

文章编号: 1007-5321(2015)02-0001-15

DOI: 10.13190/j.jbupt.2015.02.001

大数据环境下的推荐系统

孟祥武, 纪威宇, 张玉洁

(北京邮电大学 智能通信软件与多媒体北京市重点实验室, 北京 100876)

摘要: 信息过载是大数据环境下最严重的问题之一, 推荐系统作为有效缓解该问题的方法, 受到工业界和学术界越来越多的关注. 如何充分利用丰富的用户反馈、社会化网络等信息进一步提高推荐系统的性能和用户满意度, 成为大数据环境下推荐系统的主要任务. 首先, 对近几年大数据环境下的推荐系统进行了综述, 对大数据和推荐系统进行了概述, 对推荐系统在传统环境下和大数据环境下的区别进行了辨析; 然后, 根据层次化的框架对推荐系统关键技术、效用评价以及应用实践等进行了概括、比较和分析; 最后, 对大数据环境下推荐系统有待深入研究的难点和发展趋势进行了展望.

关键词: 大数据; 推荐系统; 关键技术

中图分类号: TP18

文献标志码: A

A Survey of Recommendation Systems in Big Data

MENG Xiang-wu, JI Wei-yu, ZHANG Yu-jie

(Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,
Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Information overload is one of most critical problems in big data, and recommendation systems which are powerful methods to solve this problem are coming under growing attention by industry and academia. The main task of recommendation systems in big data is to improve the performance and accuracy along with user satisfaction utilizing user feedback, social network and other information. A survey of the recommendation systems in the big data is proposed, which includes the summarization of big data and recommendation systems, the differences between the recommendation systems in traditional environment and in big data, key techniques, evaluation and typical applications according to a hierarchical framework. Finally, the prospects for future development and suggestions for possible extensions are also discussed.

Key words: big data; recommendation systems; key technology

随着移动应用的发展, 信息数据量正爆炸式增长^[1], 大数据概念受到了工业界和学术界的普遍关注. 根据估算, Google 公司在 2009 年就为美国经济贡献 540 亿美元^[2]. 由此可见, 大数据正在引发一场新的技术革命.

科技在丰富人类生活的同时, 也给人们带来了

选择的困扰——如何快速有效地从繁杂的数据中获取有价值的信息. 推荐系统作为解决“信息过载”问题的有效方法^[3], 正在发挥着显著的作用. 传统的推荐系统在处理大数据时存在的问题正在限制其性能发挥. 为了充分挖掘数据价值, 提高推荐系统的性能和实时性, 进一步有效缓解信息过载的问题,

收稿日期: 2014-12-29

基金项目: 国家自然科学基金项目(60872051); 北京市教育委员会共建项目

作者简介: 孟祥武(1966—), 男, 教授, 博士生导师, E-mail: mengxw@bupt.edu.cn.

系统和算法的扩展和改进、支撑理论的发展成为推荐系统亟待解决的问题。

当前,大数据环境下推荐系统的研究工作已经起步^[4],但是还没有形成完善的理论和技术。为了促进研究,积极探索,笔者对大数据环境下推荐系统的研究与应用进展进行综述。

1 推荐系统和大数据

1.1 传统推荐系统

传统推荐系统生成推荐项的过程中有2个重要阶段^[3]:数据预处理阶段和推荐生成阶段。在数据预处理阶段,推荐系统需要从数据中获取用户偏好;推荐生成阶段,推荐系统根据用户偏好信息,利用推荐算法,从数据集中生成用户推荐项目。

偏好获取技术^[5]是指通过跟踪、学习用户的兴趣、偏好以及性格特征等信息,实时、准确地发现不同用户对各种网络服务的需求,并对其变化做出适应和调整。传统的用户偏好获取技术通过显式或隐式的方式获取用户的偏好,主要分为启发式和建模两类。前者利用一些具有直观意义的启发式方法来获取用户需求,如最近邻算法、聚类(K-Means 算法^[6])、相似度计算等;后者通过引入机器学习技术学习一个模型,如决策树归纳^[7]、贝叶斯分类^[8]、聚类等。针对用户偏好随时间迁移的问题,研究者使用一些自适应方法,如信息增补技术、遗传算法^[9]和神经网络技术,来解决此问题。

从信息过滤的角度来看,传统的推荐系统^[3]主要分协同过滤推荐系统(CF, collaborative filtering recommendation)^[10]、基于内容推荐系统、混合推荐系统。随着移动端设备的发展,又出现了上下文感知推荐系统^[11]。

1.2 大数据的基本概念及其特点

大数据目前还没有统一的定义。一种狭义的定义:大数据是指不能装载进计算机内存存储器的数据。李国杰等^[1]对大数据的定义:一般意义上,大数据是指无法在可容忍的时间内用传统IT技术和软件工具对其进行感知、获取、处理和服务的数据集合。

大数据有多方面的特点,目前,人们将大数据的特征总结为“5V 模型”^[12],即体量大(volume)、速度快(velocity)、模式多(variety)、难辨别(veracity)和价值大密度低(value)。

1.3 传统推荐系统和大数据环境下推荐系统的异同点

大数据环境下的推荐系统是传统推荐系统的延伸,由于大数据环境比传统环境面临更加复杂的信息提供环境和数据特征,只有在充分、准确提取和预测用户在大数据环境下产生的各种数据中蕴含的用户偏好后,才能有效生成准确度更高的推荐。因此,尽管大数据环境下推荐系统的基本思想与传统推荐系统是相似的,但着重考虑大数据环境给推荐系统带来的影响:数据产生的速度更快,数据高维稀疏,内容采样渠道更多,多源数据在融合时由于结构和采集方式的不同会引入更高的噪声和冗余,数据结构比例发生变化,非结构数据、半结构数据成为主要数据,流式数据也成为常见数据类型。数据内容变得丰富,推荐系统可以采集到丰富的用户隐式反馈数据。移动网络的快速发展,促使移动应用变得丰富多彩,用户使用移动设备或登录移动应用产生丰富的移动社会化网络数据,尤其是基于位置的GPS数据成为重要的数据。以数据处理为主的诸多大数据问题使推荐系统对数据处理能力的要求更高,同时丰富的数据使得用户对推荐系统的实时性和准确性要求更高,从而使得适合传统推荐系统的方法并不能直接应用到大数据环境下的移动推荐中^[12-14],需要进行算法的改进和扩展,才能较好地满足大数据环境下推荐系统的需求。

大数据环境下推荐系统的大数据特性主要可以从以下几个方面解释^[12]:①需要处理的数据量更大,多源数据融合会引入高维稀疏性数据,且数据存在更高的冗余和噪声,因此,大数据环境下推荐系统要求更高的数据处理能力;②采集数据以用户隐式反馈数据为主,相比传统推荐系统使用的用户评分数据,隐式反馈数据的使用有更多优势,其采集成本更低,对用户的干扰更小,拥有丰富的数据量,可以缓解传统推荐系统评分数据的稀疏性对推荐系统性能的影响;③数据更新速度更快,这要求大数据环境下的推荐系统有更有效的更新计算效率,使其能满足更新数据的处理问题;④对推荐项目的准确性要求更高,在大数据环境下,丰富的数据为提高推荐的准确度提供了机遇,同时,严峻的“信息过载”问题使得推荐系统必须更加高效精准地为用户生成推荐项;⑤对推荐项目的实时性要求更高,大数据环境下,数据更新快,信息产生速度快,信息的时效性更短,因此,只有提高推荐系统的实时性,才能在有效

时间为用户推荐项目,真正意义上为用户缓解“信息过载”问题。大数据环境下推荐系统与传统推荐系统的主要差异如表 1 所示。

表 1 大数据环境下推荐系统与传统推荐系统的主要差异

项 目	大数据下推荐系统	传统推荐系统
输入数据	数据规模更大,数据稀疏性、冗余度、噪声更强	数据规模小,数据稀疏性、冗余度、噪声较小
数据类型	以隐式反馈数据为主	以显式评分数据为主
数据更新	数据更新快,以增量更新为主	一段时间更新一次,以全局精确计算更新
推荐结果	推荐结果准确性要求更高	准确度要求较低
推荐实时性	推荐实时性需求高	实时性要求较低

2 大数据环境下推荐系统关键技术

根据文献[3]提出的层次化移动推荐系统研究框架,提出了大数据环境下推荐系统基本框架,如图 1 所示。大数据环境下推荐系统框架被划分为 4 层,分别为源数据采集层、数据预处理层、推荐生成层以及效用评价层。其中,在数据预处理层把采集到的相关数据进行预处理计算,其数据处理结果作为推荐系统数学形式的输入,主要工作为用户偏好获取、社会化网络构建、上下文用户偏好获取等;推荐生成层是推荐系统的核心,在大数据环境下,该层主要任务就是引入和充分处理大数据,并且生成实时性强、精准度高以及用户满意的推荐结果,目前主要的推荐技术有大数据环境下基于矩阵分解的推荐

