

融合云环境用户情境兴趣的移动 SNS 信任推荐模型

宓 翠 陈 晶 苏妍嫒 张亚明 刘海鸥

(燕山大学 经济管理学院, 互联网+与产业发展研究中心, 河北 秦皇岛 066004)

E-mail: haiou2009@163.com

摘 要: 传统协同过滤(Collaborative Filtering, 简称 CF) 推荐面临的数据稀疏性问题是当前移动社交网络服务(Social Networking Services, 简称 SNS) 个性化推荐亟待解决的问题。在回顾相关文献的基础上, 提出融合云环境用户情境兴趣的移动 SNS 信任推荐模型。融合移动 SNS 情境相似度矩阵与信任度矩阵进行 CF 推荐, 基于此缓解数据稀疏性导致的推荐性能下降问题, 同时采用 MapReduce 并行处理方式提高大规模复杂社交网络的并行挖掘性能。实验结果表明, 融合云环境用户情境兴趣的移动 SNS 信任推荐模型较好地情境相似度与信任度融入到 CF 推荐过程中, 缓解了由于数据稀疏与信任稀疏导致的推荐精度下降问题, MapReduce 化的并行处理方法也有效提升了大规模复杂社交网络的并行推荐性能。

关键词: 云环境; 情境兴趣; 移动 SNS; 信任推荐模型

中图分类号: TP311

文献标识码: A

文章编号: 1000-1220(2018)03-0484-06

Trust Recommender Model for Mobile Social Network Integrating Cloud Environment User's Situational Interest

MI Cui, CHEN Jing, SU Yan-yuan, ZHANG Ya-ming, LIU Hai-ou

(Internet + and Industrial Development Research Center, Yanshan University, Qinhuangdao 066004, China)

Abstract: One of the key problems facing mobile SNS personalized recommendation is the data sparse and cold start problem. By reviewing the research progress, we proposed a trust recommender model for mobile SNS based on contextual interest in cloud environment. Based on combining user contextual interest similarity and trust similarity together, the model could effectively alleviate the sparse data problems. Finally, we used MapReduce to improve the large-scale complex social network parallel performance. The experimental results showed that the method smoothly introduced the interest similarity and social trust networks into traditional CF and solved the recommendation accuracy due to sparse data problems effectively. Besides, the parallel method effectively solved the recommend problem of large-scale complex social network and improved the real-time recommendation performance.

Key words: cloud environment; contextual interests; mobile social networking services; trust recommender model

1 引言

在当前的“数字爆炸时代”, 每 18 个月产生的数字信息相当于过去几千年数据量的总和。世界最大的在线影片租赁服务商 Netflix 一直致力于不断改进和完善其个性化推荐服务, 分别于 2006、2009 年推出两个“百万美元计划”以提高其推荐引擎的性能。当前 Netflix 又推出了基于海量数据进行挖掘的个性化推荐系统, 主要用于解决海量数据环境下个性化推荐所面临的复杂信息处理与即时个性化服务问题。然而, 海量数据环境下移动社会化推荐对位置的敏感性、紧急性、便携性决定了移动 SNS 推荐系统须满足一些特殊要求: 对用户兴趣变化反应快、能处理短期兴趣、能处理变化很大的项目属性、社交属性、情境敏感等, 使传统 CF 推荐技术不能直接平移到移动 SNS 中, 由此也导致移动推荐的情境信息过载、数据稀疏、推荐系统脆弱性与可扩展性、推荐结果多样性与精确性两难困境等问题远远没有得到解决, 难以满足云计算服务时代大规模、高

质量和实时性的移动 SNS 用户个性化服务需求。

因此, 本文在研究中引入云环境用户情境兴趣信息参量, 基于情境兴趣与信任关系双重视角, 提出融合云环境用户情境兴趣的移动 SNS 信任推荐模型, 深入探讨由于海量数据信息环境导致的数据稀疏与推荐精度下降的问题, 显著改善传统移动推荐的实时性和服务质量, 提高大规模社交网络个性化挖掘的深度、规模性和质量, 从而为用户提供更为精准的个性化服务, 为移动社交网络提供商和广大用户创造价值, 因此具有重要的理论意义与实际应用价值。

2 相关工作

云计算与大数据时代的来临以及数据密集型科学研究范式的兴起使“数据稀疏性”问题备受关注^[1]。由于融合了社会化网络和个性化用户偏好信息, 社会化信任推荐有效提升了传统推荐的精确度、冷启动、稀疏性与稳定性等^[2]。如国外学者 Kant Vibhor(2013)^[3]在研究推荐系统的数据稀疏性问题

