

大数据技术时代的农业信息资源共享

李晓静 臧运平

(青岛农业大学图书馆, 山东青岛 266109)

摘要: 在分析大数据时代农业信息资源特点的基础上, 考察大数据技术在农业信息资源共享方面的利用价值, 从农业信息处理的4个方面, 即信息采集—信息存储—信息挖掘—信息利用, 介绍大数据技术在农业信息资源共享中的应用。这有助于实现农业信息资源的有效开发与高效利用, 加快农业信息化的进程。

关键词: 大数据; 大数据技术; 农业信息资源; 信息资源共享

中图分类号: G253

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2013.06.009

Research on Agricultural Information Resources Sharing Based on Big Data Technology Era

Li Xiaojing, Zang Yunping

(Library of Qingdao Agricultural University, Qingdao 266109)

Abstract: The advent of the era of big data brought opportunities and challenges to the scientific research, and agricultural informatization also inevitably affected by big data, people gradually realize the importance of big data in agricultural information resources construction. Based on the analysis of characteristics of agricultural information resource in the era of big data, this paper researched the big data technology utilization value in agriculture information resource sharing, and introduced the technology of big data in the application of the agriculture information resource sharing that includes four aspects: information collection, information storage, information mining and information use. This can help us to achieve effective development and the utilization of agricultural information resources and accelerate the process of agricultural informatization.

Keywords: big data, big data technology era, agricultural information resources, information resource sharing

1 引言

随着大数据时代的来临, 数据已成为一种新的经济资产, 科研已进入了以数据为基础进行科学发现的第四范式^[1]。大数据给科学研究与实践带来了新的机遇和挑战。我国农业正在逐步转向现代农业、数字农业与信息农业。在长期的科学研究中, 人们通过观察、试验和计算等多种方式积累了大量有科学价值和实用意义的农业信息资源。这些信

息资源信息量大、学科专业广、综合性强, 是农业科学研究的基础和必不可少的资源^[2]。近年来, 研究大数据技术在农业信息资源建设中的应用越来越广泛, 取得了一定的研究成果。王儒敬分析了我国农业信息化发展面临的困难和瓶颈问题, 并提出了“加快云计算与大数据处理技术与农业应用, 开发农业大数据库管理系统”的发展策略^[2]。梁川等介绍了数据挖掘技术及其在农业上的应用, 并提出了农业数据挖掘系统的构建方案, 对于通过数据

作者简介: 李晓静(1981-), 女, 青岛农业大学图书馆馆员, 研究方向: 农业信息化与信息资源建设。

基金项目: 2010年度教育部人文社会科学研究一般项目基金资助项目“农业院校图书馆对农村提供信息服务模式的研究”(10YJA870032)。

收稿日期: 2013年8月20日。

挖掘技术来进一步提升农业信息服务水平具有一定意义^[3]。曹丽英等为了解决农业信息资源整合建设过程中的共知、共建和共享等问题,将云计算模型引入农业信息资源整合建设中,提出了云环境下农业信息资源整合的实现机制^[4],这对于解决农业信息化建设中的“信息孤岛”和“重复建设”等问题具有一定价值。此外,李光达等阐述了云环境下农业信息服务的模式,提出了云计算与农业信息服务结合的一个实践探索——农科云,对于解决农业科技信息孤岛和服务滞后等问题具有一定意义^[5]。这些研究表明,人们已经认识到大数据在农业信息资源建设中的重要性,开始借助大数据及云计算等技术解决农业信息资源建设中的关键问题,试图加快农业信息化的发展。在此基础上,一个最值得关注的问题就是,在这个“数据爆炸而知识贫乏”的时代,如何更好地将大数据技术应用于农业信息资源共享中。

为了更加确切地了解大数据时代农业信息资源建设的研究现状,笔者以“中国学术文献网络出版总库(CNKI)为数据源,检索策略为SU=('大数据'+海量数据+'big data'+extensive data)*'农业信息',检索日期为2013-07-06,进行精确检索,去重、筛选后,共获得20条结果。年度发展趋势如图1所示。从图1中可以看出,学术界对大数据在农业信息领域应用的发展趋势,2006年与2010年分别出现两次高峰,2013年又开始出现上升趋势。大数据在农业信息领域应用的研究尚处于初步发展阶段,有待进一步加深研究深度和广度。本文在分析大数据时代农业信息资源特点的基础上,考察了大数据技术在农业信息资源共享方面的利用价值,并从农业信息处理的流程分析了大数据技术在农业信息资

