

1-2-2019

## Multiple-attribute Entity Recommendation Based on Classification

Meina Song

*1.School of Computer Science, Beijing University of Posts & Telecommunications, Beijing 100876, China;  
;2.Information Network Engineering Research Center of Ministry of Education, Beijing 100876, China;*

Xuejun Zhao

*1.School of Computer Science, Beijing University of Posts & Telecommunications, Beijing 100876, China;  
;2.Information Network Engineering Research Center of Ministry of Education, Beijing 100876, China;*

Haihong E

*1.School of Computer Science, Beijing University of Posts & Telecommunications, Beijing 100876, China;  
;2.Information Network Engineering Research Center of Ministry of Education, Beijing 100876, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

---

## Multiple-attribute Entity Recommendation Based on Classification

### Abstract

**Abstract:** In the process of exploring entity recommendation, the entity containing diverse attributes has gained more and more attention. Most of the current researchers mainly select one attribute, and embody it in the related algorithms and their extensions even though the entity is combined with multiple attributes in entity recommendation. *In this paper, on the basis of the classification method, we delve into physical properties of the recommended entities, divide entity's attribute information network into multiple sub ones. In sub information network, bounded by the amount of attributes, the single attribute and even multiple attributes can be diverted into diverse paths of entity similarity, combining with entity similarity and related algorithm, where we can get the recommended results. This study not only refers to various attributes, but also improves the precision of recommendation.*

### Keywords

entity recommendation, multiple attribution, classification, similarity path

### Recommended Citation

Song Meina, Zhao Xuejun, E Haihong. Multiple-attribute Entity Recommendation Based on Classification[J]. Journal of System Simulation, 2018, 30(2): 405-413.

## 基于分类的多属性实体推荐

宋美娜<sup>1,2</sup>, 赵雪君<sup>1,2</sup>, 鄂海红<sup>1,2</sup>

(1.北京邮电大学计算机学院, 北京 100876; 2.教育部信息网络工程研究中心, 北京 100876)

**摘要:** 在面向实体推荐的探究过程中, 包含多样属性的实体获得了越来越多的关注。然而, 在面对这种复杂实体对象的推荐过程中, 当前的研究工作者大多选用其中一种属性, 利用相关算法得以实现。在基于分类处理的方法, 对推荐实体的属性进行了深入探究, 将物品属性网络划分为多个子信息网络。在子信息网络中, 以属性数目为界, 将子网络中的多属性以及单属性转化为实体相似度的多样途径, 结合实体相似度以及相关算法, 得出推荐结果。最后通过实验验证本文的算法, 引用多样属性不仅体现了推荐多样性, 还提高了推荐精度。

**关键词:** 实体推荐; 多样属性; 分类; 相似度途径

中图分类号: TP391.9

文献标识码: A

文章编号: 1004-731X (2018) 02-0405-09

DOI: 10.16182/j.issn1004731x.joss.201802005

### Multiple-attribute Entity Recommendation Based on Classification

Song Meina<sup>1,2</sup>, Zhao Xuejun<sup>1,2</sup>, E Haihong<sup>1,2</sup>

(1.School of Computer Science, Beijing University of Posts & Telecommunications, Beijing 100876, China;

2.Information Network Engineering Research Center of Ministry of Education, Beijing 100876, China)

**Abstract:** In the process of exploring entity recommendation, the entity containing diverse attributes has gained more and more attention. Most of the current researchers mainly select one attribute, and embody it in the related algorithms and their extensions even though the entity is combined with multiple attributes in entity recommendation. *In this paper, on the basis of the classification method, we delve into physical properties of the recommended entities, divide entity's attribute information network into multiple sub ones. In sub information network, bounded by the amount of attributes, the single attribute and even multiple attributes can be diverted into diverse paths of entity similarity, combining with entity similarity and related algorithm, where we can get the recommended results. This study not only refers to various attributes, but also improves the precision of recommendation.*

**Keywords:** entity recommendation; multiple attribution; classification; similarity path

## 引言

实体推荐是个性化推荐中普遍使用的推荐算法, 大量应用在用户、物品以及相关属性等多实

体的数据集中<sup>[1]</sup>。在实体推荐过程中, 通常将数据集划分为用户、物品以及属性<sup>[2]</sup>。其中, 根据属性是否为数据集直接给出, 可将其分为显性属性以及隐性属性。最为显著的显性属性关系是用户对物品的评分, 大多数的研究都是基于此进行的。然而, 用户对物品的评分数据既冗杂又稀疏, 考虑到推荐结果所需的准确性以及计算机的承载能力<sup>[3]</sup>, 推荐系统引入额外的隐性属性对相



收稿日期: 2016-01-07 修回日期: 2016-05-05;  
基金项目: 国家科技支撑计划(2014BAK15B01);  
作者简介: 宋美娜(1974-), 女, 山东, 博士, 教授, 研究方向为服务科学与服务工程、大数据; 赵雪君(1994-), 女, 重庆, 硕士, 研究方向为个性化推荐、大数据。