收稿日期: 2017-08-22 收修改稿日期: 2017-10-30 基金项目: 国家自然科学基金项目(71271186, 71671159) 资助; 教育部人文社科基金项目(17YJCZH109) 资助; 河北省自然科学基金项目(G2017203319, G2016203220, G2015203380) 资助。作者简介: 宓 翠, 女, 1979 年生, 博士研究生, 副教授, 研究方向为电子商务; 陈 晶, 女, 1994 年生, 硕士研究生, 研究方向为个性化推荐与数据挖掘; 苏妍嫒, 女, 1991 年生, 博士研究生, 研究方向为社交网络; 张亚明, 男, 1962 年生, 博士, 教授, 博士生导师, 研究方向为社交网络; 刘海鸥(通信作者), 男, 1981 年生, 博士, 副教授, 研究方向为个性化推荐。

时, 将用户间的信任关系引入到协同过滤推荐模型的构建中, 以此来提高推荐的精度; 学者 Sohn Jongsoo (2015) [4] 对原先基于内容的 CF 推荐模型进行优化, 有效缓解了推荐系统数据库的冷启动问题; 我国学者孟祥武 (2014, 2015) [5-6] 对社会网络推荐研究进行了系统述评, 提出了一种融合项目特征和移动用户信任关系、基于位置的社交网络推荐算法; 张富国 (2014) [7] 基于社交网络视角出发, 对近几年关于社交网络信任方面研究个性化推荐的相关文献进行了比较分析, 最后对研究面临难点问题今后的发展趋势进行了展望; 刘英南 (2015) [8] 在研究过程中构建了动态信任传递模型, 以此解决协同过滤稀疏性引起的“冷启动”、“新用户”、“新资源”等问题; 为了有效解决用户信任关系难以获取的问题, 于美琪 (2017) [9] 提出了基于项目评分与用户信任关系的 CF 推荐算法, 通过评分用户间的相互信任关系来挖掘用户的社会关系与偏好关系, 以此来推荐 CF 推荐。

通过对上述文献的分析发现, 当前已有部分学者基于社交网络信任来缓解数据稀疏性导致的推荐精度下降问题, 也取得了一定的成果。但移动社交网络环境下, 如何综合挖掘移动 SNS 用户在位置、时间、业务需求要素, 基于不同资源对象和情境兴趣来进行社会化推荐的研究还较为稀少。综合考虑用户历史兴趣、情境信息、信任关系缓解移动 SNS 推荐数据稀疏性问题的研究更是鲜有涉及。此外, 当前关于社交网络分析方面的研究, 大多限于几十到几百个节点, 即其所处理的网络规模较小, 因此在大规模复杂社交网络, 传统处理方法的效果不佳。对于当前海量数据而言, 难以满足其大规模网络计算的需求。云计算处理技术实现了移动推荐系统与大规模海量数据源的集成, 充分利用其跨平台、并行处理、高扩展性等优势, 可以显著增强个性化推荐系统海量数据挖掘的实时性, 增强系统的可拓展性与可靠性。如张亚明 [10] 提出的面向云环境的并行推荐方法, 显著提高了推荐算法在海量数据规模下的执行效率。鉴于此, 本文提出基于云环境用户情境兴趣与信任关系的移动 SNS 信任推荐模型, 通过融合情境兴趣相似度与用户间的信任关系, 缓解协同过滤算法的数据稀疏性, 同时采用 MapReduce 技术解决大规模复杂社会网推荐并行挖掘的难题。

3 模型描述

本节对融合用户情境兴趣的移动 SNS 信任推荐模型进行细致描述。在本文构建的移动 SNS 推荐模型中, 假设存在移动 SNS 用户组 M 、参与评价的项目组 S 、移动 SNS 用户面临的外在情境信息 $Context$ 。在诸多的移动 SNS 用户间还存在相互的信任关系, 本文设为 TU ; 本文研究的目标用户设为 u_i , 且 $u_i \in U$; 对于 u_i 未评分的项目集合, 本文设为 S_j ; 在研究过程中, 通过 CF 推荐方法预测 u_i 对 S_j 的评分, 然后按照 Top-N 规则, 将排名靠前的项目推荐给 u_i 。总体来讲, 本文模型具体描述如下。

3.1 用户—项目评分矩阵

首先, 构建推荐模型的移动 SNS 用户——项目评分矩阵。设推荐系统中移动 SNS 用户的数量为 M (其中 $M = \|U\|$)。则移动 SNS 用户的集合记为 $U = \{u_1, u_2, \dots, u_M\}$; 移动用户面向的项目集合采用 $S = \{s_1, s_2, \dots, s_N\}$ 表示, 其中 N 为项目的数量, 即 $N = \|S\|$ 。用户在使用完系统中存在的诸多项目之后, 将结合自身的切实经历与用户体验对其进行

评分。我们采用公式 (1) 来表示移动 SNS 用户——项目评分矩阵, 本文记为 $RS_{M \times N}$ 。

$$RS_{M \times N} = \begin{pmatrix} RS_{u_1 s_1} & RS_{u_1 s_2} & \cdots & RS_{u_1 s_N} \\ RS_{u_2 s_1} & RS_{u_2 s_2} & \cdots & RS_{u_2 s_N} \\ \cdots & \cdots & \cdots & \cdots \\ RS_{u_M s_1} & RS_{u_M s_2} & \cdots & RS_{u_M s_N} \end{pmatrix} \quad (1)$$

