

一种融合异构信息网络和评分矩阵的推荐新算法

张邦佐 桂欣 何涛 孙焕垚 杨晟雨 韩宇茹
东北师范大学计算机科学与技术学院, 长春 130117
(zhangbz@nenu.edu.cn)

A Novel Recommender Algorithm on Fusion Heterogeneous Information Network and Rating Matrix

Zhang Bangzuo, Gui Xin, He Tao, Sun Huanyao, Yang Chengyu, and Han Yuru
(College of Computer Science and Information Technology, Northeast Normal University, Changchun, 130117)

Abstract In the present era of big data, recommender system is an effective way to solve the problem of information overload. Heterogeneous information network provides the opportunity for recommender system to fuse more information of users and items. While the traditional recommender system based on collaborative filtering only uses the user item rating matrix, this paper puts forward a novel solution to integrate the merits of them. Firstly, establishes the heterogeneous information network based on the user and item attributes, secondly computes the similarity matrix of users and items based on meta path similarity algorithm, thirdly uses the new three methods to fuse the similarity matrixes and the user item rating matrix, finally uses the traditional recommendation techniques based on matrix factorization to predict and then merges the results. With the experiment in the international standard data set MovieLens 100K and the film properties extended from IMDB, verifies the effectiveness of the method that the use of heterogeneous information network to fuse more users and item attributes information and the rating matrix can improve the recommendation accuracy effectively.

Key words heterogeneous information network; matrix factorization; similarity matrix; meta path; collaborative filtering;

摘要 在当今的大数据时代, 推荐系统是解决信息超载的有效手段。异构信息网络为推荐系统更多地融合用户与项目信息提供了机遇, 而传统的协同过滤推荐系统仅仅使用用户-项目评分矩阵, 为了结合二者的优点, 提出了一个融合它们的新的解决方案。首先通过对推荐系统中的用户属性和项目属性建立异构信息网络, 其次使用基于元路径的相似度计算方法分别计算用户与项目的相似度矩阵, 然后使用本文提出的三种将相似度矩阵与用户-项目评分矩阵融合方法, 最后采用传统的基于矩阵分解的推荐技术进行预测及结果合并。在 MovieLens 100K 这一国际标准数据集上以及通过 IMDB 对电影属性进行扩展后, 通过实验验证了使用异构信息网络更多地引入用户与项目属性信息以及融合评分矩阵可以有效地提高推荐精度。

关键词 异构信息网络; 矩阵分解; 相似度矩阵; 元路径; 协同过滤;

中图法分类号 TP391

推荐系统是人们解决信息超载^[1]问题的有效手段^[2], 特别是在当前的大数据时代, 如何从海量的数据中准确地找到自己所需要的信息, 是一个更加紧迫

的问题。在工业界, 特别是以电子商务为代表的电子商务服务行业, 推荐系统带来了服务质量的巨大改变, 取得了巨大的成功, 也掀起了研究的热潮, 比如: 提供

收稿日期: 2014-11-4

基金项目: 吉林省产业技术研究与开发专项项目 (JF2012c016-2)

流媒体服务的Netflix声称自己2/3的用户观看来自于推荐, Amazon也宣称35%的销售来自于推荐系统。

协同过滤^[3]方法基于这样的思想:用户在过去有相同的偏好,那么他们在未来也会有相似的偏好,它也是推荐系统中研究得最早也是最多的领域,并在工业界得到了广泛的应用。协同过滤技术一般分为两类:基于记忆的和基于模型的,基于模型的方法会首先离线处理原始数据,学习相应模型,通常采用各种矩阵分解方法^[4]。矩阵分解也被认为是当前协同过滤推荐系统中效果最好的方法。纯粹的协同过滤方法的输入数据只有用户-项目评分矩阵,因此它只应用了用户对项目的历史评分。