系统、基于隐式反馈的推荐系统、基于社会化推荐系统以及组推荐系统;在效用评价层,在将推荐结果呈现给用户时,需要结合用户的反馈数据,利用准确性、实时性、新颖性、多样性等评价指标评价推荐系统的性能,并根据需求对其进行扩展、改进等。

2.1 大数据环境下偏好获取技术

用户偏好获取是推荐系统的关键技术,偏好获取的准确性,直接影响推荐系统准确性和实时性。传统的偏好获取系统因为缺乏数据,很难精准获取用户短期内偏好,无法准确捕捉用户偏好的动态变化。大数据环境下,偏好获取系统可以采集到更丰富的用户数据,如何从数据中获取用户短时期内的偏好,捕捉到用户偏好的动态变化,成为大数据环境下推荐系统的研究热点和难点^[15-17]。

传统偏好获取系统由于数据不足,只能分析用户长期偏好,大数据环境下,系统可以获得更充足的数据,以此捕捉到用户短期偏好,它能动态反映用户最新的偏好,提高用户偏好预测的准确性和实时性。Yang 等^[15]提出一种有效局部隐式反馈模型,并且使用随机梯度下降法(SGD, stochastic gradient descent)^[18]优化模型,可以高效地获取用户的短期偏好。

用户的偏好随时间迁移,大数据为研究者精确捕捉这种变化带来可能。Rafailidis 等^[16]提出用户偏好动态获取(UPD, user-preference dynamics)技术,当用户偏好发生变化时,捕获量化 UPD 率,根据 UPD 值调整用户历史偏好的权值,结合用户统计学数据,对用户-项目-时间三维矩阵进行耦合张量分解,可以很好地捕捉偏好变化过程,获取用户动态偏好。

神经网络能通过调整权值改变输出,这种性质能很好地追踪用户的兴趣变化^[5],因此很多学者关注于使用神经网络追踪动态的用户偏好。Oh 等^[17]把用户偏好用关键字向量表示,当作深层神经网络(DNN, deep neural network)的输入,用户偏好发生变化时调整 DNN 的权值,能很好地捕获用户动态偏好。

处于隐私的考虑,用户有时候不愿意透露自己的统计学数据,只能获取用户对项目的评分,因此不能很好地获取用户偏好。Bhagat 等^[19]提出了贝叶斯矩阵分解推理,联合主动和被动 2 种方式,只需要用户按要求提供的评分数据,通过主动学习算法能最大化用户私有属性的期望值,同时最小化对用户

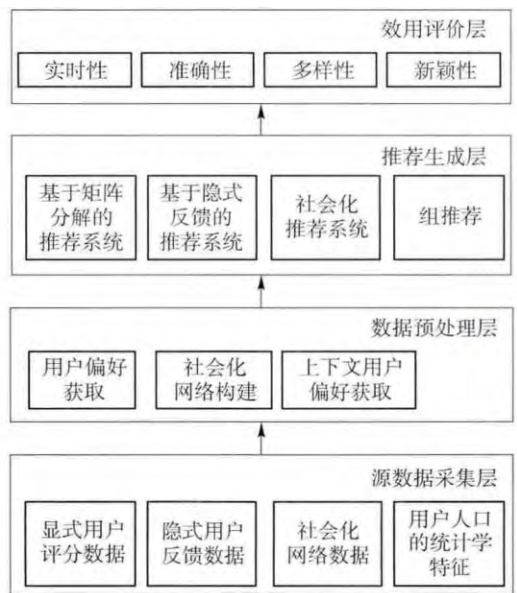


图 1 大数据环境下的推荐系统基本框架

分类的误判值。

2.2 大数据环境下基于矩阵分解的推荐系统

大数据环境下,推荐系统输入的数据有高维稀疏特性,数据存在较高的冗余和噪声,并且流数据也是一类重要的数据类型,传统的推荐系统处理这类数据时性能下降很快。研究者希望使用一类推荐系统,其能较好地处理高维稀疏数据,并且对冗余和噪声不敏感,有较低的计算复杂度,并且有良好的扩展性。

因此,近年来矩阵分解已经逐步取代传统推荐算法而成为研究热点和主流模型^[20-34],这主要归功于其准确性和扩展性。该算法的核心思想是把推荐问题转化为矩阵完全分解问题,把稀疏用户评分矩阵映射到给定的用户集合和项目集合,通过矩阵运算预测缺失评分,反映用户对项目的潜在偏好,对用户进行推荐。矩阵分解算法的优点是能降低高维数据稀疏性,对噪声和冗余不敏感,可扩展性良好,但缺点是算法解释性差,缺乏明确意义,计算复杂度^[21]。

传统的矩阵分解算法有奇异值分解(SVD, singular value decomposition)^[20-29]、非负矩阵分解(NMF, non-negative matrix factorization)^[30]、概率矩阵分解(PMF, probabilistic matrix factorization)^[34]等。这些算法的共同特点是将高维矩阵分解成为2个或多个低维矩阵的乘积形式,便于在一个低维空间研究高维数据的性质^[13]。PureSVD^[20]直接对用户评分矩阵 R 做SVD分解,未知的值用0填充,可以快速获取用户对项目预测评分。但是,SVD允许分解出现负值,有时可能缺乏解释。NMF可以保证分解所得矩阵的每个元素均是正值,这使得NMF具有直观的物理意义。PMF从概率的角度预测用户评分,假设用户和商品的特征向量矩阵都符合高斯分布,基于这个假设,把用户偏好问题转换为概率组合问题,从更深层次讨论了矩阵分解的概率解释。

大数据环境下,研究者把多种上下文因素引入建模,因此需要扩展传统的矩阵分解方法,使其能处理更高维度张量数据。Symeonidis等^[32]把位置上下文引入推荐系统,引入高维SVD(HOSVD, higher order SVD),很好地处理了用户-位置-活动三维张量。Koren等^[33]把时间上下文引入建模,提出一种SVD++算法,提高预测用户电影评分的精准度。

PMF方法也只能结合两方面信息进行二维分解。大数据环境下,为了提高推荐精准度,需要把更

多的概率因素引入建模。有研究者进行PMF的扩展研究。涂丹丹等^[22]将联合PMF(UPMF, union PMF)引入上下文广告推荐,把用户评分矩阵分解为用户、广告和网页特征矩阵的乘积。PMF建立于用户和商品特征向量矩阵都符合高斯分布的假设之上,然而,大数据环境下,数据分布通常存在多种分布可能,因此需要扩展其理论框架,使PMF能处理非高斯分布的数据。Bauer等^[23]提出了一种矩阵分解框架,能处理任意概率分布的PMF算法,提高了算法的普适性,以自动化方式取代传统数据的人工处理过程,提高了推荐系统的效率。

社会关系也是一类重要的上下文,将其引入建模,可以提高推荐的精准度。Pálovics等^[25]提出一种混合矩阵分解算法(MMF, mixing matrix factorization),计算每种潜在因素在用户决策时的权重,生成一种实时在线推荐。

大数据环境下,推荐系统能收集到几十种甚至更多的上下文数据,把这些上下文数据引入建模,能起到提高推荐精准度的效果。然而,如果全部引入上下文数据,会增加系统的计算复杂度,同时,在收集到的上下文数据中,并不是每种上下文都对用户产生积极作用,或者说有的上下文对用户特征有微小作用,这时通过过滤掉这些上下文,可以缓解分解过程的计算复杂度。Cheng等^[24]提出一种梯度推进分解机,通过贪婪算法计算受上下文影响的特征值,提高计算效率。

流式数据是大数据环境下主要的数据形式,传统的矩阵分解方法不能很好地处理流式数据的数据更新和迭代耗费问题。Díaz-Aviles等^[27]提出一种流排序矩阵分解算法(RMFX, stream ranking matrix factorization),这是一种针对流式数据的矩阵分解改进模型,算法进行了I/O扩展,只需要个性化的小缓冲器,就能满足数据流的计算需求,使用SGD优化分解过程,由于改进了更新数据的储存问题,可以较好地处理流式大数据。社会网络的用户交互流数据反映用户对项目(如商品、博客、新闻等)的偏好,但传统的推荐方法不能很好地保护用户的隐私,并且处理流数据时结果加权均根方误差(WRMSE, weighted root mean square error)^[35]较大。Isaacman等^[36]用矩阵分解的方法处理流数据,使用分布式算法,防止了集中式方法中不受信任的第三方获取用户信息,同时降低了推荐结果的WRMSE值。

传统的矩阵分解过程计算复杂度过高,当处理

高维上下文数据或流式数据时,用户特征向量矩阵和项目特征向量矩阵乘积的计算复杂度将难以忍受.为了解决这个问题,Bachrach等^[37]把用户偏好问题由内积最大化搜索问题转化成为欧几里德空间最近邻搜索问题,用PCA-Tree表示最近邻结构,用汉明距离衡量距离,降低了算法的计算复杂度.