其中, 用户 u_i 对项目 S_i 的评分本节通过 $rs_{u_i s_i}$ 来表示; 式 (1) 中矩阵 $RS_{M \times N}$ 的每一行都为用户 u_i 对评价项目 S 的具体评分, 即 u_i 的项目评分向量, 记为 $RS_{u_i} = (rs_{u_i s_1}, rs_{u_i s_2}, \dots, rs_{u_i s_N})$; 矩阵 $RS_{M \times N}$ 中的每一列是所有用户 u 对某一个项目 S_i 的评分, 记为 $RS_{s_j} = (rs_{u_1 s_j}, rs_{u_2 s_j}, \dots, rs_{u_M s_j})$ 。

3.2 移动 SNS 用户情境表示

鉴于本文的推荐模型中涉及到了用户的情境信息, 因此需精炼表述移动 SNS 用户所面临的具体情境。根据移动 SNS 用户所面临的位置、温度、时间等多种不同情境构成要素, 本文使用向量计算公式来表示 n 种不同类型的情境信息, 记为 $Context_y = (C_1, C_2, \dots, C_n)$ 。其中 C_k ($k = 1, 2, \dots, n$) 表示一种具体类型的情境, 则两个不同的情境可使用向量 $Context_x$ 和 $Context_y$ 表示。这两个情境间的相似度记为: $Sim(Context_x, Context_y)$ 。表示情境 $Context_x$ 和 $Context_y$ 在 k 类情境背景下比较的相似程度。鉴于移动 SNS 用户是在一定情境下对项目进行评分, 且该情境信息将对用户评分产生影响, 因此在评分过程中引入了移动 SNS 用户的情境信息, 由此来对传统 CF 的用户—项目评分矩阵进行扩充, 形成用户—项目—情境评分矩阵, 即将情境信息 $Context_k$ 融入到原先移动 SNS 用户的每一个项目评分 $rs_{u_i s_i}$ 中去, 实现融合情境兴趣的移动 SNS 信任推荐。

3.3 用户—信任矩阵

在推荐模型中引入移动 SNS 用户间的信任关系首先要建立信任网络、用户、信任矩阵。这里的信任网络是一个以移动 SNS 用户实体为节点、以各用户实体间存在的信任关系 (分为直接信任和间接信任) 为边而所构成的一个网络系统。这里提到的直接信任是授信用户通过直接的信息交互过程而获得了对受信用户的信任关系, 而间接信任则需要通过信任传递 (如与受信用户有直接信任关系的其他用户) 而获得的信任关系。在移动 SNS 有人网络中, 各个节点间的直接或间接信任关系通过他们的连线 (边) 来表示, 边的权重就代表节点间的信任度。公式 (2) 即为信任网络中的移动 SNS 用户—信任矩阵 $TU_{M \times N}$ 。

$$TU_{M \times N} = \begin{pmatrix} tu_{u_1 u_1} & tu_{u_1 u_2} & \cdots & tu_{u_1 u_N} \\ tu_{u_2 u_1} & tu_{u_2 u_2} & \cdots & tu_{u_2 u_N} \\ \cdots & \cdots & \cdots & \cdots \\ tu_{u_M u_1} & tu_{u_M u_2} & \cdots & tu_{u_M u_N} \end{pmatrix}_{M \times N} \quad (2)$$

其中 $tu_{u_i u_j}$ 表示用户 u_i 对 u_j 的信任度, $tu_{u_i u_j}$ 的取值范围为 $[0, 1]$ 区间的实数。当 $tu_{u_i u_j} = 1$ 时, 此时表明用户 u_i 对用户 u_j 是完全信任的; 而当 $tu_{u_i u_j}$ 取值为 0 时, 则表示另外一种极端情况, 即用户 u_i 对用户 u_j 是一种完全不信任的关系。

4 模型建立

4.1 构建情境相似度矩阵

本文融合情境兴趣的移动 SNS 信任推荐建模步骤如下:

Step 1. 首先获取移动 SNS 用户—项目评分矩阵及评分时用户所处的情境。在建立云环境用户情境兴趣的信任推荐模型时,先要建立移动 SNS 用户—项目评分矩阵 $RS_{M \times N}$,即用户 $User$ 对项目 $Item$ 的评分矩阵,其计算公式为:

$$RS_{M \times N} = \begin{pmatrix} rs_{u_1s_1} & rs_{u_1s_2} & \cdots & rs_{u_1s_N} \\ rs_{u_2s_1} & rs_{u_2s_2} & \cdots & rs_{u_2s_N} \\ \cdots & \cdots & \cdots & \cdots \\ rs_{u_Ms_1} & rs_{u_Ms_2} & \cdots & rs_{u_Ms_N} \end{pmatrix}_{M \times N} \quad (3)$$

