

# 基于大数据挖掘的高校知识管理系统<sup>①</sup>

李景奇<sup>1</sup>, 卞艺杰<sup>1,2</sup>

<sup>1</sup>(河海大学 商学院, 南京 210098)

<sup>2</sup>(江苏省“世界水谷”与水生态文明协同创新中心, 南京 211100)

**摘要:** 当前高校大数据应用面临数据集成困难、成果难以应用、知识难以管理等问题, 急需加以解决. 本文结合知识管理与软件工程方法, 首先设计了基于 CRISP-DM 的高校大数据挖掘流程; 在此基础上, 设计了高校知识管理系统的总体架构和功能模型; 然后对主要功能模型进行了详细设计, 并对其关键技术——数据集成方法、大数据挖掘技术进行了分析和设计; 最后给出了该系统的开发环境及测试情况. 该系统设计方案对高校大数据应用提供了参考和借鉴.

**关键词:** 大数据挖掘; 知识管理; 高校知识管理系统; 数据集成; 知识审核

引用格式: 李景奇, 卞艺杰. 基于大数据挖掘的高校知识管理系统. 计算机系统应用, 2017, 26(9): 54-61. <http://www.c-s-a.org.cn/1003-3254/5984.html>

## Big Data Mining-Based University Knowledge Management System

LI Jing-Qi<sup>1</sup>, BIAN Yi-Jie<sup>1,2</sup>

<sup>1</sup>(Business School, Hohai University, Nanjing 210098, China)

<sup>2</sup>(Jiangsu Provincial Collaborative Innovation Center of World Water Valley and Water Ecological civilization, Nanjing 211100, China)

**Abstract:** The current application of big data in universities is facing many problems such as difficulties in data integration, results application and knowledge management, which need to be resolved urgently. Combined with the method of knowledge management and software engineering, this paper firstly designs the process of university big data mining based on CRISP-DM. On this basis, it designs the overall structure of the university knowledge management system and the function models, and then it designs the main function models in detail. Next, it analyzes the key technologies of data integration and big data mining. Finally, it gives the development environment and test of the system. The design scheme of the system provides a reference for the application of big data in universities.

**Key words:** big data mining; knowledge management; university knowledge management system; data integration; knowledge audit

## 1 引言

### 1.1 高校大数据应用及其不足

随着高校数字校园的建设, 业务管理信息系统日益增多, 使用日益频繁, 产生的业务数据及日志数据指数级增加. 另一方面, 大数据技术在社会经济各领域取得的成果显示出其巨大的应用价值. 如何应用大数据技术, 分析高校数字校园中产生的巨量数据, 发现其中

蕴藏的规律, 服务于广大的师生, 成为高校大数据应用的主要问题.

大数据技术在高校中有很多的应用场景, 无论高校管理者还是信息管理部门, 亦或是普通教师和学生, 都希望能够及时从数据中获取到更加精细、更加准确的决策信息. 但目前来看, 高校大数据应用存在着许多不足, 主要表现在:

<sup>①</sup> 基金项目: 河海大学高等教育科学研究项目(201612010)

收稿时间: 2017-01-03; 采用时间: 2017-02-15

(1) 数据集成困难. 高校大数据存在于各个业务系统之中, 在管理上, 各部门之间数据管理各自为政, 共享困难; 在技术上, 数据集成要处理各种各样的数据接口, 这些接口或为数据库接口、或为消息, 不一而足.

(2) 成果难以应用. 大数据挖掘的成果如何应用, 如何把成果传递给更多的校内用户, 如何评价应用效果, 这些都是高校大数据应用面临的最直接的问题.

(3) 知识难以管理. 大数据挖掘产生的知识如果不能直接应用于高校工作实践中, 如何进行保存和管理, 谁来负责, 放在哪里. 这些问题都需要加以解决.

因此, 数据集成、知识管理及应用是高校大数据应用的关键. 高校大数据应用需要从知识管理的角度, 从知识发现、知识存储、知识迁移、知识应用四个方面, 结合大数据技术, 来实现对高校大数据的挖掘和应用.

## 1.2 知识管理系统及其应用

知识管理系统有许多定义, 美国学者阿肖克基于一种集成的观点, 认为用于管理机构知识流程的系统称为知识管理系统(Knowledge Management System, KMS)<sup>[1]</sup>, KMS 支持机构流程包括知识创造、知识存储/检索、知识转移和知识应用. 组织中的知识管理系统的核心驱动力来源于那些与改善质量管理过程相关联的组织获取、分享和运用知识的能力.