另一种解决矩阵分解计算复杂度高的方法是进行并行化扩展,这种方法可以通过集群来提高推荐算法的计算能力,已有的成熟并行化框架或计算平台(如Hadoop或Cloud)为算法并行化提供了平台基础,因此,有大量研究者研究矩阵分解算法的并行化扩展.目前,2种主要的并行化优化方法是SGD^[18]和交替最小二乘法(ALS, alternating least squares)^[38].

SGD^[18]是一种潜在完全因素分解方法,其计算复杂度低,通过有效的并行化扩展^[39-40],SGD能充分利用集群和多核的计算效率.然而,SGD并行化存在一些挑战:在运算过程中,随着数据量的增加,因为并行化数据分享引起的读写冲突更加严重,这限制了计算设备的性能.因此,SGD的并行化扩展成为研究热点.为了缓解读写冲突问题,Gemulla等^[41]提出分布式SGD(DSGD, distributed SGD),该算法将矩阵 R 划分成块,通过调整更新的序列,避免了并行更新带来的重写冲突,但是这种方法没有考虑到每个分块中用户偏好评分数据的数量,虽然容易实现,但是可能会引起计算节点上的负载倾斜.Zhuang等^[39]在DSGD的基础上,优化了负载平衡,提出快速并行化SGD.已有的SGD方法都不能很好地根据输入数据的一些固有性质处理大数据,且大部分改进方法都采用了大容量的同步处理模式,即在集群中所有计算节点的本地副本同步更新共享数据,之后定期同步更新.不同的改进方案在算法收敛速度和计算节点通信成本间进行权衡,选择适合的更新频率,目前还没有统一成熟的结论.Petro-ni等^[42]采用基于图分割的思想,提出一种异步图SGD(GASGD, graph asynchronous SGD),把输入数据当作是由用户或项目为节点,用户对项目的评分为边的图,使用贪心分割算法,使计算节点间实现基本负载均衡.

ALS^[38]是另一种基于潜在因素的矩阵分解计算方法,有良好的并行化性能,但本身计算复杂度高,研究者直接把ALS方法集成到并行化平台,利用集群计算矩阵分解问题,其成本低,有良好的工

业化应用.Schelter等^[26]用ALS方法解决低秩矩阵分解问题,利用Map-Reduce编程实现并行化应用,取得较为良好的集群计算性能.

2.3 大数据环境下隐式反馈数据的推荐系统

大数据环境下,隐式反馈数据(如用户视频点击、浏览网页、转发微博、购买商品等行为数据)是主要的输入数据形式,这类数据不需要用户投入更多的精力,同时也不会影响用户正常生活,收集成本低、应用场景广泛,数据规模也更大^[15],而用户评分数据只有非常稀疏的数据量^[14].这些条件决定了在大数据环境下,基于隐式反馈数据的推荐系统将成为推荐系统的主要形式之一.

传统的推荐系统忽视了大量的隐式反馈信息,而只关注于分析用户评分数据,这不仅浪费了宝贵的大数据资源,更限制了大数据环境下推荐系统的发展^[14].

隐式反馈数据分为“选择”和“未选择”两类,其中“选择”数据数量较少,该数据能直接反映用户偏好,“未选择”数据数量众多,却不能直接解释为用户不喜欢,而是无法确定用户偏好^[14].目前,研究者主要使用正隐式反馈数据,如Pálovics等^[25]使用用户收听音乐的行为数据,而浪费了大量用户未收听音乐的数据.针对该问题,印鉴等^[14]提出一种隐式反馈推荐模型(IFRM, implicit feedback recommendation model),将推荐任务转化为用户选择行为发生概率的最大化问题,达到直接对隐式反馈数据进行建模的目的,这样既利用了“未选择”信息,又避免引入负例的同时引入噪声,提升了推荐质量.同时,借鉴了降维方法解决高维稀疏数据的噪声问题,进一步采用分桶的并行化隐式反馈模型p-IFRM,提高了算法的效率.隐式信任数据是一种由用户间交互行为反映的用户关系,Fazali等^[43]使用隐式信任数据预测用户信任值评分,实验证明与采用用户评分数据获取结果相似,但其数据采集成本更低,有很好的应用前景.

相比于用户评分数据,隐式反馈数据能直观反映用户的行为偏好.大数据环境下,丰富的隐式反馈数据使得短期局部用户偏好的捕捉也成为可能,当前短期偏好可以较好预测用户未来一个时间段内的偏好,生成实时性强的推荐结果.Yang等^[15]提出基于局域隐式反馈大数据的推荐算法,模型利用局部和全局的隐式反馈数据,基于用户未来短期内的音乐偏好受到当前用户偏好影响的思想^[44],把用户

时间划分为多个时间切片,在每个时间切片内,综合考虑用户的上下文环境(如休息、工作或跑步)对用户歌曲选择的影响,根据当前时间切片内获取的用户音乐偏好预测下一个时间切片内用户的音乐偏好,进而为用户准确推荐歌曲,并且使用SGD优化算法,提高算法实时性,同时调节时间切片的粒度,从而获取用户长期稳定偏好和用户短期易变偏好。

传统的推荐方法在处理评分数据时有良好性能,但隐式反馈数据没有直接的评分,不同于基于评分预测的方法,直接的基于排序的方法在处理隐式反馈数据时有更好的效果。Zhao等^[45]把微博中提取的用户对商品反馈信息加入排序算法,取得了良好的电子商务推荐效果。但是传统的排序方法目标函数最小化需要付出很大代价,需要在目标采样上牺牲一定精度来改善算法的计算效率^[46],而大数据时这种牺牲往往不能容忍。有研究者认为,在大数据环境下,数据采样的方式不再重要,甚至不需要采样。基于这样的思想,Takács等^[38]提出RankALS,该算法不进行采样,直接对排序目标函数最小化,提高了数据处理效率。

2.4 大数据环境下社会化推荐系统

社会化网络的成熟以及通信网络的快速发展,形成了社会化网络信息的大数据化,这类数据蕴含丰富价值,反映用户偏好。为了缓解大数据环境下用户评分数据的稀疏性,大量研究者利用丰富的社会化网络数据,把用户社会关系、信任关系等引入推荐建模,进一步提高推荐系统精准度。Guo等^[47]把社会化网络用户信任关系引入用户偏好建模,联合传统用户评分数据,共同提高了推荐精准度。

社会化网络形成的丰富数据如何快速处理是社会化推荐系统的主要难点。大数据环境下,社会化网络间的用户交互产生了流式大数据,传统的社会化推荐系统无法较好地解决流式数据的处理和更新问题。针对该问题,Díaz-Aviles等^[27]采用选择采样策略,在内存中选择建立一个个性化缓冲区,该区域只用于加载和更新模型数据,而不需要将产生的全部流式数据一次性加载到内存空间,这缓解了推荐系统的储存压力,同时引入矩阵分解方式快速处理流式数据,是大数据环境下处理流式数据的有效方法。在数据形式上,大数据环境下社会化网络平台给研究者提供了更丰富的隐式反馈数据,如用户间行为体现出的信任关系数据,这种数据更客观地反映出用户间可靠的信任关系。Fazeli等^[43]量化用户

间信任关系,将其引入社会化数据矩阵分解过程中,其数据采集成本更低,却获得了和用户显式信任评分数据相似的推荐精准度。

随着移动设备和移动网络的快速发展,设备GPS和各种加速器提供了基于用户个人行为的感知数据,这些数据形成了用户轨迹的大数据形式。Zheng等^[48]引入用户轨迹数据GPS trajectories,该数据由用户历史位置组成。如何很好处理这类数据,进一步提高推荐的精准度和个性化,成为大数据环境下社会化推荐系统的难点和热点。大量研究者关注位置数据的处理和分析。Zheng等^[49]为处理GPS轨迹数据,提出基于树的分层图,然后提出超文本诱导的主题搜索(HITS, hypertext induced topic search)。为了感知GPS数据中用户行为,Zheng等^[50]提出一种基于监督学习的方法。

移动社交网络成为主要的移动网络应用,并以此产生了大量涉及用户位置、活动的海量数据内容。有研究者处理这类数据,以提高推荐系统性能。移动网络用户发表微博时标注自己的位置信息,Hu等^[51]提出一种基于空间主题(ST, spatial topic)建模,对微博主题进行空间建模,使用稀疏编码技术缩短数据学习过程时间,提高了推荐的性能。位置数据多数没有直观评分数据,因此基于排序的算法有较好性能,Zheng等^[31]引入基于排序的协同张量和矩阵分解模型。有研究者使用用户-位置-活动张量表示用户位置数据,为了从中提取用户偏好,Symeonidis等^[32]引入HOSVD分解用户张量数据,发掘用户、位置和活动数据间的潜在语义关联,为了缓解数据累积过程中的计算复杂度问题,采取增量计算方法更新数据(包含用户、位置及活动信息)。