其中,式(3)中的 $rs_{u_i s_j}$ 表示用户 u_i 对项目 s_j 的评分情况, $\overline{rs_{u_i}}$ 代表 u_i 的平均评分, $\overline{rs_{u_j}}$ 则代表 u_j 的平均评分情况。

Step 2. 计算目标用户与其他用户的相似度 $sim(u_i, u_j)$ 。本文在计算时采用 Pearson 相关系数度量公式,由此计算目标用户 u_i 和用户 u_j 的相似度,如式(4)所示。

$$sim(u_i, u_j) = \frac{\sum_{s_k \in US_{u_i s_j}} (rs_{u_i s_k} - \overline{rs_{u_i}}) (rs_{u_j s_k} - \overline{rs_{u_j}})}{\sqrt{\sum_{s_k \in US_{u_i s_j}} (rs_{u_i s_k} - \overline{rs_{u_i}})^2} \sqrt{\sum_{s_k \in US_{u_i s_j}} (rs_{u_j s_k} - \overline{rs_{u_j}})^2}} \quad (4)$$

其中,目标用户 u_i 和用户 u_j 两者相似度使用 $sim(u_i, u_j)$ 为表示, μ_i 和 u_j 两者共同评分过的项目本文使用 $US_{u_i s_j}$ 表示, u_i 和 u_j 的平均评分本文使用 $\overline{rs_{u_i}}$ 和 $\overline{rs_{u_j}}$ 表示。

需要指出的是, $US_{u_i s_j}$ 的大小十分重要。若 u_i 和 u_j 共同评分过的项目集合 $US_{u_i s_j}$ 的模越大,说明目标用户 u_i 和 u_j 之间的相似度越高。换言之,倘若 $sim(u_i, u_j) = sim(u_i, u_k)$, 但 $|US_{u_i s_j}| > |US_{u_i s_k}|$, 则目标用户 u_i 和用户 u_k 的相似度要低于目标用户 u_i 和 u_j 的相似度。鉴于传统的 Pearson 度量公式(4)在计算目标用户 u_i 和用户 u_k 的相似性时未能考虑这方面因素的影响,本文借鉴 Zhang YM^[11] 提出的情境化信息推荐机制,引入 $|US_{u_i s_j}|$ 来对式(4)进行修正,如式(5)所示:

$$sim'(u_i, u_j) = \frac{1}{1 + e^{-\frac{|US_{u_i s_j}|}{2}}} \times sim(u_i, u_j) \quad (5)$$

其中,式(5)中的 $\frac{1}{1 + e^{-\frac{|US_{u_i s_j}|}{2}}}$ 是 Sigmoid 函数 $\frac{1}{1 + e^{-x}}$ 的变形。当 $|US_{u_i s_j}|$ 无限大时,则 $\frac{1}{1 + e^{-\frac{|US_{u_i s_j}|}{2}}}$ 无限接近于 1。

Step 3. 根据情境修正相似性度量公式。当目标用户的情境信息与已评分用户的评分情境相似度 $Sim_i(ItemContext, ItemRatingContext)$ 较大时,该已评分用户就获得了较高的推荐权值。记为 $k = Sim(ItemContext, ItemRatingContext)$, 本文参考相似性度量公式对本文式(5)进行修正,由此得出式(6):

$$sim''(u_i, u_j) = c \times sim'(u_i, u_j) + (1 - c) \times k \times sim'(u_i, u_j) \quad (6)$$

其中,式(6)中的 c 为可调系数,其大小隶属于 $[0, 1]$ 区间。可以看出,当 $c = 1$ 时,式(6)的计算结果相当于没有考虑用户的情境因素,即还是传统的 CF 推荐算法。

4.2 构建用户信任度矩阵

在移动社交网络信任网络中,鉴于许多 SNS 用户间可能不存在或者存在不明显的潜在信任关系,由此会使得推荐的信任度矩阵非常稀疏,导致信任度计算的困难。因此,在计算移动 SNS 用户间信任度的过程中,引入用户间信任关系所具

有的传递特征,即根据信任传递规则来描述移动 SNS 用户的间接信任关系,计算得出信任度,以此进行 CF 推荐。

Step 1. 计算直接信任度 $dt(a, b)$ 。在移动 SNS 信任关系网络中,查看是否有直线连接节点 a 与 b ,存在则直接信任值 $dt(a, b) =$ 权值 $w(a, b)$;不存在则直接信任值 $dt(a, b) = 0$ 。权值 $w(a, b)$ 的计算公式为:

$$w(a, b) = \frac{\sum_{u \in R^T} t_{a, b} r_{b, i}}{\sum_{u \in R^T} t_{a, b}} \quad (7)$$

其中, i 为待评价等级的目标物品, a 为目标用户, α 为给定阈值, $t_{a, b}$ 为目标用户 a 对用户 b 的信任值; $r_{b, i}$ 为用户 b 对物品 i 的评价等级; R^T 为信任值 $t_{a, b}$ 大于或等于 α 且对物品 i 进行过评价的用户集。

