

# 基于 Hadoop 的企业知识管理系统<sup>①</sup>

李 锋, 贾茂想, 涂如男

(江苏科技大学 经济管理学院, 镇江 212003)

通讯作者: 李 锋, E-mail: [lf@just.edu.cn](mailto:lf@just.edu.cn)

**摘 要:** 基于 Hadoop 的知识管理系统可以增强企业海量知识存储与处理的能力, 实现面向管理决策的知识发现。通过研究 Hadoop 平台及其特点, 分析了基于 Hadoop 的知识管理系统的优势, 构建了基于 Hadoop 的企业知识管理框架模型, 并设计了基于 Hadoop 的知识管理系统的逻辑架构, 最后利用基于 MapReduce 的协同过滤算法对知识推荐模块进行了详细设计。应用 Hadoop 等大数据技术及推荐系统技术构建知识管理系统既能实现知识管理的个性化和智能化, 又能满足企业知识管理系统的扩展性和经济性的需求。

**关键词:** 知识管理系统; HDFS; MapReduce; 协同过滤

引用格式: 李锋, 贾茂想, 涂如男. 基于 Hadoop 的企业知识管理系统. 计算机系统应用, 2018, 27(8): 63-69. <http://www.c-s-a.org.cn/1003-3254/6485.html>

## Enterprise Knowledge Management System Based on Hadoop

LI Feng, JIA Mao-Xiang, TU Ru-Nan

(School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

**Abstract:** Hadoop-based knowledge management systems can enhance corporate ability to store and process massive knowledge and realize knowledge discovery oriented to management decisions. Through studies on Hadoop platform and its characteristics, the advantages of Hadoop-based knowledge management systems were analyzed, a Hadoop-based corporate knowledge management framework model was built, and a logical framework was designed for Hadoop-based knowledge management systems. Eventually, detailed design was done on the knowledge recommendation module by applying MapReduce-based collaborative filtering algorithm. With knowledge management systems being built based on Hadoop and other big data technologies and recommendation system technologies, knowledge management can be personalized and intelligentized, and the scalability and economic demands of corporate knowledge management systems can be met as well.

**Key words:** knowledge management system; HDFS; MapReduce; collaborative filtering

大数据时代的到来, 大规模生产、分析和应用数据成为了新的发展趋势。在企业长期的生产和运营过程中, 产生了大量的有价值的的数据, 对于这些数据的收集和分析, 可以提升企业管理并获得竞争优势<sup>[1]</sup>。随着企业产生的半结构化和非结构化数据逐年增加, 如何

存储和分析数据、如何挖掘和共享知识, 成为了面临的主要问题。通过提升内部知识管理能力, 知识管理系统有助于建立开放和协作的生态系统、利用内外部知识流, 从而提高创新能力<sup>[2]</sup>。因此, 许多企业也开始建设知识管理系统, 以管理组织学习和业务知识<sup>[3]</sup>。然而, 对

① 基金项目: 国家社科基金青年项目(13CGL069); 教育部人文社科研究青年基金(13YJC880098); 国家自然科学基金面上项目(7177116, 71471091, 71271119)

Foundation item: Youth Project of National Social Science Foundation (13CGL069); Youth Project of Humanities Social Science Foundation of Ministry of Education of China (13YJC880098); The General Program of National Natural Science Foundation of China (7177116, 71471091, 71271119)

收稿时间: 2017-11-28; 修改时间: 2017-12-21; 采用时间: 2018-01-19; csa 在线出版时间: 2018-07-28

于 PB 级别的非结构化数据而言, 现有的知识管理系统仅靠传统的数据库技术已不能满足用户多样化的需求, 还需要结合分布式计算和存储技术. 解决大数据环境下的知识存储、知识共享、知识挖掘、知识检索与推荐的问题, 企业需要构建一种基于大数据技术的知识管理系统, 从而支持知识创新和知识应用.