<http://www.china-simulation.com>

关评分算法进行改进，以此实现实体推荐。

从推荐这一最终目的出发，在为各实体进行合理推荐的过程中，首先从用户对物品的评分入手，通过矩阵分解(MF)等方法<sup>[4]</sup>可以较好的获取相关用户对物品的非数值形式或者空缺的评分，然而基于该评分的推荐结果过度依赖于主观评分，从而忽略了用户或物品等实体的自身属性。因此，将实体属性引入推荐系统中，扩展推荐的背景信息，有利于提高推荐结果的精确度和多样性。

目前，大量研究<sup>[5]</sup>对用户的隐性属性——社交关系进行了深度探讨：通常引入相似度来刻画社交关系，并以此联合用户评分数据，对用户进行全面推荐。然而，由于在推荐系统中实体物品以及相关属性的占有比率远超于用户及其相关性<sup>[6]</sup>，涉及物品(及其属性)的关系网络相较社交关系网络更具有研究意义，所以本文将会着重探索物品(及其属性)的关系网络(简称物品属性网络)。

针对物品属性网络的研究，通常会引入其一种属性，对相似度途径进行模拟、度量<sup>[7-9]</sup>；然而，实际物品属性网络包含多样属性，出于仿真准确性的考虑，处理多样实体属性已成为发展的必要趋势。然而，如何高效统一的处理多样属性具有一定的难度。因此，本文针对多样属性的不统一特性<sup>[10]</sup>，对多样属性按照分类的方法，划分为子属性网络，完成预处理；然后，按照属性数目划分子属性网络，用单一属性以及多样属性来构造相似度度量，完成相似度途径的研究探讨；最后结合实体相似度以及相关的算法<sup>[11]</sup>，将本文的多样相似度途径带入，得到最终推荐结果。

通过上述工作，本文的主要研究成果如下：

1、在实体推荐系统中，对实体以及相关属性进行分析，针对多样属性不统一的特点，按照分类的方法，进行属性预处理。

2、在子属性网络中，基于途径等价的理念，依据属性数目等特点，构造多样相似度途径。

3、结合实体相似度以及相关的算法，在本文实验部分带入上文的多样相似度途径，与先前算

法结果进行对比，多属性相似度途径的实验效果较优。

因此，在面对物品的多样属性时，通过将物品属性网络划分为子属性网络，在子属性网络中构造多样相似度途径，对推荐算法进行探究。

## 1 引用方法

### 1.1 相似度途径

相似度刻画实体间的关系，而其相关的计算过程是通过相似度途径呈现的，见表1。相似度<sup>[12]</sup>数值的大小代表着实体关系的好坏，实体关系的好坏可通过相似度途径的复杂程度来进行反映。

表1 相似度途径  
Tab. 1 Path of similarity

途径	语义
$P-R$	$p_i \in P$ 拥有属性 $r_i \in R$
$R-R$	属性 $r_i, r_j \in R$ 间的关联
$P-R-P$	$\{p_i, p_j\}$ 共有属性 $r_i \in R$
$P-R-RR-R-P$	$\{p_i, p_j\}$ 的共有属性 $RR$ 可以为两个不同属性 $R$ 所习成

相似度途径由实体与属性、属性与属性间的关系构成。譬如，实体拥有属性，可以表述为实体—属性关系；两种属性间有关联，可以表述为属性—属性关系；更为复杂的途径可以由实体—属性关系以及属性—属性关系进行共同表述<sup>[8]</sup>。

不妨在电影属性网络中，考虑上述途径的具体语义，其中  $p_1, p_2 \in P$  分别代表不同的电影， $r_1, r_2 \in R$  分别代表电影类型以及关键字。如途径  $P-R, movie-genre$  表明电影的电影类型属性  $r_i \in R$ 。

如途径  $R-R, genre-keyword$  表明电影类型属性  $r_i \in R$  与关键字属性  $r_2 \in R$  之间的关系。

单属性途径  $P-R-P$ ， $r_i \in R$  是相同类型实体  $P$  下的实体  $\{p_1, p_2\}$  间的链接。例如  $movie-genre-movie$ ，表示电影  $\{p_1, p_2\}$  具有相同的电影类型  $r_1$ ，据此可以向对  $p_1$  这种电影类型感兴趣的推荐  $p_2$ 。

多属性途径  $P-R-RR-R-P$  中， $r_i \in R$  是  $p_i \in P$

和  $r_2 \in RR$  之间的联系, 相同类型实体  $P$  下的实体  $\{p_1, p_2\}$  间的属性  $r_1 \in R$  通过共有属性  $r_2 \in RR$  链接。同理 movie-genre-keyword-genre-movie, 表示电影  $\{p_1, p_2\}$  拥有不同的电影类型  $r_1$ , 但这两种电影类型具有相同的关键词  $r_2$ 。

