评价一个特定商品时,其所给出评分的平均值;5) 目标商品所属的类别。

3.1.3 度量标准

实验选取在推荐系统评价中经常使用的 2 个度量标准来比较不同推荐模型的性能:平均绝对误差(MAE)和均方根误差(RMSE)。式(14)和(15)分别给出两者的定义:

$$MAE = \frac{1}{N} \sum_{r=1}^N |R_r - R'_r| \tag{14}$$

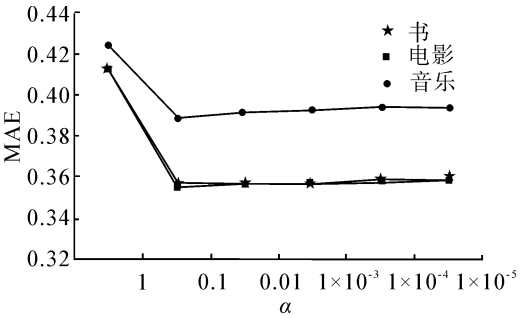
$$RMSE = \sqrt{\frac{1}{N} \sum_{r=1}^N (R_r - R'_r)^2} \tag{15}$$

式中: N 表示预测的总次数, R_r 表示一个项目的真实评分, R'_r 表示相应的预测评分。其值越小,表明推荐模型的性能越高。

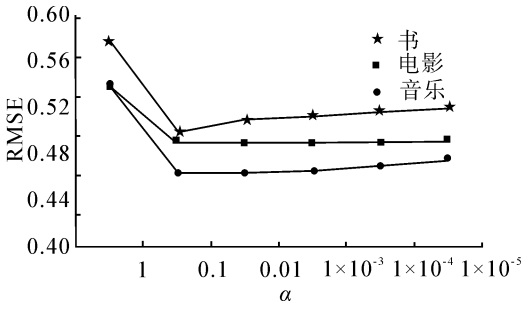
3.2 实验结论

首先使用豆瓣网数据集来说明 CS 算法中不同参数值的选取对推荐性能的影响。经过交叉验证之

后得到规则化常量 $\lambda = 0.1$ 。图 3 给出了当数据集的不同子集(如书数据、电影数据、音乐数据)被应用时,CS 推荐模型的性能如何随着参数值 α 的变化而变化,参数 α 决定了有多少社会网络信息量被整合进 CS 推荐模型中(见式 11)。实验中设置在求解矩阵因式分解模型中潜在特征向量的维数为 10,迭代求解次数为 20。后续实验会给出这些变量的变化如何影响基于矩阵因式分解的推荐模型的性能。从图 3 可以看出随着 α 值的增大,MAE 和 RMSE 的值首先减少,接下来当到达一定阈值时(大约在 $\alpha = 0.1$ 处)其值变得相对稳定(只是轻微下降)。因此可能得出社交网络信息可以有效改善推荐质量的结论,并且 $\alpha = 0.001$ 是一个合适的阈值来很好地平衡用户—项目评分矩阵和社交网络信息。



(a) MAE 实验结果



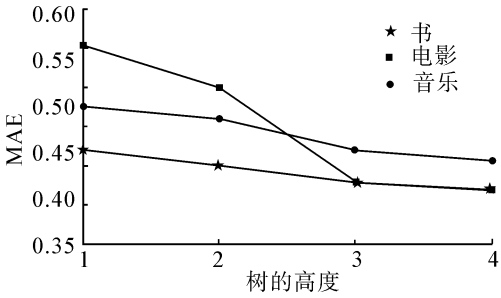
(b) RMSE 实验结果

图 3 参数 α 的影响 ($\lambda = 0.1$, 维数 = 10, 迭代次数 = 20)

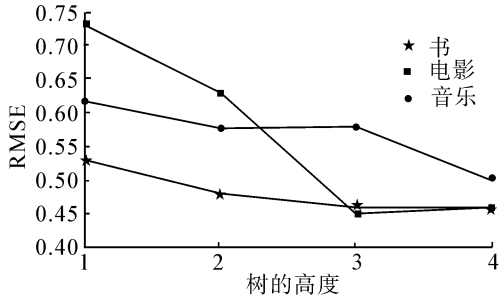
Fig.3 Impact of α ($\lambda = 0.1$, dimensionality = 10, iteration = 20)

接下来,验证上下文信息数量对推荐性能的影响。这一点可以通过控制决策树的高度来实现。也就是说,如果设树的高度为 1,则只有一种类型的上下文信息在树的划分时被使用;如果设树的高度为 4,表示所有的上下文信息都被应用到推荐系统中来。图 4 给出了不同数量上下文信息的实验结果。

从图 4 中可以看出在所有情况下,上下文信息越多则会产生越高的推荐精度,即 MAE 和 RMSE 的值越小。实验结果表明上下文信息很大程度上改善了推荐系统的性能,另一方面,从实验结果中可以看出本文所选取的上下文信息是非常有用的。