知识管理系统在社会组织中的应用非常广泛. Witold Staniszki 等<sup>[2]</sup>为欧盟构建了“结构化基金项目知识门户”. 该知识门户的目标是支持组织和个人参与到项目建议书的起草过程中, 以获得尽可能多的符合欧洲委员会标准的高质量的合格建议书. 该门户支持并实现专家交互的部分自动化功能, 同时提供对项目建议书的起草过程中产生的知识进行记录、检验和传播的工具. Hind Benbya 等<sup>[3]</sup>认为企业门户最重要的功能是支持知识管理应用. 在门户技术的支持下, 再结合社会文化和管理方法, 才能取得知识管理在企业中应用的成功. 杨晓明等<sup>[4]</sup>从知识管理的过程出发提出了学校教育知识管理模型, 认为用户可以从教育资源、实践知识或经验知识中获得知识, 并借助一定的工具技术进行开发挖掘, 使知识存储到知识库中, 然后再借助共享平台(如实时交流、知识评估、微博、博客和论坛等)使知识得到共享应用, 最后运用激励机制, 借助知识整合系统进行知识创新.

高校是一个复杂的社会运行组织, 知识是其最核心的基因. 科研是知识的创造, 教学是知识的传承, 管

理和社会服务是知识的应用. 因此知识管理系统在高校中具有特殊的意义. 但在实践中, 由于知识的定义过于广泛, 在高校中所需要管理的知识远远大于其管理能力. 正因为如此, 大数据技术为高校的知识管理带来了新的动力.

基于大数据挖掘的高校知识管理系统(Big data mining-based University Knowledge Management System, BUKMS)是基于大数据技术, 集成各种数据挖掘算法, 通过数据挖掘项目管理, 支持高校知识管理的信息系统. 该系统基于 CRISP-DM(Cross-Industry Standard Process for Data Mining)数据挖掘过程模型, 通过 J2EE 技术实现. 该系统能够供高校教育管理人员、教师及教育研究人员直接使用, 提供实时或离线大数据挖掘服务.

## 2 高校知识管理系统设计

### 2.1 BUKMS 流程设计

在数据挖掘领域, CRISP-DM 是应用最广泛的、事实上的工业标准<sup>[5]</sup>. CRISP-DM 模型提供了一个数据挖掘项目的生命周期的总体描述. 它描述了一个项目的各个阶段、各阶段的任务以及这些任务之间的关系. 它把数据挖掘项目的生命周期分为六个阶段组成, 分别为业务理解、数据理解、数据准备、建模、评价和部署<sup>[6]</sup>.

基于 CRISP-DM 模型, 高校大数据挖掘流程根据业务功能分为四个阶段: 背景评估、数据采集、模型探究、知识应用. 每一阶段都有相应的功能模块来实现. 具体业务功能流程如图 1 所示.

背景评估主要包括价值描述、目标设定、数据来源和算法分析. 价值描述用于确定项目本身的应用价值; 目标设定确定大数据挖掘目标; 数据来源分析数据可能从何处获得; 算法分析则确定为实现目标所需要使用的算法.

数据采集包括三种方式: 人工采集、数据集成、Web 数据. 人工采集的方式包括问卷调查、档案调查以及文献调研. 数据集成方式则是从关系型数据库中获取数据. Web 数据则是从管理系统、网络平台及系统日志中主动下载或抓取数据.

模型探究时, 首先选择算法, 对数据进行处理、建模, 得出相应的结论, 然后进行评估, 如果不满意, 就要重新选择算法, 直至得到满意的结果.

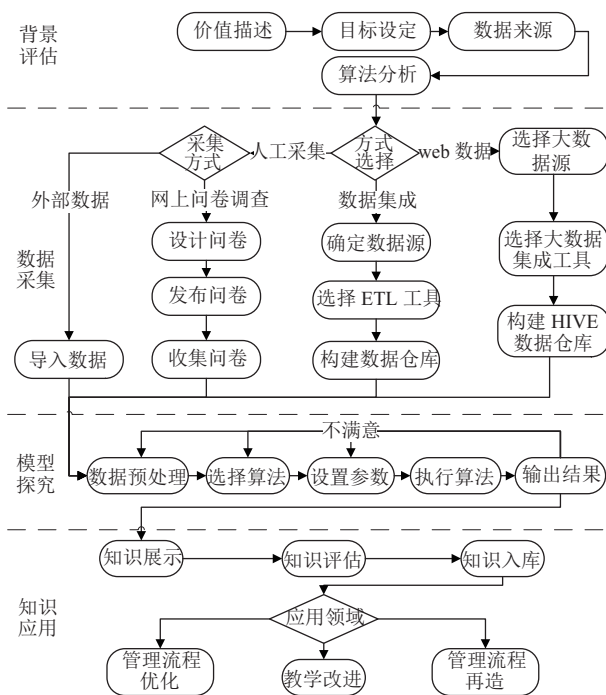


图1 BUKMS 业务功能流程图

知识应用主要包括知识展示、知识审核、知识入库、管理流程优化、管理流程再造、教学改进。知识应用根据输出结果,进行意义描述,并以知识的形式展示出来,或存储于知识管理系统中,供他人利用;或直接应用于高校管理业务流程中;或直接应用于教学设计之中,发挥其作用。然后根据应用效果,作进一步的数据挖掘。

