

耦合社会信任信息的矩阵分解协同过滤模型

蒋伟*, 秦志光

(电子科技大学信息与软件工程学院 成都 610054)

【摘要】在过去的十年中,协同过滤(CF)推荐系统已经取得了巨大的成功。然而,用户-物品矩阵的稀疏性和冷启动问题仍然是一个挑战。在线社交网络的出现,为推荐系统提供了大量社交网络信任信息,从而为解决这一问题提供了契机。该文基于矩阵分解协同过滤方法,提出了一种集成用户信任信息的模型。该方法利用用户信任信息对用户隐因子进行修正,采用自编码器来提取用户和物品隐特征向量的初始化特征,并针对社交网络中的信任关系提出了信任群组的检测算法。大规模的真实数据集上进行的广泛的实验表明,该模型与相关算法对比,不但能有效缓解冷启动,而且取得了更好的推荐性能。

关键词 协同过滤; 矩阵分解; 推荐系统; 辅助信息; 社交网络; 信任感知

中图分类号 TP311 **文献标志码** A **doi**:10.3969/j.issn.1001-0548.2019.03.018

A Matrix Factorization Collaborative Filtering Model with Trust Information

JIANG Wei* and QIN Zhi-guang

(School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract Collaborative filtering (CF) recommender system has been a most successful recommendation model in the past decade. However, the sparseness of user-item matrix and cold-start problem still remain the challenges. The emergence of online social networking provides a great deal of social trust information for recommender systems, thus providing an opportunity to solve these problems. In this paper, based on matrix factorization collaborative filtering method, a model of integrating user trust information is proposed. This method uses trust information of users to amend the user latent factors and employs an auto-encoder to extract the initialization features of user and item latent feature vectors. And then a trust group detection algorithm is proposed for the trust relationship in the social network. Extensive experiments on real data sets show that the proposed model can not only effectively alleviate cold start, but also achieve better recommendation performance than the compared algorithms.

Key words collaborative filtering; matrix factorization; recommender system; side information; social network; trust-aware

虽然协同过滤推荐系统已经在学术界广泛应用,并在Amazon和Ebay等公司得到应用,但大多数技术都有一些固有的弱点。众所周知的一个问题是冷启动和数据稀疏性。商用推荐系统可用评分的密度一般小于1%^[1],传统的推荐算法不能提供很好的推荐性能,尤其是纯粹挖掘用户物品评分矩阵的协同过滤推荐,无法处理只对少量物品进行评分的用户。

现实生活中,由于稳定而持久的社会关系约束,人们更愿意与朋友分享个人意见,而且更信任朋友的建议,因此在购买产品或消费服务(如电影、音乐、书籍、事物等)之前,常常向社交网络中的朋友寻求

建议,并且其爱好很容易受朋友的影响。因此,近年来社交网络的快速发展,为提高推荐性能、解决用户端稀疏性及冷启动问题提供了契机。在社交网络中,用户可以创建并分享内容、标注内容、评分评论、加入社区、与朋友联接,用户成了这些社交网络的核心。有了社交网络,可以在不侵犯隐私的情况下轻易取得其中的朋友信任关系,并可以把社交网络中的兴趣网络(Twitter)或熟人关系网(例如Facebook、LinkedIn等)作为用户偏好的来源。例如,由于社交兴趣,用户可能阅读某一特定主题的新闻;由于社会信任,用户可能喜欢她好友在Facebook上推荐的电影。

1 集成社会信任的混合协同过滤研究现状

1.1 基于社会关系的推荐

基于社会网络的推荐方法, 假定一个用户在社会网络中具有直接或间接的社会关系, 这种方法通常采用用户的评分来推荐。社会评分网络是社会网络的一种。用户可以创建与其他用户的社交关系, 也可以对一些物品进行评分。当用户之间共享大量相同的物品及评分信息时, 协同过滤最为有效, 但对于解决冷启动、评分稀疏性等问题, 这种基于相似性的方法, 就不可能找到类似的用户, 其推荐效果很差。然而, 基于社会网络的推荐利用用户之间的大量联系信息来解决这种相似性问题, 从而为解决冷启动、评分稀疏性问题提供了有效的手段。