Step 2. 计算推荐信任度 $ct(a, b)$ 。在移动 SNS 信任关系网络中寻找下述路径:起点为 a 、终点是 b 的路径,其子路径 $a'b'$ 为相连的点,且对于与 a' 相连的任何点(非 b') x 存在边 $a'x$, $a'b'$ 的权值 $W(a'b') > W(a'x)$,据此计算推荐信任度,如式(8):

$$ct(a, b) : ct(a, b) = \prod_{x, y \in U} w(x, y) \quad (8)$$

Step 3. 计算 a 与 b 的信任度 $Trust(a, b)$: $Trust(a, b) = \alpha \times dt(a, b) + (1 - \alpha) \times ct(a, b)$ 。 α 为 $[0, 1]$ 的参数,代表 dt 占 $Trust$ 的比重。选择信任网络中与 a 点直接信任度最大的点 x ,其推荐信任度更为可信;按照信任传递规则寻找与 x 直接信任度最大的点 y ,反复直至找到 b 点。

4.3 融合情境与信任度矩阵

在融合用户情境兴趣的过程中,本文将传统 CF 与信任推荐相融合的推荐思想,将情境兴趣相似度矩阵与移动 SNS 用户信任度矩阵进行融合,由此计算信任度与情境兴趣相似度相融合的一个综合取值,最终进行协同过滤推荐。本文情境兴趣矩阵与信任度关系矩阵融合的公式为式(9):

$$Weight = \begin{cases} 0 & \text{if } Sim(i, j) = 0 \text{ and } Trust(i, j) = 0 \\ Trust(i, j) & \text{if } Sim(i, j) = 0 \text{ and } Trust(i, j) > 0 \\ \frac{2 \times Sim(i, j) \times Trust(i, j)}{Sim(i, j) + Trust(i, j)} & \text{if } Sim(i, j) > 0 \text{ and } Trust(i, j) > 0 \end{cases} \quad (9)$$

其中, μ_i 和 u_j 间的相似度采用 $Sim(i, j)$ 表示, μ_i 和 u_j 的信任度记为 $Trust(i, j)$, $Weight$ 为合并后的综合值。

4.4 产生推荐

本文采用中心加权的计算方法,将推荐的权值 $Weight(i, j)$ 替换为融合移动 SNS 用户情境相似度与用户间信任度的一个复合值来进行推荐,计算方法如式(10)所示。

$$P_{u, i} = \overline{R_u} + \frac{\sum_{i=1}^m (R_v(i) - \overline{R_v}) \times Weight(u, v)}{\sum_{j=1}^m Weight(u, v)} \quad (10)$$

其中,式(10)中的 m 表示相近邻居的个数, $R_u(i)$ 表示用户 u 对项目 i 的评分值, $\overline{R_u}$ 则为平均评分值,同理可得 $\overline{R_v}$ 为用户 v 所获得的平均评分值, $Weight(u, v)$ 为 u 对 v 的推荐权值。最后,按照式(10)给出最后的 Top-N 推荐结果。

4.5 Mapreduce 化的云计算并行推荐

为了提高大规模复杂社交网络环境下的推荐性能,本节在 Hadoop 云环境实现模型的 MapReduce 化并行推荐,具体

分为以下四个 MapReduce 处理过程.

4.5.1 Mapreduce 化的向量生成过程

Step 1. 根据输入的信任信息向量,由此生成输出结果,最后表达为(Long, String) 键值对的形式.

Step 2. 根据 MapReduce 中的 Map 函数进行数据处理,将输入信息转化为用户 ID、项目 ID 及对应的偏好值,生成全新的键值对:(用户 ID, 项目 ID, 用户偏好).

Step 3. 通过全局搜索获取所有信息映射项目 Hadoop 集合.

Step 4. 根据 MapReduce 的 Reduce 函数构建偏好值向量,基于此生成(用户 ID, 用户向量) 形式的键值对.

4.5.2 Mapreduce 化的共生矩阵计算过程

Step 1. 将上一个 MapReduce 输出值(用户 ID, 用户向量) 作为 MapReduce 处理的输入.

Step 2. 按照 Map 函数的运算规则,计算共生矩阵所包含的各元素的对应值.

Step 3. 根据上一步骤找出每个项目 ID 与其他所有项目 Hadoop 同时出现的键值对.

Step 4. 按照 MapReduce 的 Reduce 函数,计算 $dt(a, b)$ 给出的键值对的共生向量.

4.5.3 Mapreduce 化的矩阵融合过程

Step 1. Map1 函数以共生矩阵键值对(项目 ID, 共生向量) 为函数输入,将其转化为存储共生向量和偏好值向量的数据结构,最后生成键值对的形式,作为下一步骤的输出.

Step 2. Map2 函数以偏好值向量键值对(用户 ID, 用户向量) 为输入,按照 Hadoop 函数生成如下形式的键值对:(项目 ID, 用户 ID, 偏好值).