### 1 国内外研究现状

知识管理系统 (Knowledge Management System, KMS) 指的是一系列应用于管理组织知识的信息系统. 它是一种建立在 IT 系统之上, 用来支持和强化组织中知识创造、存储、传播和应用过程的信息系统<sup>[4]</sup>. 而企业知识管理系统 (Enterprise Knowledge Management System, EKMS) 有狭义和广义之分, 广义的 EKMS 将企业整体看作一个知识处理系统, 它是企业模型的抽象和企业知识管理方面的视图; 狭义的 EKMS 是指支持企业知识管理过程的信息系统<sup>[5]</sup>. 简而言之, 知识管理系统是为了实现知识创新和知识共享并对知识进行管理的一种工具. 近年来, 关于知识管理系统的研究取得了丰厚的理论成果和实践经验, 其中文献<sup>[6]</sup>从快速应对复杂变化的应用环境、支持知识管理生命周期过程、满足用户个性化需求的角度, 提出了一种敏捷知识管理系统. 文献<sup>[7]</sup>从用户行为动机的角度出发, 将用户创新的理念融入到知识管理系统中, 探讨了基于用户创新模式知识管理系统. 文献<sup>[8]</sup>构建了基于本体的知识管理系统框架并实现了部分检索模块. 文献<sup>[9]</sup>结合云计算的战略理念, 研究了基于云计算的企业 KMS 再造的体系. 文献<sup>[10]</sup>结合 Web 2.0 的特点, 构建

了面向服务的知识管理系统四层模型.

目前大多数的研究主要是从知识管理的理念出发, 进行相关理论以及模型的研究. 而对于大数据技术在知识管理系统中的应用研究较少且不够深入, 论文从大数据时代处理数据的思维入手, 突出用户行为模式和人机交互的因素, 着眼于海量知识的检索 (知识需求为导向)、个性化知识推荐 (知识浏览为导向)、虚拟社区知识共享 (知识交流为导向)、知识挖掘 (知识创新为导向) 以及情报服务 (知识应用为导向) 等功能, 利用大数据相关技术, 构建基于 Hadoop 的企业知识管理系统 (Hadoop-based Enterprise Knowledge Management System, H\_EKMS). 从而满足用户个性化和智能化的需求, 达到知识转移、应用、共享与创新的目的.

### 2 H\_EKMS 的优势

#### 2.1 Hadoop 体系结构

Hadoop 的核心是 HDFS 和 MapReduce. HDFS (Hadoop Distributed File System) 是一个分布式文件系统, HDFS 以流式数据访问模式来存储超大文件, 运行于商用硬件集群上<sup>[11]</sup>. 通过实现冗余来保障数据的安全存储, 具有高容错性的特点. MapReduce 是一种建立在 HDFS 上的分布式计算框架, 可用于数据处理的编程模型, 具有良好的扩展性和高容错性<sup>[12]</sup>. MapReduce 任务过程分为 Map 阶段和 Reduce 阶段, 每个阶段都是以键值对 (key/value) 作为输入和输出 (见图 1), 其一般形式为:

$$\text{Map: } (K1, V1) \rightarrow \text{list}(K2, V2)$$

$$\text{Reduce: } (K2, \text{list}(V2)) \rightarrow \text{list}(K3, V3)$$

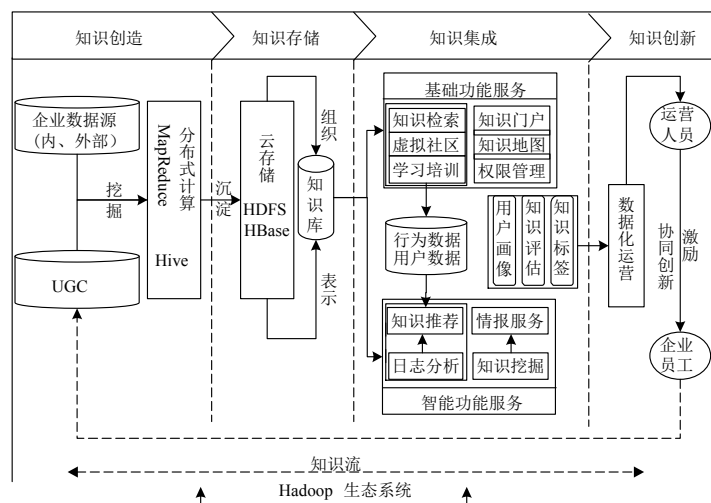


图 1 基于 Hadoop 的企业知识管理框架模型

## 2.2 H\_EKMS 的优势

随着 EKMS 的访问量与知识量的增加,传统的 EKMS 已无法满足用户的需求.另外,对于 EKMS 而言,长期积累的数据与知识是企业非常宝贵的财富,如何应用这些数据和知识来提升企业管理是未来知识管理系统应当解决的问题.大数据与云计算技术的发展改变了企业的知识管理模式和用户的行爲观念,对 EKMS 的功能和架构的设计提供了新的思路.而 Hadoop 以其开源、适合处理海量数据集的特点成为了研究大数据的重要平台,满足了面向企业的信息系统的扩展性和经济性的需求.与传统的 EKMS 相比, H\_EKMS 具有如下几个优势:

### (1) 满足用户个性化的需求

大数据背景下信息过载十分严重,而每个用户的精力有限,如何在繁杂的信息中找到自己感兴趣的内容是亟待的问题.然而传统的 KMS 用户交互性差,不能够为用户提供更加个性化的服务,造成了 KMS 使用率不高的现象,不能够达到企业构建 KMS 的目的. H\_EKMS 利用推荐系统技术解决信息过载的问题,能够为用户提供个性化知识服务,进一步增加用户粘性和增强用户使用 KMS 活跃度,从而真正发挥 KMS 作用.

### (2) 提升知识存储与分析的性能

伴随着数据的爆炸式增长,除了数据规模的增大,数据的复杂度也在增加(半结构化和非结构化数据增多),其多样性、价值密度低和超大规模的特点与传统的数据库和数据仓库的数据存在着很大的差异,而传统的 EKMS 实现的主要是在线事务处理(OLTP),已不能满足大数据处理的需求.由于 H\_EKMS 底层采用的是 HDFS 进行知识的存储,知识文件被分割成不同的块(block)存储在不同的数据节点(datanode)之上,每个块的副本数可以根据实际需求进行设定,从而实现冗余存储,保证数据存储的安全性.后台采用的是大规模的部署 Hadoop 的服务器组成的集群,从而提供了海量的数据存储空间和高性能的计算能力. Hadoop 能够在集群的每台机器之间动态地移动数据,将作业分散到不同节点执行,并保证各个节点的动态平衡,因此处理速度非常快.另外 MapReduce 具有数据本地化的特性,从而减少数据传输的时间开销,实现高性能的计算. MapReduce 编程模型支持多数统计分析和机器学习的算法,采用分布式并行处理的策略,加快了数据处理的速度和效率,而 Mahout 构建了一个快速创建可伸

缩、高性能的机器学习应用程序的环境,对于大规模数据挖掘有着良好的支持.目前 MapReduce 应用的计算场景有很多,主要包括流量统计、用户行为分析、推荐引擎、数据挖掘分类器等.

### (3) 实现面向管理决策的知识发现

知识发现是提取有用信息和隐藏信息的过程,从大规模数据中发现潜在的、有用的知识,旨在服务于管理决策.传统的知识管理系统更多地考虑对显性知识的管理,而这些显性知识只是冰山一隅.在大数据时代,知识管理的范畴扩大了,基于 Hadoop 的 EKMS 不局限于现有知识的管理,而是更加关注商务智能的过程,将知识的采集的来源丰富化(企业内部大数据和企业外部大数据),通过知识发现解决企业数据充足而知识匮乏的窘境,把大量的原始数据转换成有价值的、便于利用的知识,从而帮助企业在激烈的市场竞争中获得战略性优势.

### (4) 提高系统的可扩展性

Hadoop 提供了一个高度可扩展的共享存储和分析平台,它可以存储和分发横跨数百个并行操作的廉价的服务器数据集群.传统的 EKMS 主要依赖于关系型数据库系统,对于处理大数据而言,以纵向扩展来打造超级计算机显然是不可取的,其扩展性不够友好. Hadoop 则实现横向扩展,能够在成百上千的数据节点上运行应用程序,具有良好的扩展性和弹性.企业可以根据自身的规模以及用户访问量等实际情况部署集群,可以灵活地增加和减少集群节点,且不会影响 KMS 的运转.