### 2.2 BUKMS 总体架构

BUKMS 以大数据挖掘平台为基础,以知识门户为核心,搭建支持高校核心业务的综合型知识管理系统。其总体架构主要包括高校知识门户、大数据挖掘平台、高校知识库和知识管理服务。具体如图2所示。

在总体架构中,大数据挖掘平台是高校知识管理系统的的核心基础平台。它有三个数据来源,即业务数据库、数据仓库和日志文档。业务数据库存放的是各个管理业务系统数据,可以通过接口协议直接连接。数据仓库则是依据时间线建立的主题式数据库,用于存储各类统计性数据。日志文档包括应用服务器、网络服务器、系统服务器、业务系统的运行和访问日志。

大数据挖掘平台包括六个模块:数据挖掘工具、数据挖掘模型、数据挖掘算法、数据挖掘项目管理、数据集成工具、Hadoop 软件系统。各个模块相互独立,数据项目管理对其余模块进行管理。数据集成工具、

数据挖掘工具、模型、算法既可以自主开发,也可以选择既有产品。Hadoop 软件系统是大数据处理平台,可以通过自主编程,实现对大数据的分布式处理。

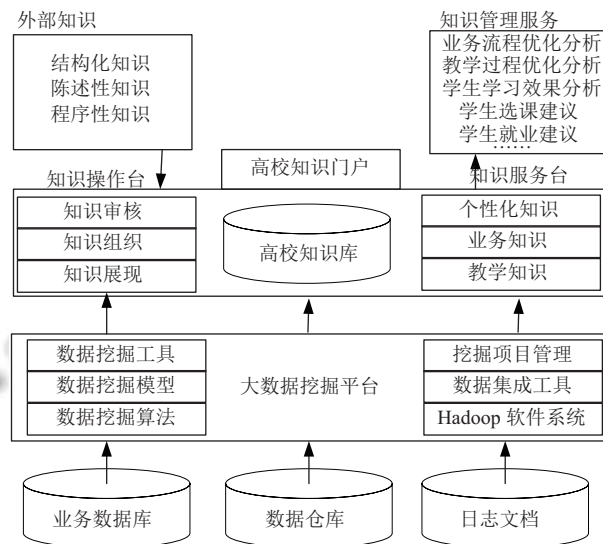


图2 BUKMS 系统总体架构图

大数据挖掘平台的输出结果转化为知识,与外部知识一起支撑了上层知识库,知识库向外提供知识门户服务。知识门户在高校知识库的支撑下,提供知识操作台和知识服务台。

知识操作台是对知识的审核及入库提供功能支撑,包括知识审核、知识分类、知识入库等功能。知识操作台的知识来源是大数据挖掘平台以及外部知识,外部知识分为三类:结构化知识、陈述性知识、程序性知识。结构化知识是指各类知识模型,能够表达相互因果关系,能够对业务进行决策支持的知识。陈述性知识是指规则、规范类文档,用于说明业务的类型、功能等。程序性知识是对业务过程的解释,包括业务流程、设计案例等。

知识服务台向用户提供知识管理服务,包括业务流程优化分析、教学过程优化分析、学习效果优化分析、办事效率分析、学生选课建议、学生就业建议等个性化知识、业务知识和教学知识。

### 2.3 BUKMS 功能模型

BUKMS 功能目标包括两个方面的内容:大数据挖掘与知识管理。大数据挖掘可以看作是知识管理过程中的知识发现过程,但在大数据时代,大数据挖掘的地位日益重要。因此系统设计的核心是实现大数据挖掘,

并在此基础上推进知识存储、转移与应用. BUKMS 包括系统管理、知识发现、知识存储、知识呈现、知识应用五大模块. 其业务功能结构如图3所示.