(a) MAE 实验结果



(b) RMSE 实验结果

图 4 上下文信息数量的影响 ($\lambda = 0.1$, 维数 = 10, 迭代次数 = 20)

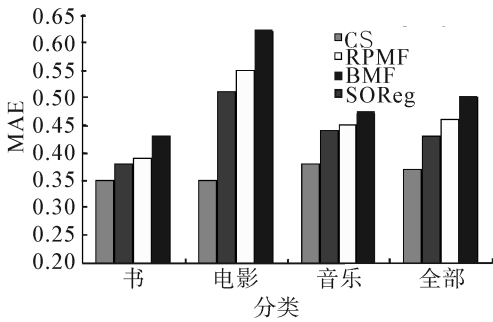
Fig.4 Impact of number of contextual information ($\lambda = 0.1$, dimensionality = 10, iteration = 20)

最后,将 CS 推荐系统和其他推荐系统的性能在豆瓣网数据集上做对比实验。在做对比实验之前,需要决定 2 个重要的参数的取值,即潜在特征向量的维度和基于矩阵因式分解模型的迭代次数。首先固定迭代次数为 10,观察潜在特征向量在不同维度下的 MAE 取值,如表 1 所示。发现随着维度的增加 MAE 的值在减少,这意味着随着维度的增加将会产生更高的推荐。但是当维度增加到大约 10 时,推荐质量的改进甚小。因此在实验中,为推荐算法的潜在特征向量维度设置为 10。同理,本文为所有基于矩阵因式分解模型的迭代次数设置为 20,因为更多的迭代次数并没有降低 MAE 的值,反而会产生更高的开销。

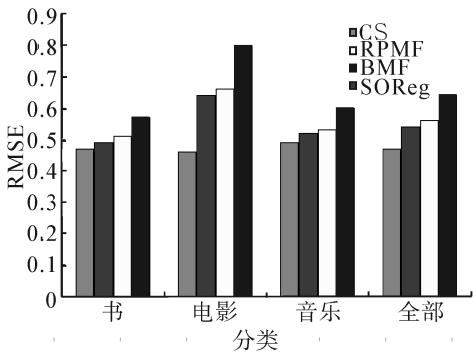
参数一量确定,下面就分别使用书数据、电影数据、音乐数据和整个豆瓣网数据来比较不同推荐模型的推荐性能。表 2 给出了对比结果。从图 5 可以看出,本文提出的 CS 推荐模型所有的实验数据中都比其他推荐模型更加精确。所有基于矩阵因式分解的推荐模型都明显优于传统的基于项目和基于用户的协同过滤算法,这表明了矩阵因式分解技术在推荐领域的优势。实验结果也表明综合考虑上下文信息和社会网络信息比只考虑某一种信息类型的推荐模型(如 SoReg 和 RPMF)具有更高的推荐质量。

表 2 向量维度与迭代次数对推荐质量的影响(MAE)
Table 2 The effect of dimension and iteration on recommendation

隐层 节点数	向量维度							迭代次数						
	2	4	6	8	10	12	14	0	5	10	20	30	40	50
书籍	0.4	0.38	0.37	0.38	0.37	0.37	0.37	0.5	0.4	0.36	0.35	0.35	0.36	0.36
电影	0.5	0.38	0.37	0.36	0.35	0.35	0.35	0.27	0.38	0.36	0.35	0.35	0.36	0.35
音乐	0.45	0.44	0.43	0.42	0.42	0.42	0.43	0.6	0.45	0.4	0.38	0.38	0.37	0.39



(a) MAE 实验结果



(b) RMSE 实验结果

图 5 不同推荐模型的性能比较

Fig.5 Performance comparison for various model

4 结束语

本文提出的 CS 算法是一个将上下文信息和社交网络信息相结合推荐算法,大大提高了推荐质量。该算法首先使用随机决策树算法基于不同的上下文信息对原始用户评分矩阵进行划分。划分后生成的子矩阵中包含相似上下文的评分,因此彼此间具有更大的影响。矩阵因式分解技术被应用到子矩阵中用来预测缺失评分。为了有效联合社会网络信息,CS 算法引进额外的社会规则术语通过学习他朋友的品味来推断用户对一个项目的喜好。为了识别有相似品味的朋友,提出了一个基于上下文信息的皮尔逊系数来衡量用户与朋友的相似性。在 2 个真实数据集上进行的验证,实验结果表明 CS 推荐算法明显优于传统的基于上下文信息和基于社交网络的推荐模型。

参考文献:

[1] ZHAO Du, FU Xiaolong. Scalable and explainable friend recommendation in campus social network system[J]. Frontier and Future Development of Information Technology in Medicine and Education, 2014, 269(1): 457-466.

[2] CHEN Cheng, MAO Chengjie, TANG Yong. Personalized recommendation based on implicit social network of researchers [J]. Pervasive Computing and the Networked World, 2013, 7719(10): 97-107.

