

一种基于云模型的社交网络推荐系统评分 预测方法

肖云鹏,孙华超,戴天骥,李 茜,李 瞰
(重庆邮电大学网络与信息安全技术重庆市工程实验室,重庆 400065)

摘 要: 本文针对评分预测中用户评分主观性及评分数据稀疏带来的预测不准确问题,围绕社交推荐的特点,设计实现一种社交网络评分预测方法. 首先,针对评分主观性问题,引入并优化相关云模型理论,提出采用综合云模型生成评分标准并转化用户评分的方法. 其次,针对预测不准确问题,通过引入隶属度达到数据降维和目标用户定位的作用,同时考虑到社交网络用户关系对评分结果的影响,分别利用社交关系及相似群体建立两个评分预测模型,并基于高斯变换融合两部分预测结果生成预测评分. 实验表明,该方案不仅克服了用户评分主观性,同时有效改善了用户评分数据稀疏情况下传统预测方法准确度偏差的问题.

关键词: 社交网络; 推荐系统; 评分预测; 云模型

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2018)07-1762-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.07.032

A Rating Prediction Method Based on Cloud Model in Social Recommendation System

XIAO Yun-peng, SUN Hua-chao, DAI Tian-ji, LI Qian, LI Tun
(Chongqing Engineering Laboratory of Internet and Information Security,
Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Focusing on the issues of rating prediction such as subjectivity of user rating and inaccurate prediction caused by rating sparsity, a rating prediction method is proposed by introducing the characteristics of social recommendation. Firstly, aiming at the subjective of user rating, we introduce and optimize the cloud model theory. Then, a method to generate rating standard by synthetical cloud model and transform user rating under the standard is proposed. Secondly, to deal the problem of inaccurate score prediction caused by data sparseness, data dimension reduction and target user location are achieved by introducing membership degree. And taking into account that user rating can be affected by their social relationship, we try to learn two rating prediction models by respectively using social relationships and similar groups. Finally, the rating value is obtained by using Gauss transform to combine the two prediction models. Experimental results show that our method not only overcomes subjectivity of user rating, but also alleviates the poor accuracy caused by rating sparsity problem in traditional rating prediction methods.

Key words: social network; recommendation system; rating prediction; cloud model

1 引言

互联网时代推荐系统可以有效帮助用户进行信息过滤,同时满足用户对信息的个性化需求. 推荐技术最初是从协同过滤发展而来的,协同过滤主要处理的是评分预测问题^[1,2]. 目前,协同过滤的发展已从单纯利用用户和物品信息预测评分到加入辅助信息来提高模型的预测精度^[3]. 现有的评分预测方法及其改

进型尽管在预测精度上取得了较好的效果,但仍然存在着评分主观性以及评分数据稀疏导致的预测不准确问题^[4-6]. 为了消除评分主观性的影响,研究者们对此进行了研究. 文献[7]在隐语义模型的优化函数中加入用户偏置参数和物品偏置参数. 文献[8]针对同行评审系统中,评分主观性造成的不公平最终决定,提出了一种被称作 CONCERT 的上下文感知审稿人交叉分配方法. 上述文献虽对评分做出了调整,但

收稿日期:2017-08-07;修回日期:2018-02-10;责任编辑:李勇锋

基金项目: 国家973重点基础研究发展计划(No. 2013CB329606);国家自然科学基金(No. 61772098);重庆市重点研发项目(No. cstc2017zdcy-zdyf0299, No. cstc2017zdcy-zdyf0436);重庆市基础科学与前沿技术研究项目(No. cstc2017jcyjAX0099)

都没从原始评分数据入手,因此所做出的调整有限.进一步,针对评分数据稀疏造成的预测不准确问题,研究者们提出了矩阵分解、机器学习等与其他模型融合的方法,有效的改善了评分数据稀疏带来的问题^[9].其中,云模^[10]型由于其可以定性的表示一个概念的特性,以及具有定量数值与定性概念相互转换的能力,而被应用到推荐系统中.文献[11]引入云模型概念,在整体层面上考虑用户相似度的计算,克服了传统基于向量的相似度比较方法严格匹配对象属性的不足.在此基础上,文献[12]对其相似度计算进一步改进,通过比较正态云模型的期望曲线实现相似度的定量表示.以上两种方法在一定程度上缓解了评分稀疏带来的问题,但其从整体层面上考虑用户相似度,损失了一定的评分信息.