现有的社会推荐系统基本上都是基于协同过滤的。参照传统的协同过滤推荐系统分类方法, 基于协同过滤的社会推荐系统可以分为两类: 基于矩阵分解(MF)的社会推荐方法和基于邻域的社会推荐方法。基于矩阵分解的社会推荐方法集成用户社会信任信息与用户-物品反馈历史信息(例如评分、点击、购买、收视率等)以提高传统的矩阵分解推荐系统的准确性。基于邻域的社会推荐包括社会网络的遍历(social network traversal, SNT)方法和最近邻(nearest neighbor, NN)方法。SNT方法对社交网络中的直接或间接链接的邻近用户进行遍历, 然后为用户生成推荐。最近邻方法将传统的CF邻域与社会邻域相结合, 预测物品的评分或生成推荐物品列表。

目前各种各样的网络站点, 例如电子商务以及在线社会网络平台, 都在收集用户社会关系信息。尤其是Facebook、Twitter和YouTube等在线社交网络平台, 为人们的交流以及虚拟社区的构建提供了新的途径。这些社交平台不仅使用户分享观点变得更容易, 而且还可以作为开发推荐系统的数据支撑和运营平台。

1.2 基于信任的矩阵分解推荐算法

近年来, 在矩阵分解协同过滤方法基础上集成社会信任信息, 成为社会信任推荐方法研究的热点。依据模型建立方式, 集成社会信任信息的矩阵分解模型可以分为两类: 1) 基于共享用户特征矩阵的方法, 这类方法同时分解评分矩阵及社交关系矩阵, 并共享用户特征矩阵; 2) 基于用户特征矩阵重表示的方法, 这类方法在采用评分矩阵计算用户特征矩阵时, 受关联用户信任信息的约束。

1) 基于共享用户特征矩阵的方法

SoRec模型^[2]是这类方法中最早的模型。它在评分矩阵分解的基础上, 进一步集成了用户与用户之间的关注关系(表示为社交信任矩阵 \mathbf{S})。SoRec在优化函数层面, 把用户-物品评分矩阵分解和用户-用户社交信任矩阵分解线性叠加在一起, 这两个矩阵分解共享相同的用户隐式空间。该模型如图1所示, 其中评分矩阵 \mathbf{R} 分解成物品和用户隐式空间 \mathbf{P} 、 \mathbf{Q} , 社交信任矩阵 \mathbf{S} 分解为用户和用户隐式空间 \mathbf{Q} 、 \mathbf{Z} , 两个矩阵分解共享相同的用户隐式空间 \mathbf{Q} 。

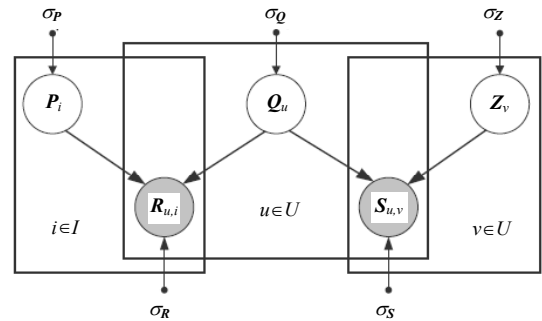


图1 SoRec模型

优化表达式为:

$$\sum_{(u,i) \text{ obs.}} (\mathbf{R}_{u,i} - \hat{\mathbf{R}}_{u,i})^2 + \sum_{(u,v) \text{ obs.}} (\mathbf{S}_{u,v}^* - \hat{\mathbf{S}}_{u,v}^*)^2 + \lambda(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2 + \|\mathbf{Z}\|_F^2)$$

SoRec之后, 其他基于共享用户特征矩阵的方法被陆续提出。文献[3]的TrustMF模型, 是基于信任与被信任关系, 将用户映射到信任者特征向量、被信任者特征向量这两个不同的特征向量; 并限制这两个特征向量与用户特征向量等价。为了放松这种等价强约束条件, 文献[4]又在TrustMF基础上, 采用概率思想, 提出了TrustPMF。