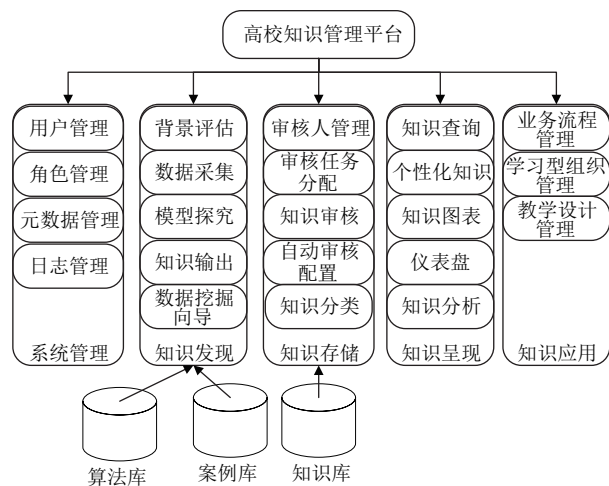


图3 BUKMS 平台业务功能结构图

系统管理包括用户管理、角色管理、元数据管理、日志管理等功能. 主要任务是平台系统功能的支持及用户管理.

知识发现包括背景评估、数据采集、模型探究、知识输出、数据挖掘向导等功能. 这一部分功能是该平台的核心. 知识发现模块实现大数据挖掘的全过程. 从功能上来看, 知识发现模块实现大数据挖掘项目管理的全过程, 即对于一个高校大数据挖掘项目, 从背景评估一直到知识应用这一整个流程, 平台提供一个向导性的管理功能. 用户通过这一向导功能, 实现对自己所采集的数据进行挖掘分析工作.

知识存储包括审核人管理、审核任务分配、知识审核、自动审核配置、知识分类等功能. 其主要任务是实现知识入库, 能够判断知识, 并以良好的分类形式保存起来. 知识存储是高校知识管理系统的重要功能, 其核心是构建知识库, 围绕知识库的入库、出库管理, 实现知识的有效存储.

知识呈现包括知识查询、个性化知识、知识图表、仪表盘、知识分析等功能. 其主要任务是为用户提供个性化的查询界面, 使用户很方便地管理数据、执行数据挖掘、查看结果. 知识呈现以知识门户的形式, 将与用户相关的知识自动集中到用户界面, 帮助用户确定数据所包含的知识内涵, 也为其提供相关的知识, 供其决策参考.

知识应用是高校知识管理系统的主要模块. 如何把大数据挖掘成果应用于高校各项管理和教学工作中, 是其根本的价值所在. 在高校大数据挖掘中, 知识应用主要包括业务流程管理、学习型组织管理和教学设计管理等功能. 业务流程管理支持用户建立业务流程, 管理流程中存在的问题. 学习型组织管理支持用户管理学习型组织, 分析组织问题. 教学设计管理功能则支持教师在教学过程中发现教学问题, 加以分析, 并形成新的教学策略.

### 3 知识应用模块功能设计

#### 3.1 业务流程管理

业务流程管理包括业务流程图展示、业务流程图管理、业务流程跟踪、业务库管理功能. 业务流程图展示将高校中的各个业务流程以图形化的方式展现出来, 便于管理和理解. 业务流程图中各个业务均来源于业务库. 业务流程跟踪功能则根据业务功能所对应的数据库及服务器, 通过调用大数据挖掘服务从业务数据库中提取数据, 分析处理, 为业务流程的再造和优化提供数据支持. 业务库管理则是把各个细分业务添加保存到业务库中, 供业务流程图构建时使用, 如图4所示.

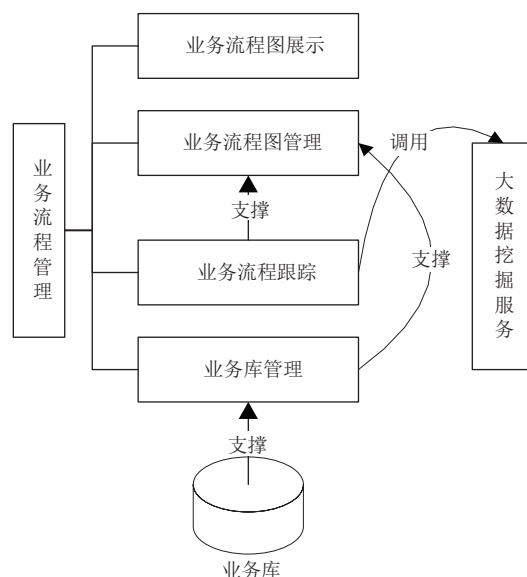


图4 业务流程管理模型功能结构图

#### 3.2 学习型组织管理

学习型组织管理功能主要包括学习型组织建设、组织问题定义、问题决策支持、以及问题基模管理.