通过实体以及属性间的相互关联关系, 我们能够在物品属性网络中生成相似度途径。同时, 在相似度途径的指导下, 实体相似度的计算可以较好的在各种算法中得以解决。

### 1.2 实体相似度

实体相似度的计算是实体推荐的关键步骤, 该相似度可以在用户对物品评分的偏差上进行一定程度的修正。

在文献[11]中, 实体相似度被首次提出, 并适用于异构网络, 因此实体相似度是进行实体推荐的重要工具, 也是提高推荐精度的重要方法。在实体途径可逆的前提下, 实体相似度定义在实体  $i$  和  $j$  上:

$$S_{ij}^{(l)} = \frac{2 \times |P_{i \sim j}^{(l)}|}{|P_{i \sim i}^{(l)}| + |P_{j \sim j}^{(l)}|} \quad (1)$$

方程(1)表明  $S_{ij}^{(l)}$  由两部分所决定:

(1) 连通性  $P_{i \sim j}^{(l)}$ : 实体  $i$  和  $j$  在实体途径  $l$  各维属性的向量乘积所定义;

(2) 能见性  $P_{i \sim i}^{(l)}$  (或者  $P_{j \sim j}^{(l)}$ ): 实体  $i$  (或者  $j$ ) 在实体途径  $l$  各维属性的自身向量相乘所定义。

### 1.3 meta 途径的实体推荐算法

根据社交网络划分方法<sup>[13]</sup>, 实体的参数设置对矩阵分解算法进行改进。本文通过对实体相似度的多重途径设置, 运用 meta 途径的实体推荐算法<sup>[7]</sup>, 实现了本文算法。

$$\min_{U, V, \Theta} \sum_{i, j} (Y \otimes (R_{ij} - U_i V_j))^2 + \lambda_0 (\|U\|^2 + \|V\|^2) + \lambda_1 \sum_{j, k} S_{jk} \|V_j - V_k\|^2 + \lambda_2 \|\Theta\|^2 \quad (2)$$

$$s.t. U \geq 0, V \geq 0, S_{jk} = \sum_{l=0}^L \theta_l S_{jk}^{(l)}, \sum_{l=0}^L \theta_l = 1$$

根据式(2), 算法按照式(3)、(4)求取各梯度, 随

后使用随机梯度算法一步步极小化目标函数(H)。

$$\frac{\partial H}{\partial U_i} = \sum_j Y_{ij} \times (R_{ij} - U_i V_j^T) V_j + \lambda_0 U_i \quad (3)$$

$$\frac{\partial H}{\partial V_j} = \sum_i Y_{ij} \times (R_{ij} - U_i V_j^T) U_i + \lambda_1 \sum_{j \neq k} S_{jk} (V_j - V_k) \quad (4)$$

根据公式(3)、(4)求取的梯度, 利用优化算法对最终算法进行求解。“其中,  $Y$  元素为评分矩阵的指示矩阵,  $Y_{ij}$  元素为指示矩阵的元素, 当  $R_{ij}$  不为 0 时,  $Y_{ij}$  为 1; 其余情况下,  $Y_{ij}$  为 0;  $R$  评分矩阵,  $S_{jk}^{(l)}$  在第  $l$  条途径上, 物品  $j, k$  的相似度,  $S_{jk}$  基于途径的相似度矩阵中物品  $j, k$  的相似度,  $U$  矩阵分解后的用户因子矩阵,  $V$  矩阵分解后的物品因子矩阵,  $\theta$  基于途径的相似度的权重,  $\Theta$  为  $\theta$  参数集,  $i, j$  物品因子矩阵的第  $i, j$  列,  $l$  基于途径的相似度对应的第  $l$  条途径,  $\lambda_0$  防止  $U, V$  过拟合,  $\lambda_0, \lambda_2$  正则化系数。

## 2 本文算法

基于物品的实体推荐将从两方面入手对推荐算法进行分析: 用户的个人兴趣评分以及物品属性网络<sup>[14]</sup>。在用户对物品属性的分析过程中, 通常情况下, 物品属性是作为对评分信息的补充在算法中得以体现。然而, 在数据爆发的时代, 物品属性的易获得以及客观性使其具有较高的研究价值。

随着大数据时代<sup>[15-17]</sup>的到来, 物品信息呈爆炸式增长, 对现实生活中的物品进行分类是非常必要的。目前最为显著的个性化推荐领域, 其涉及刻画用户画像, 研究者需要采集大数据量的用户生活中方方面面的信息(支付能力, 家庭状况、受教育情况以及爱好领域等), 虽然选择一种属性就可以较好的对用户进行评估, 但这无法构成长久有效的用户评估方案。

在物品属性网络中, 物品属性并不是以个体的形式出现, 其间存在着很多关联<sup>[18]</sup>。比如在电影属性网络中, 属性关键字以及属性类型在很大

程度是并列关系，属性演员和属性导演很大程度上来源于一个流派，属于合作关系。对类型科幻、关键字未来感兴趣的用户大多会对史蒂文·斯皮尔伯格以及其御用演员哈里森·福特感兴趣。此外，物品属性网络中属性大多不统一，比如上文所提到的属性关键字以及属性类型、属性演员和属性导演这两类属性在客观分析中会采用完全不同的评价标准，通常采用枚举法对前者进行分析，对于后者采用排序得分。考虑到属性间相关性以及不统一性，本文将物品属性网络按照关联性划分为子属性网络；再基于该子属性网络内部的信任关系，构造相关的相似度途径，以属性数目为界，对子网络中的多属性以及单属性进行分门别类的分析。该方法在数据量大的情形下，能更优的做到基于分类的多属性实体推荐。