2) 基于用户特征矩阵重表示的方法

文献[5]提出了STE(social trust ensemble)方法, 这个方法与SoRec不一样, 它是在评分预测函数上把评分矩阵分解方法和社交矩阵分解线性组合在一起, 得到用户对物品的评分。这种方法不仅考虑了用户自身的偏好对于评分的影响, 而且考虑了用户的朋友对其评分的影响。用户 u 对物品 i 的评分预测为(α 表示用户的朋友对于其评分的影响权重):

$$\hat{\mathbf{R}}_{u,i} = r_m + \alpha \mathbf{Q}_u \mathbf{P}_i^T + (1 - \alpha) \sum_{v \in F_u^+} \mathbf{S}_{u,v} \mathbf{Q}_v \mathbf{P}_i^T$$

文献[6]提出的SocialMF方法, 在RMSE(root mean squared error)指标上, 该方法的推荐效果要好于SoRec和STE模型。相比这两个模型, SocialMF模型解决了社交网络中的信任传播问题。一个用户的

隐特征向量既依赖于其直接的邻近用户, 又可通过网络进行传播, 控制依赖性的权重随着距离的增大而逐渐衰减。该方法通过在矩阵分解模型的优化函数中添加社会正则约束项来实现用户隐特征向其朋友隐特征逼近。此方法先对用户 u 的朋友 v 的隐特征 \mathbf{Q}_v 用信任度加权, 然后再与用户 u 的隐特征 \mathbf{Q}_u 求差值, 最后用这个差值来构成约束项, 其表达式为:

$$\sum_{(u,i)\text{obs}} (\mathbf{R}_{u,i} - \hat{\mathbf{R}}_{u,i})^2 + \beta \sum_{\text{all } u} \left\| \left(\mathbf{Q}_u - \sum_{v \in F_u^+} s_{u,v}^* \mathbf{Q}_v \right) \right\|_F^2 + \lambda (\| \mathbf{P} \|_F^2 + \| \mathbf{Q} \|_F^2)$$

文献[7]提出了SoReg, 它是一种基于相似性的社会正则约束方法。可以采用皮尔逊相关系数(Pearson correlation coefficient, PCC)或者向量空间相似度(vector space similarity, VSS)来求用户之间的相似性。不同于文献[6]中的基于平均的正则化项。它基于SocialMF提出了另一种社会正则化约束项。该方法先求出用户 u 的隐特征 \mathbf{Q}_u 与朋友 v 的隐特征 \mathbf{Q}_v 的差值的二阶范数, 然后再用 u 和 v 之间的相似度来加权。SocialMF方法通常称为基于平均的方法, 而这里的方法称为基于个体的方法。其表达式为:

$$\beta \sum_{\text{all } u} \sum_{v \in F_u^+} \text{sim}(u,v) \| \mathbf{Q}_u - \mathbf{Q}_v \|_F^2$$

用户偏好具有多样性, 用户可能同时隶属于多个不同的社群, 因此, 文献[8]提出重叠社群正则化推荐模型MFC。文献[9]基于现有模型假定社交关系具有同构关系这一猜想, 提出了基于社交维度的推荐模型SoDimRec。这里的社交维度表示一组兴趣爱好相似的用户集合, 这些用户之间不要求具有社交关系。该方法采用传统社群发现算法识别用户的社交维度, 并根据社交网络的同质性及异构性, 假定每一个社交维度对应一个偏好向量, 属于某社交维度的用户, 其特征向量应与该社交维度偏好向量相似。

文献[10]提出LoCo模型来解决用户冷启动。该模型通过分析解决冷启动问题的典型方法, 包括社交近邻方法、联合矩阵分解、特征映射矩阵分解, 从线性回归的角度提出这种低秩冷启动模型。

2 模型的提出

2.1 社会信任度的计算

在社交网络中, 用户之间的关系是通过用户之间的链接来体现的。有可能是通用的社交网络, 如Facebook; 也有可能是一个特定领域的社交网络, 比如电影推荐领域Flixster和各种产品推荐的Epinions。把社交网络表示为有向图 $G=(U,F)$, 其