David A. Garvin<sup>[7]</sup>提出了“学习型组织”的快速判定

标准. 这一判定标准以 5 个问题的形式出现, 用以识别不同的行为. (1) 组织有没有明确的学习行动计划? (2) 组织能否自由地讨论不和谐的信息? (3) 组织能否避免不犯同样的错误? (4) 当关键员工离开时, 组织是否失去了重要的知识? (5) 组织是否基于自己的知识采取行动?

因此一个学习型组织可以归纳为 5 方面的要求: 学习行动计划、组织保障制度、知识管理制度、经验交流制度、知识转化制度. 学习行动计划用于描述组织当前的知识基础以及学习的目标; 组织保障制度则是要保障组织能够自由地讨论各种信息, 包括不利信息; 经验交流制度则有利于使组织能够对成功的经验和失败的教训进行总结讨论, 使其成为组织记忆的一部分, 从而避免同样错误的发生; 知识管理制度则有利于随时收集关键员工的知识, 避免因员工流失造成的知识流失; 知识转化制度则应鼓励员工应用知识采取行动, 提高工作效率.

因此学习型组织建设主要是围绕学习型组织的五个方面进行. 组织问题定义就是要把组织中存在的问题归纳出来, 并将其归为某类问题基模. 问题基模是依据彼得·圣吉所定义的八种基模为基础, 并可以自定义添加. 问题决策支持是根据预先定义好的问题基模分析步骤, 通过调用大数据挖掘程序, 对相关数据进行挖掘分析, 以支持更好的解决问题. 如图 5 所示.

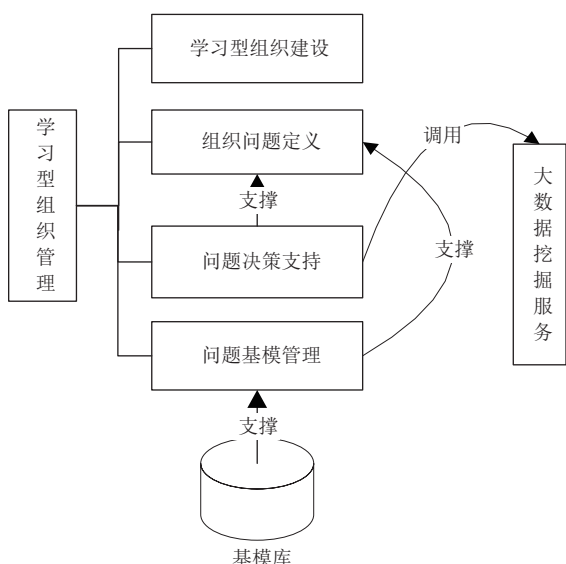


图 5 学习型组织管理功能结构图

### 3.3 教学设计管理

教学设计管理的任务是对数据挖掘知识在教学过

程中加以应用. 教学设计管理从教学模式出发, 通过构建教学模式库, 提供对教学的支持. 教师在教学中, 通过参考相应的教学模式, 可以构建自己独特的教学过程模型. 同时, 通过教学过程跟踪, 收集学生网上学习行为, 然后对此进行大数据挖掘, 则能发现学习问题和学习规律, 协调双方行为, 提高教学效果. 具体如图 6 所示.