2 相关知识

2.1 云模型

定义 1 正向正态云发生器(FCG),是从定性到定量的映射,它可以将定性的云特征值($Ex_1 \cdots Ex_i | En_1 \cdots En_i | He_1 \cdots He_i$) ($i = 1, 2, \dots, n$) 转化为定量的云滴分布,云滴用序对 (x_i, u_i) 表示.

定义 2 逆向正态云发生器(RCG),是从定量到定

性的转换,它可以将一定数量的多维数字集合 $X = \{x_j | x_j = (x_{j1}, x_{j2}, \dots, x_{ji})\}$ 转换为以数字特征($Ex_1 \cdots Ex_i | En_1 \cdots En_i | He_1 \cdots He_i$)表示的概念,其中 $i = 1, 2, \dots, n$.

2.2 综合云模型

定义 3 综合云模型(SC),一种可用于将两个同一层次的概念合并为一个较高层次的综合概念的模型,给定两个一维云 $C_1(Ex_1, En_1, He_1)$ 和 $C_2(Ex_2, En_2, He_2)$,其通过 $SC(C_1, C_2)$ 合并后生成较高层次概念的父云为 $C(Ex, En, He)$.

3 基于云模型的评分预测方法

3.1 方法概述

针对评分标准不统一及预测不准确问题,首先,利用云模型定量数值到定性概念转换的特点,设计一种评分转换方法,使每个用户评分在同一个标准下.其次,在隶属度相关理论的基础上,考虑到社交因素的影响,分别利用好友群体及相似群体构建两个预测模型,并对其使用高斯变换进行融合构成综合评分预测方法,最后通过对评分转换方法逆向还原得到最终预测评分,其中相似群体通过 KMeans 算法聚类获取.整体系统框图如图 1 所示.