Step 3. 按照全局搜索策略,计算得出每个项目 ID 的共生向量以及偏好值向量.

Step 4. 按照 Reduce 函数,计算移动 SNS 用户对每个项目的预测评分,最后以(用户 ID, R) 的形式输出.

4.5.4 Mapreduce 化的推荐产生过程

Step 1. 本次 Map 函数的输入为上一个 MapReduce 的输出(用户 ID, R).

Step 2. 根据全局搜索策略,计算并输出用户对每个待选项目的预测评分值.

Step 3. 将 Reduce 函数所有的预测评分值进行排序,给出最后的 Top-N 推荐结果.

5 实验结果及分析

分布式 Hadoop 云环境可提高大规模社交网络环境下的并行处理性能.因此,本文将实验环境配置在一个使用 9 台服务器搭建的 Hadoop 计算机集群上,其中一台为 NameNode,其余为 DataNode.9 台服务器硬件配置为:Lenovo 服务器一台,其中内存为 4G,硬盘 1T,CPU 主频 2.8G.其他 8 台服务器采用了普通 PC 机,内存 2G,硬盘 320G,CPU 主频 2.6G.其中,Hadoop 云环境的 NameNode 为 Lenovo 服务器,命名为 Hadoop;将其余 8 台 PC 机作为 Hadoop 的 DataNode,具体命名为 hadoop1-hadoop8.

5.1 数据集

本文实验抓取了移动应用商城类网站上最常被下载的移动 SNS 服务,以此作为本文推荐系统的“项目”数据集.鉴于

云环境用户情境兴趣的移动 SNS 信任推荐模型中涉及到移动 SNS 信任关系,本文以河北省高校 5 所在校大学生为研究对象,这些大学生可在该类网站上进行一些信息共享操作,如分享项目评分及购后评价等;同时,用户还可将其他同学加入到个人的信任列表中,建立自己的“社交网络”.据此,可根据本文模型可计算大学生用户间的信任关系(信任度取值范围为 $[0, 1]$).鉴于本文研究涉及显式的情境相关移动用户评分,进一步邀请被调查者将相关调查信息发送给 5 位好友用户,邀请其对上述移动 SNS 服务中可能感兴趣或者不感兴趣的移动 SNS 服务项目进行评分:测试用户除对移动 SNS 网络服务进行打分外(打分取值为 1-5,其中 1 表示非常不喜欢,5 表示非常喜欢),还需给出其对相关移动网络服务评分的特定“时间”情境信息,如评分时为清晨、上午、中午、下午、傍晚或者夜里.最后获得了 253 位用户对上百种移动 SNS 应用服务的评分数据集,其中不含情境信息的评分记录 4074 条,含有情境相关信息的评论 2467 条;此外,该数据集还包含用户间信任关系信息,列出了存在直接信任关系的用户 ID,将此作为实验测试部分的数据集.

5.2 评价标准

本文实验评价标准主要包括两个方面,首先测试云环境下基于 Hadoop 的并行挖掘有无提升模型计算的性能,其次是测试模型对数据稀疏性导致推荐精度下降的缓解情况.

作为并行计算测试的评价指标,加速比主要用来对比单机与并行计算两种不同环境下特定算法运行所耗费的时长,其计算方法为单机运行时间与并行运行时间的比值,具体公式表示为: $S = T(1) / T(N)$.其中, $T(1)$ 为算法在单机环境下的运行时间, $T(N)$ 为多机并行处理的时间,两者的比值结果即为加速比.此外,为了对比 Hadoop 云环境下不同数目节点对并行计算结果的影响,还引入了相对加速比指标进行对比: $S_{\text{相对}} = T(\text{单 DataNode}) / T(\text{多 DataNode})$.其中, $S_{\text{相对}}$ 表示相对加速比,单 DataNode 集群运行的时间表示为 $T(\text{单 DataNode})$,多 DataNode 集群运行时间为 $T(\text{多 DataNode})$.

在推荐性能测试方面,本文采用平均绝对偏差 MAE、用户覆盖率 Coverage 以及 $P(u) @ N^{[12]}$.其中,MAE 表明预测评分与实际评分间的偏差程度,MAE 的计算公式为: $MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N}$.覆盖率指标主要衡量对数据集的覆盖范围,本文采用 Coverage 表示,计算公式为:

$$Coverage = \frac{1}{n} \sum_{i=1}^n comp(p_i) \quad (11)$$

其中,测试实验执行次数使用 n 表示,算法计算得出的用户预测评级使用 p_i 表示.而 $P(u) @ N$ 则表示推荐列表中与用户 u 符合的用户需求项目数与 N 的比值,计算公式为: $P(u) @ N = (\text{relevant items in top } n \text{ items for } u) / N$.