中 U 是用户集合, 且 $|U|=u_0$, F 是朋友关系链接集合。社交网络信息由矩阵 $\mathbf{S} \in \mathbf{R}^{u_0 \times u_0}$ 表示。每个用户 u 都有一个集合 F_u^+ 表示其信任或关注的直接邻居, 同时存在另外一个集合 F_u^- , 该集合中的用户对 u 有信任或关注关系。用户 u 与用户 v 直接、加权的社交关系(如用户 u 信任/知道/关注用户 v)是由一个正值 $s_{u,v} \in (0,1]$ 来代表。如果不存在关系或未观察到某社交关系时 $s_{u,v} = s_m$, 通常 $s_m = 0$ 。社交权重 $s_{u,v}$ 可以解释为用户在社交网络中信任或知道用户 v 的程度, 它可能是用户 u 关于用户 v 的明确反馈(例如通过投票), 或者从隐式反馈中推断(例如交互/通信的程度)。通常情况下社会信任 $s_{u,v}$ 是正的, 在某些特殊情况下, 它也可以采取负值, 以明确表示两个用户的兴趣相互矛盾。图2是一个社交网络的简单示例。其中展示了6个用户之间的社交网络, 其中每个用户有一组朋友。每个有向的关系链接都由正的信任值加权。本文中的术语“信任网络”和“社交网络”表述的意思相同, 并且假设信任值都是正值。

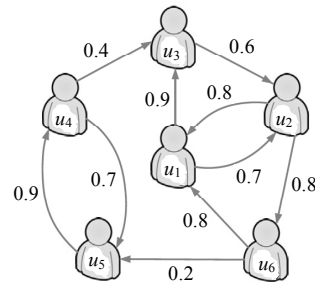


图2 社交网络示例

用矩阵 $[T_{u,v}]_{m \times m}$ 表示信任值, 矩阵中的元素 $T_{u,v}$ 表示社会关系中用户 u 对用户 v 的信任程度, 通常是非对称的。人们总是希望需求朋友的推荐, 但是对一个用户来说, 来自他/她所信任的朋友的推荐, 并不总是适合于他/她, 因为他/她可能具有不同的偏好。所以用偏好相似度来对信任程度计算公式进行修正:

$$T_{u,v} = \frac{\text{sim}(u,v) \times \text{trust}(u,v)}{\sum_{s \in N(u)} \text{sim}(u,s) \times \text{trust}(u,s)} \left(\sum_{v \in N(u)} T_{u,v} = 1 \right)$$

式中, 这里 $\text{sim}(u,s)$ 表示是用户 u 和用户 s 之间的相似度, 可以通过计算他们的隐因子向量 \mathbf{P}_u 和 \mathbf{P}_s 之间的cosine相似度来得到, $\text{trust}(u,s)$ 是用户 u 分配给用户 s 的信任值, 在没有明确给出信任值的情况下, 如果用户 u 链接到用户 s 则其值为1, 否则为0; $N(u)$ 是用户 u 直接链接出去的用户集合。对于被间接链接的用户, 通过乘法规则传播信任, 从而计算出对

他/她的信任程度。如果有多条信任路径, 选择最短的路径进行传播。

2.2 集成社会信任的模型

因为人们相信朋友的口味, 所以总是求助于朋友去看电影、听音乐或是推荐书等。有时, 为了做出决定, 用户会征求许多朋友的意见。这里基于这种假设来提出基于矩阵分解技术的社会推荐模型。

正如上述分析, 在为用户 u 进行推荐时, 其信任的所有朋友 $N(u)$ 的偏好也应该有所考虑。为了实现这一点, 把社会信任集成到矩阵分解协同过滤模型中, 具体是通过社会信任信息来调整用户的隐因子。建模如下:

$$\hat{r}_{u,i} = \mu + b_u + b_i + \left(\mathbf{P}_u + \alpha \sum_{v \in N(u)} \mathbf{T}_{u,v} \mathbf{x}_{u,v} \right)^T \mathbf{Q}_i$$

$$L = \frac{1}{2} \sum_{(u,i) \in R} (r_{u,i} - \hat{r}_{u,i})^2 + \frac{1}{2} (\lambda_1 b_u^2 + \lambda_2 b_i^2 + \lambda_3 \|\mathbf{P}_u\|^2 + \lambda_4 \|\mathbf{Q}_i\|^2)$$

式中, $N(u)$ 是用户 u 的朋友集合; $\mathbf{T}_{u,v}$ 是相似性函数, 不同的朋友具有各自的相似性。如果用户 v 与用户 u 非常相似, 如 $\mathbf{T}_{u,v} = 0.90$, 那么用户 v 就对用户 u 的偏好特征的修正多一些。另一方面, 如果两个用户不相似, 如 $\mathbf{T}_{u,v} = 0.1$, 那么用户 v 的修正就少一些。