基于社会化网络的推荐系统的主要衡量标准是均根方误差(RMSE, root mean square error),RMSE最小化是研究者主要关心的问题^[52]。社会化网络数据中隐式反馈数据占很大比例,因此基于排序的推荐系统有广阔的应用前景,然而基于排序的推荐系统(如k-排序推荐系统)RMSE值较高,这影响了推荐系统的性能。针对该问题,Yang等^[52]在社会化信任加强矩阵分解模型的基础上,把最近邻引入信任网络的潜在特征空间,显著降低了推荐系统的RMSE值。

为了处理社会化网络数据,大量研究者考虑把推荐系统集成在成熟可靠的并行化计算平台,利用集群高效处理数据,同时降低成本和技术难点,因此

并行化的方法成为工业界和学术界都关注的方法。Zhang 等^[53]对用户使用的品牌进行特征提取,引入 Aprior 算法对提取到的特征值进行聚类,获取拥有最多用户的类似品牌,采取一种基于回归的分布式迭代随机阈值算法(DISTA, distributed iterative shrinkage-threshold algorithm),该方法能随机优化数据训练过程,把 Aprior 算法集成到 Hadoop 平台,可以快速定义潜在推荐品牌。

2.5 大数据环境下组推荐系统

大数据环境下,为每位用户生成个性化推荐需要进行大量计算,成本高。在一些特定领域,如家庭电视节目、餐馆等内容的推荐上,根据用户间行为相似性,将其按照偏好划分为不同层次的组,为用户进行群组推荐,可以用最小的经济代价,极大程度满足用户的需求,同时缓解推荐系统的计算复杂度。因此,组推荐系统成为大数据环境下主要的推荐应用之一。传统的组推荐算法缺乏少数用户群偏好数据,群成员间的权重缺乏区分,通常简单地采用等权重方式融合偏好,不能突出高影响力群成员在组中重要作用,群成员间的层次不明显,偏好融合时可能出现偏差,这些问题限制了组推荐应用的发展。

为了缓解组推荐系统群成员间影响力缺乏区分、层次不明显的问题,Wang 等^[54]提出主题建模方法,根据主题划分子群,分别计算组成员在所在主题下的影响特征矩阵,衡量其对该群组权重,按照层次划分群成员作用,使群偏好精度有一定提升。

为了缓解组推荐系统的数据稀疏性问题,研究者引入基于位置的社会化网络数据(LBSN, location-based social network),这些数据涉及用户位置和活动信息,反映用户在时空上的偏好。Purushotham 等^[55]使用分层次的贝叶斯网络,利用主题模型学习 LBSN 的用户活动数据,利用矩阵分解技术快速获取群成员偏好,进一步生成群组偏好,有较好的实时性。

由于缺乏成员偏好数据,传统的推荐系统只能通过协同诱导方式生成群偏好,这种情况极可能生成令大多群成员都不满意的族群推荐项目^[56]。大数据环境下,研究者期望采用丰富隐式反馈大数据来弥补传统组推荐的数据缺失。基于排序的推荐系统在处理隐式反馈数据时有良好性能,研究者把丰富的用户偏好信息引入到用户群偏好的生成过程中,减少因缺少群成员用户信息而引入的不满足实

际情况的群组推荐项在组排序推荐中的数量,提高推荐准确性。Naamani-Dery 等^[56]提出一种启发式的迭代组偏好获取方法,在群成员决策融合阶段最小化用户偏好诱导效果,减少了不满足条件的组推荐项,使生成的组推荐项可以更好地反映群组偏好,最精确时合适的推荐项数量占排序列表的 90% 以上。

2.6 大数据环境下推荐算法扩展

大数据环境下,推荐系统有更高的数据处理和时效性需求,为此,大量研究者关注于传统推荐算法的扩展,希望通过扩展使推荐系统能较好地完成任务。目前,主要的扩展方法如图 2 所示。

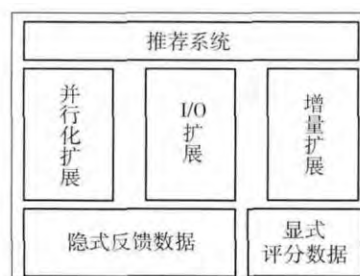


图 2 算法的扩展方式

传统的串行计算模式很难满足大数据环境下推荐系统的计算需求,而成熟的并行化计算框架(如 Hadoop^[57]或 Cloud 平台^[58])使集群计算成为可能,通过对传统算法进行并行化编程^[59],则可以有效地提高推荐系统的计算效率。有大量研究者进行算法的并行化扩展,Katkar 等^[59]对朴素贝叶斯分类进行了并行化扩展;Ferreira 等^[57]使用 Map-Reduce 编程,在 Hadoop 平台上实现了高维大数据的聚类算法。对于实时性较强的推荐内容,研究者利用 Cloud 平台实现数据的快速处理,Yu 等^[58]把旅游推荐系统集成到 Cloud 平台,有较好的实时性;Walunj 等^[60]把电子商务推荐系统集成到 Mahout 框架上;Xu 等^[61]把新闻推荐系统集成到 Hadoop 平台;Gong 等^[62]把电子商务推荐系统集成到 Cloud 平台等。

传统推荐算法 I/O 过程存在一些限制,需要一次性将数据加载到内存中,或者频繁通过调用 I/O 使用数据,这些步骤降低了算法的 I/O 效率,甚至使算法无法处理大规模数据。针对传统算法一些限制,研究者对推荐系统进行 I/O 扩展^[63-64],提高算法 I/O 效率,使推荐系统能处理大规模数据。Kant 等^[65]提出一种构建规范性决策树算法(CCDT, con-

traction canonical decision tree algorithms) 通过特征值聚类获得分支相对较少的密集型决策树,再引入以交叉验证为基础的删减方法最大化减少不理想的决策树分支,在大数据环境下依然能迅速形成规范型决策树. Franco-Arcega 等^[63]提出一种快速决策树算法,能缓解大数据下决策树生成难以收敛的问题. Xue 等^[64]用压缩的 I/O 改进算法,能较大程度改善大数据环境下的聚类效率.

大数据环境下,数据累积速度快,传统推荐算法的全局精确计算更新思想无法满足大数据环境下推荐系统的数据更新需求^[12],增量更新将逐步取代全局精确更新方式,研究者对推荐系统进行增量计算扩展以缓解数据更新处理时的计算复杂度. Yang 等^[66]对决策树进行增量计算,在处理大数据时,不仅能提高计算效率,对于大数据中存在的噪声也有良好的抵抗性.

2.7 大数据环境下推荐系统的效用评价

效用评价对于检验推荐系统的性能和发现其存在的问题十分重要,是推荐系统不可或缺的步骤,而数据集和效用评价指标是 2 个重要因素.

2.7.1 大数据环境下推荐系统的相关数据集

目前,大数据环境下的推荐系统使用的数据集有 2 种,即真实数据集^[25 36-38 51 56 67-68]和模拟数据集^[27 60]. 其中真实数据集使用更广泛,但为了模拟流数据^[27],也需要使用一定的规则,利用真实数据和一些模拟流规则生成模拟流式数据.

Twitter 数据集^[27 51 68]是一家著名的在线社交网络公司的数据集合,有丰富的用户交流信息和统计学信息,配合一定的模拟规则,则可以仿真大数据流,进行流式数据的推荐研究.

Netflix 数据集^[26 36 38 42 56]来自于美国大型 DVD 租赁公司,有大量用户对电影的评分和评价数据集,是分析大数据环境下用户电影推荐的数据源之一.

Yahoo! Music 数据集^[15 37-38 42 56]来源于雅虎音乐,有丰富的用户对音乐的评价、评分、切换歌曲等相关数据,是研究大数据环境下用户音乐推荐的数据源之一.

Epinions 数据集^[43]包含 49 290 名用户对 139 738 个不同项目的评分、664 824 条评价和 487 181 条信任关系状态.

GPS trajectories 数据集^[48-50]来源于微软亚洲研究院 Geolife 项目,搜集了 182 名用户在 5 年时间内(2007 年 4 月到 2012 年 8 月)的 GPS 轨迹数据,用

户位置信息由经度、纬度、海拔、时间信息组成,数据库中包含 17 621 条轨迹数据,其总距离为 292 951 km,全时长 50 176 h,数据密度高,时间间隔为 1 ~ 5 s,相邻点距离为 5 ~ 10 m.

除此之外还有一些数据集,如 Last.fm 数据集^[14-16 36]、MovieLens 数据集^[20 42]、Xbox Movies 数据集^[37]、Yelp 数据集^[51]、Gowalla 数据集^[55 67],以及豆瓣^[14]、Sina 微博^[45 54]等.