本文针对物品属性网络的研究，首先依据属性间的关联，采取分类的方法对物品属性进行子属性网络划分；然后在子属性网络中，将属性与相似度途径进行转化：以属性数目为界，对多属性以及单属性在相似度途径构建；最后，结合实体相似度以及引用算法，对属性相似度以及推荐结果进行求取。

## 2.1 子属性网络划分

推荐系统所涉及的实体属性网络中，基于用户的属性网络内存在着信任关系子网络，因此可以通过用户间的信任关系，对用户群体进行划分<sup>[10]</sup>；在可靠的群体划分关系下，当面对全新的推荐对象时，采用群体中一人信任，全体信任；反之亦然。推荐的信任机制，会使得推荐结果更为准确高效。同理，在基于物品属性网络的推荐系统研究中，实验研究以物品属性间的信任关系为基准，将物品属性进行划分，构造出相关的子属性网络。考虑到仿真过程中，物品数量往往多于用户数量，如若以物品间的信任关系为参照，分类数目会过多，无法获取立竿见影的效果，同时，将会因过多的子属性网络，导致冗余过多，无法对

用户评分进行持续的分析处理。本文将物品属性网络将会依照物品的关联性以及统一性，划分为子属性网络。

本文针对物品属性网络的特点，基于属性相关性的准则，对物品属性进行分类，形成子属性网络。比如在电影属性网络中(包含关键字，导演，演员，国家，摄影师，制片人以及类型等)<sup>[8]</sup>，属性关键字以及属性类型在很大程度上是并列关系，属性演员和属性导演很大程度上来源于一个流派，属于合作关系。对类型科幻、关键字未来感兴趣的用户大多会对史蒂文·斯皮尔伯格以及其御用演员哈里森·福特感兴趣。此外，物品属性网络中属性存在不统一性，比如上文所提到的属性国家、属性摄影师和属性制片人这两类属性在客观分析中会采用完全不同的评价标准，通常采用枚举法对属性国家进行分析，而对于属性摄影师以及制片人属性则会采用排序得分的方法。基于此，本文的电影属性分为三大子属性网络：电影缩影(电影关键字，电影类型)，电影人(导演，演员，摄影师以及制片人)以及电影出品国家。根据图的概念，电影属性图中每一个内置子图都对应一个子属性网络。

通过本文划分后的物品属性网络，在研究过程中，可着眼于子属性网络，由小及大。物品属性网络中，属性间关系可以用矩阵  $P$  所表示，同理，划分后的子属性网络  $c$  中的属性关系可用  $P^c$  表示。比如，电影缩影这一子属性网络中包含类型属性和关键字属性，可直接通过求取电影在电影缩影上的关系，得到两电影间的关系；不同于未采用子属性网络时，需要对每一个属性(或者某一个属性)的关系进行分别求取。通过该子属性网络，直接赋值的物品属性  $u$  以及属性  $v$  间的关系  $P_{u \sim v}^c$ ，且满足以下几种关系：

$$(1) \text{ 对称性: } P_{u \sim v}^c = P_{v \sim u}^c$$

$$(2) \text{ 极大性: } P_{u \sim v}^c \in [0, 1], P_{u \sim u}^c = 1$$

通过对物品  $u$  以及  $v$  间在子属性网络上直接赋值得到的关系  $P_{u \sim v}^c$ ，从而形成子属性网络的关系

矩阵  $P^c$ , 对于推荐系统有着很好的推进作用。本文利用该关系矩阵结合实体相似度, 助力实体推荐。下文将会就关系矩阵转化为相似度途径进行进一步探究。

## 2.2 相似度途径

由于物品属性网络本身的复杂性, 以及研究关注点着重在其最终的相似度算法的选择上, 大多数算法通常随机选取相似度途径, 或者从数值上模拟出多种单一属性的最优化结果来担当相似度途径统一尺度, 并没有充分利用多种属性间的关系来对相似度途径的求取进行更好的指导; 同时也没有对单一属性下的途径相似度进行优劣比较。然而在实际生产中, 相似度的优劣比对会对实体推荐算法提供一定的指导意义。

通过上文子属性网络的构造, 将得到子属性网络中按照属性数目划分, 分为单属性子网络以及多属性子网络。然后在利用子属性网络以及其关系矩阵构建相似度途径的过程中, 使得相似度求取过程较为精准、便利、多样。因此, 本文对单/多属性构成的相似度途径(简称单/多属性途径)进行逐步探究。

### 2.2.1 单属性途径

在单属性子网络中, 实体属性通常以独立的形式引入相似度算法。如图 1 实体的多重属性图所示, 当实体 A 对实体 B 进行推荐的过程中, ①实体 A 拥有属性 A, ③实体 B 拥有属性 A, 由于属性 A 不与其他属性产生关联, ②实体 A、B 通过属性 A 所在的相似度计算之后, 可以进行实体推荐。