无论是基于用户的协同过滤还是基于项目的协同过滤,其中最近邻的计算都只依赖于用户-项目评分矩阵。由于评分矩阵一般都非常稀疏,因此协同过滤的主要挑战就是要用相对较少的有效评分来得到准确的预测。如何更好的理解和利用用户与项目信息是协同过滤算法一种有效的扩展方法^[5]。作为大数据挖掘工具的异构信息网络(Heterogeneous Information Networks)^[6]为实现这一方法提供了有力保证。

以经典的电影推荐为例,在实际生活中,人们喜欢某部电影,不仅会看其他用户对电影的评分,也会考虑到电影的演员、导演等属性,电影厂家也会有意地将这些作为卖点。因此考虑用户和电影的多个属性,构建异构信息网络,通过选择不同的属性就会产生不同的联系路径,各个路径都对推荐结果有一定的贡献,合并这些结果可以实现更好的推荐。

目前基于异构信息网络的推荐工作还处于起步阶段,通常仅仅考虑用户与项目的交互关系所形成的异构信息网络,却放弃了用户-项目评分矩阵这一协同过滤推荐中最重要的信息来源。本文的独特之处,即是基于异构信息网络中元路径选择学习多个用户与项目相似度,并与用户-项目评分矩阵融合实现更好的推荐。工作针对经典的电影推荐情景,并在MovieLens100K这一常用数据集上,以平均绝对误差这一常用的评测标准通过实验证明了方法的有效性,提高了推荐质量。

本文工作主要分为五个部分,第一部分介绍相关工作;第二部分介绍异构信息网络及其建立;第三部分为算法的详细描述及复杂性分析;第四部分为实验设计及结果分析与讨论;最后,在第五部分对全文进行总结。

1. 相关工作

在学术界,自1992年出现了第一篇关于协同过滤的论文^[3]以来,推荐系统一直保持着很高的研究热度,并逐渐发展成为一门独立的学科,美国计算机学会(ACM)从2007年开始还设立了推荐系统年会。

推荐系统一般由推荐对象(项目)、用户和推荐算法构成。其中推荐算法是推荐系统中核心的部分,在很大程度上决定了推荐性能的优劣。推荐系统一般包括基于协同过滤的推荐、基于内容的推荐、基于知识的推荐、基于网络结构的推荐,以及混合推荐方法等^[4]。

异构信息网络是随着社会网络的发展而出现的,一个较新的研究领域^[6],也称为异构社会网络(Heterogeneous Social Networks)或者多关系社会网络(Multi-Relational Social Network)^[7]。异构信息网络通过网络模式指定了对象集合上的类型约束和对象间的关系约束。这些约束使得异构信息网络是半结构化的,从而指导人们更好地去探索网络语义。异构信息网络可以从许多互联的社会的、科学的、工程的到商业应用的大规模数据集上进行构建^[8],也可以在电子商务中比如Amazon和eBay,在线电影数据库如IMDB(the Internet Movie Database),和各种数据库上进行应用。

异构信息网络的出现,为推荐系统提供了新的机遇,并为融合多种属性及属性组合的推荐技术提供了一种新的思路和信息集成框架,为实现更好的推荐提供了理论基础,这是推荐系统研究中新的挑战 and 机遇。

学术视频网站videlectures.net第一次将异构信息网络与推荐系统相结合^[9],考虑了学术视频所附带的讲演、作者、会议、用户之间构成的异构信息网络,目前该网站已经成为了世界上最大的学术视频网站。第一个真正使用异构信息网络进行推荐系统构建的工作是HeteRecom^[10],作者针对IMDB数据集,为每部电影节点选定演员、导演和电影类型三个文本属性构建网络,在KDD2012上介绍了系统原型构建方法。针对特定的用途和特定的上下文信息,OptRank^[11]将异构信息网络用于解决标签(tag)推荐问题;HeteroMF^[12]算法利用异构信息网络考虑了实体类型和每个实体的上下文依赖关系,并通过转换矩阵与矩阵分解技术相结合,针对冷启动和某一上下文占主导的情景改进了推荐效果;Liu等^[13]将全文信息用于引文推荐。针对经典情景的推荐问题,Yu等^[14]采用异构信息网络考虑了项目之间的相似度关系,并将用户与项目评分情况作为隐式反馈信息构建异构信息网