3 用自编码器初始化

在大部分的矩阵分解中, 用户和物品的隐特征因子都是用随机值进行初始化的。这些算法收敛到局部最优将依赖于初始点。由于相似的用户(例如对物品评分值相似的用户)具有相似的用户特征向量, 因此, 如果特征向量的初始值更准确地描述了用户之间或物品之间的相似性, 则将可以训练出更优化的模型^[11]。因此, 本文通过从原始评价矩阵中提取特征来获得初始值, 并采用自动编码器来学习用户和项目的隐特征的初始值(\mathbf{P} 和 \mathbf{Q})。因为矩阵分解是非凸优化问题, 所以不能保证可以得到最优隐特征矩阵 \mathbf{P} 和 \mathbf{Q} ^[12], 而且根据 \mathbf{P} 和 \mathbf{Q} 的不同初值, 会收敛到不同的局部最优解, 但如果初始值设置适当, 结果可以接近最优。

压缩自动编码器CAE(contractive autoencoder)^[13]是生成特征表示的一种有效方法。CAE对小的扰动具有鲁棒性, 所以使用CAE作为预训练算法以得到用户和物品隐特征值的初始值。

4 信任群组的检测

目前, 群组检测算法大多基于网络节点之间的可达性来识别。信任群组nt-clique(n trust clique)检测算法基于n-clique^[14], 在社交网络中使用用户之间的信任关系来检测群组, 且该算法可以检测重叠的群组。首先重新定义文献[14]中的n-clique概念, 然后再定义社会信任网络nt-clique, 最后详细论述nt-clique群组检测算法。

4.1 信任群组的定义

1) n-clique的定义

给定一个网络 G , n -clique是一个最大子图, 其中每对节点之间的最大距离不超过 n , 这里的距离是在原始网络中定义的, 即:

$$\text{dist}(v_i, v_j) \leq n \quad \forall v_i, v_j \in G$$

2) 质量函数的定义

由质量函数来定义划分好坏, 得分高的划分就是好的划分。然而, 一个划分是否优于另一个划分, 取决于信任群组的具体定义和所采用的质量函数。对于图的任何划分 P , 质量函数定义为:

$$Q(P) = \sum_{C_i \in P} q(C_i)$$

式中, C_i 是划分 P 的第 i 个聚类; $q(C_i)$ 是划分中聚类 C_i 的质量函数。这里, 划分的质量是由各个聚类质量的和来定义, 所以质量函数 Q 是加性的。聚类的质量函数 $q(C_i)$ 定义为:

$$q(C_i) = \frac{1}{|P|} \frac{\sum_{u,v \in C_i} T_{u,v}}{\sum_{u \in C_i, v \notin C_i} T_{u,v}}$$

式中, $|P|$ 是划分 P 中聚类的个数。 $Q(P)$ 的最大值意味着同一个聚类中的用户之间的信任度要大于此聚类中的用户与其他聚类中用户之间的信任度。同一个聚类中的用户的偏好更接近, 而不同聚类中用户之间偏好差异更大。

3) nt-clique定义

一个nt-clique是一个信任网络 G_T 的子图, 其中每一对顶点之间的最大距离不大于 n 。 G_T 中的聚类应该保证全局质量函数 $Q(P)$ 取得最大值。

4.2 信任群组检测算法

这是一个NP问题, 因此, 使用最优化方法来近似。该算法以文献[15]为基础, 并结合了贪婪算法^[16]和模拟退火算法。贪婪算法总是在每个步骤中做出最好的选择, 但这并不能保证算法达到全局最优解, 所以再结合模拟退火算法来实现全局优化的概率化

过程。通过将这两种方法结合起来,希望得到一种快速、概率的全局最优信任群组检测算法。

nt-clique算法的步骤如下:

1) 检测算法总共运行 l 轮迭代,在每一轮迭代中产生划分 P 。最开始划分 P 被初始化为用户集合 U ,也就是每一个用户都构成一个单独的聚类,把它标记为 C_i (clique群),其中 $i=1,2,\dots,n_c$, n_c 是类的数量。 J 是 C_i 连接的顶点的标号集合。首先令 $i=1$ 。

