

基于机会协作的农业物联网大数据处理系统设计

袁 芬^{1*}, 徐从富²

(1. 浙江长征职业技术学院 计算机与信息技术系, 杭州 310023; 2. 浙江大学 计算机科学与技术学院, 杭州 310027)

(* 通信作者电子邮箱 YuanFenlaoshi@163.com)

摘 要:针对在农业物联网中通信环境复杂和大数据处理效率低的问题,提出了基于机会协作的自适应大数据处理机制。根据农业物联网应用需求,综合考虑了农业产业环境对无线数据传输的影响建立了跨层交互分析模型,结合机会协作机制和大数据处理需求,设计了大数据处理方案。实验分析和测试表明,所提大数据处理方案在系统吞吐率、可靠性和系统处理性能等方面,表现出了比传统协作机制和数据处理方案更优越的性能。

关键词:农业物联网;大数据;机会协作;系统设计

中图分类号: TP393 **文献标志码:** A

Big data processing system based on opportunistic cooperation for agricultural Internet of things

YUAN Fen¹, XU Congfu²

(1. Department of Computer and Information Technology, Zhejiang Changzheng Vocational and Technical College, Hangzhou Zhejiang 310023, China;

2. College of Computer Science and Technology, Zhejiang University, Hangzhou Zhejiang 310027, China)

Abstract: Aiming at the complex communication environment and low efficiency of big data processing in agricultural Internet Of Things (IOT), a big data processing mechanism was proposed based on the adaptive collaborative opportunities. According to the requirements of agricultural application and the impact of agricultural environment on wireless data transmission, a cross-layer interaction analysis model was established, which was combined with opportunities for collaborative mechanisms and big data processing requirements. Then the design of a large data processing was proposed. Experimental analysis and testing show that the proposed big data processing scheme has better system throughput, reliability, and system processing performance than traditional coordination mechanism and data processing programs.

Key words: agricultural Internet of things; big data; opportunistic cooperation; system design

0 引言

随着无线通信技术,特别是物联网技术及其应用平台的研究与开发,通过物联网应用技术为现代化农业发展提供重要支撑和平台,成为现代工业界、农业界和学术界研究的热点之一^[1]。农业物联网技术不仅可以有效加强农产品质量安全的监督和预警^[2-3],而且能够有效改善农业综合决策和管理水平。此外,基于物联网应用技术平台可以升级传统农业流程,对农作物的用药、灌溉、施肥及其相关畜牧业等实现实时监控,提高资源利用率,减少对环境的污染,特别是加强农作物疾病监控与常见传染性疫情预警,在保证农作物增产率的同时提高生产经济效益,从而实现农业现代化高效经营和可持续发展。

同时在农业物联网中,由于监测数据种类多,数据量大^[4-5],如何提供有效地进行大数据传输和大数据处理成为重要挑战之一。因此,为了使得物联网技术能够更好地为现代化智慧农业提供技术支撑,研究农业物联网中的大数据处理系统具有重要意义。

Choo 等^[6]为了解决大数据处理的复杂度和冗余度问题提出了一种可视化数据处理方案。文献[7]基于三维数字地

球平台软件 Skyline,有效地提高检索速度并降低输出成本。针对大数据的处理复杂度问题,Jelinek 等^[8]基于价值链战略,特别是商业模式创新结合市场导向的理论模型和基于资源的观点,通过加强搜索目的性,从而增强大数据分析在物联网等应用中的长期可持续性以便适应不断变化的环境。为了在智能电网中将无线网络技术、物理系统和云技术相结合,文献[9]提出了一种可扩展的软件平台。该平台具有可以动态响应需求的应用程序以便执行智能需求管理,缓解智能电网高峰负荷同时提供了自适应信息集成技术。Biem 等^[10]针对描述流时间序列分析和管理问题,提出一种基于 IBM 的 InfoSphere 流平台建立解决方案,具有异常检测、建模、平滑、预测和跟踪时间序列分析等功能。Kraska 等^[11]针对大数据和云计算相结合中遇到的问题,研究了开发传统的数据库系统分解的三层体系架构。元开元等^[12]针对实时采集和处理城市高速车辆大数据应用中遇到的问题,提出了一种大规模数据实时处理方法,根据数据处理系统的控制参数实现流水线阶段化操作,提高硬件资源利用率,并改造了应用平台中的数据结构、I/O 策略等算法,进一步改善数据处理系统的性能。文献[13]综述了物联网平台下实时大数据信息的交换背景以及数据库等方面的交换方法,在此基础上提出