络,并借助排序学习得到最终推荐列表,之后作者又将该方法应用于实体推荐^[15],这两个工作的评估标准采用的是信息检索中的基于排序结果的平均倒数排序(Mean Reciprocal Rank, MRR)准则。

以上工作,在推荐过程中均未考虑用户-项目评分矩阵这一协同过滤推荐中最重要的信息来源,针对的推荐问题各不相同,测评标准也不一样。本文工作除了基于异构信息网络中元路径选择学习多个用户与项目相似度,更重要的是与用户-项目评分矩阵相融合来实现更好的推荐,在MovieLens这一最常用数据集上,针对电影推荐这一经典的推荐情景,以通常的评测标准验证了提出方法的有效性。

2 异构信息网络简介

异构信息网络是一种特殊的信息网络,它包含多种形式的对象类型,也包含多种类型的关系。信息网络^[8]通常定义如下:

定义 1. 信息网络(Information Network)是一个有向图 $G=(V, E)$,它具有一个对象类型映射函数 $\tau: V \rightarrow A$ 和一个链接类型映射函数 $\phi: E \rightarrow R$,其中每一个对象 $v \in V$ 属于一个特定的对象类型 $\tau(v) \in A$,每一个链接 $e \in E$ 属于一个特定的关系 $\phi(e) \in R$,如果两个链接属于同一个关系类型,那么两个链接具有同样的开始对象类型和终止对象类型。

与传统的网络定义不同之处在于,该定义明确区分了网络中的对象类型和关系类型。当对象类型 $|A| > 1$ 或者关系类型 $|R| > 1$ 时,该网络称为异构信息网络,否则称为同构信息网络。

对于一个给定的复杂的异构信息网络,为了更好地理解网络中的对象类型和链接类型,必须提供它的元描述。因此要定义描述网络元结构的网络模式^[8]。

定义 2. 网络模式(Network Schema)记作 $TG=(A, R)$,是异构网络 $G=(V, E)$ 的元模板,它具有对象类型映射 $\tau: V \rightarrow A$ 和链接类型映射 $\phi: E \rightarrow R$,是定义在对象类型 A 和链接类型 R 上的有向图。

相对同构信息网络而言,异构信息网络更加接近现实世界,具有更强的描述能力,现实世界的许多网络都可以很顺利地转换为半结构化的异构信息网络。在经典的MovieLens100K数据集中,为了进一步完善电影的常用属性,本文从IMDB数据库中对电影属性进行了进一步的扩充。本文使用的用户与电影的异构网络模式分别如图1和图2,由于图中的边都是双向的,所以没有标出其方向。