2) 假设合并类 C_i 和顶点 $v_j \in J$,计算相应质量函数的变化量 ΔQ 。当 j 扫描完 C_i 的所有邻接顶点,找到类 C_i 具有最大的质量函数变化量 ΔQ_{\max} 。

3) 如果 $\Delta Q_{\max} > 0$,接受它并且置 $C_j = t$ 。否则以一定概率接受它(类似模拟退火,这里把接受概率设置成 0.4)。

4) 如果 $i+1 < n_c$,则 $i=i+1$ 并且回到步骤2); 否则回到步骤5)。

5) 如果迭代次数为 l ,则运行了 l 轮得到了 l 个不同的划分结果。最后选择具有最大质量函数值的划分作为最终的结果。如果迭代次数小于 l ,作如下操作:当 i 遍历完所有的信任群组后,得到一个新的划分 P ,然后用相同的标记来合并群组;重置标记并更新信任群组的数量 n_c (通常 n_c 逐渐减小);迭代数量增加1;回到步骤2)。

5 实验与分析

5.1 数据集

本文采用开放的数据集 Flixster 评估模型的性能。Flixster 是一个社交网络服务,用户可以向其朋友列表添加一些用户创建一个社交网络,也可以对电影进行评分。它是一种广泛应用于评价 CF 算法的电影数据集。数据集包括评分(评分在区间 $[1,2,\dots,5]$ 内)和信任关系。数据集的详细统计数据汇总如表 1 所示。

表1 Flixster数据集统计数据

统计项	数值
用户数	1 049 455
物品数	4 923 59
评分数	8 238 597
信任关系数	26 771 123

5.2 评价指标

1) 均方根误差 RMSE: 均方根误差用于度量预测准确度,本文用来度量推荐算法预测评分与用户实际评分的相似程度,定义为:

$$\text{RMSE} = \sqrt{\sum_{u,i} (r_{u,i} - \hat{r}_{u,i})^2 / N}$$

式中, $r_{u,i}$ 是实际评分值; $\hat{r}_{u,i}$ 是预测的评分值; N 表示测试数据集中评分的数量。

2) 覆盖率: 它可以用于度量,被预测过的用户或物品数量在整个用户或物品数量中的百分比,也叫召回率(recall)。当数据集高度稀疏时,一些推荐模型可能无法预测测试数据中的所有评分,本文的算法目标之一就是缓解该问题,覆盖率的提高程度可以用来衡量该问题的解决程度。覆盖率定义为:

$$R = S/N$$

式中, R 是覆盖率或召回率; N 表示被测试用户-物品对的数量; S 能做出预测评分的用户-物品对的数量。通常覆盖率需要结合准确率来评价推荐性能,因为推荐系统不能仅仅为了提高覆盖率,而不顾推荐的准确率。本文采用 F -值来衡量准确率。

3) F -Measure: 它是精度和召回率的调和平均值,是测试准确性的一种度量。首先把 RMSE 转换成 $[0,1]$ 区间的精度,即:

$$P = 1 - \text{RMSE}/4$$

式中, P 代表精度度量。因为数据评分集中在区间 $[1,5]$,最大的错误值为 4,故对 RMSE 除以 4 以变换到区间 $[0,1]$ 中。然后, F -Measure 定义为:

$$F - \text{Measure} = (2PR)/(P + R)$$

5.3 实验结果

5.3.1 初始化评估

1) 初始化的影响

通常,用户、物品的隐特性初始化方法有:随机初始化、零初始化、 k -均值初始化和规范化切割(normalized-cut, Ncut)初始化^[17]。而模型采用压缩编码器(contractive autoencoder, CAE)来初始化。将初始化方法与这 4 种方法进行了比较。对于这个学习模型,设置 $K=60$,使用 2t-clique ($n=2$)。图 3 表明,CAE 初始化方法可以提供比其他方法更好的推荐性能。其主要原因是,该初始化方法比其他初始化方法更能反映用户和物品相似性上的原始关系。

2) 维数的影响

使用预训练提取特征和降维。维数是影响算法的一个重要因素,因为它是矩阵分解的因子个数。使用网格搜索法搜索模型参数的最佳值。设置 P, Q 的超参数 λ_3, λ_4 均为 0.1, P, Q 的学习率为 0.01,用 2t-cliques 社区检测。图 4 表明,随着维数的增加, RMSE 迅速下降,其原因是随着维度的增加,模型的容量逐渐增大,建模能力越强。然而, RMSE 值开始上升,当维数增加到一定值(60)时,由于过度拟合, RMSE 曲线开始上升。