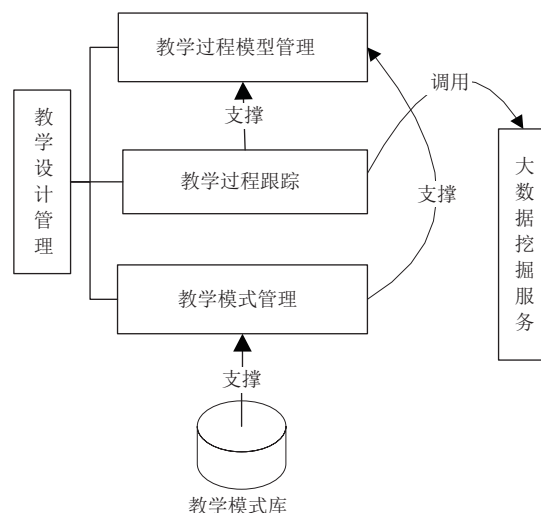


图 6 教学设计管理功能结构图

## 4 关键技术

### 4.1 数据集成技术

高校大数据来源复杂, 如何能够对各种数据进行有效集成, 将大数据挖掘所需要数据归集到系统中来, 是高校知识管理系统所面临的基本问题. 数据集成的关键是其易用性, 能够很方便地使用集成各种数据. 基于此, 必须对数据集成进行分类设计. 首先, 对于用户缺乏相关数据, 可以提供网络调查工具, 通过网络问卷的形式, 收集数据; 其次, 对于文件型数据, 可以通过文件导入方式, 直接将数据导入到系统数据库中, 然后进行分析; 再次, 对于能够提供数据库接口, 则通过 ETL 工具(如 ODI/kettle)或 java/nodejs 等程序设计语言, 实现数据集成. 特别是 nodejs, 能够使用户简单地通过第三方组件库, 很方便地实现数据库之间的数据交换, 也能实现数据文件的读写. 最后, 对于需要通过 Hadoop 大数据平台进行处理的, 则可以通过 sqoop 工具将关系型数据库(如 MySQL, Oracle 等)中的数据导入到 Hadoop 的 HDFS 中, 也可以将 HDFS 的数据导出到关系型数据库中.

通过调研,当前数据集成的主流工具包括以下产品: ODI、KETTLE、Sqoop、Nodejs<sup>[8-11]</sup>。

ODI 能够实现关系型数据库的实时及定时集成操作,能够进行全量集成及增量集成,数据集成效率高,处理数据量大,常用于基于 Oracle 数据库的数据集成。但数据转换过程中对复杂业务较难实现。Kettle 是一款开源数据集成工具,能够实现关系数据库的实时及定时集成操作,能够在不同类型数据库间实现数据共享,数据集成高效稳定。但同样它在数据转换过程中也难以实现复杂业务。Sqoop 是基于 Hadoop 平台,实现 Hadoop 平台与关系数据库的数据交换的开源工具,但不适用于关系数据库之间数据共享。Nodejs 是基于 Chrome V8 引擎的 JavaScript 运行环境和工具,基于扩展库, nodejs 可以轻松实现数据库及文件之间的数据共享。通过简单编程,可以插入复杂业务,能够实现数据集成过程的可视化。但其性能不太稳定,对大批量数据处理不够理想。

基于高校大数据挖掘平台的数据集成需求,通过对数据集成工具的分析,这些工具均可在实际中应用。具体数据集成方式的选择示意图如图 7。

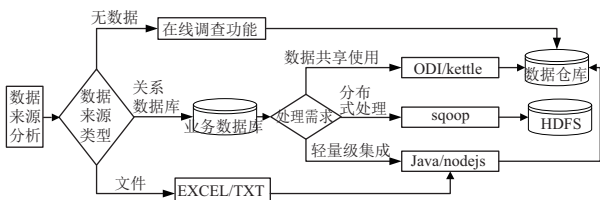


图 7 数据集成工具选择示意图

#### 4.2 大数据挖掘技术

对于大数据挖掘,该系统采用在线和离线两种方式。在线方式中,用户在提交数据文件后,系统自动调用大数据挖掘方法,完成数据挖掘工作,返回数据挖掘结果。离线方式下,需要在系统中提交相关项目资料,而在线下利用大数据挖掘工具来完成数据挖掘工作。目前的大数据挖掘平台主要包括两大类平台: Hadoop 与 Spark。

Hadoop 平台生态完善,支持资源丰富。不足之处在于数据处理编程复杂,并且大数据的算法、呈现都需要基于 Hadoop 的第三方软件支持。Spark 包括四大模块: Spark SQL-RDD(数据执行的基本单元)、MLlib(机器学习)、Graphx(图计算)、Spark Streaming(实时处

理)<sup>[12]</sup>。Spark 是基于内存的编程模型,它可以把中间的迭代过程不放在磁盘中,数据直接在内存中执行,极大地提高了执行速度。但它没有底层数据存储平台,一般仍要基于 Hadoop 平台来实现。

通过简单对比,可以发现 Hadoop 生态较 Spark 为完整,并且可用资源较多,有利于编程人员开发构建。不过 Spark 近年来发展也很快,可以视需要在 Hadoop 平台上进一步扩展。

BUKMS 系统基于 Hadoop 平台进行开发部署,具体技术架构如图 8 所示。BUKMS 系统技术架构可以分为五个层次,分别为展示层、业务层、数据建模层、数据处理层和数据存储层。