## 3 H\_EKMS 构建

### 3.1 基于 Hadoop 的企业知识管理框架模型

本文以数据化运营为视角,依托 Hadoop 生态系统的相关大数据技术,构建了基于 Hadoop 的企业知识管理框架模型,该模型由人、技术、知识源和功能服务构成(见图 1),突出“数据驱动管理创新、管理与技术并重”的理念,整个模型框架建立在 Hadoop 生态系统之上,对于数据的存储、分析有着良好的支持.以下主要从知识创造、知识存储、知识集成和知识创新 4 个方面进行阐述.

#### (1) 知识创造

在数据大爆炸时代,全球的数据量呈现不断增长的趋势.大数据背景下知识管理生命周期的各个环节变得复杂了,其中知识的采集不在局限在企业内部,而



是延伸到了企业的外部. 企业的外部大数据包括产业链数据、互联网数据、竞争情报数据和 UGC 数据, 这些数据不仅蕴含着巨大的商业价值, 而且还是企业知识发现和知识积累的重要基础. 但是这些数据具有量大、多样和价值密度低等特性, 依靠传统的数据库技术实现存储和在线事务处理 (OLTP) 并不能够满足企业需求, 引入 MapReduce 编程模型和 Hive 数据仓库分别实现大数据的分布式计算与在线分析处理 (OLAP), 可以轻松达到企业大数据挖掘的目的, 进而为知识的储备与应用奠定基础.

### (2) 知识存储

知识创造阶段的输出是经过分析处理之后形成的具有一定价值和规范的知识与数据, 但是仍具有多样和量大的特性. 一方面随着企业规模的增大和持续化运转, 这些数据和知识的积累量增大的速度会越来越快, 另一方面对于存储的安全性和操作的便捷性要求也越来越高. 云计算为实现云存储提供了技术支撑, 通过 HDFS 实现分布式存储, 既满足了知识量大的需求又满足了存储的安全性需求, 而 HBase 是构建在 HDFS 之上的面向列的分布式数据库, 支持实时访问超大规模数据集, 为企业的大数据应用提供了较好的解决方案.

### (3) 知识集成

知识集成是通过 H\_EKMS 实现的, 是知识应用和知识创新的基础. 主要包括基础功能和智能功能, 其中基础功能主要包括: 知识门户、知识地图、权限管理、知识检索、虚拟社区、学习培训等. 知识门户是统一展示知识的平台, 是企业或组织内部具体用户与知识管理系统软件进行人机交互的接口, 用于知识的发布与共享, 并且集成了知识管理的各个功能模块, 使得企业各部门员工之间知识交流和信息共享更加的流畅. 知识地图是具有索引的功能, 将显性的知识以图表、文字等形式表现出来, 能够很方便地按照知识地图的指引找到所需要的知识. 权限管理是为了系统的安全考虑, 将系统的用户划分成不同的级别进行权限控制. 知识检索是知识管理系统的最基础的功能之一, H\_EKMS 是建立在大数据技术基础之上的, 能够快速精准地响应用户的检索操作, 从而可以提升用户体验. 虚拟社区是知识管理系统用户进行知识交流和知识共享的重要平台, 用户可以通过虚拟社区平台回答、提问、推荐和贡献知识, 真正使隐性的经验类知识得到挖掘与传播, 有利于促进形成知识共享、知识创新的文化氛围. 学习培训是系统用户进行知识补给和推动

内部知识传播的平台, 提供了学习的随时随地性, 既能够增强员工的基本技能, 又能够为企业节省大量的人力和财力.

智能功能是 H\_EKMS 有别于其他传统的 EKMS 的特色功能, 主要包括知识推荐、日志分析、知识挖掘和情报服务. 知识推荐是综合运用机器学习相关的算法, 通过分析基础功能服务模块的知识检索、虚拟社区、学习培训所产生的行为数据、用户生成内容 (User Genetated Content, UGC) 以及知识元数据, 挖掘用户的需求、兴趣以及行为模式, 从而实现个性化知识推荐. 伴随着知识管理系统的运转会产生大量的日志数据, 这些数据既包括系统运行过程中的运行状态, 又包括用户的一些访问数据, 日志分析功能模块可以对这些数据进行分析处理, 比如通过用户搜索的关键词日志分析可以得出用户的搜索偏好, 通过分析系统的错误信息日志可以进行异常报警, 从而可以了解系统的安全状况以及用户使用系统的行为模式, 进而为知识推荐奠定基础. 竞争情报已成为企业在长期发展过程中的核心竞争力之一, 尤其是在大数据时代, 企业若能够从海量繁杂的信息资源中挖掘出情报, 便会率先掌握和预测商机, 从而可以帮助企业进行科学决策和战略制定. 知识挖掘功能模块可以从数据资源中提炼知识, 促使隐性知识到显示知识的转换. 知识挖掘是知识管理的重要环节, 是知识创新的基础, 从海量的数据中提取结构化的知识, 将其运用到企业的生产和管理的过程中, 可以提升企业的管理效率. 另外知识挖掘可以实现搜索引擎的智能化, 通过知识挖掘构建知识图谱, 促使知识更加系统化、全面化以及知识之间的相互联系, 可以使用户的搜索更加便利. 最后, 知识挖掘也可以为企业的情报服务提供参考信息, 形成一个良性循环.