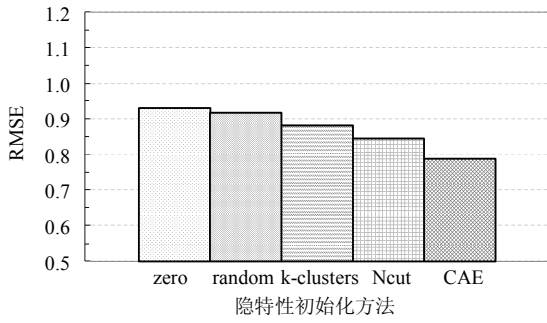


图3 初始化的影响

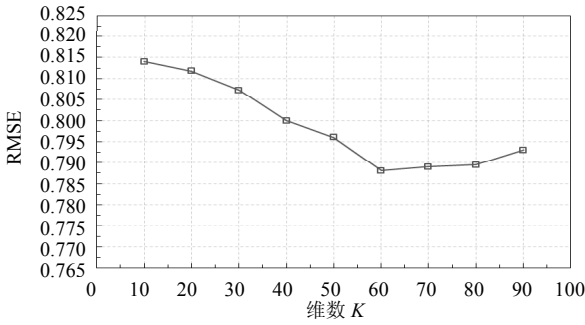


图4 维数的影响

5.3.2 社群检测中社群大小的影响

Clique的大小 n 决定了近邻数量。对于测试数据集, 设置 $K = 60$, n 从1到5变化。图5显示clique大小 n 对预测精度的影响。当 n 增加时, RMSE逐渐下降; 然而, 当达到一定的阈值(大约为2), RMSE开始增加。导致这种情况的原因是: 1) 当 n 相对小时, clique有助于对缺失评分进行预测; 2) 当 n 大于一定的阈值, 由于需要参考的近邻太多, 很难提取准确的特征。

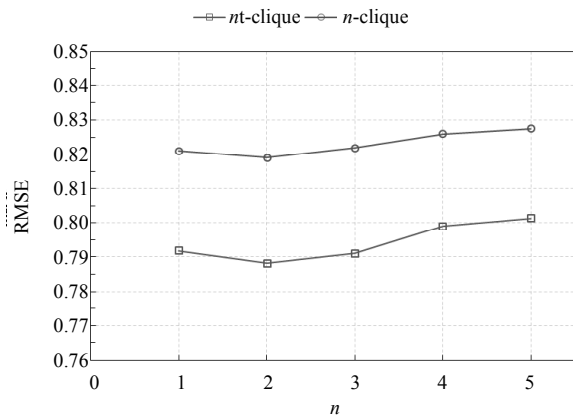


图5 clique大小的影响

5.3.3 综合比较

用以下最新的模型评估提出算法的性能。

1) TidalTrust^[18]: 采用信任推断算法预测评分。

2) MoleTrust^[19]: 使用信任度量来寻找相似的用户, 并通过信任边来传播和预测信任。

3) BMF^[20]: 是基本的MF模型, 没有使用社会网络。

4) STE^[5]: 在MF模型的用户特征中集成了其朋友的特征。

5) SocialMF^[6]: 在矩阵分解模型中集成了信任传播。

6) GWNMF^[21]: 基于图规整的非负矩阵分解。

使用网格搜索来获取这些模型的最佳性能, 其结果如表2所示。

表2 综合比较

指标	Tidal-Trust	Mole-Trust	BMF	STE	Social-MF	GWNMF	Our Model
RMSE	0.834 5	0.825 2	0.815 3	0.802 3	0.797 7	0.796 1	0.788 1
Coverage/%	75.54	76.39	86.33	89.43	89.93	92.13	94.78
F-Measure	0.773 4	0.779 4	0.830 5	0.845 8	0.847 9	0.856 7	0.873 3

实验在3个评价指标(RMSE、覆盖率、F值)上比较了上述模型的推荐性能。这里仍然将维数设置为60。根据表2, 可以得出以下结论: 该模型在3个评估指标上优于本文实验中的其他模型。