源共享中的应用,从而为农业信息资源共享的实现提供建议和思路。

2 农业信息资源的特点与内容

农业有其自身的一些特点,如自然界作物品种复杂、病虫害发生频繁、与气候变化密切相关,许多知识还未被人们认识,然而农业领域所积累的信息之多是其他任何一个领域都无法相比的,因此农业信息也呈现出量大、多维、动态、不完全、不确定等各种特性,而且农业信息的内容包涵了化学、生物、自然资源、生态环境、大气、植物、动物等多种学科。随着大数据时代的到来,农业信息呈海量增长的趋势越来越明显,同时信息载体与传播方式也呈现多元化的趋势。

在不同的发展阶段,农业信息资源的内容虽然有所不同,但基本上都包含以下几部分的内容:农业政策法规、农业科学技术信息、农业市场供求信息、农业生产相关信息等。这些资源大都分布在农业科研院所、高校、科技厅、农业厅及其下设的农业管理部门、农业技术推广单位以及涉农企业等。这些农业信息资源的主要类型有:农业类图书、期刊、报纸等,农业类科技报告、政府出版物、会议文献、学位论文、专利文献、标准文献等特种文献,农业类数据库、农业类OA资源以及免费的网络农业信息资源等。

据统计,我国目前涉农网站有4万多个,多半是国家及地方政府建立的农业信息服务网站,但是存在着资源重复建设的问题。由于农业信息资源缺乏统一的资源表达与操作标准,信息异构、分散等现象普遍,因此出现了农业信息资源“碎片化”的现象。农业信息资源共享是解决农业信息资源“碎

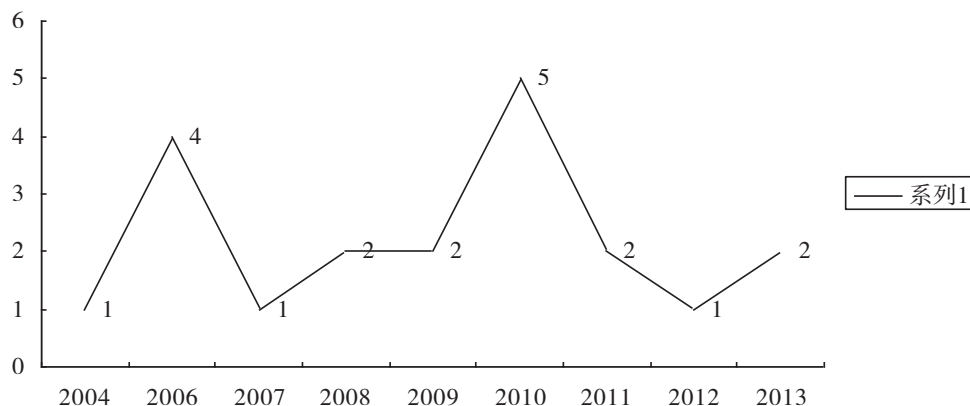


图1 研究发展趋势图

片化”问题的一个新视角。

3 大数据技术的特点

大数据处理与传统数据处理的区别如图2所示。大数据处理技术包含了数据的获取、存储、挖掘及可视化等各种技术。

常用的大数据获取技术有传感器技术、Web2.0技术、RFID技术等。在数据存储中，传统的网络存储系统采用集中的存储服务器存放数据，已不能满足大规模存储的需要。大数据存储采用的是分布式

网络存储系统，储存的非结构化数据管理系统中的各种原始数据。大数据挖掘技术就是运用确定的数据挖掘模型，从已有数据中提取出用户所需要的知识。大数据可视化技术集成了多种数据挖掘算法和可视化工具，帮助用户直观地、实时地发掘和理解大量数据背后的知识。