### (4) 知识创新

通过对用户数据及其行为数据分析建立员工用户画像, 并对知识价值、知识量及知识标签进行有效地评估和展示, 一方面有利于企业知识管理人员及时了解实施知识管理的成效, 并对实施知识管理的绩效评价提供参考; 另一方面有利于企业管理人员根据用户特征有的放矢地进行激励, 并充分利用 WEB 2.0 技术促进用户生成内容, 进而促进知识转移与共享, 进一步提升企业创新能力.

## 3.2 H\_EKMS 设计

设计结合以上对知识管理框架模型的分析, 设计

了 H\_EKMS 的逻辑架构, 见图 2. 整个系统的架构可分为应用层和技术支撑层两大逻辑单元. 具体分析如下:

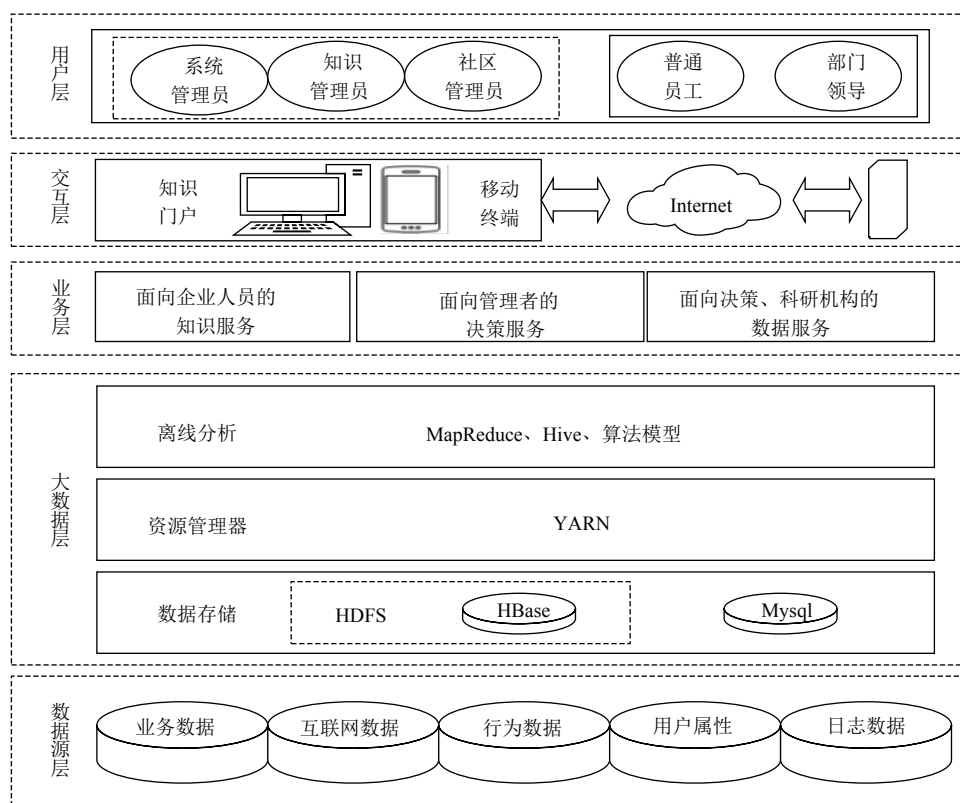


图 2 H\_EKMS 逻辑架构

### (1) 技术支撑层

技术支撑层是 H\_EKMS 的保障, 主要由数据源层和大数据层组成. 其中数据的来源主要分为企业外部大数据和企业内部大数据, 外部大数据包括互联网大数据和行为数据, 企业内部大数据主要包括业务数据、用户属性数据和系统运转过程中产生的日志数据. 这些数据蕴含了大量的客户信息、企业竞争性情报和产业链信息, 对这些数据的深入挖掘, 一方面可以为企业管理决策提供有用的知识储备, 帮助企业精准定位商务价值和市场空间, 另一方面可以促进产业链的协同创新. 大数据层为整个系统提供分布式存储和分布式计算的技术支持, 其中数据存储是构建在分布式文件系统之上的, 传统的关系型数据库能够提供比较完整的 ACID 操作, 对于查询、连接以及索引有着很好的支持, 功能非常强大. 与关系型数据库相比, HBase 更适合海量数据的存储和处理以及高并发访问, 并且具有很好的扩展性. 资源管理器 YARN 可以支撑多种计算框架, 为以后的系统扩展提供了很大的便捷. 离线

分析处理主要是通过 MapReduce 编程模型和 Hive 数  
据仓库实现分布式计算.