5.3 实验结果分析

5.3.1 并行推荐技术的性能比较

在测试推荐模型不同节点的加速比时,首先对选定的测试数据集进行数据分配,按照数据量大小定义为数据集 D1、D2、D3、D4,其包含的数据分别为 1000、2000、3000 和 4000 条用户评分的数据集.在此基础上,本文测试 Hadoop 云环境下上述数据集的并行运行时间,Hadoop 运行的节点数量分别选

择 2、4、6、8 四种不同的情况,记录下各数据集在不同节点下的运行时间,具体如图 1 所示。

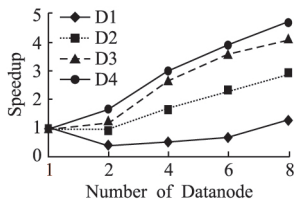


图 1 Hadoop 云环境下算法并行推荐的加速比效果图
Fig. 1 Speedup of parallel algorithm in Hadoop environment

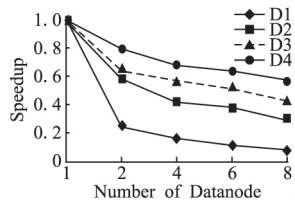


图 2 Hadoop 云环境下并行推荐的扩展比效果图
Fig. 2 Spread ratio of parallel algorithm in Hadoop environment

根据图 1 的测试结果可以看出,当测试数据集为 D1 时,本文推荐模型在 Hadoop 环境下运行的加速比会随着节点数的增加而呈现出先降后升的变化;而当测试数据集为 D2-D4 时,加速比在同数量节点情况下随数据量的增长而加快,在相同数据量的情况下会随节点数的增加而上升。根据图 2 的扩展比测试结果可以看出,D1 数据集测试出的扩展比曲线下降趋势非常明显,而 D2-D4 数据集测试时的扩展比虽然也呈现出下降趋势,但随节点数的增加而呈现出较为平稳的下降趋势。从图 2 的对比结果还可以看出,在小数据量如 D1 的测试情况下,推荐模型在 Hadoop 环境下的运行效率较低,没有很好的体现并行计算的优势,究其原因,主要是 Hadoop 环境下推荐系统的作业启动时间以及交互时间较长,而真正用于并行计算的时间则相对较短,因此综合在一起的最终运行时间较长,很难体现 Hadoop 集群并行计算的优势。但随着测试数据量的增加,Hadoop 环境下推荐系统启动与交互的时间比例会逐步缩小,而更多的比例则主要用于系统的并行计算。因此,测试数据集越大,算法并行计算的速度优势就越明显,并行处理的效率也越趋向于稳定。

5.3.2 推荐性能比较

为了对体现本文推荐模型性能方面的优越性,本文选择与传统协同过滤、传统的情境上下文预过滤方法进行对比,三类推荐方法的 MAE、Coverage 以及 $P(u)@N$ 比较具体如图 3 与图 4 所示。

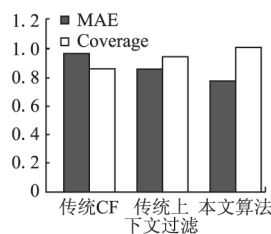


图 3 三类方法的 MAE 与 Coverage 比较
Fig. 3 Comparison of MAE and Coverage of three algorithms

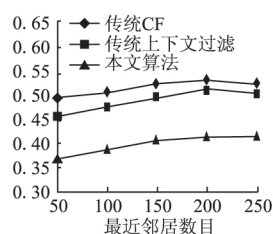


图 4 三类方法的 $P(u)@N$ 比较
Fig. 4 Comparison of $P(u)@N$ of three algorithms

从图 3 可以看出,实验条件下传统 CF 推荐方法没有使 MAE 减小,而传统的情境上下文预过滤方法以及本文模型的 MAE 相比于传统 CF 有所降低,表明引入情境的推荐取得了更优效果。而在数据稀疏情况下,传统情境上下文预过滤方法的 MAE 较高,这是因为加入上下文情境后的用户-评价矩阵

被进一步稀释,使得推荐算法在启动阶段很难找到相似用户,由此导致精准度的下降。本文模型更好地将兴趣相似度与社交网络信任度融入到推荐过程中,改善了由于数据稀疏而导致的最近邻居搜索难题,因此在启动时比传统 CF 以及传统的情境上下文预过滤的推荐性能更优。此外,本文模型的 Coverage 指标也高于传统 CF 与传统情境上下文过滤方法,这是因为情境兴趣相似度与信任关系相互融合的推荐模型在选择移动 SNS 用户最近邻居的过程中,根据移动 SNS 用户间的动态信任关系对最近邻居进行了筛选,由此选择出了可信的相似用户;而在计算评分阶段,又进一步对用户情境相似性与用户信任关系进行了加权计算,因此在很大程度上加强了推荐结果与目标用户情境的相似性。由此可见,情境兴趣相似度与信任度融合的推荐模型在一定程度上扩大了最近邻居的选取范围,即增大了对邻居用户的覆盖率,有助于提升社会化推荐的精度。而图 4 对三种算法 $P(u)@N$ 的比较也进一步表明,在不同最近邻居数目的情况下,本文模型与传统 CF 以及传统的上下文过滤推荐算法相比具有更高的 $P(u)@N$ 准确率,因此获得了更高的推荐质量与精度。