运用大数据技术从庞大的非结构化数据中分析潜在的有价值的信息，是农业信息资源共享的基础，其价值可总结为如下两个方面：(1)利用大数据技术分析农业用户的需求与偏好，从而开发出更有针对性的农业信息产品，更高效地实现农业信息资源的共建共享。(2)利用大数据技术分析资源共享体系构建过程中的数据，能够更好地预测可能遇到的障碍，防范于未然，实现资源共享的目标。

4 大数据技术在农业信息资源共享中的应用

农业信息资源共享首先就是对现有农业信息资源的整合与处理。就农业信息本身而言，信息整合与处理可以分为以下几个流程：农业信息的采集、存储、挖掘与利用，如图3所示。各流程都有其特定的要求，将大数据相关技术应用于各个流程，有利于数据处理的高效性与可用性，满足大数据时代

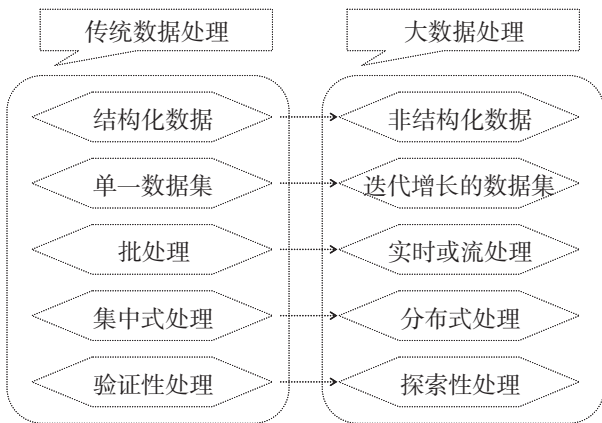


图2 大数据处理与传统数据处理的区别

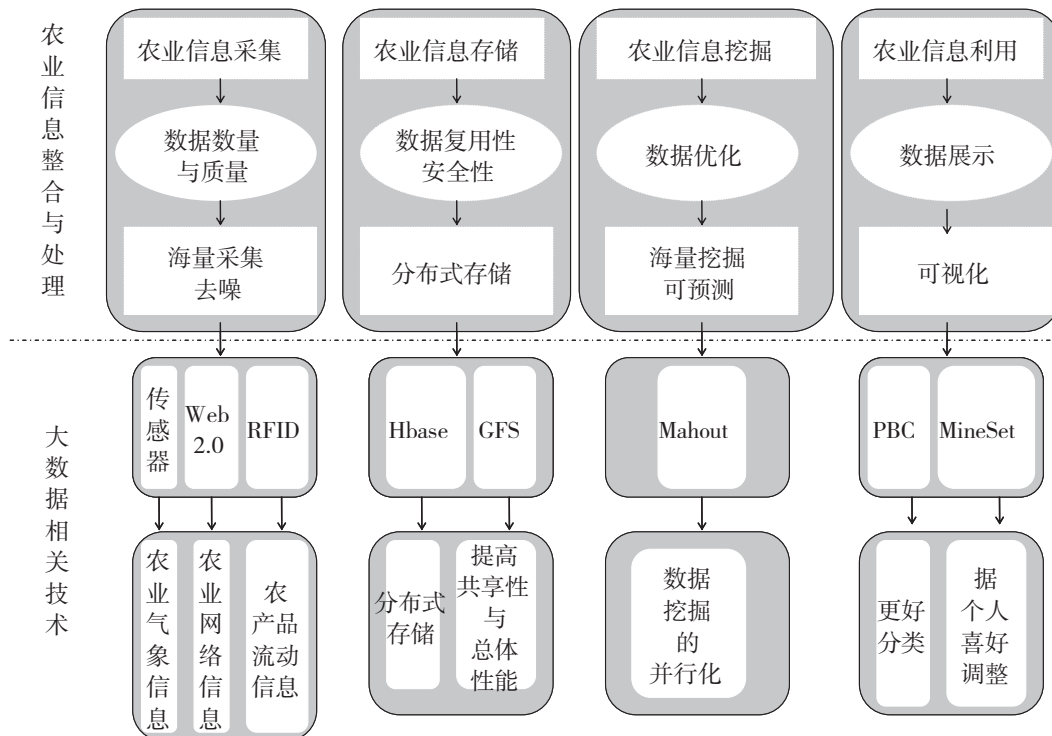


图3 基于大数据技术的农业信息资源共享流程图

农业信息资源丰富、类型多样等造成的信息处理高要求^[6]。

4.1 农业信息采集

农业信息采集包括数据的获取、抽取、清洗。农业信息资源呈现出海量增长、类型繁多、内容多样等特点,而且这些数据都以不同的形式被接收和记录,信息收集面临着海量与复杂采集的挑战,对大数据的获取主要是完成对元数据的选择。获取的数据可能具有多种结构和类型,而数据抽取的过程就是将复杂数据转换为单一的或便于处理的过程。此外,获取的这些数据并不全具有价值,也存在我们不需要的数据,甚至存在部分错误干扰数据。通过大数据技术去噪,对大数据进行“清洗”^[7]。采集农业信息的大数据技术主要有应用传感器、Web2.0技术及RFID等。

近年来,传感器迅速普及,大量的数据不断被传感器接收,为农业大数据的获取提供了有力保障。无线传感器可以采集设备的数字信号,然后通过无线传感器网络传输到监控中心的无线网关,直接送入计算机,进行分析处理。这种技术适合应用于农业气象信息的获取,可以实现对气象环境的全面检测,包括温湿度的检测、光照度的检测等。

Web2.0是指网络上用户的身份由“读者”进化成了“作者”与“共建人员”,由被动地接收网络信息向主动创造网络信息发展。随着博客、百科全书及社交网络的发展,大量的网页点击与交流成为农业大数据的来源之一,获取这些农业网络信息,可以极大地丰富农业信息资源的内容。