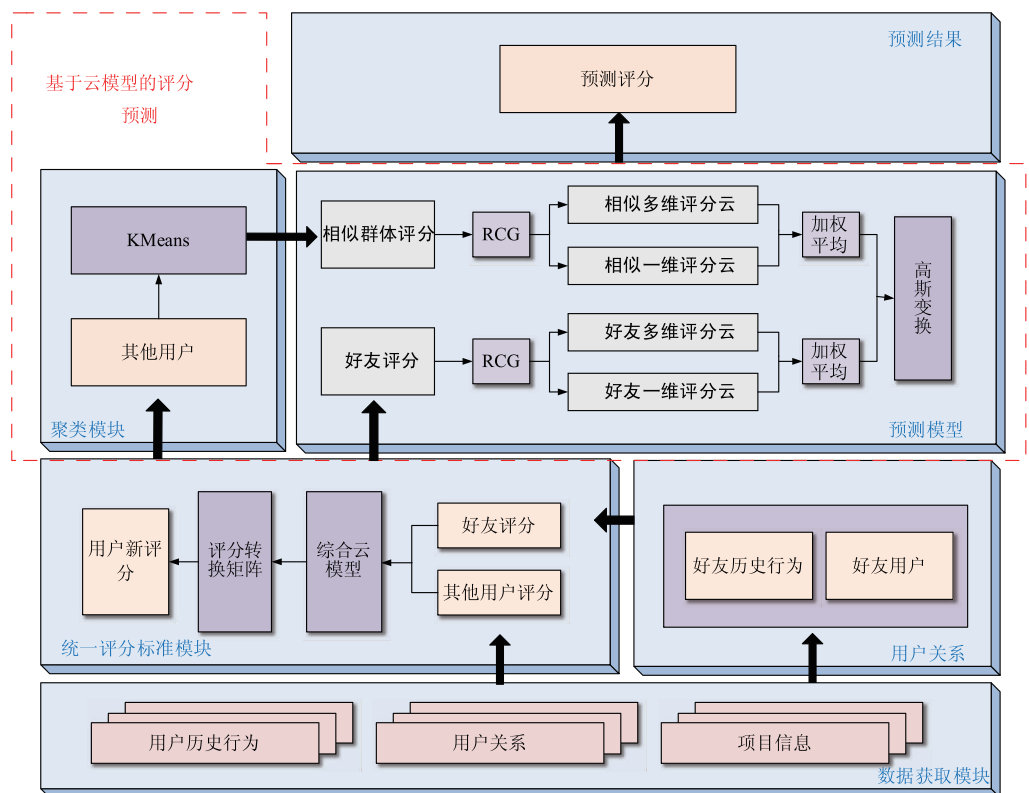


图1 整体系统框图

3.2 关键问题

3.2.1 评分转换方法

本文首先通过逆向云发生器对每个用户生成相应的评分云,表示其个人的评分尺度,其次,利用综合云模型合并用户评分云生成总体评分标准,最后转换用户评分.

定义 4 云评分转换矩阵(CRCM),是一种用户评分在整体标准云下转换评分的过程,假设标准云的数字特征为 (Ex, En, He) ,若存在一组评分向量 $\mathbf{X} = (x_1, x_2, \dots, x_i)$,评分云为 $C_x(Ex_x, En_x, He_x)$, x_i 在云 C_x 中的隶属度为 u_i ,则在标准云中隶属度为 u_i 的随机变量 x'_i 为新的评分,转换公式如下:

$$x'_i = \begin{cases} Ex + \sqrt{-2 * \ln u_i} En', & x_i > Ex_x \\ Ex - \sqrt{-2 * \ln u_i} En', & x_i < Ex_x \end{cases} \quad (1)$$

其中 $En' = NORM(En, He^2)$.

3.2.2 综合评分预测方法

本文通过引入隶属度达到数据降维和目标用户定位的作用,评分预测的方法源于一个假设,即若目标用户对其相似群体存在一个隶属度,则短期内,目标用户的评分在其相似用户的评分中有一个相同的隶属度.

定义 5 云评分预测模型(CRP),一种可以进行评分预测的云滴发生器,结构如图 2 所示.

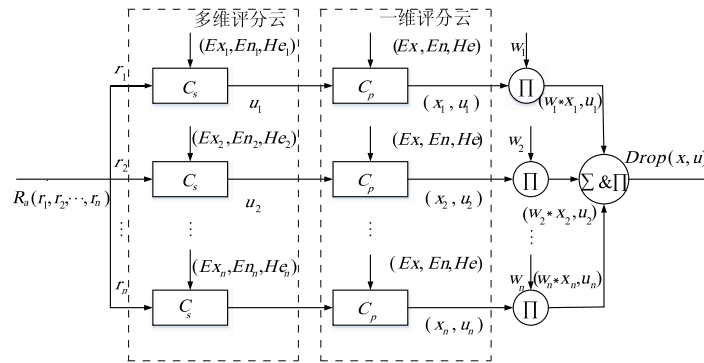


图2 云评分预测模型框图

令 $\mathbf{R}_s = \{R_1, R_2, \dots, R_n\}$ 、 $\mathbf{R}(r_1, r_2, \dots, r_k)$ 分别为相似群体对目标用户历史评分项目和待预测项目的评分数据,对其通过逆向云发生器生成多维评分云 $C_s(Ex_1 \dots Ex_n | En_1 \dots En_n | He_1 \dots He_n)$ 以及一维评分云 $C(Ex, En, He)$.

计算目标用户历史评分 $\mathbf{R}_a(r_1, r_2, \dots, r_n)$ 在多维评分云 C_s 中的隶属度,式中 $i = 1, 2, \dots, n$.

$$\mathbf{U} = \{u_i = e^{-\frac{(r_i - Ex_i)^2}{2 * En_i^2}}, En_i' = NORM(En_i, He_i^2)\} \quad (2)$$

计算一维评分云 C 在隶属度 u_i 下的云滴

$$\mathbf{X} = \left\{ x_i = \begin{cases} Ex + \sqrt{-2 * \ln u_i} En', & r_i < Ex_i \\ Ex - \sqrt{-2 * \ln u_i} En', & r_i > Ex_i \end{cases} \right\} \quad (3)$$

另 w_i 为目标用户的第 i 个历史评分项目与待预测项目的共同被评分数量,则预测云滴 (x, u) 产生方式如下:

$$Drop(x, u) = \left\{ (x, u) \mid x = \frac{\sum_{i=1}^n w_i * x_i}{\sum_{i=1}^n w_i}, u = \prod_{i=1}^n u_i \right\} \quad (4)$$