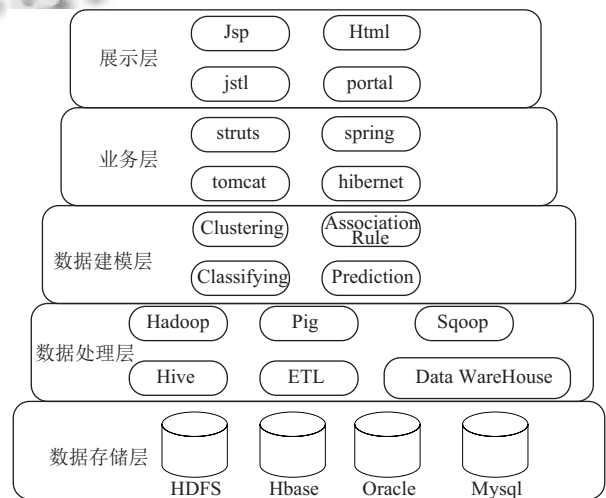


图 8 BUKMS 系统技术架构图

数据存储层负责数据的底层存储,存储采用关系数据库和 NoSQL 数据库两种类型,分别用于处理不同的业务需求。数据存储层包括多种存储技术,主要有 HDFS、Hbase、Oracle 以及 Mysql。HDFS 是 Hadoop 平台的文件存储系统, Hbase 主要用于存储大规模的日志数据。Oracle 用于业务数据存储。Mysql 用于构建 J2EE 平台,承担 BUKMS 系统的数据管理。

数据处理层是负责所有数据的持久存储、处理。它从各个数据源获取数据,数据经过抽取、清洗、加载等转换操作后,进入数据仓库。该层包括 Hadoop、Pig、Sqoop、Hive、ETL、Data Warehouse 等支撑软件。

数据建模层的任务是执行挖掘算法进行数据挖掘。数据建模任务主要包括聚类、分类、关联规则与预测四种。

业务层负责具体的事务处理。业务层主要包括大

数据挖掘项目的业务处理过程,根据大数据处理业务处理流程来进行。

展示层是将数据挖掘结果以图形、报表等形式进行展示的功能。展示层主要从数据展现上,为用户提供详细的、个性化的、可视化的数据分析信息,使用户能够容易地理解数据挖掘所带来的成果。

## 5 系统开发环境及测试

### 5.1 系统开发环境

该研究基于 IBM V7000 存储及 VMWARE5.1 软件,搭建了如图 9 所示的大数据挖掘平台。WEB 服务器提供 J2EE 知识管理门户服务, Hadoop 集群提供大数据处理环境。WEB 服务器使用 windows2008 R2 系统, mysql5.0 数据库。Hadoop 集群采用一主三辅架构,即一台 MASTER 命名节点,三台 SLAVE 数据节点,通过 MASTER,分配 Map-Reduce 任务,由 SLAVE 数据节点执行。四台服务器均采用 centOS6.0 64 位系统, 4G 内存, 100G 硬盘空间。数据库服务器提供数据仓库支持,其它业务系统数据库通过 ETL 工具集成到数据库服务器中。行为日志服务器用于存储业务服务器的日志文件,通过工具将文件调入到大数据平台中。

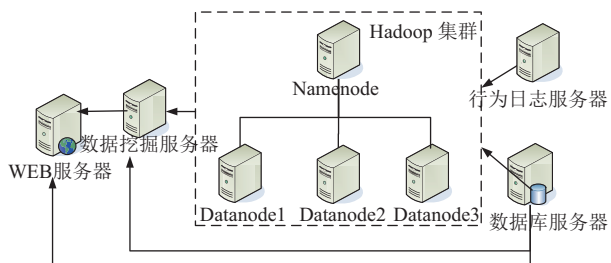


图 9 BUKMS 系统硬件平台结构图

### 5.2 实验案例: 远程教育管理信息系统访问次数与校外学习点应用情况分析

该实验基于远程教育管理信息系统进行,通过分析其访问数据,获取访问量较大的地区信息,然后有针对性地对该地区学习点进行培训和意见征集,改进远程教育线上管理流程。

首先,在知识管理门户中,建立远程教育管理流程,主要包括学籍管理和教学管理两大类。学籍管理主要包括学生录取信息、注册信息、学籍信息、学籍异动信息的管理。教学管理主要包括课程信息和成绩信息的管理。这些管理流程都与学习点有着直接关系。

然后,根据业务流程问题,建立大数据挖掘项目。根据大数据挖掘流程,从背景评估、数据采集、模型探究、知识应用四个方面进行设计和实施。

#### 5.2.1 背景评估

(1) 项目价值: 希望通过本项目,从中可以发现各地区管理员及学生的访问特征,以针对性地进行调研,改进远程教育管理流程。

(2) 项目目标: 根据应用服务器日志,发现各地区学习点的访问特征。

(3) 数据来源: 从应用服务器中获取日志文件。

(4) 算法选择: K-MEANS。

(5) 工具选择: Hadoop 集群, WEKA3.6。

#### 5.2.2 数据采集

本例从应用服务器中获取日志文件,该日志文件大小 660 M,日志记录为 620 万条,记录了该系统自 2011 年 11 月至 2014 年 11 月近 3 年的访问记录。