RFID(无线射频),是一种利用无线射频通信方式实现的非接触式自动识别技术,它以无人干预、远距离读取等诸多优势受到全球信息化领域的关注^[8]。这种技术已经被广泛应用于农产品供应链领域,通过嵌入电子标签,可以追踪农产品销售系统中产品的流动情况,从而实时掌握农产品流动信息,更好地为农产品加工、销售等企业的决策提供真实可靠的依据。

4.2 农业信息存储

农业信息存储过程中最关键的一点就是保证数据的复用性与安全性,即经过“清洗”后的数据的有效存取是至关重要的。如果将这些数据随便放入农业数据仓库中,就有可能造成访问受阻,从而致使数据难以复用。传统的网络存储系统已不能满足大规模存储的需要。分布式网络存储系统适合用来

储存非结构化数据管理系统中的各种原始数据。

Hadoop的HBase就是一种分布式、面向列的存储系统。它最大的特点就是能够解决传统数据库无法解决的超大规模与分布式的特点。因目前农业传感器分布广、数目多的特点,我们可以采用HBase构建集群存储管理系统,分布式存储传感器的数据^[9]。

谷歌开发的GFS文件系统,是基于共享、网络存储的分布式文件系统,它的自动多数据备份存储增加了数据存储的安全性与可靠性。该存储系统有利于提高农业存储资源的共享性和总体性能,可以满足大容量、高性能的存储要求。

4.3 农业信息挖掘

如何更好地挖掘农业信息资源的价值,满足不同类型农业用户的需求,是农业信息资源共享的关键环节。农业数据挖掘是指从农业数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的过程,也是数据优化的过程。如使用数据挖掘技术对农业信息资源中的气象信息(温度、湿度、风力等)进行挖掘,就能及早发现灾害天气的到来,又如通过对农产品价格走势的数据挖掘,就能对将来的价格进行预测。

随着信息产业的发展,数据挖掘技术已经进入了第5代,即基于云计算的数据挖掘,其关注的重点就是实现传统数据挖掘算法的并行化。如Hadoop项目中的Mahout子项目就是致力于数据挖掘的并行化,提供了大量的基于Hadoop采用并行实现的机器学习和知识发现算法^[10]。农业信息资源因其信息量大,学科专业广,内容涵盖广,不同的农业类用户对农业信息资源的需求有其特殊性,大数据挖掘技术能处理农业信息资源的这些特点,通用的数据挖掘工具可以做多种模式的挖掘,我们可以根据农业用户的需求选择合适的数据挖掘工具。