定义 6 高斯变换(GT),一种融合两个预测云滴 (x_s, u_s) 和 (x_f, u_f) 形成评分 r 的方法.

$$r = \frac{x_s \sqrt{-2 \ln u_f} + x_f \sqrt{-2 \ln u_s}}{\sqrt{-2 \ln u_f} + \sqrt{-2 \ln u_s}} \quad (5)$$

用好友群体替换定义 5 中的相似群体即可获得 (x_f, u_f) , (x_s, u_s) 则为相似群体下的云滴.

4 实验与分析

4.1 实验环境

实验使用的测试数据是通过使用豆瓣站点提供的 API 获取.数据集记录了 702 个用户及其关注好友对 4003 部电影的评分,共包含 30 万条数据,每个用户至少对 20 部电影进行了评分,每个电影至少被 20 个用户评价过,评分范围是 1 到 5.

为了证明本方法的有效性,我们与当前主流的评分预测方法奇异值分解(SVD++)进行横向比较,同时与文献[12]中基于云模型的协同过滤算法 ECM 进行纵向比较,本文方法则被称为 CM,最后对实验结果进行相应的分析,实验使用的评价指标是在推荐系统评分预测中常用的均方根误差(RMSE)和平均绝对误差(MAE).

4.2 统一评分标准对比分析

本节实验采用在三种不同方法下对比统一评分标准和未统一评分标准的均方根误差(RMSE)和平均绝对误差(MAE),对比结果如图 3~8 所示:

从图 3~8 中可以看出,三种评分预测方法在统一

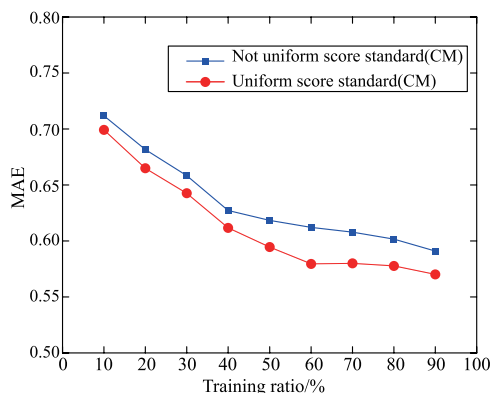


图3 CM方法的MAE对比

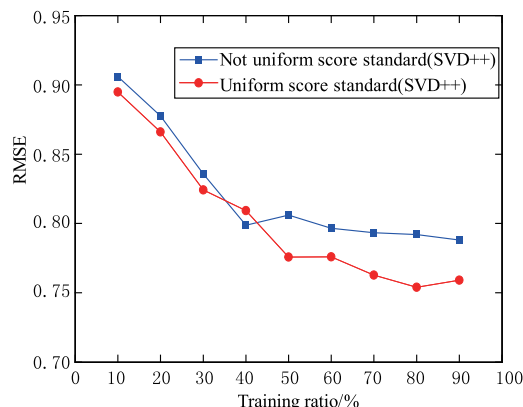


图6 SVD++方法的RMSE对比

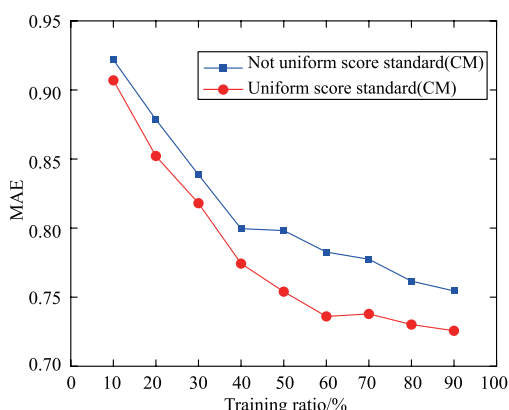


图4 CM方法的RMSE对比

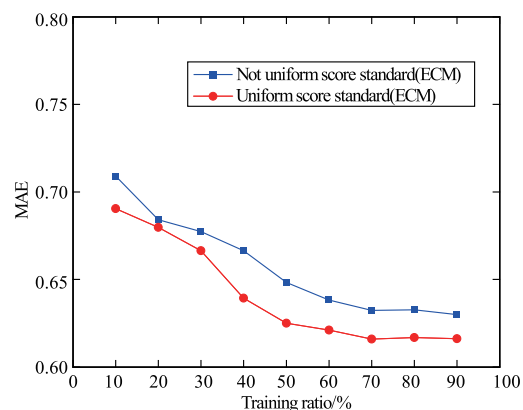


图7 ECM方法的MAE对比

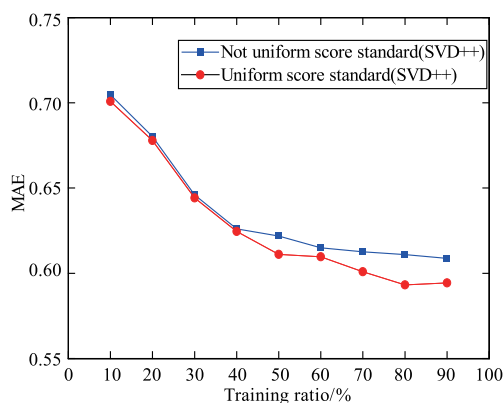


图5 SVD++方法的MAE对比

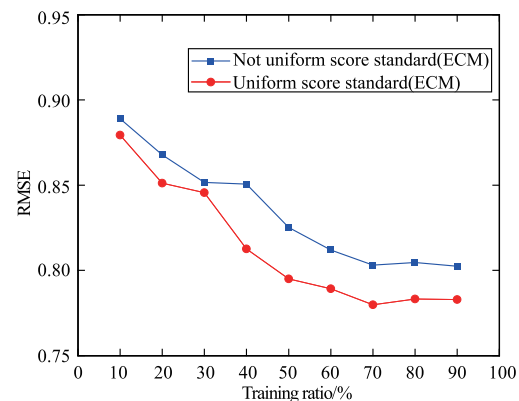


图8 ECM方法的RMSE对比

评分标准后的 RMSE 和 MAE 整体上均优于统一评分标准前的。图中横坐标为训练集比例,纵坐标为评价指标。对于不同的方法,统一评分标准对 RMSE 和 MAE 的影响也不同,统一评分标准在 UBCF 方法中取得了最优的效果,这是因为在统一评分标准下可以更为准确的发现与目标用户兴趣相似而非评分习惯相似的群体。对于 SVD ++ 算法提升不甚明显,这是因为算法本身就考虑了用户主观性的影响。由上述分析可以得出结论,统一评分标准可以有有效的提高评分预测的准确度,但提升的效果因方法的不同而不同。

4.3 不同稀疏度下的预测误差对比分析

本节实验通过改变训练集占评分数据的总数,观测三种方法在不同稀疏度下的 MAE 及 RMSE,实验结果如图 9 ~ 12 所示。

图 9 ~ 12 展示了三种方法在不同稀疏度下预测准确度的实验对比,从图 9 ~ 12 中可以看出,本文所提出的方法在 MAE、RMSE 上的表现整体上都要优于其他几种方法,其中,在和 ECM 算法的对比上,提高的程度较为明显,但相较于 SVD ++ 算法提高的效果相差不大,尤其当训练集小于 30% 时,这是因为在使用逆向云发生器在云