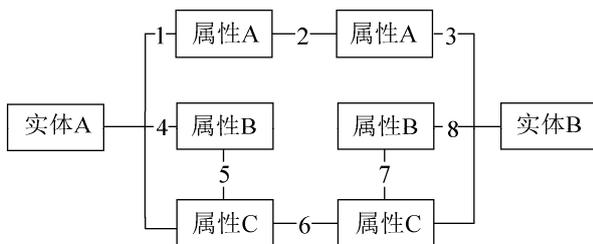


图 1 实体的多重属性图

Fig. 1 Multiple attribute graph of entity

在上述的实体推荐描述过程中, 由于属性 A 参与推荐, 可以实现实体—属性—实体这一单属性途径。以电影为例, 在国家的单属性子网络中, 当电影 A 在对电影 B 进行推荐的过程中, 使用了实体相似度的算法, 国家属性会完成实体相似度上的比对, 完成实体推荐。该单属性途径为电影—国家—电影。

事实上, 除了单属性子网络, 多属性子网络中也可采用单属性途径。在电影缩影这一多属性子网络中, 存在着电影—关键字—电影以及电影—类型—电影等单属性途径。单属性途径间的优劣比较会对后文的多属性途径, 乃至实体推荐算法提供一定的指导意义。

### 2.2.2 多属性途径

基于子网络划分的统一性及相关性原则, 多属性子网络中的属性间是相互关联的, 并非完全独立的<sup>[19]</sup>。比如图 1 实体的多重属性图所示, ④实体 A 拥有属性 B, ⑧实体 B 拥有属性 B, ⑤⑦表明属性 B 与属性 C 是一种承接关系(或其他相关关系), 在这种情形下, 不论实体 A、B 中是否包含属性 C, ⑥实体 A、B 通过属性 C 联合属性 B 所在的相似度计算之后, 可以进行实体推荐。以如下的图 1 为例, 当实体 A 在对实体 B 进行推荐的过程中, 所形成的相似度途径将会涉及 B、C 多种属性, 即为多属性途径。其中, 属性 B 为推荐链接属性, 以属性 C 作为推荐因子属性, 经过属性 C 的相似度计算之后, 可以进行实体推荐。

在上述的途径描述过程中, 由于属性 C 在属性 B 为链接的基础上参与推荐, 可以实现实体—属性 B—属性 C—属性 B—实体这一推荐途径。以电影缩影这一多属性网络为例, 当电影 1 在对电影 2 进行推荐的过程中, 在属性类型以及属性关键字这些多种属性上使用了相似度算法时, 属性类型以及属性关键字会作为相似度的途径进行推荐, 完成多属性途径电影—类型—关键字—类型—电影的推荐。

## 2.3 算法途径设计

在实际的相似度算法应用中，需要对上述的两种途径进行合理使用。考虑到单属性子网络中属性途径的不可选择性，本文着重讨论多属性子网络。如图2的多属性子网络中，当实体对属性A进行调用时，由于属性A的独立性，仅形成如图2右的实体—属性A—实体单属性途径；当实体对属性B或者属性C进行调用时，出于属性B与属性C的关联性，可形成如图2右实体—属性B—属性C—属性B—实体这一多属性途径。

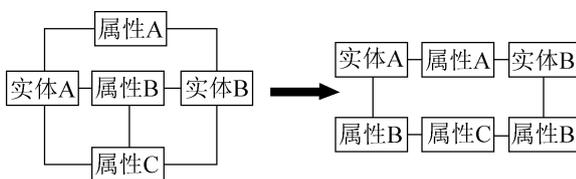


图2 多属性途径形成过程

Fig. 2 Multi attribute path formation process

通过2.1节中的单属性途径在多属性子网络中的扩展应用，多属性子网络中，不光可以引用多属性途径，还可将其多种属性拆分为多个单属性途径。本文针对多属性子网络的算法中，可对其单属性途径进行实验，将实验效果好的属性作为推荐因子属性，并引入实体相似度及引用算法，进行多属性途径对应的实验。

在多属性子网络中，不妨以实体的两种属性为例，首先对算法中的实体—属性1—实体以及实体—属性2—实体两种单属性途径进行优劣对比，在前者性能更佳的情况下，可以采用实体—属性1—属性2—属性1—实体进行多属性途径的实体推荐。

以电影为例，在电影缩影这一多属性子网络中，首先将电影—关键字—电影以及电影—类型—电影这两种单属性途径进行优劣对比，电影—关键字—电影这一途径的实验效果优于电影—类型—电影这一途径的情形下，本文以多属性途径电影—关键字—类型—关键字—电影为途径进行多属性途径实验；反之选用电影—类型—关键字—类型—电影。

本文的多属性途径简化了原先各属性的相似

度分开求解再进行优化的算法过程，然而，如何求取基于本文单属性途径以及多属性途径的实体相似度仍需讨论。

单属性途径：原先当实体仅针对一个属性时，将数据集中相应的数据调用出来，直接求取实体相似度，而目前由于单属性途径着重实体属性间的关联，因此对实属性网络  $P^C$  经过实体相似度求取  $S^C$  之后，进行如下处理， $S^C = S^C \times (S^C)^T$ 。