在异构信息网络中,两个对象能够通过不同的属性相连,这些不同的属性路径代表了不同的意义,因

此两个对象的相似度依赖于在异构信息网络中的搜索路径。

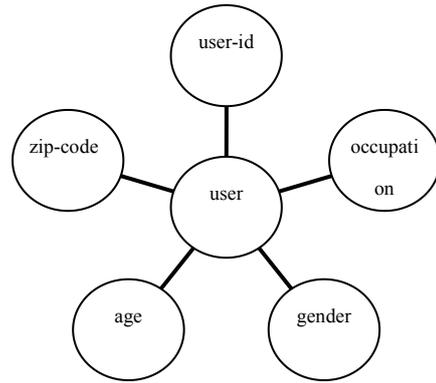


图1 根据 MovieLens 数据集建立的用户网络模式

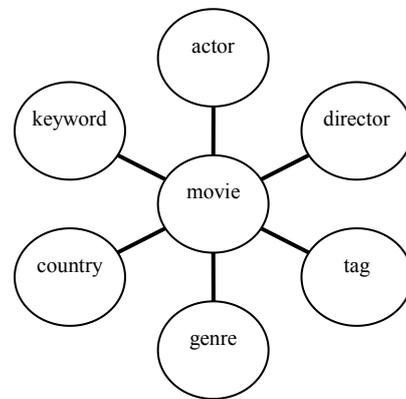


图2 根据 MovieLens 和 IMDB 数据集建立的电影网络模式

定义 3. 元路径^[16]是定义在网络模式 $TG=(A, R)$ 中的一条路径, $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_l} A_{l+1}$ 表示 A_1 和 A_{l+1} 之间的复合关系 R , 其中 $R = R_1 \circ R_2 \circ \dots \circ R_l$ 表示关系上的复合运算符,路径 P 的长度 $|P|$ 为在路径 P 中的关系个数,即 l 。

3 基于异构信息网络的推荐

为了充分利用用户和项目的属性信息,本文提出了基于异构信息网络的推荐方法。首先用基于元路径选择的相似度计算方法分别得出用户和项目的相似度矩阵,然后利用矩阵相乘方法将用户相似度矩阵、用户-项目评分矩阵、项目相似度矩阵进行融合,得到用户与项目之间的内在关系,之后再利用基于协同过滤的矩阵分解方法进行预测,最后进行结果合并,获得最终预测结果。

在一个有 m 个用户, n 个项目的协同过滤推荐系统中, 首先定义如下记号:

$$US = \{u_1, u_2, \dots, u_m\}, u_i \text{ 代表系统用户};$$

$$WS = \{w_1, w_2, \dots, w_n\}, w_i \text{ 代表系统项目};$$

$UA = \{p_1, p_2, \dots, p_k\}$, p_i 代表参与推荐过程的用户某一方面的属性, k 是用户属性数;

$WA = \{q_1, q_2, \dots, q_t\}$, q_i 代表参与推荐过程的项目某一方面的属性, t 是项目属性数;

$R_i \in R^{n \times m}$, 表示用户-项目评分矩阵, 是一个 n 行 m 列的矩阵。

3.1 基于元路径的相似度^[16]计算

定义 4. 基于元路径的相似度: 如果有元路径为 $P(A_1 A_2 \dots A_i \dots A_{i+1})$, 其中 A_i 表示项目的属性实例, 定义 R_i 为 A_i 到 A_{i+1} 上的关系, 即在 A_i 和 A_{i+1} 之间有一条边相连。则可以得到任意对象 s, t (s 是属性 A_i 的对象, t 是属性 A_{i+1} 的对象) 之间的相似度计算公式定义为:

$$HeteSim(s, t | R_1 \circ R_2 \circ \dots \circ R_l) = \frac{1}{|O(s | R_1)| |I(t | R_l)|}$$

$$\sum_{i=1}^{|O(s | R_1)|} \sum_{j=1}^{|I(t | R_l)|} HeteSim(O_i(s | R_1), I_j(t | R_l) | R_2 \circ \dots \circ R_{l-1})$$

(1)

其中函数 $O(R)$ 和 $I(R)$ 分别表示关系 R 的出度和入度。公式 (1) 是递归定义的, 可以在异构信息网络上通过图搜索来实现。

这种相似度总是基于某一选定的元路径的, 即针对某一电影 (movie), 则可以从电影到演员 (actor) 再到电影的路径探索下去, 比如如下路径:

movie-(actor-movie-actor)-movie;

同样, 用户之间的相似度也可以按照用户属性进行计算。 r 通常取自然数, 如果取值较大将会得到一个较长的路径, 但是由于相似度都是小于等于 1 的数值, 所以值会很快衰减, 由于实际计算的限制也不需要取得太长。注意到元路径通常有多条, 因此通过元路径的相似度计算可以考虑对象的多个属性, 这也符合现实生活中的直观。