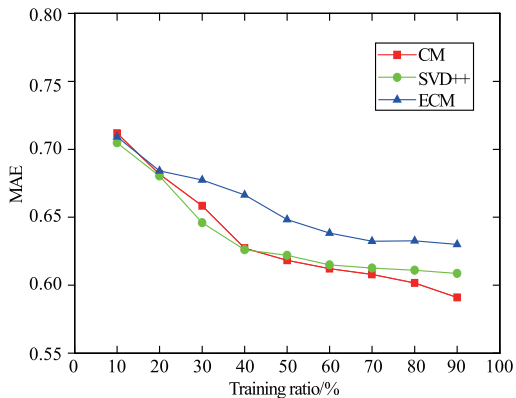


图9 未统一评分标准下三种方法的MAE对比图

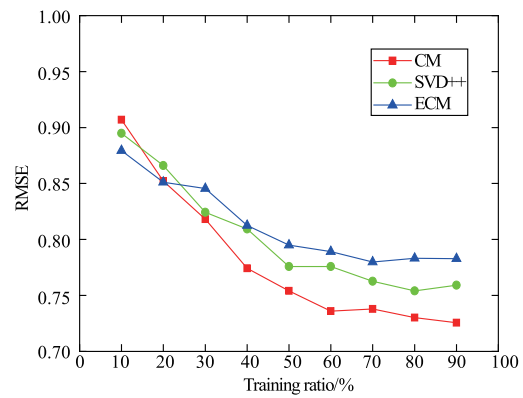


图12 统一评分标准下三种方法的RMSE对比图

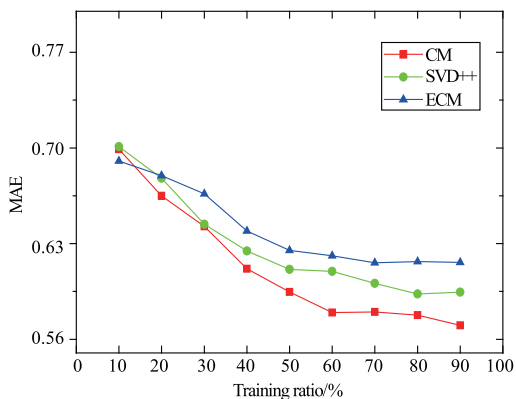


图10 统一评分标准下三种方法的MAE对比图

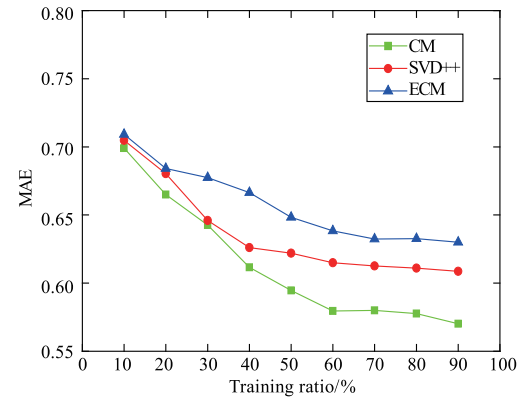


图13 三种方法MAE对比图

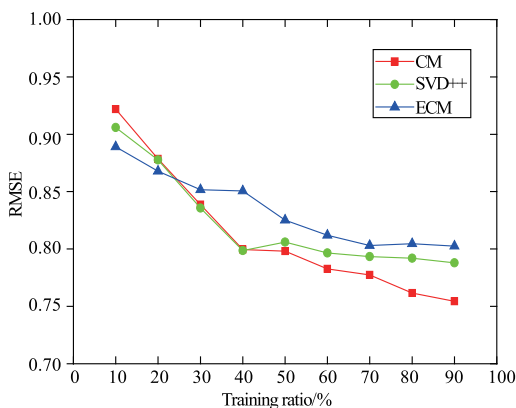


图11 未统一评分标准下三种方法的RMSE对比图

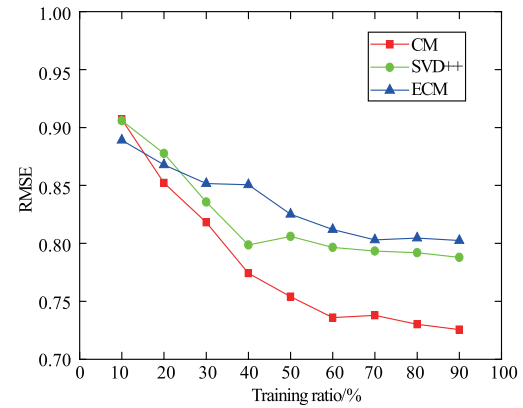


图14 三种方法RMSE对比图

滴量极端稀少时仍然会产生一定的误差,从而造成目标用户对评分云隶属度计算的误差. 总体来说,在缓解评分数据稀疏提高预测准确度方面,本方法整体上高于其他两种方法,在不同稀疏度下,提高幅度略有差别.

4.4 基于云模型的评分预测实验对比

本节实验将从整体上对比三种方法在不同稀疏度下的预测准确率,实验结果如图13~14所示.