2.7.2 大数据环境下推荐系统的效用评价指标

朱郁筱等^[35]对推荐系统的评价指标进行了综述,其主要划分为 4 个方面,分别是准确度指标、基于排序加权的指标、覆盖率指标以及多样性和新颖性指标. 其中,主要的准确度指标有平均绝对误差(MAE, mean absolute error)^[43]、平均正确率均值(MAP, mean average precision)^[14 49]、平均百分比排序(MPR, mean percentage ranking)^[14]、RMSE^[15 43]、AUC 指标、平均排序分、precious^[45]、recall^[27 43];主要的加权排序指标有折扣累积利润(DCG, discounted cumulative gain)^[43 45 49]和排序偏差准确率;主要的覆盖率指标有预测覆盖率、推荐覆盖率和种类覆盖率;主要的新颖性指标有 UE(unexpectedness),多样性没有较为统一的评价指标. 对于组推荐,有准确性指标如 Avg. Accuracy^[55]、Avg. RMSE^[55]、GMAE(group MAE)^[54]等.

传统的评价指标很多是针对显式评分的推荐系统,而大数据环境下,隐式评分(如点击、购买等)的应用成为推荐系统的主流^[41],因此,需要新的评价指标来衡量大数据环境下的推荐系统.

对于个性化搜索,推荐结果是否良好的评判指标是用户愿意停留在推荐内容上的持续时间. 根据内容不同,用户点击停留行为是不同的,因此, Yi 等^[69]提出了个性化停留时间,能很好反映用户的个性化推荐性能.

大数据环境下,推荐系统的多样性和新颖性受到更多关注,传统推荐算法缺乏系统的衡量指标, Vargas 等^[70]根据多样性和新颖性提出了 3 个指标:选择、发现和相关性.

3 大数据环境下推荐系统应用进展

大数据环境下,推荐系统的个性化、普适性和高效性使其具有广阔的应用前景,本节就推荐系统的应用进展进行了分析、总结,表 2 列举了推荐系统的一些典型案例.

表 2 典型的大数据环境下推荐系统应用分类

应用领域	典型应用	推荐方法
电子商务推荐	Zhao 等 ^[45]	METIS
	Zhang ^[53]	DISTA
	Walunj 等 ^[60]	CF/Mahout
基于位置推荐	Ge 等 ^[28]	PMF/LPMF/MMMF
	Zheng 等 ^[49]	HITS
	Hu 等 ^[51]	ST
	Lu 等 ^[67]	PTR +
新闻推荐	Oh 等 ^[17]	DNN
	Diaz-Aviles ^[27]	RMFX
	Xu 等 ^[61]	CF/Hadoop
广告推荐	Tu 等 ^[22]	JPMF
	Kant 等 ^[65]	CCDT
电影/音乐推荐	Pálovics 等 ^[25]	SGD-MMF
	Schelter 等 ^[26]	ALS-LRMF
	Takács 等 ^[38]	ALS-Rank
	Petroni 等 ^[42]	GASGD

3.1 大数据环境下的电子商务推荐

根据 BIA/Kelsey 的研究评估,至 2017 年美国的社会广告投入将超过 30 亿美元^[53],因此电子商务领域有广阔的发展前景,电子商务推荐作为向用户提供合理产品的重要线上推荐应用,有重要的发展应用潜力。

目前,有大量研究者关注电子商务推荐应用的研究,其中有研究者把丰富的社会化网络信息引入推荐建模过程。Zhang 等^[53]把用户描述文件、历史活动等社会网络信息引入用户偏好建模,精确模型预测社会品牌的用户偏好,寻找品牌的目标用户,该方法有效提高目标用户选择的准确性。Zhao 等^[45]提出一个先进的产品推荐系统 METIS,该系统能从用户新浪微博中探测用户购物意图,根据用户微博和在线评论建立用户配置文件,根据配置文件中提取的用户统计学信息匹配生成产品推荐,该系统有良好的实时性。

为了提高推荐系统实时性,降低应用成本,Walunj 等^[60]在开源分布式推荐框架 Mahout 上集成一系列推荐算法,该系统架设简单,成本较低,有良好的实时性,并且有完善的评估方法和一系列成熟的推荐算法提供,可以满足不同情况下的电子商务推荐。Gong 等^[62]在云计算框架上集成电子商务推荐系统,建立云平台下的电子商务服务,有良好的商

用性。

3.2 大数据环境下基于位置的推荐

随着 LBSN 的发展,产生大量位置相关数据,这些数据反映用户位置上下文的偏好,能为用户推荐基于位置的项目和服务,如旅游路线、餐馆、旅馆、热点景观的推荐等。近年来基于位置的推荐产生越来越高的价值,如滴滴打车就是借助位置服务产生的价值链,形成了高价值的互联网公司,同时基于位置的大数据推荐也成为了大数据环境下重要的推荐应用。

位置数据反映了用户在位置上下文下的偏好。Hu 等^[51]利用位置大数据为用户推荐旅游地点,如分析了纽约市的 Twitter 用户数据,以及菲尼克斯市的 Yelp 数据,比传统的几种旅游推荐有明显的性能提升。

传统的一些旅游推荐系统,根据用户输入的限制条件对推荐项进行过滤,推荐粒度较为粗糙,有研究者引入更多信息对用户建模。考虑用户在规划旅游时通常考虑时间因素, Lu 等^[67]把时间因素引入建模,提出一种个性化旅游推荐(PTR, personalized trip recommendation),景点的预测评分由用户旅游偏好和用户时间限制因素共同决定,进一步提出并行化扩展系统 PTR+,从更精确粒度提高用户旅游推荐精准度。不同于其他实体产品,用户旅游决策还受到花费因素的影响, Ge 等^[28]提出了一类基于花费的旅游推荐,用户的花费包括财务上和時間上的花费 2 种,参考用户的花费能力估计用户旅游偏好,为用户制定适合自身经济条件的旅游项目。

GPS 轨迹数据反映用户时空上行为偏好,被应用到基于位置的推荐系统,有良好的精准度。Zheng 等^[49]提出了一种关于 GPS 轨迹数据的旅游轨迹推荐系统,该系统基于两点考虑:①在区域内轨迹路线长的用户,对该区域更熟悉;②在一区域内多用户轨迹中出现频繁的位置,是该区域的热点位置。该系统有良好的推荐效果。

3.3 大数据环境下的新闻推荐

大数据时代,信息爆炸式产生,新闻生命周期更短,用户渴望快速获取自己感兴趣的新闻内容,了解相关信息和知识,提高自身竞争力,因此,新闻推荐成为重要的大数据推荐应用之一。

不同于其他商品,新闻有很强的实时性,热点变化快,普通的推荐系统无法实时追踪用户对新闻偏好的变化,因此需要一类自适应的方法,能实时学习

用户偏好。Oh 等^[17]把 DNN 引入用户偏好建模,用自适应方法追踪新闻热点变化,能准确地捕捉用户新闻偏好,为用户提供准确的新闻。

有研究者采集到丰富的在线社交网络数据,建立用户个人描述文件,对用户偏好建模。Lee 等^[68]抽取 Twitter 用户的微博、转发以及标签数据,建立用户个人配置文件,根据用户偏好为用户提供个性化的新闻推荐服务。

用户在使用社会化网络时产生大量实时数据,持续进入社交网络的实时数据更新着流行事件,反映新闻热点变化。Diaz-Aviles 等^[27]分析 Twitter 社会数据流,捕捉 Twitter 全局趋势,并根据该趋势为用户推荐新闻。

当前很多新闻推荐系统数据来源于单一新闻源,大数据环境中,多个新闻源组成异质性源点,为了能追溯新闻热点话题,需要能处理这种异质性大数据。Xu 等^[61]融合多个信息源数据,分析多源数据追踪新闻热点话题,提高了新闻推荐的精准度。

3.4 大数据环境下的广告推荐

在线广告有良好的经济效益,有客观的经济价值和研究意义,是推荐系统的重要应用之一。然而,在大数据环境下,在线广告推荐出现了一些新的挑战:处理数据规模大、高元分类特征的兴起、未知的特征值等。针对这些问题,Kant 等^[65]划分在线广告为保证支付和无保证支付两类情况,把决策树引入在线广告应用中,分析用户偏好与其浏览页面和广告内容间的高阶关联,该应用有良好的推荐性能。

传统的广告推荐方法很多关注用户点击率,但是,很多误点和错点的数据引入增加推荐的错误率。针对传统的上下文推荐算法通常对长尾特性的数据预测准确性不高的问题,涂丹丹等^[22]把 UPMF 方法引入上下文广告推荐,分析用户感兴趣且与浏览网页内容相关的广告,该方法结合用户浏览信息、用户点击广告信息和网页与广告关联等信息,提高了广告推荐的精准度和性能。

3.5 大数据环境下的电影/音乐推荐

目前,互联网上有大量的电影/音乐资源,这些多媒体资源丰富了用户的日常生活。然而,对电影/音乐等多媒体资源内容本身的分析比较困难,大规模数量的影音资源进一步增加了推荐系统的分析负担。如何从大规模数据中提取用户对电影/音乐的偏好并向其推荐,是大数据环境下推荐系统在电影/音乐领域应用需要研究的问题。