### (2) 应用层

就管理层面而言, 业务层主要分为三大板块, 主要分为面向企业人员的知识服务、面向管理者的决策服务及面向决策和科研机构的数据服务. H\_EKMS 秉承技术与管理并重的理念, 有效解决目前企业知识管理形同虚设的窘境, 更加强调知识应用的重要性, 通过为不同的使用人员提供不同的服务, 可以促进知识应用到实处, 比如通过为科研机构提供数据服务, 促进产学研协同创新, 可以为企业的发展提供良好的思路, 从而可以加快企业创新. 就技术层面而言, 业务层采用 B/S 架构模式, 选用 SSH 框架进行开发, 实现分层处理. 从用户的角度出发, 根据不同的权限设定, 系统共分了管理员和普通用户两种角色, 管理员包括系统管理员、知识管理员和社区管理员, 普通用户包括部门领导和普通员工. 其中系统管理员负责知识维护、公告以及问卷的发布与删除以及知识统计等工作, 知识管理员

具有新建、修改和删除的权限,社区管理员主要负责论坛板块的管理.普通用户是系统的主要受众群体,根据不同的职能层级,划分不同的权限.交互层提供了用户使用系统的接口,用户可以通过PC端或者移动终端的浏览器来访问H\_EKMS,真正实现随时随地的知识共享与应用.应用层是系统的功能模块,也是系统的核心组件,具有灵活的扩展性.

### 3.3 H\_EKMS 知识推荐模块设计与实现

#### (1) H\_EKMS 知识推荐模块设计

EKMS 不同于电商网站,由于电商网站的用户群体非常庞大,所以推荐系统一般采用基于物品的协同过滤算法(ItemCF),这样可以进一步提高计算的速度.考虑到企业员工人数相对固定而知识量逐渐增加的问题,本文选择基于用户的协同过滤(UserCF)方法作为系统的推荐算法,主要使用企业员工与系统交互时所产生的反馈数据完成整个推荐过程.基于用户的协同过滤算法基本思想是通过计算用户的相似度来找到其近邻,然后根据近邻对其未评分的物品的评分进行预测,最后将预测评分较高的物品推荐给当前用户.基本推荐流程见图3.

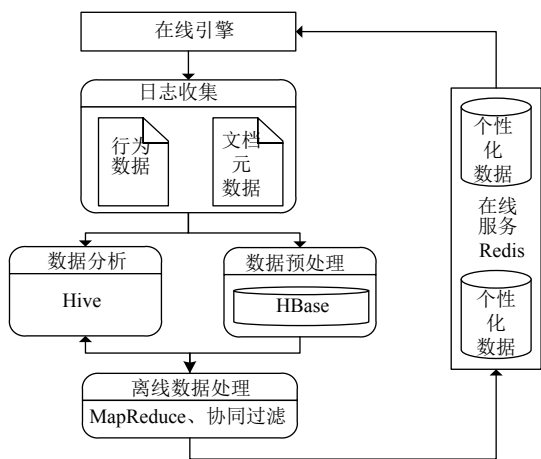


图3 基于Hadoop的EKMS知识推荐流程

首先,计算用户的评分,得到User-Item评分矩阵,由于并不是所有用户都会对所阅读的文档进行评分,所以得到的是一个稀疏度较高矩阵.