通过基于元路径选择的相似度计算后, 对于用户来说可以得到用户间关于属性 p_i 的相似度矩阵,

$$U(p_i) \in R^{m \times m} (0 < i \leq k),$$

对于项目来说也可以得到项目间关于属性 q_i 的项目相似度矩阵 $W(q_i) \in R^{n \times n} (0 < i \leq t)$ 。

3.2 与用户-项目相似度矩阵的融合

对于协同过滤推荐系统来说, 用户相似度矩阵是一个 m 行 m 列的方阵, 同样项目相似度矩阵是一个 n 行 n 列的方阵, 同时注意到用户-项目的评分矩阵是一个 m 行 n 列的矩阵。因此本文提出通过在用户-项目评分矩阵左乘用户相似度矩阵或者右乘项目相似度矩阵的方法来实现融合用户与项目属性的目的, 三种信息融合方式分别为:

融合方式 1: (用户-项目评分矩阵) \times (项目相似度矩阵)

融合方式 2: (用户相似度矩阵) \times (用户-项目评分矩阵)

融合方式 3: (用户相似度矩阵) \times (用户-项目评分矩阵) \times (项目相似度矩阵)

Yu 等首次提出融合方式 1, 但是在计算时为了与网络结构结合却将用户-项目评分矩阵转换为 0/1 矩阵^[14], 因此不能与本文方法进行对比。融合方式 3 示例如图 3。

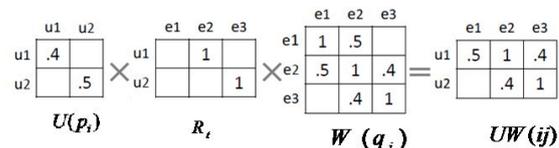


图 3. 用户、项目相似度矩阵与用户-项目评分矩阵融合示例

经过以上几种融合方法, 可以得到如下多个用户-项目关系矩阵:

$$UW(ij) \in R^{n \times m} (0 < i \leq k, 0 < j \leq t)$$

3.3 基于矩阵分解的推荐

2009 年 Netflix 竞赛表明, 矩阵分解方法在协同过滤推荐系统中非常有效。矩阵分解方法通常被归类为基于模型的协同过滤, 传统的矩阵分解方法针对的是用户-项目评分矩阵, 而本文中针对的是与用户或项目相似度矩阵融合后的 $UW(ij)$ 矩阵。本文主要使用了两个最著名也是最常用的矩阵分解算法:

1. 奇异值分解 (Singular Value Decomposition, SVD)^[17], 是矩阵分析中正正规矩阵酉对角化的推广。

为了控制过拟合, 它将一个矩阵分解成用户和项目两个特征值基向量。为了缩短过程, 使用了梯度下降法, 在对特征进行合适的选择之后, 可以实现精确度比较高的预测。

2. 非负矩阵分解 (Non-negative Matrix Factorization, NMF)^[18], 也是将矩阵分解为用户和项目基向量, 但是它存在一个约束条件: 所有的低秩矩阵必须是非负数, 该算法用了乘法更新原则来减小实际值和评估值的欧氏距离或者相对熵。

3.4 推荐结果的合并

针对多种元路径选择和多种矩阵融合方法, 经过矩阵分解方法运算后会得到多个结果预测矩阵, 因此需要进一步合并多个结果, 得到最终的预测结果。本文主要选取了较简单的取多个结果平均值的合并方法^[19], 经过合并之后就可以得到最终的预测结果。

3.5 算法复杂性分析

对于基于元路径的相似度计算方法, 如果采用宽度优先的单向搜索, 复杂度是 $O(n * m * l)$, 如果采用双向搜索复杂度是 $O(n^3)$; 融合方法需要进行矩阵乘法, 复杂度是 $O(n^3)$; 两种矩阵分解预测算法复杂度接近, 均为 $O(n^3)$; 推荐结果合并方法的复杂度为 $O(n^2)$ 。因此综合考虑整体算法的复杂度为 $O(n^3)$ 。可见使用异构信息网络进行矩阵融合并不会增加整体的时间复杂度。

传统的协同过滤推荐方法根本不考虑用户和项目的属性信息, 如用户的年龄、职业等人口统计学信息, 以及项目丰富的描述信息。通过引入异构信息网络, 可以系统化地考虑这些信息, 通过元路径选择, 可以计算用户以及项目之间基于元路径的相似度, 从而引入更多的背景信息, 有助于减少预测误差, 提高精度。