4.4 农业信息展示与利用

如何将经过数据挖掘后的农业信息清晰明朗地呈现给农业用户是农业信息资源共享的最终表现。以图形化、图像化甚至动画化的形式将农业信息资源展现出来更能满足农业用户个性化、多样化的需求。

如PBC的新版本引入了一些分类算法,通过与用户的交互,更好地得出分类结果。又如MineSet 2.6使用了6种可视化工具来表现数据,对同一个

(下转第58页)

- [8] 王凌华.大型科学仪器设备共享平台建设发展研究[J].中国科技纵横,2011(21):381-382.
- [9] 王丽萍.提高大型科学仪器设备利用率的探讨[J].长沙铁道学院学报:社会科学版,2009(2):38-39.
- [10] 国家科技基础条件平台中心.关于召开2012年国家科技基础条件平台绩效考核评审会的通知[EB/OL].[2012-12-21].http://www.nstic.gov.cn/showContent.jsp?page=1346220832468.
- [11] 黄国梁,刘金屏.区域性大型科学仪器设备共建共享服务平台管理机制探索[J].太原科技,2008(6):12-13,17.
- [12] 李鸿宾.打造大型科学仪器设备共享服务平台[J].中国科技成果,2007(22):24-25.
- [13] 彭洁,田杰.推进科学仪器设备协作共享的公权、私权探讨[J].现代科学仪器,2006(6):120-122.
- [14] 孙绪华.关于促进我国大型科学仪器设备共享的思考[J].实验技术与管理,2006,11(23):10-14.
- [15] 陆建玲,秦娟,夏春阳,等.建设大型科学仪器设备资源共享服务平台的实践与思考[J].科技管理研究,2007,27(9):89-91.

(上接第44页)

数据挖掘结果可以不同形式表示,用户可以任意放大、旋转、移动图形,并可从不同的角度观看,也可以根据个人喜好调整最终效果,以便更好地理解。这些大数据可视化技术能够帮助实现农业信息资源更加生动、多样化地呈现给用户,从而提高信息资源的展示效果与利用率。

5 结语

大数据技术能够为农业信息资源共建共享提供方便有效的手段,利用大数据相关技术可实现区域乃至全国范围的农业信息资源共享,促进农业信息资源与农业用户的有效衔接,满足农业用户的需求,是实现农业信息资源有效开发与高效利用,加快农业信息化进程的重要途径。但是,在大数据技术应用过程中会遇到各种问题,如政策法规问题、知识产权问题、技术难题等,需要今后进一步探讨解决。

参考文献

- [1] 李广建,杨林.大数据视角下的情报研究与情报研究技术[J].图书与情报,2012(6):1-8.
- [2] 王敬儒.我国农业信息化发展的瓶颈与应对策略思考[J].中国科学院院刊,2013(5):337-343.
- [3] 梁川,王文生,谢能付.农业信息资源上数据挖掘的应用[J].中国农学通报,2009,25(11):243-247.
- [4] 曹丽英,张晓贤,赵月玲,等.云计算在农业信息资源整合模式中的应用[J].中国农机化,2012(3):141-144.
- [5] 李光达,郑怀国,谭翠萍,等.基于云计算的农业信息服务研究[J].安徽农业科学,2011,39(27):16959-16961.
- [6] 尚明华,秦磊磊,王风云,等.山东省农业信息服务平台构建思路与前景展望[J].农业网络信息,2006(2):48-51.
- [7] 窦万春,江澄.大数据应用的技术体系及潜在问题[J].中兴通讯技术,2013(6):1-15.
- [8] 易芙宇.RFID:企业的数据采集员[J].中国科技信息,2005(18):51-52.
- [9] 周利珍,陈庆奎.基于HBase的农业无线传感信息存储系统[J].计算机系统应用,2012,21(8):6-9.
- [10] 杨宸铸.基于HADOOP的数据挖掘研究[M].重庆:重庆大学出版社,2010:20-21.