多属性途径：首先分析多样属性间的关系，根据上文的方法，确定推荐因子属性以及推荐链接属性；将因子属性网络  $P_1^C$  划分为各个属性向量，依据属性间的对应关系，将其与链接属性网络  $P_2^C$  中相关属性进行匹配，完成向量间的乘法运算。最后，经过上文的整合，得到实属性网络  $P^C$ ，经过实体相似度求取  $S^C$  之后，进行如下处理， $S^C = S^C \times (S^C)^T$ 。

## 3 实验

### 3.1 实验数据集

本文实验采用 Grouplens 研究中心的标准数据集 MovieLens-100k，评分数据集的范围为文献 [1,5]；数据包含用户对电影的评分，电影关键字以及电影类型等电影相关信息，满足本文相似度途径的设计要求。为了数据集的有效性，Grouplens 研究组的学者们将从 MovieLens 网站中获取到的数据模块进行了预处理，将数据集中评分数量少于 20 的用户进行了剔除。

我们使用了评分数据集 MovieLens-100k 这一数据集，本文在原数据集的基础上，对用户评分低于 25 的用户进行了删减，最终本文的评分数据包含用户 706 名，电影 100 023 部，类型 18 类，关键字 342 个。本文在推荐算法的精度以及推荐的效率方面，本算法表现通过预测精度 MAE(平均绝对偏差)进行评估<sup>[20]</sup>。其中，MAE 公式如下所示：

$$MAE = \frac{1}{T} \sum_{i,j} |\overline{R_{ij}} - R_{ij}| \quad (5)$$

式中:  $T$  为用户  $U$  在系统中评分物品的个数;  $R_{ij}$  是用户对物品的实际评分;  $\bar{R}_{ij}$  是推荐系统对物品的预测得分。

### 3.2 实验结果

本实验中, 采取测试集 40%, 60%, 80% 比例进行比对, 试验, 同时, 矩阵分解的维度  $d=5, 10, 15, 20$ 。本算法的算法结果除了本文的单属性途径以及多属性途径的对比外, 本文的算法也与经典概率矩阵分解算法(PMF)以及初始算法进行比对。为了方便实验分析, 本文将采用类型、关键字以及关键字-类型-关键字这几种缩写来代替电影-类型-电影、电影-关键字-电影以及电影-关键字-类型-关键字-电影这几种相似度途径对应的算法。表 2 中呈现了 40%、60% 以及 80% 测试集下的实验结果。

从表二的实验结果中, 可以看出关键字-类型-关键字实验效果上, 相较于经典概率矩阵分解算法, 精度总体提升了约 3% 左右; 相较于类型算法, 精度总体提升了约 2% 左右; 相较于关键字算法, 精度总体提升了约 1%。与初始算法相对比, 实验精度大部分提升了 1%。同时, 矩阵分解维度越高, 测试集比例越高, 实验精度越高。联合上述各表, 初始算法在测试集为 40%、60% 以及 80% 情况下进行对比, 由于 80% 实验效果最好, 此处给出 80% 测试集下的算法比较(图 3、4 中横坐标中 5、10、15 以及 20 代表矩阵分解的维度, 40%, 60%, 80% 为测试集的比例):

通过上述各途径的算法对比图, 显而易见, 整

体算法随着测试集比例的增大, 在矩阵分解维数一定的统一算法下, 基于单属性途径以及多属性途径的本文算法 MAE 值基本逐渐降低, 算法表现越来越好, 且都优于 PMF 算法; 通常在单属性途径下, 实验效果途径电影-关键字-电影优于原始途径优于电影-类型-电影, 虽然偶有单属性途径电影-关键字-电影在测试集为 60%, 80% 的情形下, 实验效果会优于电影-关键字-类型-关键字-电影这种多属性途径。接下来, 通过对图 4 单属性途径实验的分析, 给出选用电影-关键字-类型-关键字-电影这一多属性途径的理由。

总体而言, 电影-类型-电影以及电影-关键字-电影这两大单属性途径以及原始途径的算法表现尚可, 无论矩阵分解维数以及实验测试集比例的大小, MAE 的数值落在 [0.7, 0.9], 预测准确度较高。从单属性途径这一角度, 在两个单属性途径间, 电影-关键字-电影途径效果较之电影-类型-电影的效果更好, 很大程度的缘由是关键字数量远超类型的一个数量集; 相较而言, 原始途径的算法由于是单一途径相似度的数值优化结果, 其实实验效果较好, 但也未能有很大的突破。

在上述单属性途径相关的相似度算法, 可以看出所有算法结果都是有意义的, 而且在随着矩阵分解维度变化的趋势是几近一致的。同时, 前文的两种单属性途径对比, 可以得到电影-关键字-电影的效果优于电影-类型-电影这一单一属性途径, 因此在多属性途径中, 选用电影-关键字-类型-关键字-电影。

表 2 40%、60% 以及 80% 测试集的实验结果  
Tab. 2 Experimental results of 40%, 60%, and 80% test sets