4 实验及结果

4.1 数据集

本文实验采用美国明尼苏达大学的 MovieLens 100K 数据集。该数据集共包含 943 个用户对 1682 部电影的 100 000 个评分, 评分采用 5 分制。为了提供更多的电影数据属性信息, 本文结合了互联网电影数据集 IMDB, 提取了更多的电影相关对象。为方便评

估和实验结果的可重复性, 本文评分数据采用 MovieLens100K 中已经划分好的 5 个数据集 (U1 到 U5), 它们都随机抽取 80% 的数据作为训练集, 20% 的数据留作测试集。

4.2 评价指标

评价指标使用通常的平均绝对误差 (Mean Absolute Error, MAE), 它表示对于测试数据偏离真实值的误差。MAE 越小则误差越小, 实验越准确。

4.3 实验设置

在计算基于元路径的相似度时分别选取了路径长度为 2 ($r=1$)、4 ($r=2$) 两种情况, 分别记为 $|P|=2, 4$;

在矩阵融合时分别选取了上述三种融合方法, 分别记为 $M=1, 2, 3$ 。

4.4 实验结果及分析

本文使用了 SVD 和 NMF 这两种方法基于矩阵分解的协同过滤推荐算法, 为了对比研究, 本文给出了不同元路径长度以及不同矩阵融合方法的运行结果, 分别见图 4 和图 5。

使用 SVD 方法在两种元路径长度和三种融合方法组合下的结果曲线如图 4。可以看到, 在使用 SVD 方法进行推荐的情况下, 元路径越长推荐效果越好, 元路径长度为 4 时的 MAE 值都明显小于元路径长度为 2 的情况, 取得了显著的推荐效果。从融合方法来看, 在使用 SVD 方法时, 分别使用了三种融合方法, 能有效提高推荐效果, 即融合项目相似度矩阵优于融合用户相似度矩阵, 同时融合用户和项目相似度矩阵优于仅融合项目相似度矩阵。虽然在元路径长度为 2 时, 同时融合用户相似度矩阵和项目相似度矩阵与仅融合项目相似度矩阵效果十分接近, 但在长度为 4 时拉开了距离, 其 MAE 值明显低于其它两种融合方法, 显示了同时融合两种相似度矩阵方法的优越。

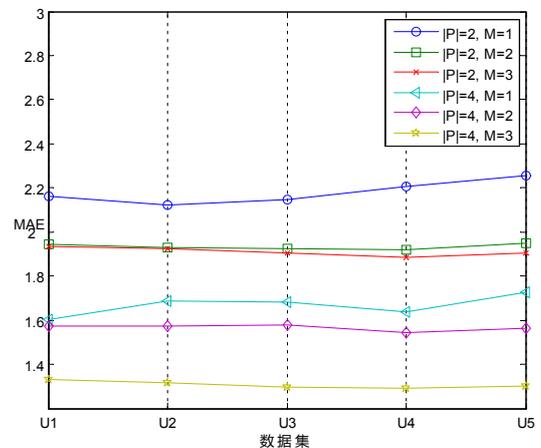


图 4 基于 SVD 方法不同的元路径长度和融合方法的比较

由图5可以看到,在使用NMF方法推荐的前提下,看到了和图4一致的更加明显的结果。从元路径长度来看,元路径越长预测效果越好。从融合方法来看,融合项目相似度矩阵优于融合用户相似度矩阵,虽然在元路径长度为2时在U4数据集上反而略低,但从平均效果上来看前者明显优于后者。同时融合用户相似度矩阵和项目相似度矩阵又优于仅融合项目相似度矩阵。