收稿日期:2014-01-24;修回日期:2014-03-07。 基金项目:国家自然科学基金资助项目(61101235)。

作者简介:袁芬(1981-),女,浙江杭州人,讲师,硕士,主要研究方向:大数据处理; 徐从富(1975-),男,浙江开化人,副教授,博士,主要研究方向:信息融合。

了适用于物联网的多种信息交换改进策略,包括:临时表技术、内存映射文件方法、文件缓冲策略等。文献[14]在车载网络平台下,提出一种适用于大数据传输的路由增强机制,基于同向行驶车辆间的拓扑稳定特性从而延长车辆间的数据传输时间,结合擦除编码技术进一步改善传输可靠性。

在已有成果的基础上针对农业物联网平台的大数据处理问题,本文设计了一种适用于农业物联网的基于机会协作的大数据自适应处理系统记为OCA。本文的主要工作包括:1)基于机会协作技术建立了数据处理模型;2)分析了农业物联网平台下端到端的大数据传输可靠性和处理复杂度规律,据此建立自适应处理系统。

1 农业物联网模型

为了有效提高农业信息化水平,改善农业综合决策,降低农业生产成本,利用物联网技术,设计一个农业物联网模型。该模型包括:采集环境数据、自适应控制农机设备、人员及机械作业实时监控等功能,使得物联网技术能够有机地结合农业生产链,构建“智慧”农业园。

该物联网模型是集成多功能的智能化控制系统,采用分层设计思想是基于跨层交互的模式,以模块为单位实现不同功能并设计交互逻辑结构,从而使得该系统具有一定的合理性、交互性和可扩展性,具体架构如图1所示。



图1 农业物联网模型架构

在该模型中设计了4层架构,包括应用层、大数据层、协作层和物理层。其中,应用层可以通过无线局域网,3G/4G移动通信网络等方式为用户提供服务,用户可以通过PC或手机等设备访问系统获得服务。大数据层提供了4个模块,分别是传感器采集数据处理模块、射频识别(Radio Frequency Identification, RFID)标签数据处理模块、多样性应用数据处理模块和自适应控制数据处理模块。协作层主要是通过机会协作通信技术为上层应用提供大数据传输可靠性保障。物理层根据数据采集设备的不同主要分为由不同类型的传感器构成的异构传感网和能够采集不同类型数据的RFID标签组成。为了实现数据共享和自适应智能控制,在系统中设计了跨层交互接口。

此外,为了提高数据挖掘效率和精度,在大数据层设立存放不同类型数据的数据库,以便实现针对特定数据库的多条件检索。同时针对不同类型和来源的大数据,该系统采用了单表检索和多表检索相结合的联合检索模式。对联合检索得到的结果再使用清洗、过滤和排序等技术处理,然后通过协作通信层的机会协作技术传输给用户。用户可以访问不同类型数据库实现多样性检索的流程如图2所示,该层的各类型数据库均可提供全文检索功能,以便为用户提供人性化的交互