电影/音乐推荐系统能采集到丰富的隐式反馈数据,如用户点击电影/音乐、用户切换歌曲和电影等。对于隐式反馈数据,推荐系统只能间接从数据中推测用户偏好,这时传统的方法需要在系统准确性和目标函数抽样计算效率两者间进行权衡,Takács 等^[38]用一种无采样的个性化排序方法处理电影/音乐推荐中的隐式反馈数据。

为了提高推荐系统的计算能力,研究者会把算法移植到 Hadoop 平台,利用集群完成大数据环境下的推荐。Schelter 等^[26]利用 Hadoop 平台处理电影评分数据,在大数据环境下有良好的计算能力。

Isaacman 等^[36]利用社交用户间的交互数据预测用户对内容的评分,算法在保护用户隐私的同时,较好地利用了电影评分数据。Pálovics 等^[25]对歌手、粉丝和音乐作品间关系进行分析,并预测粉丝对歌曲的评分。

Petroni 等^[42]用一种均衡负载的方法提高了推荐算法的效率,使其能快速有效地处理大数据环境下的电影评分数据。

Yang 等^[15]利用用户对音乐的偏好易发生变化的特性,发掘隐式反馈数据中用户的短期易变偏好,能快速生成音乐推荐。

Bachrach 等^[37]在运行时间和推荐质量间进行权衡,用一种转换方法加快了推荐生成的效率,在电影/音乐真实数据集上有很好的推荐性能。

Naamani-Dery 等^[56]使用启发式的迭代算法消除大多数推荐项,提高了真实电影数据集上的组推荐的准确性。

4 大数据环境下推荐系统展望

大数据时代,推荐系统有采集更丰富数据的机会和能力,因而处理信息过载和面向用户提供个性化服务的能力增强,研究和应用取得明显进展。但是,作为大数据环境下的一个重要的研究领域,其中还有很多值得深入研究并可能取得成果的研究内容,其中主要包括以下几个方向。

1) 数据稀疏性和冷启动问题

传统推荐系统的准确性和覆盖率都与用户评分数据的稀疏性有关,大数据环境下,评分数据的稀疏性更加严峻。为了解决这个问题,有研究者使用隐式反馈数据^[15,43]或者社会化网络数据^[47]来缓解推荐系统数据稀疏性问题,提高推荐精准度。

大数据环境下,上下文数据高维稀疏,因此也会

影响上下文感知推荐系统的性能。为了缓解稀疏性,研究者使用矩阵分解^[20-34]、PCA^[13, 37]等主流降维算法,把高维稀疏数据映射到低维密集空间。

对于新用户的冷启动问题,丁伟峰等^[20]提出了一种最大化参数变化的主动采样方法,能在 PureSVD 模型的基础上,选取最大化模型参数变化的样本,基于贪婪算法提出一种快速的近似采样算法,能在可接受的时间内得到采样列表,可以提高学习新用户偏好的效率。使用多异质性源点数据,也有助于缓解推荐系统的冷启动问题^[61]。

2) 大数据处理与增量计算问题

大数据环境下,数据生成速度快,多源数据融合引入更多的噪声和冗余,传统推荐算法的精确计算方式计算复杂度过高,Mayer-Schönberger 等^[71]提出的近似计算方式将逐步成为大数据环境下的主要计算方法,大数据本身可以提高算法的精准度。增量更新逐渐成为一种主要的数据更新方式,当出现新的用户或项目数据时,只对新加入的项目以及产生关联的边进行更新计算,并对原有推荐结果进行微调,而不是对全部数据进行更新计算。每隔一段时间,就用自适应方法消除局部计算引入的误差,令推荐结果的偏差不上升^[71]。

对于大数据流式数据,研究者改进算法消除相关限制,令数据不需要一次性加载进入内存,只使用个性化的小缓冲,这种方法每次只采样新数据进行计算,降低了计算复杂度^[27]。

3) 社会化推荐

推荐系统的研究人员早期就发现了“口碑效应”^[3]的有效性。社会网络中存在多种数据,如用户注册的统计学数据、交流产生的信息数据,也包括用户间的信任关系,以及丰富的上下文数据(如位置大数据等),这些数据可以提高推荐系统的精准度。

大量学者研究如何改进算法^[27, 52],充分利用社会化网络数据提高推荐系统的性能,缓解数据稀疏性带来的系统性能下降。Yang 等^[52]提出社会信任增强的矩阵分解模型,就把社会化网络用户间的信任关系引入推荐系统建模,以此来提高推荐系统性能。Diaz-Aviles 等^[27]改进了推荐系统的 I/O 过程,用一个小的缓冲暂存社会网络流数据,解决了社会化网络大数据的输入和储存问题,并使用矩阵分解处理这些数据,提高了推荐系统的实时性和精准度。

移动设备和移动网络的发展与进步,使得系统可以采集和感知移动用户空间行为信息,形成移动

上下文的大数据化,其中最主要的上下文是用户的位置信息。基于位置的上下文信息反映用户在不同空间场景下的行为偏好,大量研究者把位置数据引入推荐建模,以此获取更加精确的用户偏好,并为用户提供一系列基于位置的服务和应用^[31-32, 51, 55]。

4) 推荐系统的多样性和新颖性

多样性和准确性是重要的衡量推荐系统精准度的性质,通常需要在两者间进行权衡,很难兼顾。传统的推荐系统研究中准确性是主要追求的性能,但是商家往往更希望提高推荐系统的多样性,因为根据“长尾特性”,多样性好的推荐系统能给用户推荐范围更广泛的产品,产生更加丰厚的经济价值,但目前的方法还是通过牺牲一定的准确性来提高推荐系统的多样性。Vargas 等^[72]使用反转用户-物品评分矩阵,利用概率调节推荐项目的准确性和多样性权重。Garcin 等^[73]使用概率最近邻算法,通过给予邻居不同权重,提高推荐多样性,但这类算法无法同时提高准确性和多样性。

研究者通过从大数据中深度挖掘项目、用户以及项目和用户间的深层次关系,寻找用户和项目间的相关性,从“因果关系挖掘”思想逐步转换成大数据的“相关关系挖掘”思想,直接通过挖掘出的相关性向用户进行个性化项目的推荐,提高推荐系统的多样性,同时保持较高的推荐准确性^[71]。因此,大数据环境下,如何进行深度数据挖掘和相关性知识发现,是未来推荐领域重点研究方向之一。

5) 用户的隐私保护和安全性问题

用户隐私和安全性问题是推荐系统发展的关键问题。推荐系统通过分析用户的人口统计学信息、行为数据、上下文信息等,从中挖掘用户偏好为用户生成精准推荐。但与此同时可能泄露用户的安全和隐私信息,大数据环境下,大规模用户数据包含更多用户隐私和安全信息,如何通过分布式的方式取代传统的集中式数据获取,在不改变用户行为习惯,不危害用户隐私安全的前提下,充分利用用户大数据生成精准推荐项目,成为大数据环境下推荐系统用户隐私保护和安全性研究的热点和难点^[19, 36, 74-75]。

5 结束语

为了缓解更加严峻的“信息过载”问题,推荐系统受到工业界和学术界越来越多的关注。大数据环境下,数据规模更大,更新速度更快,数据类型更多,传统的推荐系统无法直接满足对大数据环境数据的

处理需求,因此在相同的框架下,提出了大数据环境下的推荐系统,其对大规模数据处理能力的需求更高,对推荐结果的准确性和实时性要求也更高。同时,大规模数据也为进一步提高推荐系统的准确性提供了机遇。目前,采集的主要用户数据是隐式反馈数据,相比于传统推荐系统主要输入数据——用户评分数据,隐式反馈数据数量大、成本低,同时对用户干扰小,其中从移动网络中采集到的移动社会化网络数据,尤其是用户位置数据,有很大的使用价值。在一些领域,组推荐以其低成本、高效率的特性有良好的应用。同时,大数据环境下的推荐系统应用领域还有很多亟待解决的问题,例如,如何利用大数据缓解推荐结果多样性,如何在保护用户隐私安全的同时充分利用大数据带来的价值等。因此,大数据环境下推荐系统仍然有重要的研究意义和巨大的应用价值。

参考文献:

- [1] 李国杰,程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域—大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6): 647-657.
Li Guojie, Cheng Xueqi. Research status and scientific thinking of big data [J]. Bulletin of Chinese Academy of Sciences, 2012, 27(6): 647-657.
- [2] Labrinidis A, Jagadish H V. Challenges and opportunities with big data[J]. Proceedings of the VLDB Endowment, 2012, 5(12): 2032-2033.
- [3] 孟祥武,胡勋,王立才,等. 移动推荐系统及其应用[J]. 软件学报, 2013, 24(1): 91-108.
Meng Xiangwu, Hu Xun, Wang Licai, et al. Mobile recommender systems and their applications [J]. Journal of Software, 2013, 24(1): 91-108.
- [4] Ye Tao, Bickson D, Yan Qiang. Second workshop on large-scale recommender systems: research and best practice [C] // 8th ACM Conference on Recommender Systems, 2014 ACM. Silicon Valley: ACM Press, 2014: 385-386.
- [5] 孟祥武,王凡,史艳翠,等. 移动用户需求获取技术及其应用[J]. 软件学报, 2014, 25(3): 439-456.
Meng Xiangwu, Wang Fan, Shi Yancui, et al. Mobile user requirements acquisition techniques and their applications [J]. Journal of Software, 2014, 25(3): 439-456.
- [6] 王鑫,黄忠义. 网络资源中基于 K-Means 聚类的个性化推荐[J]. 北京邮电大学学报, 2014, 37(S1): 120-124.
Wang Xin, Huang Zhongyi. Network resource personalized recommendation based on K-Means clustering [J]. Journal of Beijing University of Posts and Telecommunications, 2014, 37(S1): 120-124.
- [7] Hong Jongyi, Suh E H, Kim J, et al. Contextaware system for proactive personalized service based on context history [J]. Expert Systems with Applications, 2009, 36(4): 7448-7457.
- [8] Pessemier T D, Deryckere T, Martens L. Extending the Bayesian classifier to a context-aware recommender system for mobile devices [C] // Internet and Web Applications and Services (ICIW), 2010 Fifth International Conference on IEEE. Barcelona, Spain: IEEE Press, 2010: 242-247.
- [9] Shahabi C, Chen Yishin. An adaptive recommendation system without explicit acquisition of user relevance feedback [J]. Distributed and Parallel Databases, 2003, 14(2): 173-192.
- [10] 鄂海红,宋美娜,李川,等. 结合时间上下文挖掘学习兴趣的协同过滤推荐算法[J]. 北京邮电大学学报, 2014, 37(6): 49-53.
E Haihong, Song Meina, Li Chuan, et al. A collaborative filtering recommendation algorithm with time context for learning interest mining [J]. Journal of Beijing University of Posts and Telecommunications, 2014, 37(6): 49-53.
- [11] 王立才,孟祥武,张玉洁. 上下文感知推荐系统[J]. 软件学报, 2012, 23(1): 1-20.
Wang Licai, Meng Xiangwu, Zhang Yujie. Context-aware recommender systems [J]. Journal of Software, 2012, 23(1): 1-20.
- [12] 程学旗,靳小龙,王元卓,等. 大数据系统和分析技术综述[J]. 软件学报, 2014, 25(9): 1889-1908.
Cheng Xueqi, Jin Xiaolong, Wang Yuanzhuo, et al. Survey on big data system and analytic technology [J]. Journal of Software, 2014, 25(9): 1889-1908.
- [13] 何清,李宁,罗文娟,等. 大数据环境下的机器学习算法综述[J]. 模式识别与人工智能, 2014, 27(4): 327-336.
He Qing, Li Ning, Luo Wenjuan, et al. A survey of machine learning algorithms for big data [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(4): 327-336.
- [14] 印鉴,王智圣,李琪,等. 基于大规模隐式反馈的个性化推荐[J]. 软件学报, 2014, 25(9): 1953-1966.
Yin Jian, Wang Zhisheng, Li Qi, et al. Personalized recommendation based on largescale implicit feedback

- [J]. *Journal of Software*, 2014, 25 (9): 1953-1966.
- [15] Yang Diyi, Chen Tianqi, Zhang Weinan, et al. Local implicit feedback mining for music recommendation [C]// the 6th ACM Conference on Recommender Systems, 2012 ACM. Dublin: ACM Press, 2012: 91-98.
- [16] Rafailidis D, Nanopoulos A. Modeling the dynamics of user preferences in coupled tensor factorization [C]// the 8th ACM Conference on Recommender Systems, 2014 ACM. Silicon Valley: ACM Press, 2014: 321-324.
- [17] Oh K J, Lee W J, Lim C G, et al. Personalized news recommendation using classified keywords to capture user preference [C]// 16th Advanced Communication Technology (ICACT), 2014 International Conference on IEEE. South Korea: IEEE Press, 2014: 1283-1287.
- [18] Takács G, Pilászy I, Németh B, et al. Scalable collaborative filtering approaches for large recommender systems [J]. *The Journal of Machine Learning Research*, 2009, 10(12): 623-656.
- [19] Bhagat S, Weinsberg U, Ioannidis S, et al. Recommending with an agenda: active learning of private attributes using matrix factorization [J]. *ArXiv Preprint ArXiv*, 2013: 1311-1321.
- [20] 丁伟峰, 郑小林, 陈德人. 基于 PureSVD 模型的协同过滤主动采样 [J]. *北京邮电大学学报*, 2013, 36 (4): 23-26.
Ding Weifeng, Zheng Xiaolin, Chen Deren. Active sampling based on PureSVD model for collaborative filtering [J]. *Journal of Beijing University of Posts and Telecommunications*, 2013, 36(4): 23-26.
- [21] Dror G, Koenigstein N, Koren Y, et al. The Yahoo! music dataset and KDD-Cup'11 [C]// 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2011 ACM SIGKDD. San Diego, CA: ACM Press, 2012: 8-18.
- [22] 涂丹丹, 舒承椿, 余海燕. 基于联合概率矩阵分解的上下文广告推荐算法 [J]. *软件学报*, 2013, 24 (3): 454-464.
Tu Dandan, Shu Chengchun, Yu Haiyan. Using unified probabilistic matrix factorization for contextual advertisement recommendation [J]. *Journal of Software*, 2013, 24(3): 454-464.
- [23] Bauer J, Nanopoulos A. A framework for matrix factorization based on general distributions [C]// Proceedings of the 8th ACM Conference on Recommender Systems. Silicon Valley: ACM Press, 2014: 249-256.
- [24] Cheng Chen, Xia Fen, Zhang Tong, et al. Gradient boosting factorization machines [C]// Proceedings of the 8th ACM Conference on Recommender Systems. Silicon Valley: ACM Press, 2014: 265-272.
- [25] Pálóvics R, Benczúr A A, Kocsis L, et al. Exploiting temporal influence in online recommendation [C]// Proceedings of the 8th ACM Conference on Recommender Systems. Silicon Valley: ACM Press, 2014: 273-280.
- [26] Schelter S, Boden C, Schenck M, et al. Distributed matrix factorization with mapreduce using a series of broadcast-joins [C]// Proceedings of the 7th ACM Conference on Recommender Systems. Hong Kong: ACM Press, 2013: 281-284.
- [27] Diaz-Aviles E, Drumond L, Schmidt-Thieme L, et al. Real-time top-*n* recommendation in social streams [C]// Proceedings of the 6th ACM Conference on Recommender Systems. Dublin: ACM Press, 2012: 59-66.
- [28] Ge Yong, Xiong Hui, Tuzhilin A, et al. Costaware collaborative filtering for travel tour recommendations [J]. *ACM Transactions on Information Systems (TOIS)*, 2014, 32(1): 4-28.
- [29] Golub G, Kahan K. Calculating the singular values and pseudo-inverse of a matrix [J]. *Journal of the Society for Industrial and Applied Mathematics*, 1965, 2(2): 205-224.
- [30] Lee D D, Seung H. Algorithms for non-negative matrix factorization [C]// 13th Advances in Neural Information Processing Systems, NIPS 2000. Denver, USA: MIT Press, 2000: 556-562.
- [31] Zheng V W, Zheng Yu, Xie Xing, et al. Towards mobile intelligence: learning from GPS history data for collaborative recommendation [J]. *Artificial Intelligence*, 2012, 184(6): 17-37.
- [32] Symeonidis P, Papadimitriou A, Manolopoulos Y, et al. Geo-social recommendations based on incremental tensor reduction and local path traversal [C]// Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks. Chicago, USA: ACM Press, 2011: 89-96.
- [33] Koren Y. Collaborative filtering with temporal dynamics [J]. *Communications of the ACM*, 2010, 53(4): 89-97.
- [34] Salakhutdinov R, Mnih A. Probabilistic matrix factorization [C]// 20th Advances in Neural Information Processing Systems, NIPS 2007. Vancouver, Canada: MIT Press, 2007, 20(3): 432-451.
- [35] 朱郁筱, 吕琳媛. 推荐系统评价指标综述 [J]. *电子科技大学学报*, 2012, 41(2): 163-175.