从图13~14中可以看出,在不同稀疏度下,我们的方法均明显优于 UBCF、ECM 方法,相较于 SVD++ 方法,当训练集大于 40% 时,仍然有着一定的优势,说明

本文基于云模型的评分预测方法在正常情况下对评分预测效果有着明显的改善.

5 总结

本文就推荐系统评分预测中遇到的用户评分主观性及预测不准确问题进行研究分析,围绕社交推荐的特点,引入并优化相关云模型理论,提出一种基于云模型的社交网络评分预测方法. 我们对统一评分标准及评分稀疏方面进行实验对比表明,本文的评分预测算法在不同稀疏度下,均能明显提高评分预测的准确度,统一评分标准

在正常情况下也能显著的优化评分预测的效果.

参考文献

- [1] LIU H, HU Z, MIAN A, et al. A new user similarity model to improve the accuracy of collaborative filtering [J]. Knowledge-Based Systems, 2014, 56(3): 156 – 166.
- [2] HU L K, SUN A X, LIU Y. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction [A]. Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval [C]. New York: ACM, 2014. 345 – 354.
- [3] 郑麟, 朱福喜, 等. 基于属性提升与局部采样的推荐评分预测 [J]. 计算机学报, 2016, 39(8): 1501 – 1514. ZHENG Lin, ZHU Fu-xi, et al. Recommendation rating prediction based on attribute boosting with partial sampling [J]. Chinese Journal of Computers, 2016, 39(8): 1501 – 1514. (in Chinese)
- [4] PARK Y, PARK S, JUNG W, et al. Reversed CF: a fast collaborative filtering algorithm using a k-nearest neighbor graph [J]. Expert Systems with Applications, 2015, 42(8): 4022 – 4028.
- [5] LUO X, ZHOU M C, LI S, et al. A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method [J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(3): 579 – 592.
- [6] ALQADAH F, REDDY C K, HU J, et al. Biclustering neighborhood-based collaborative filtering method for top-n recommender systems [J]. Knowledge and Information Systems, 2015, 44(2): 475 – 491.
- [7] KOREN Y. Factor in the neighbors [J]. ACM Transactions on Knowledge Discovery from Data, 2010, 4(1): 1 – 24.
- [8] LI L, WANG Y, LIU G, et al. Context-aware reviewer assignment for trust enhanced peer review [J]. Plos One, 2015, 10(6): e0130493.
- [9] 王俊, 李石君, 杨莎, 等. 一种新的用于跨领域推荐的迁移学习模型 [J]. 计算机学报, 2017, 40(10): 2367 – 2380. WANG Jun, LI Shi-jun, YANG Sha, et al. A new transfer learning model for cross-domain recommendation [J]. Chinese Journal of Computers, 2017, 40(10): 2367 – 2380. (in Chinese)
- [10] 李德毅, 刘常昱, 等. 不确定性人工智能 [J]. 软件学报, 2004, 15(11): 1583 – 1594. LI De-yi, LIU Chang-yu, et al. Artificial intelligence with uncertainty [J]. Journal of Software, 2004, 15(11): 1583 – 1594. (in Chinese)
- [11] 张光卫, 李德毅, 李鹏, 等. 基于云模型的协同过滤推荐算法 [J]. 软件学报, 2007, 18(10): 2403 – 2411. ZHANG Guang-wei, LI De-yi, LI Peng, et al. A collaborative filtering recommendation algorithm based on cloud model [J]. Journal of Software, 2007, 18(10): 2403 – 2411. (in Chinese)
- [12] 李海林, 郭崇慧, 邱望仁. 正态云模型相似度计算方法 [J]. 电子学报, 2011, 39(11): 2561 – 2567. LI Hai-lin, GUO Chong-hui, QIU Wang-ren. Similarity measurement between normal cloud models [J]. Acta Electronica Sinica, 2011, 39(11): 2561 – 2567. (in Chinese)

作者简介



肖云鹏 男. 1979 年 8 月出生, 安徽蚌埠人. 副教授、硕士生导师, 主要研究方向为社交网络、机器学习.
E-mail: xiaoyunpeng@cqupt.edu.cn



孙华超 男. 1989 年 3 月出生, 江苏徐州人. 现为重庆邮电大学硕士研究生, 主要研究方向为推荐系统.