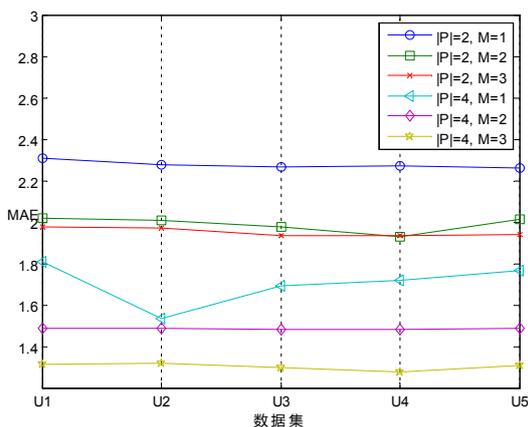


图5 基于NMF方法不同的元路径长度和融合方法的比较

综合图4和图5数据发现使用SVD和NMF进行协同过滤推荐MAE值相差不大,效果也差不多,以后我们将考虑更多的矩阵分解方法,测试它们在融合用户和项目相似度情况下的执行情况,这也是我们未来的工作之一。

在图4和图5中当元路径长度为2时,使用融合方法2和3的MAE值比较接近,可能和我们的数据集扩充方式有关,这两种方法都使用了电影相似度矩阵,在原始MovieLens 100K数据集中电影属性相对较少,所以我们通过IMDB数据集又提取了更多的电影属性,这样在元路径较短时用户属性与电影属性在异构信息网络中几乎独立,不发生联系,扩充之后由于用户属性数量明显少于电影属性,所以在融合中使用电影相似度矩阵会对结果造成较大的影响,从而使既融合电影相似度矩阵又融合用户相似度矩阵方法来说效果不明显。而当元路径长度增加之后,在异构信息网络中用户属性和电影属性发生相互联系,这样属性数量的影响就不明显了。这也给我们一个启发,为了达到更好的推荐效果,以后需要关注如下两个问题,即如何更多地挖掘用户和项目属性,以及如何选取合理的元路径长度,这也是我们未来可能的工作。我们也将选取更大的数据集进行实验进一步验证我们的结果。

5. 结束语

本文针对当今的大数据时代更多地关注数据之间的网络结构关系这一特性,使用异构信息网络这一有利工具针对经典的推荐问题进行了研究。通过对推荐系统中的用户属性和项目属性建立异构信息网络,使用基于元路径的相似度计算方法分别计算用户与项目的相似度矩阵,特别是使用提出的三种与用户-项目评分矩阵融合的方法,再结合传统的基于矩阵分解的推荐技术进行预测,及对结果进行合并,在MovieLens100K这一国际标准数据集上以及通过IMDB对电影属性进行扩展之后,使用经典的评测方法实验验证了本文方法的有效性。因此使用异构信息网络更多地引入用户与项目属性信息,并通过融合用户-项目评分矩阵,可以有效地利用数据的内在信息,从而较好地提高了推荐精度。

致谢 感谢美国UIUC的Xiao Yu博士关于数据预处理问题的耐心解答;感谢北京邮电大学石川博士关于异构信息网络相似度计算方法的指导。同时感谢评审专家提供的宝贵意见和建议!

参考文献

- [1] 许海玲, 吴潇, 李晓东, 阎保平. 互联网推荐系统比较研究. 软件学报, 2009(2): 350-362
- [2] Linyuan Lu, Matus Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, Tao Zhou. Recommender systems. Physics Reports Vol. 519 (1), PP. 1-49 (2012)
- [3] D. Goldberg, D. Nichols, B.M. Oki, and D. Terry. Using Collaborative Filtering to Weave an Information Tapestry. Comm. ACM, 1992, 35(12):61-70.
- [4] Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich. Recommender Systems: An Introduction. Cambridge University Press, 2011.
- [5] Gediminas Adomavicius, Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2005, 17(6):734-749.
- [6] Jiawei Han. Mining heterogeneous information networks by exploring the power of links. Discovery Science, Springer, 2009.
- [7] Deng Cai, Zheng Shao, Xiaofei He, Xifeng Yan and Jiawei Han. Mining Hidden Community in Heterogeneous Social Networks. 2005 ACM-SIGKDD Workshop on Link Discovery: Issues, Approaches and Applications (LinkKDD'05), Chicago, IL, Aug. 2005.
- [8] Yizhou Sun, Jiawei Han. Mining Heterogeneous Information