形式和可视化大数据分析结果。

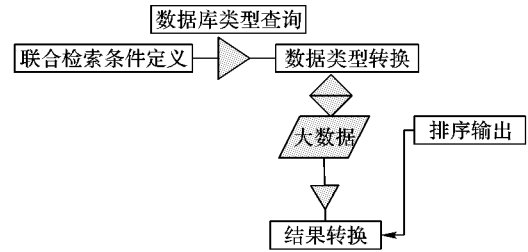


图2 大数据挖掘流程

2 基于机会协作的大数据处理

2.1 机会协作机制

农业物联网中,物理层各类通信终端组成通信子网,采用机会协作通信方式转发数据。协作通信过程中,初始状态选择考虑距离较近或信道状态较好的节点作为下一跳转发节点。在此无线通信过程中,因为节点具有随机移动性,从目的节点方向走向发送节点过程中,也会收到无线信号即该数据分组。如图3所示,源节点S向汇聚节点H发送数据分组,由于H节点的随机移动距离S节点越来越远导致数据接收失败,而此时R₁节点正在靠近S节点,以一定的概率作为下一跳中继转发节点,同时R₂节点正以2倍于R₁节点的速度靠近S节点,有一定的机会优于R₁成为最佳中继节点帮助转发数据分组到达目的节点H。

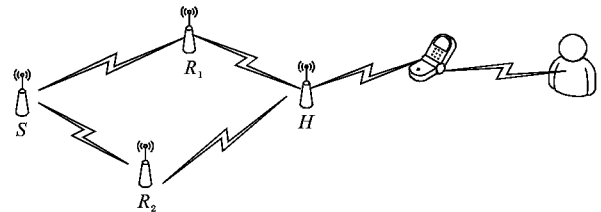


图3 机会协作模式

由此可见,R₁和R₂在移动过程会监听到其他节点发送或转发的数据分组,如果可以正确解码其他节点发送的数据或广播的无线信号,便可以以一定的概率成为中继节点提高数据传输可靠性。此外,物联网系统中,用户可通过手机终端直接访问汇聚节点进行自适应实时智能控制。

源节点S向汇聚节点H发送无线信号的通信距离如式(1)所示:

$$y_{SH} = \sqrt{(x_S - x_H)^2 + (y_S - y_H)^2} \quad (1)$$

任意2个节点间通信距离也可由式(1)求得。某节点靠近源节点时移动感知距离可由式(2)求出,某节点远离源节点时的移动距离可由式(3)求解。

$$L_1 = vt \quad (2)$$

$$L_2 = v\varphi \quad (3)$$

其中: t 表示源节点发送无线信号的时间间隔, φ 表示接收节点无线广播的时间间隔。根据 L_1 和 L_2 的分析结果,利用机会原理可得到节点移动速度二阶导,机会转发数据概率和数据分组正确接收率分别如式(4)、(5)和(6)所示:

$$\delta = \pi - \arcsin \frac{L_1^2 + L_2^2 - y_{SH}^2}{2L_1L_2} \quad (4)$$

$$p = \frac{\sqrt{L_1^2 + L_2^2 - 2L_1L_2}}{y_{SH}^2 \arctan \delta} \quad (5)$$

$$P_{\text{bit}} = 1 - \frac{L_1^2 + L_2^2 - 2L_1L_2}{y_{SH}^2 \arccos p} \quad (6)$$

此外,源节点 S 或其他转发节点的发射功率可结合式(4)、(5)和(6)求得,如式(7)所示:

$$\sigma = \sqrt{\frac{L_1^2 + L_2^2}{2\delta p(1 - P_{\text{bit}})}} \quad (7)$$