	40% 测试集结果				60% 测试集结果				80% 测试集结果			
	$d=5$	$d=10$	$d=15$	$d=20$	$d=5$	$d=10$	$d=15$	$d=20$	$d=5$	$d=10$	$d=15$	$d=20$
PMF	0.914	0.897	0.876	0.872	0.889	0.865	0.847	0.836	0.841	0.829	0.799	0.779
原始	0.894	0.862	0.868	0.834	0.867	0.842	0.837	0.816	0.810	0.805	0.785	0.783
类型	0.915	0.887	0.864	0.847	0.879	0.856	0.841	0.824	0.823	0.813	0.807	0.795
关键字	0.882	0.873	0.865	0.852	0.857	0.849	0.839	0.823	0.815	0.798	0.774	0.769
关键字-类型- 关键字	0.873	0.861	0.846	0.843	0.857	0.856	0.837	0.812	0.808	0.796	0.781	0.775

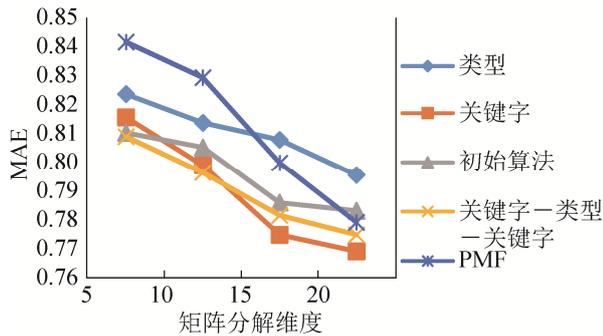


图3 80%测试集各途径下的算法比较

Fig. 3 Comparison of algorithms for 80% test sets

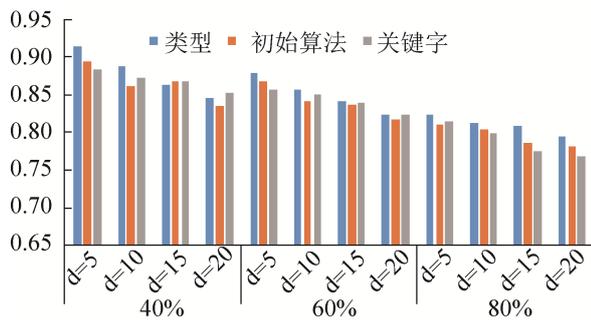


图4 单属性途径实验结果对比

Fig. 4 Comparison of experimental results of single attribute pathway

通过上文的分析,针对多属性网络的探究,首先针对单属性途径的探究,本文采用电影—类型—电影以及电影—关键字—电影两种途径来进行探究,得出优劣之分是很有必要的;可以挑选更为优秀的单属性途径中对应的属性作为多属性途径的首要元素,需要,比如在电影—关键字—电影优于电影—类型—电影的情形下,选取电影—关键字—类型—关键字—电影这一多属性途径。因此,推荐过程选用实体相似度乃至各种相似度时,可以先对简单的单属性途径进行优劣对比后,接着对复杂的多属性途径进行探索。

## 4 结论

本文出于实体推荐中的实体多样属性的相关性以及不统一性,对实体属性网络中的属性进行分类,生成单属性子网络以及多属性子网络。本文实验出于普适性的要求,在多属性子网络中,选用 movielens-100k 数据集。基于多样属性的复杂性,

首先对单属性途径的实体算法结果进行比较,再选用合适的多属性途径,同时,将这些多样途径的算法结果与初始算法进行比较。实验结果表明,相较于初始算法结果,本文针对子属性网络的多属性途径是有效的。然而,由于本文在针对多属性途径的探究过程中,仅使用了三维属性,在一些情形下,单维属性比其效果更甚。在未来的工作中,本文研究者会就数据更为丰富的数据集,属性更为复杂的属性网络,对其划分成的子属性网络间的关系进行更深的研究,以期符合大数据的要求,同时,期望实验推荐效果能够得到提升。

## 参考文献:

- [1] Shi C, Liu J, Zhuang F, et al. Integrating heterogeneous information via flexible regularization framework for recommendation[J]. Knowledge and Information Systems (S0219-1377), 2016, 49(3): 835-859.
- [2] Ma H. An experimental study on implicit social recommendation[C]//Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2013: 73-82.
- [3] 黄仁, 孟婷婷. 个性化推荐算法综述[J]. 中小企业管理与科技, 2015(8): 271.  
Huang Ren, Meng tingting. Summary of personalized recommendation algorithm[J]. Small and medium enterprise management and technology, 2015(8): 271.
- [4] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. Computer, 2009 (8): 30-37.
- [5] Gu Q, Zhou J, Ding C. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs[C]//Proceedings of the 2010 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2010: 199-210.
- [6] Davis D, Lichtenwalter R, Chawla N V. Multi-relational link prediction in heterogeneous information networks [C]//Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on. IEEE, 2011: 281-288.
- [7] Yu X, Ren X, Gu Q, et al. Collaborative filtering with entity similarity regularization in heterogeneous information networks[J]. IJCAI HINA(S1045-0823), 2013.
- [8] Yu X, Ma H, Hsu B J P, et al. On building entity