其次,根据得到User-Item评分矩阵,通过User<sub>i</sub>和User<sub>j</sub>对所有文档的评分来计算用户的相似度,获得当前用户的近邻.常用的相似度计算方法主要包括Jaccard公式,余弦相似度,皮尔逊相关系数等<sup>[13]</sup>.本文选择余弦相似度来计算用户的相似性,计算公式为:

$$sim(x,y) = \frac{\sum_{i \in I_{xy}} r_{xi} r_{yi}}{\sqrt{\sum_{i \in I_x} r_{xi}^2} \sqrt{\sum_{j \in I_y} r_{yj}^2}}$$

其中,  $I_{xy}$  表示同时被用户  $x$  和用户  $y$  都评分的文档集合,  $r_{xi}$  表示用户  $x$  对文档  $i$  的评分,  $r_{yi}$  表示用户  $y$  对于文档  $i$  的评分,  $I_x$  和  $I_y$  分别表示用户  $x$  和用户  $y$  的所评分文档集合.

最后,计算推荐结果.由上一步可以得到当前用户的近邻,根据近邻对当前用户未评分的文档的评分进行加权求和即可得到当前用户对文档的预测评分.计算公式为:

$$pred_{xi} = \bar{r}_x + \frac{\sum_{y \in N_u} sim(x,y)(r_{yi} - \bar{r}_y)}{\sum_{y \in N_u} |sim(x,y)|}$$

其中,  $pred_{xi}$  代表用户  $x$  对未评分的文档  $i$  的预测评分,  $\bar{r}_x$  代表用户  $x$  对所评分的文档的平均评分,  $\bar{r}_y$  代表用户  $y$  对所评分的文档的平均评分,  $N_u$  代表近邻集合.

通过以上分析,编写MapReduce程序实现基于用户的协同过滤算法,可以有效地解决传统的协同过滤算法扩展性的问题,对评分数据进行离线计算,即可得到推荐的文档及相应的预测评分的推荐列表,将推荐列表存储在HBase或者其他数据库中,提供前端在线的个性化服务,至此也就完成了整个推荐过程.

#### (2) 实验数据与开发环境

本文通过虚拟化平台VMware Workstation 12 Pro搭建Hadoop集群环境, Linux操作系统选择ubuntu-12.04.5, Hadoop版本为Hadoop-2.7.1, JDK版本为JDK1.8, 开发工具为Eclipse, Hadoop集群由5台虚拟主机组成,采用完全分布式模式,其中s0为NameNode, S1-S3为DataNode, S4为辅助NameNode对基于MapReduce的UserCF算法的测试采用Book-Crossing数据集,该数据集包括图书社区的278858个用户对271379本书的评分信息,本文共抽取82690条记录进行模拟实验,包括7874个用户的评分信息.

#### (3) 结果分析

由(2)可知, S1-S3为DataNode, 本文将数据集划分为4份,分别选取1968、3937、5905和7874个用户的评分数据,在不同的集群规模上进行计算(选取的DataNode依次为S1, S1和S2, S1、S2和S3).采用加速比进行性能评价,其计算公式为:



$$S_p = \frac{T_s}{T_n}$$

其中,  $T_s$  表示固定规模数据集下在单个 DataNode 的运行时间,  $T_n$  表示固定规模数据集下在  $n$  个 DataNode 的运行时间. 具体结果见图 4.

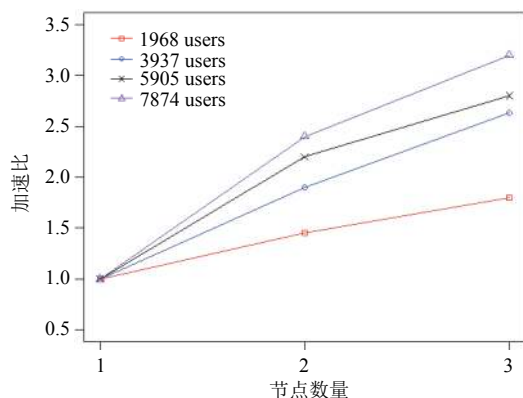


图 4 UserCF 加速比

如图 4 所示,随着 DataNode 节点的增加,加速比大致呈现出线性增加的趋势,能够进一步说明对于相同数量的数据集而言,通过增加计算节点的数量,可以提升数据处理的效率.可以看出当数据量较小时,加速比的变化趋势相对平缓,说明对于较小的数据集而言,通过增加计算节点的数量并不能够显著提升计算效率,这也进一步说明 Hadoop 更加适合处理规模较大的数据集.由于用户的评分数据十分稀疏,所以在 Hadoop 平台上实现基于用户的协同过滤算法可以充分发挥分布式并行计算的优势,从而有效解决算法的扩展性问题.