[3] ZHONG Erheng, FAN Wei, YANG Qiang. Contextual collaborative filtering via hierarchical matrix factorization[C]// Proceedings of the Siam International Conference on Data Mining. Texas, USA, 2012: 744-755.

[4] RENDLE S, GANTNER Z, FREUDENTHALER C. Fast context-aware recommendations with factorization machines [C]//Proceedings of the 34th International ACM SIGIR Conference. Beijing, China, 2011: 230-241.

[5] BALTRUNAS L, LUDWIG B, RICCI F. Matrix factorization techniques for context aware recommendation[C]//Proceedings of the Fifth ACM Conference on Recommender Systems. Chicago, USA, 2011: 301-304.

[6] YANG Xiwang, STECK H, LIU Yong. Circle-based recommendation in online social networks[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China, 2012: 1267-1275.

[7] NOEL J, SANNER S, TRAN K. New objective functions for social collaborative filtering[C]//Proceedings of the 21th International Conference on World Wide Web. Lyon, France, 2012: 859-868.

[8] XU Bin, BU Jiajun, CHEN Chun. An exploration of improving collaborative recommender systems via user-item subgroups[C]//Proceedings of the 21st International Conference on World Wide Web. Lyon, France, 2012: 21-31.

[9] LIU Qiang, WANG Chengwei, XU Congfu. A modified PMF model incorporating implicit item associations [C]//Proceedings of the International Conference on Tools with Artificial Intelligence. Athens, Greece, 2012: 7-9.

[10] JIANG Ming, CUI Peng, LIU Rui. Social contextual recommendation[C]//Proceedings of the 21st ACM Interna-

tional Conference on Information and Knowledge Management. Maui, USA, 2012: 566-578.

[11] MA Hao, ZHOU Dengyong, LIU Chao. Recommender systems with social regularization [C]//Proceedings of the Fourth ACM International Conference on Web Search And Data Mining. Hong Kong, China, 2011: 287-296.

[12] MA Hao, YANG Haixuan, LYU M R. Sorec: social recommendation using probabilistic matrix factorization [C]// Proceedings of the 17th ACM Conference on Information and Knowledge Management. [S.l.], USA, 2008: 931-940.

作者简介:



李慧,女,1979 年生,讲师,博士研究生,主要研究方向为智能信息处理、社会网络分析、计算机技术及应用。主持并完成江苏省自然科学基金 1 项,参与国家自然科学基金 2 项,出版专著 1 部,发表学术论文 20 余篇。



马小平,男,1961 年生,教授,博士,主要研究方向为控制理论及应用、计算机技术及应用。主持并完成多项科研项目,其中国家“863”项目 1 项、国家自然科学基金项目 2 项、江苏省自然科学基金项目 3 项、江苏省高校基础研究项目 2 项、大型企业横向科研项目 20 余项。



胡云,女,1978 年生,副教授,主要研究方向为复杂网络分析理论及应用、数据挖掘、多 Agent 系统。主持并完成科研项目多项,其中国家自然科学基金项目 1 项、江苏省自然科学基金项目 1 项,参与出版专著 1 部,发表学术论文 10 余篇。

2015 中国计算机视觉大会 (CCF CCCV)

CCF Chinese Conference on Computer Vision 2015

中国计算机视觉大会 (CCF Chinese Conference on Computer Vision, CCF CCCV) 是由中国计算机学会 (CCF) 主办, CCF 计算机视觉专业委员会协办的计算机视觉领域盛会。该系列会议每 2 年举行一次, 是国内计算机视觉领域最主要的学术活动, 旨在为从事计算机视觉领域的学生、老师和工业界研究人员提供一个学科互动交流平台, 促进领域内的学术交流以及学术界与工业界之间的交流, 提高国内视觉领域的研究水平。大会将邀请多位知名专家学者做主题报告, 分享计算机视觉领域最新最热的理论和方法, 使与会者接触到学科最前沿的研究工作和团队, 使得研究人员分享计算机视觉领域的研究成果、创新思想、最新研究进展。

2015 年中国计算机学会中国计算机视觉大会 (CCCV2015) 将于 2015 年 9 月 18~20 日在陕西省西安市举行, 由西安电子科技大学计算机学院、电子工程学院和综合业务网理论及关键技术国家重点实验室联合承办。本次会议将汇聚从事计算机视觉理论与应用研究的人员, 广泛开展学术交流, 研讨发展战略, 共同促进计算机视觉相关理论、技术及应用的发展。

组织机构

指导委员会

- 谭铁牛 中科院自动化所
- 陈熙霖 中科院计算所
- 王 涛 爱奇艺公司

- 查红彬 北京大学
- 赖剑煌 中山大学
- 王 亮 中科院自动化所

大会主席

- 谭铁牛 中科院自动化所

- 高新波 西安电子科技大学

程序委员会主席

- 查红彬 北京大学
- 王 亮 中科院自动化所

- 陈熙霖 中科院计算所
- 苗启广 西安电子科技大学计算机学院