#### 5.2.3 模型探究

步骤 1. 数据预处理。将日志文件传输到 Hadoop 集群,然后执行 map-reduce 过程,统计每个 IP 地址的访问数量,输出到 MYSQL 数据库,共统计出 44152 个 IP 地址,最高访问数量 127961,最低为 1。然后再通过程序将 IP 地址对应的地区查询出来,输出到对应的字段。数据表结构包括访问 IP、访问数量、所在地区(即 IP 登记地址)。然后按地区进行分类汇总,共获得有效记录 1476 条,地区访问量最高为 361479,最低为 10。

步骤 2. 选择算法。使用 k-means 算法进行聚类。

步骤 3. 设置聚类参数 numClusters 分别为 2、3、5、10,参数形式如:

```
weka.clusterers.SimpleKMeans -N 2-A "weka.core.EuclideanDistance -R first-last" -I 500 -S 100.
```

步骤 4. 执行算法。结果(聚为 10 类时)如图 10 所示。

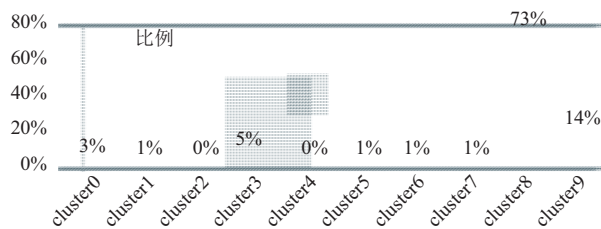


图 10 访问次数聚类结果图

### 5.2.4 知识应用

(1) 知识展示: 根据实验情况, 形成实验报告, 并在知识管理门户中可以查看.

(2) 知识评估: 从本次挖掘结果看, 聚为 10 类实例分布较为合理, 对于学习点的选择有一定的参考价值.

(3) 管理流程优化建议: 选取访问次数在 2789 次以上(即前 7 类)的地区学习点进行应用系统的使用意见和建议征集, 而加强在 cluster8 与 cluster9 中的学习点进行应用系统使用培训.

## 6 结语

基于大数据挖掘的高校知识管理系统将大数据技术与高校知识管理有效地结合起来, 解决了高校数据集成困难、成果难以应用、知识难以管理的问题. 这些问题的解决有利于提高高校知识发现能力和知识应用水平, 提高校园管理和教学的智慧化水平. 在应用层面, 它有效地结合了业务流程管理、学习型组织管理、教学设计管理等高校核心业务应用, 具体而针对性地提出了问题解决方案, 为高校的管理人员和教师利用大数据挖掘带来便利, 使他们能够自主挖掘数据信息, 提高管理水平, 提高教学效率.

### 参考文献

1 阿肖克·贾夏帕拉. 知识管理: 一种集成方法. 安小米译. 北

京: 中国人民大学出版社, 2013: 222–226.

- 2 Staniszki W, Staniszki E. Intelligent agent-based expert interactions in a knowledge management portal. Traunmüller R. Electronic Government. Berlin, Heidelberg, Germany. Springer. 2003. 296–299.
- 3 Benbya H, Passiante G, Belbaly N A. Corporate portal: A tool for knowledge management synchronization. International Journal of Information Management, 2004, 24(3): 201–220. [doi: 10.1016/j.ijinfomgt.2003.12.012]
- 4 杨晓明, 李小聪, 王时进. 学校教育知识管理系统框架设计及其实现思路. 中国教育信息化·高教职教, 2010, (6): 47–49.
- 5 Shearer C. The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing, 2000, 5(4): 13–22.
- 6 Azevedo A, Santos MF. KDD, semma and CRISP-DM: A parallel overview. IADIS European Conference Data Mining 2008. Amsterdam, the Netherlands. 2008. 182–185.
- 7 戴维 A. 加尔文. 学习型组织行动纲领. 邱昭良译. 北京: 机械工业出版社, 2004.
- 8 崔有文, 周金海. 基于 KETTLE 的数据集成研究. 计算机技术与发展, 2015, 25(4): 153–157.
- 9 于金良, 朱志祥, 梁小江. 一种基于 Sqoop 的数据交换系统. 物联网技术, 2016, 6(3): 35–37.
- 10 林瑶. web 实时数据同步研究[硕士学位论文]. 北京: 北京邮电大学, 2015.
- 11 孙玮. 基于 ODI 技术搭建高校数字化校园公共数据平台. 软件工程师, 2014, 17(7): 56–57, 44.
- 12 李文栋. 基于 Spark 的大数据挖掘技术的研究与实现[硕士学位论文]. 济南: 山东大学, 2015.