- Networks: Principles and Methodologies. Morgan & Claypool, 2012.
- [9] Miha Grear, Nada Lavrac. A Methodology for Mining Document-Enriched Heterogeneous Information Networks. *Discovery Science 2011*: 107-121.
- [10] Chuan Shi, Chong Zhou, Xiangnan Kong, Philip Yu and Gang Liu. HeteRecom: A Semantic Recommendation System in Heterogeneous Networks. In: *Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'12) (system demo)*, Beijing, China, 2012.
- [11] Wei Feng, Jianyong Wang. Incorporating Heterogeneous Information for Personalized Tag Recommendation in Social Tagging Systems. *Proc. the 18th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Beijing, China, Aug. 12 - Aug. 16, 2012. PP: 1276-1284.
- [12] Mohsen Jamali, Laks Lakshmanan. HeteroMF: recommendation in heterogeneous information networks using context dependent factor models. *Proceedings of the 22nd international conference on World Wide Web (WWW 2013)*. May 13-17, 2013, Rio de Janeiro, Brazil. Pages 643-654.
- [13] Xiaozhong Liu, Yingying Yu, Chun Guo, Yizhou Sun, and Liangcai Gao. Full-Text based Context-Rich Heterogeneous Network Mining Approach for Citation Recommendation. *ACM/IEEE Joint Conference on Digital Libraries (JCDL'14)*, London, 2014.
- [14] Xiao Yu, Xiang Ren, Yizhou Sun, Bradley Sturt, Urvashi Khandelwal, Quanquan Gu, Brandon Norick, and Jiawei Han. HeteRec: Entity Recommendation in Heterogeneous Information Networks with Implicit User Feedback. *Proc. of 2013 ACM Int. Conf. Series on Recommendation Systems (RecSys'13)*, Hong Kong, Oct. 2013
- [15] Xiao Yu, Xiang Ren, Quanquan Gu, Yizhou Sun and Jiawei Han. Collaborative Filtering with Entity Similarity Regularization in Heterogeneous Information Networks. *Proc. of IJCAI-13 HINA workshop (IJCAI-HINA'13)*, Beijing, China, Aug. 2013.
- [16] Chuan Shi, Xiangnan Kong, Philip S. Yu, Sihong Xie and Bin Wu. Relevance Search in Heterogeneous Networks. In: *Proceedings of the 15th International Conference on Extending Database Technology (EDBT'12)*, Berlin, Germany, 2012.
- [17] B.M. Sarwar, G. Karypis, J.A. Konstan, and J.Reidl. Application of dimensionality reduction in recommender system - a case study. In *ACM WebKDD 2000 Web Mining for E-Commerce Workshop*, 2000.
- [18] Daniel D. Lee and H. Sebastian Seung. Learning the parts of Objects by Non-negative Matrix Factorization. *Letters to Nature*, Vol. 401, 1999.
- [19] Judith Masthoff. *Group Recommender Systems: Combining Individual Models*. Springer US, 2011, pp 677-702.
- 张邦佐**, 男, 1971年生, 副教授, 研究方向: 推荐系统, 数据库, web 智能。2009年获得吉林大学博士学位。2011年美国伊利诺伊大学芝加哥分校访问学者, 发表论文30多篇。
- 桂欣**, 女, 1992年生, 本科生, 研究方向: 智能系统, 推荐系统。
- 何涛**, 男, 1992年生, 本科生, 研究方向: 数据库, 推荐系统。
- 孙焕焱**, 男, 1992年生, 本科生, 研究方向: 推荐系统。
- 杨晟雨**, 男, 1992年生, 本科生, 研究方向: 推荐系统。
- 韩宇茹**, 女, 1994年生, 本科生, 研究方向: 教育技术学。

本文校对负责人:张邦佐, 13304318265,
zhangbz@nenu.edu.cn.