#### 4 结论与展望

在大数据技术的背景下,通过对 KMS 架构再思考,结合开源的、有庞大社区支持的 Hadoop 云平台以及 Hadoop 生态圈的其他大数据技术,以用户行为模式和人机交互为出发点,以知识的共享、应用和创新为落脚点,构建了更加个性化和智能化的 H\_EKMS,可以有效解决企业人员使用 KMS 积极性不高的窘境,并提高企业海量数据的存储于处理的能力.着重分析基于 Hadoop 的知识管理框架模型和系统的逻辑架构,考虑到知识大爆炸、信息过载的问题,提出了推荐系统在 EKMS 的应用.最后,详细分析知识推荐功能模块实现思路 and 关键技术,采用了分布式 MapReduce 程序实现基于用户的协同过滤算法,有效地解决协同过滤算法

带来的扩展性问题.本文构建的 EKMS 是基于 Hadoop 2.0 的,主要采用 Mapreduce 实现数据的批处理,而 Hadoop 2.0 的资源管理系统 YARN (Yet Another Resource Negotiator) 可以支持多种计算框架,如 Spark, Storm 等,随着企业知识管理的业务场景不断丰富,对计算框架进行整合,实现低延迟大数据处理有待进一步研究.

#### 参考文献

- Li BA. Knowledge management based on big data processing. *Information Technology Journal*, 2014, 13(7): 1415–1418. [doi: 10.3923/itj.2014.1415.1418]
- Tseng SM. The effects of information technology on knowledge management systems. *Expert Systems with Applications*, 2008, 35(1-2): 150–160. [doi: 10.1016/j.eswa.2007.06.011]
- Santoro G, Vrontis D, Thrassou A, *et al.* The internet of things: Building a knowledge management system for open innovation and knowledge management capacity. *Technological Forecasting and Social Change*, 2017. [doi: 10.1016/j.techfore.2017.02.034]
- Alavi M, Leidner DE. 知识管理和知识管理系统: 概念基础和研究课题. 郑文全, 译. *管理世界*, 2012, (5): 157–169.
- 常志平, 蒋馥. 企业知识管理系统的柔性战略分析. *科学与科学技术管理*, 2002, 23(9): 89–92.
- 张明宝, 马静, 米传民. 敏捷知识管理系统的设计与实现. *情报杂志*, 2009, 28(2): 35–39.
- 张庆华, 张庆普. 基于用户创新模式的企业知识管理系统研究. *情报杂志*, 2011, 30(6): 126–129.
- Zhang JS, Zhao W, Xie G, *et al.* Ontology-based knowledge management system and application. *Procedia Engineering*, 2011, 15: 1021–1029. [doi: 10.1016/j.proeng.2011.08.189]
- 叶春森, 汪传雷, 梁雯. 基于云计算的企业知识管理系统再造体系研究. *情报理论与实践*, 2016, 39(3): 80–84.
- 王伟军, 刘艳芬, 王显彬, 等. 基于 Web2.0 的知识管理系统的设计. *情报理论与实践*, 2009, 32(6): 97–101.
- Shvachko K, Kuang HR, Radia S, *et al.* The hadoop distributed file system. *Proceedings of the IEEE 26th Symposium on Mass Storage Systems and Technologies*. Incline Village, NV, USA. 2010. 1–10.
- White T. *Hadoop: The Definitive Guide*. 4th ed. O'Reilly Media, Inc., 2015.
- Nilashi M, Bagherifard K, Ibrahim O, *et al.* Collaborative filtering recommender systems. *Research Journal of Applied Sciences, Engineering & Technology*, 2013, 5(16): 4168–4182.