2.2 大数据处理机制与方法

基于前文建立的物联网机会协作通信模型,本节建立多样性大数据处理机制。基于数据类型感知特性建立各类型数据库映射机制,基于流水线提高数据处理性能,如图4所示。

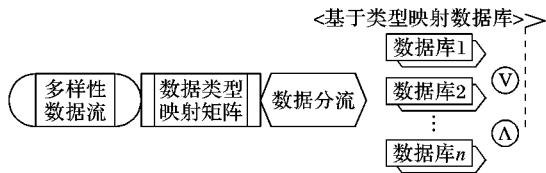


图4 大数据处理流程

在实际大数据处理系统中,因为难以准确预测用户即将检索的数据库及数据映射,因此在本文所提出大数据处理系统OCA中,基于进程实现数据处理流水作业。不同数据处理作业之间建立缓冲区,其中存放了多样性数据映射Hash二维矩阵,以便进行排序和最优化处理。同时,在数据联合检索过程中考虑了大数据多类型映射的局部性以及数据作业流交互方式。于是,OCA可以实现终端内外数据替换时快速查找到检索频次高、检索次序合理、交互间隔代价最小的数据映射表项以及Hash元素。此处,作业流间的大数据处理缓冲区中存放的缓存大数据映射确定为首选匹配项,基于机会协作可靠传输的数据分组从链路层映射到大数据层各类数据库中的Hash表项并初始化相应参数,按照大数据各类型参数综合评价更新Hash表项为下一次服务提供支撑。

根据多节点发起请求缓冲原理和作业流间缓冲区处理机制,可得到作业流大数据处理吞吐率并由式(8)得到。

$$\mu = \frac{kP_{\text{bit}} \arctan \delta}{\delta p} \quad (8)$$

其中 k 表示用户发起大数据处理请求个数。

基于机会协作的大数据处理算法如下所述:

输入:作业流映射 i ;

输出:大数据处理速度 v 。

1) $i = 1, p = 0, P_{\text{bit}} = 0$ 。

2) 一次作业流大数据处理结束后,根据式(4)结合式(5)和式(6)得到 p, P_{bit} , 然后更新Hash矩阵和作业流编号 i 。

3) 如果 $\mu < 0.5$ 则进行新一轮机会协作中继节点选择和数据流最优化处理。

4) 否则,保护现场继续进行数据转发和作业流优化处理。

5) 如果作业流编号 $i > n$,则大数据处理结束;否则重复第2)~4)步操作。

在上述算法中,首先初始化Hash矩阵和大数据作业流编号以及节点移动速度。其次,及时更新作业流间缓冲区Hash表项和移动速度二阶导,实时监测大数据处理吞吐率,如果吞吐率小于0.5则需要更新数据库映射;如果大数据处理速度二阶导和吞吐率可以保持一定优化状态,则依据用户需求逐步按作业流进行数据处理。

另外,通过构件(封装多源数据源的查询模块)来实现异

构、多源大数据集成处理方法,为了实现多源、异构数据大数据集成和共享,减少不同数据类型转换代码数量,先用XML作为数据转换格式,把不同的大数据对象优化成统一的全局数据,再利用这些不同构件访问多源异构大数据。

在分析以上异构、多源大数据集成处理机制和方法基础上,为了实现农业物联网的大数据挖掘任务,更好提高大数据处理系统的性能,提出基于B/S结构云计算服务模式的数据挖掘设计方案,主要包括1个数据库群、1个服务器群、大数据管理子系统、大数据挖掘子系统和挖掘算法子系统等5个部分。

3 系统性能测试

通过2组实验对机会协作的大数据传输机制的可靠性以及农业物联网的大数据处理系统的性能进行验证与测试。

第一组实验采用数学分析和仿真验证的方法从丢包率、吞吐量和能效对比分析了所提出的基于机会协作的OCA机制和传统的协作通信机制,参数设置与环境变量如表1所示,结果如图5所示。