- Zhu Yuxiao, Lü Linyuan. Evaluation metrics for recommender systems [J]. Journal of University of Electronic Science and Technology of China, 2012, 41(2): 163-175.
- [36] Isaacman S, Ioannidis S, Chaintreau A, et al. Distributed rating prediction in user generated content streams [C]//5th ACM Conference on Recommender Systems, 2011 ACM. Chicago: ACM Press, 2011: 69-76.
- [37] Bachrach Y, Finkelstein Y, Gilad-Bachrach R, et al. Speeding up the Xbox recommender system using a Euclidean transformation for inner-product spaces [C]//8th ACM Conference on Recommender Systems, 2014 ACM. Silicon Valley: ACM Press, 2014: 250-257.
- [38] Takács G, Tikk D. Alternating least squares for personalized ranking [C]//6th ACM Conference on Recommender Systems, 2012 ACM. Dublin: ACM Press, 2012: 83-90.
- [39] Zhuang Yong, Chin Weisheng, Juan Yuchin, et al. A fast parallel SGD for matrix factorization in shared memory systems [C]//7th ACM Conference on Recommender Systems, 2013 ACM. Hong Kong: ACM Press, 2013: 249-256.
- [40] Recht B, Ré C. Parallel stochastic gradient algorithms for large-scale matrix completion [J]. Mathematical Programming Computation, 2013, 5(2): 201-226.
- [41] Gemulla R, Nijkamp E, Haas P J, et al. Largescale matrix factorization with distributed stochastic gradient descent [C]//17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011 ACM. San Diego: ACM Press, 2011: 69-77.
- [42] Petroni F, Quercioni L. GASGD: stochastic gradient descent for distributed asynchronous matrix completion via graph partitioning [C]//the 8th ACM Conference on Recommender Systems. Silicon Valley: ACM, 2014: 241-248.
- [43] Fazeli S, Loni B, Bellogin A, et al. Implicit vs explicit trust in social matrix factorization [C]//8th ACM Conference on Recommender Systems, 2014 ACM. Silicon Valley: ACM Press, 2014: 317-320.
- [44] McGinty L, Reilly J. On the evolution of critiquing recommenders [M]. US: Springer, 2011: 419-453.
- [45] Zhao Xin Wayne, Guo Yanwei, He Yulan, et al. We know what you want to buy: a demographic-based system for product recommendation on microblogs [C]//20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014 ACM. New York: ACM Press, 2014: 1935-1944.
- [46] Töschner A, Jahrer M. Collaborative filtering ensemble for ranking [J]. Journal of Machine Learning Research W&CP, 2012, 18: 61-74.
- [47] Guo Guibing. Integrating trust and similarity to ameliorate the data sparsity and cold start for recommender systems [C]//7th ACM Conference on Recommender Systems, 2013 ACM. Hong Kong: ACM Press, 2013: 451-454.
- [48] Zheng Yu, Xie Xing, Ma Weiying. GeoLife: a collaborative social networking service among user, location and trajectory [J]. IEEE Data Eng. Bull, 2010, 33(2): 32-39.
- [49] Zheng Yu, Zhang Lizhu, Xie Xing, et al. Mining interesting locations and travel sequences from GPS trajectories [C]//18th International Conference on World Wide Web, 2009 ACM. Spain: ACM Press, 2009: 791-800.
- [50] Zheng Yu, Li Quannan, Chen Yukun, et al. Understanding mobility based on GPS data [C]//10th International Conference on Ubiquitous Computing. Seoul, 2008 ACM. South Korea: ACM Press, 2008: 312-321.
- [51] Hu Bo, Ester M. Spatial topic modeling in online social media for location recommendation [C]//7th ACM Conference on Recommender Systems, 2013 ACM. Hong Kong: ACM Press, 2013: 25-32.
- [52] Yang Xiwang, Steck H, Guo Yang, et al. On top- k recommendation using social networks [C]//6th ACM Conference on Recommender Systems, 2012 ACM. Dublin: ACM Press, 2012: 67-74.
- [53] Zhang Kunpeng, Ouksel A, Fan Shaokun, et al. Scalable audience targeted models for brand advertising on social networks [C]//8th ACM Conference on Recommender Systems, 2014 ACM. Silicon Valley: ACM Press, 2014: 341-344.
- [54] Wang Jing, Zhao Hui. Social group recommendation using topic models [J]. Journal of Chemical and Pharmaceutical Research, 2014, 6(7): 679-684.
- [55] Purushotham S, Kuo C C J, Shahabdeen J, et al. Collaborative group-activity recommendation in location-based social networks [C]//3rd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, 2014 ACM. Dallas: ACM Press, 2014: 8-15.
- [56] Naamani-Dery L, Kalech M, Rokach L, et al. Preference elicitation for narrowing the recommended list for groups [C]//8th ACM Conference on Recommender Systems, 2014 ACM. Silicon Valley: ACM Press, 2014: 333-336.

- [57] Ferreira C R L , Traina J C , Machado T A J , et al. Clustering very large multi-dimensional datasets with mapreduce [C] // 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining , 2011 ACM. San Diego: ACM Press , 2011: 690-698.
- [58] Yu Yuantse , Huang Chungming , Lee Yuntz. An intelligent touring system based on mobile social network and cloud computing for travel recommendation [C] // 28th International Conference on Advanced Information Networking and Applications Workshops (AINA) , 2014 IEEE. Victoria , Canada: IEEE Press , 2014: 19-24.
- [59] Katkar V D , Kulkarni S V. A novel parallel implementation of naive Bayesian classifier for big data [C] // 2013 International Conference on Green Computing , Communication and Conservation of Energy (ICGCE) , 2013 IEEE. India: IEEE Press , 2013: 847-852.
- [60] Walunj S G , Sadafale K. An online recommendation system for e-commerce based on apache mahout framework [C] // 2013 Annual Conference on Computers and People Research , 2013 ACM. Cincinnati: ACM Press , 2013: 153-158.
- [61] Xu Shengwu , Xia Zhengyou. Hot news recommendation system across heterogenous websites using Hadoop [C] // Advanced Materials Research , 2014 TTP. Switzerland: TTP Press , 2014 , 989: 4704-4707.
- [62] Gong Songjie , Xu Jiongbo. Electronic commerce personalized recommendation model under cloud computing environment [J]. Applied Mechanics and Materials , 2014 , 513(2) : 639-642.
- [63] Franco-Arcega A , Carrasco-Ochoa J A , Sanchez-Diaz G , et al. Building fast decision trees from large training sets [J]. Intelligent Data Analysis , 2012 , 16 (4) : 649-664.
- [64] Xue Zhenghua , Shen Geng , Li Jianhui , et al. ComPres-sion-aware I/O performance analysis for big data clustering [C] // 1st International Workshop on Big Data , Streams and Heterogeneous Source Mining: Algorithms , Systems , Programming Models and Applications , 2012 ACM. Beijing , China: ACM Press , 2012: 45-52.
- [65] Kalyanakrishnan S , Singh D , Kant R. On building decision trees from large-scale data in applications of on-line advertising [C] // 23rd ACM International Conference on Information and Knowledge Management (CIKM) , 2014 ACM. Shanghai: ACM Press , 2014: 669-678.
- [66] Yang Hang , Fong Simon. Incrementally optimized decision tree for noisy big data [C] // 1st International Workshop on Big Data , Streams and Heterogeneous Source Mining: Algorithms , Systems , Programming Models and Applications , 2012 ACM. Beijing: ACM Press , 2012: 36-44.
- [67] Lu E H C , Chen Chinyu , Tseng V S. Personalized trip recommendation with multiple constraints by mining user check-in behaviors [C] // 20th International Conference on Advances in Geographic Information Systems , 2012 ACM. California: ACM Press , 2012: 209-218.
- [68] Lee W J , Oh K J , Lim C G , et al. User profile extraction from Twitter for personalized news recommendation [C] // 16th International Conference on Advanced Communication Technology , 2014 IEEE. Korea: IEEE Press , 2014: 779-783.
- [69] Yi Xing , Hong Liangjie , Zhong Erheng , et al. Beyond clicks: dwell time for personalization [C] // 8th ACM Conference on Recommender Systems , 2014 ACM. Silicon Valley: ACM Press , 2014: 113-120.
- [70] Vargas S , Castells P. Rank and relevance in novelty and diversity metrics for recommender systems [C] // 5th ACM Conference on Recommender Systems , 2011 ACM. Chicago: ACM Press , 2011: 109-116.
- [71] Mayer-Schönberger V , Cukier K. Big data: arevolution that will transform how we live , work , and think [M]. US: Houghton Mifflin Harcourt , 2013: 1-261.
- [72] Vargas S , Castells P. Improving sales diversity by recommending users to items [C] // 8th ACM Conference on Recommender Systems , 2014 ACM. Silicon Valley: ACM Press , 2014: 145-152.
- [73] Garcin F , Faltings B , Donatsch O , et al. Offline and online evaluation of news recommender systems at swissinfo. ch [C] // 8th ACM Conference on Recommender Systems , 2014 ACM. Silicon Valley: ACM Press , 2014: 169-176.
- [74] Becchetti L , Bergamini L , Colesanti U M , et al. A lightweight privacy preserving SMS-based recommendation system for mobile users [J]. Knowledge and Information Systems , 2014 , 40(1) : 49-77.
- [75] 霍峥 , 孟小峰. 轨迹隐私保护技术研究 [J]. 计算机学报 , 2011 , 34(10) : 1820-1830.
- 霍峥 , 孟小峰. A survey of trajectory privacy-preserving techniques [J]. Chinese Journal of Computer , 2011 , 34(10) : 1820-1830.