6 结论

针对移动 SNS 推荐存在的数据稀疏问题,在 CF 推荐过程中系统引入了用户情境兴趣与 SNS 信任关系,提出融合云环境用户情境兴趣的移动 SNS 信任推荐模型。本文贡献主要包括以下两方面:

- ①在传统协同过滤推荐模型中融入移动 SNS 用户情境兴趣与 SNS 信任关系,基于此构建了融合情境兴趣的移动 SNS 信任推荐模型,较好地情境相似度与信任度融入到协同过滤推荐过程中,实验结果表明,模型获得了更高的准确率和更低的误差值,有效缓解了数据稀疏问题。
- ②提出面向云环境的移动 SNS 推荐并行处理方法,基于并行加速比与相对加速比的评价结果显示,本文方法有效提高了协同过滤在海量数据规模下的执行效率,系统的可扩展性有所提升。本文后续研究将探讨稀疏信任信息环境下的移动 SNS 推荐问题,同时考虑云环境用户不同情境语义关联关系,引入具有语义关系的情境标签对云环境用户进行信任聚类,由此提高移动 SNS 推荐系统的准确率。

References:

- [1] Bobadilla J, Ortega F, Hernando A, et al. A collaborative filtering approach to mitigate the new user cold start problem [J]. Knowledge-Based Systems 2012 26(1): 225-238.
- [2] Chen Chien Chin, Wan Yu-hao, Chung Mengchieh, et al. An effective recommendation method for cold start new users using trust and distrust networks [J]. Information Science 2013 224(3): 19-36.
- [3] Kant Vibhor, Bharadwaj Kamal K. Fuzzy computational models of trust and distrust for enhanced recommendations [J]. International Journal of Intelligent Systems 2013 28(4): 332-365.
- [4] Sohn Jong-Soo, Bae Un-Bong, Chung In-Jeong. Contents recommendation method using social network analysis [J]. Wireless Personal Communications 2015 73(4): 1529-1546.
- [5] Hu Xun, Meng Xiang-wu, Zhang Yu-jie, et al. Recommendation Algorithm combining item features and trust relationship of mobile

- users[J]. Journal of Software 2014 25(8):1817-1830.
- [6] Liu Shu-dong, Meng Xiang-wu. Recommender systems in location-based social networks[J]. Chinese Journal of Computers 2015 38(2):322-336.
- [7] Zhang Fu-guo. Survey of online social network based personalized recommendation[J]. Journal of Chinese Computers Systems 2014, 35(7):1470-1476.
- [8] Liu Ying-nan, Xie Jin-kui, Zhang Jia-li et al. Recommendation algorithm based on trust in social network[J]. Journal of Chinese Computers Systems 2015 36(6):1165-1170.
- [9] Yu Mei-qi, Kuang Li, Wo Bin et al. APP recommendation combing item features and social network[J]. Journal of Chinese Computers Systems 2017 38(2):310-313.
- [10] Zhang Y M. Recommending approach for network community marketing based on user interest map in cloud environment[J]. Journal of Intelligence 2013 31(3):184-188.
- [11] Zhang Ya-ming. A contextualized collaborative filtering recommendation model for mobile information service in cloud environment[J]. Journal of the China Society for Scientific and Technical Information 2012 33(5):508-519.
- [12] Ren Xing-yi, Song Mei-na, Song Jun-de. Context-aware point-of-interest recommendation in location-based social networks[J]. Chinese Journal of Computers 2017 40(4):824-841.

附中文参考文献:

- [5] 胡 勋, 孟祥武, 张玉洁, 等. 一种融合项目特征和移动用户信任关系的推荐算法[J]. 软件学报 2014 25(8):1817-1830.
- [6] 刘树栋, 孟祥武. 基于位置的社会化网络推荐系统[J]. 计算机学报 2015 38(2):322-336.
- [7] 张富国. 基于社交网络的个性化推荐技术[J]. 小型微型计算机系统 2014 35(7):1470-1476.
- [8] 刘英南, 谢瑾奎, 张家利, 等. 社交网络中基于信任的推荐算法[J]. 小型微型计算机系统 2015 36(6):1165-1170.
- [9] 于美琪, 邝 砾, 吴 斌, 等. 融合社会网络和项目特征的移动应用推荐[J]. 小型微型计算机系统 2017 38(2):310-313.
- [11] 张亚明. 面向云环境的移动信息服务情景化协同过滤推荐[J]. 情报学报 2012 33(5):508-519.
- [12] 任星怡, 宋美娜, 宋俊德. 基于位置社交网络的上下文感知的兴趣点推荐[J]. 计算机学报 2017 40(4):824-841.