6 结束语

本文首先分析了基于社会信任信息的协同过滤研究现状, 并重点论述了矩阵分解算法中集成社会信任信息的方法。然后, 提出了集成社会信任关系的矩阵分解协同过滤算法。在该算法中, 采用了压缩自动编码器来提取用户和物品的初始化隐特征向量, 并且提出一个在信任网络中检测信任群组的算法。在实际数据集上的实验表明, 本文的模型取得了比其他相关模型更好的覆盖率和推荐精度, 从而验证了本文模型确实能够有效地缓解冷启动、提高准确性。

参 考 文 献

[1] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the International Conference on World Wide Web. New York : ACM Press, 2001: 285-295.

[2] MA Hao, YANG Hai-xuan, LYU M R, et al. Sorec: Social recommendation using probabilistic matrix factorization [C]//Proceedings of the ACM Conference on Information and Knowledge Management. California: ACM Press, 2008: 265-275.

[3] YANG Bo, LEI Yu, LIU Da-yong, et al. Social collaborative filtering by trust[C]//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Beijing: AAAI Press, 2013: 2747-2753.

[4] YANG Bo, LEI Yu, LIU Da-Yong, et al. Social collaborative

- filtering by trust[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1633-1647.
- [5] MA Hao, KING I, LYU M R. Learning to recommend with social trust ensemble[C]//Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval. Boston: ACM Press, 2009: 19-23.
- [6] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks[C]//Proceedings of the ACM Conference on Recommender Systems. Barcelona: ACM Press, 2010: 26-30.
- [7] MA Hao, ZHOU Deng-yong, LIU Chao, et al. Recommender systems with social regularization[C]//ACM International Conference on Web Search and Data Mining. Hong Kong, China: ACM Press, 2011: 287-296.
- [8] LI Hui, WU Ding-ming, Tang Wen-bin, et al. Overlapping community regularization for rating prediction in social recommender systems[C]//Proceedings of the 9th ACM Conference on Recommender Systems. Vienna: ACM Press, 2015: 27-34.
- [9] TANG Ji-liang, WANG Su-hang, HU Xia, et al. Recommendation with social dimensions[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Honolulu: AAAI Press, 2016: 251-257.
- [10] SEDHAIN S, MENON A K, SANNER S, et al. Low-Rank linear cold-start recommendation from social data[C]//Proceedings of the 31th AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2017: 1502-1508.
- [11] HIDASI B, TIKK D. Initializing matrix factorization methods on implicit feedback databases[J]. Journal of Universal Computer Science, 2013, 19(12): 1834-1853.
- [12] REZAEI M, BOOSTANI R, REZAEI M. An efficient initialization method for nonnegative matrix factorization [J]. Journal of Applied Sciences, 2011, 11(2): 354-359.
- [13] RIFAI S, VINCENT P, MULLER X, et al. Contractive auto-encoders: Explicit invariance during feature extraction [C]//Proceedings of the International Conference on Machine Learning. Washington: IMLS, 2011: 833-840.
- [14] FORTUNATO S. Community detection in graphs[J]. Physics Reports, 2009, 486(3): 75-174.
- [15] ZHOU Zhen, WANG Wei, WANG Liang. Community detection based on an improved modularity[M]. Berlin: Springer Berlin Heidelberg, 2012: 638-645.
- [16] BLONDEL V D, GUILAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics, 2008(10): 155-168.
- [17] ZHANG He, YANG Zhi-rong, OJA E. Improving cluster analysis by co-initializations[J]. Pattern Recognition Letters, 2014, 45(11): 71-77.
- [18] MASSA P, AVESANI P. Trust metrics in recommender systems[M]. London: Springer London, 2009: 259-285.
- [19] MASSA P, AVESANI P. Trust-aware recommender systems[C]//Proceedings of the ACM Conference on Recommender Systems. Minneapolis: ACM Press, 2007: 17-24.
- [20] KANNAN R, ISHTEVA M, PARK H. Bounded matrix factorization for recommender system[J]. Knowledge & Information Systems, 2014, 39(3): 491-511.
- [21] GU Quan-quan, ZHOU Jie, DING C. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs[C]//Proceedings of the SIAM International Conference on Data Mining. Ohio: SIAM Press, 2010: 199-210.

编辑 刘飞阳