- recommender systems using user click log and freebase knowledge[C]//Proceedings of the 7th ACM international conference on Web search and data mining. ACM, 2014: 263-272.
- [9] Zheng V W, Zheng Y, Xie X, et al. Towards mobile intelligence: Learning from GPS history data for collaborative recommendation[J]. Artificial Intelligence (S0004-3702), 2012, 184: 17-37.
- [10] Yang X, Steck H, Liu Y. Circle-based recommendation in online social networks[C]// Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012: 1267-1275.
- [11] Sun Y, Han J, Yan X, et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks[J]. Proceedings of the VLDB Endowment (S2150-8097), 2011, 4(11): 992-1003.
- [12] 荣辉桂, 火生旭, 胡春华, 等. 基于用户相似度的协同过滤推荐算法[J]. 通信学报, 2014(2): 16-24. DOI:10.3969/j.issn.1000-436x.2014.02.003.
- Rong Huigui, Huo Shengxu, Hu Chunhua, et al. Collaborative filtering recommendation algorithm based on user similarity[J]. Journal of Communication, 2014(2): 16-24. DOI:10.3969/j.issn.1000-436x. 2014. 02. 003.
- [13] Ji M, Han J, Danilevsky M. Ranking-based classification of heterogeneous information networks[C]// Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011: 1298-1306.
- [14] Qian X, Feng H, Zhao G, et al. Personalized recommendation combining user interest and social circle [J]. Knowledge and Data Engineering, IEEE Transactions on (S1041-4347), 2014, 26(7): 1763-1777.
- [15] 宫夏屹, 李伯虎. 大数据平台技术综述[J]. 系统仿真学报, 2014, 26(3): 489-496.
- Gong Xiayi, Li Bohu. Technical review of big data platform technology[J]. Journal of System Simulation 2014, 26 (3): 489-496.
- [16] 张慷. 手机用户画像在大数据平台的实现方案[J]. 信息通信, 2014(2): 266. DOI:10.3969/j.issn. 1673- 1131. 2014.02.185.
- Zhang Kang. Mobile phone user portrait in the scheme of big data platform[J]. Journal of Information Communication, 2014, (2): 266. DOI: 10.3969/j.issn. 1673-1131. 2014. 02. 185.
- [17] 陶雪娇, 胡晓峰, 刘洋, 等. 大数据研究综述[C]//第八届全国仿真器学术年会论文集. 2013: 142-146.
- Tao Xuejiao, Hu Xiaofeng, Liu Yang, et al. A summary of large data research on [C]// Eighth National emulator academic annual conference. 2013: 142-146.
- [18] 胡堰, 彭启民, 胡晓惠, 等. 一种基于隐语义概率模型的个性化 Web 服务推荐方法[J]. 计算机研究与发展, 2014, 51(8): 1781-1793. DOI:10.7544/issn1000-1239. 2014. 20130024.
- Hu Yan, Peng Qimin, Hu Xiaohui, et al. A personalized recommended method of Web service based on the latent semantic probabilistic model[J]. Computer Research and Development, 2014, 51 (8): 1781-1793. DOI: 10.7544/ issn1000-1239. 2014. 20130024.
- [19] Sun Y, Han J, Aggarwal C C, et al. When will it happen? relationship prediction in heterogeneous information networks[C]// Proceedings of the fifth ACM international conference on Web search and data mining. ACM, 2012: 663-672.
- [20] 刘建国, 周涛, 郭强, 等. 个性化推荐系统评价方法综述[J]. 复杂系统与复杂性科学, 2009, 6(3): 1-10. DOI: 10.3969/J.Issn.1672-3813. 2009.03.001.
- Liu Jianguo, Zhou Tao, Guo Qiang, et al. Review of evaluation methods for personalized recommendation system [J]. complex system and complexity science, 2009, 6(3): 1-10. DOI: 10.3969/J.Issn. 1672-3813.2009.03.001.

(上接第 404 页)

- [5] 王三民, 王宝树. 贝叶斯网络在战术态势评估中的应用[J]. 系统工程与电子技术, 2004(11): 84-87, 143.
- Wang Sanmin, Wang Baoshu. Application of Bayesian networks in tactical Situation assessment[J]. Systems Engineering and Electronics, 2004(11): 84-87, 143.
- [6] 史志富, 张安. 贝叶斯网络理论及其在军事系统中的应用[M]. 北京: 国防工业出版社, 2012: 32-68.
- Shi Zhifu, Zhang An. Bayesian Network Theory and Its Application in Military System[M]. Beijing: National Defend Industry Press, 2012: 32-68.
- [7] 赵晓辉, 姚佩阳, 张鹏. 动态贝叶斯网络在战场态势估计中的应用[J]. 光电与控制, 2010, 17(1): 44-47.
- Zhao Xiaohui, Yao Peiyang, Zhang Peng. Application of Dynamic Bayesian Network in Battlefield Situation Assessment[J]. Electronics Optics & Control, 2010, 17(1): 44-47.