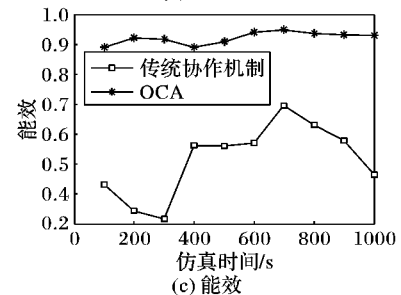
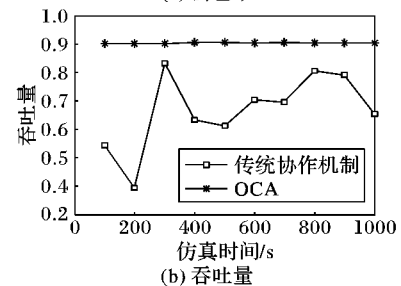
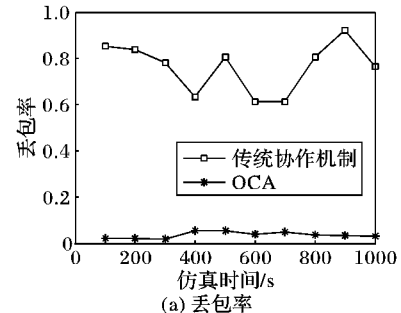


图5 大数据传输性能测试

从图5中可以看出,所提基于机会协作的大数据传输机制在可靠性、能效和吞吐量与传统协作机制相比具有明显的优越性,这主要取决于移动感知的候选节点机会式参与大数据转发。

第二组实验实时监测系统大数据传输流持续1000 min,以便对OCA与传统物联网从读写性能和检索命中率2个方面进行基准测试其系统性能,结果如图6所示。测试结果表明从大数据处理过程中的读写性能和联合检索命中率而言所提OCA均优于传统物联网系统中的处理机制。

表1 参数设置

实验参数	值
仿真时间	1 000 s
拓扑大小	1800 m × 1000 m
移动节点数	100
数据分组大小	1 500 B
暂停时间	0 ~ 5 s
移动模型	随机移动
节点移动速度	5 ~ 10 m/s

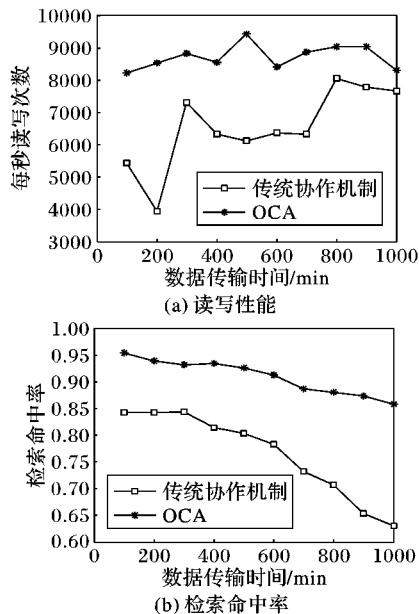


图6 大数据处理性能测试

4 结语

农业物联网中的节点移动性、无线通信环境的复杂性、高误码率和大数据多样性等特点使得为其提供可靠的数据传输和高效的大数据处理性能成为物联网应用的新挑战。本文首先针对农业物联网应用平台,提出了一种基于跨层交互的网络分析模型;其次,分析了节点移动性与协作通信特点,提出基于机会协作的大数据可靠传输机制,然后设计了基于机会协作的适用于农业物联网的大数据处理系统;最后从传输可靠性和系统大数据处理性能两个方面证明本文所提 OCA 系统比传统协作机制和物联网大数据处理机制更有效。

参考文献:

- [1] YAN X, WANG W, LIANG J. Application mode construction of Internet of things for facility agriculture in Beijing[J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(4): 149 - 154. (阎晓军, 王维瑞, 梁建平. 北京市设施农业物联网应用模式构建[J]. 农业工程学报, 2012, 28(4): 149 - 154.)
- [2] ZHAO X. Study on key technologies in agriculture Internet of things [J]. Taiwan Agricultural Research, 2011, 113(6): 103 - 107. (赵湘宁. 农业物联网中关键技术研究进展[J]. 台湾农业探索, 2011, 113(6): 103 - 107.)
- [3] CHEN Y. Discussing on relation between agricultural Internet of things and agriculture complex large system intelligent control [J]. Agriculture Network Information, 2012(2): 8 - 12. (陈一飞. 农业复杂大系统的智能控制与农业物联网关系探讨[J]. 农业网络信息, 2012(2): 8 - 12.)
- [4] WIGAN M R, CLARKE R. Big data's big unintended consequences [J]. Computer, 2013, 46(6): 46 - 53.
- [5] MENG X, CI X. Big data management: concepts, techniques and challenges [J]. Journal of Computer Research and Development, 2013, 50(1): 146 - 169. (孟小峰, 慈祥. 大数据管理: 概念、技术与挑战[J]. 计算机研究与发展, 2013, 50(1): 146 - 169.)
- [6] CHOO J, PARK H. Customizing computational methods for visual analytics with big data [J]. IEEE Computer Graphics and Applications, 2013, 33(4): 22 - 28.
- [7] HAN X X, LI J Z, YANG D H, et al. Efficient skyline computation on big data [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(11): 2521 - 2535.
- [8] JELINEK M, BERGEY P. Innovation as the strategic driver of sustainability: big data knowledge for profit and survival [J]. IEEE Engineering Management Review, 2013, 41(2): 14 - 22.
- [9] SIMMHAN Y, AMAN S, KUMBHARE A, et al. Cloud-based software platform for big data analytics in smart grids [J]. Computing in Science and Engineering, 2013, 15(4): 38 - 47.
- [10] BIEM A, FENG H, RIABOV A V, et al. Real-time analysis and management of big time-series data [J]. IBM Journal of Research and Development, 2013, 57(3/4): 1 - 12.
- [11] KRASKA T, TRUSHKOWSKY B. The new database architectures [J]. IEEE Internet Computing, 2013, 17(3): 72 - 75.
- [12] YUAN K, ZHAO Z, FANG J, et al. Real-time processing for high speed data stream over large scale data [J]. Chinese Journal of Computers, 2012, 35(3): 477 - 490. (元开元, 赵卓峰, 房俊, 等. 针对高速数据流的大规模数据实时处理方法[J]. 计算机学报, 2012, 35(3): 477 - 490.)
- [13] JIA X. Big data real-time information exchange strategy research based on the Internet of things [J]. E-Government, 2011(4): 16 - 20. (贾晓丰. 基于物联网的大数据量实时信息交换策略研究[J]. 电子政务, 2011(4): 16 - 20.)
- [14] TIAN R, SUN L, LIU Y, et al. COBRA: A collaboration based reinforcement mechanism for mass transmission in VANETs [J]. Journal of Computer Research and Development, 2009, 46(12): 2076 - 2084. (田锐, 孙利民, 刘燕, 等. COBRA: 车载网络中基于协作的大数据传输增强机制[J]. 计算机研究与发展, 2009, 46(12): 2076 - 2084.)
- [11] Association for the Advancement of Medical Instrumentation . ANSI/AAMI EC57: 2012, Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms [S]. Arlington: Association for the Advancement of Medical Instrumentation, 2012: 1 - 22.
- [12] SHENSA M. The discrete wavelet transform: wedding the a trous and Mallat algorithms [J]. IEEE Transactions on Signal Processing, 1992, 40(10): 2464 - 2482.
- [13] PRTTools [EB/OL]. [2013-10-20]. <http://prtools.org/>.
- [14] LIU X, YAN C, HU Z. Arrhythmia classification based on mathematical morphology and support vector machine [J]. Journal of Computer Applications, 2013, 33(4): 1173 - 1175. (刘雄飞, 晏晨伟, 胡志坤. 基于数学形态学及支持向量机的早搏失常识别[J]. 计算机应用, 2013, 33(4): 1173 - 1175.)
- [15] TONG J, DONG J. Electrocardiogram classification using combined classifiers [J]. Journal of Computer Applications, 2010, 30(4): 1225 - 1228. (童佳斐, 董军. 分类器组合在心电图分类中的应用[J]. 计算机应用, 2010, 30(4): 1225 - 1228.)

(上接第 2135 页)