

无线传感器网络中多类型数据融合研究综述*

赵继军, 魏忠诚, 李志华, 刘昊, 连彬

(河北工程大学信息与电气工程学院, 河北邯郸 056038)

摘要: 探讨多类型数据融合面临的新问题, 从分析相关性出发, 探讨多类型数据中可供融合的关系, 提出数据类型属性相关的概念, 并综述基于相关性的数据融合研究现状。给出了多类型数据融合研究的一般流程, 按流程综述了相关文献, 对各种技术进行了分析与对比, 并指出了研究的难点、突破口以及未来的研究重点。

关键词: 无线传感器网络; 多类型数据融合; 预测算法; 相关性

中图分类号: TP393

文献标志码: A

文章编号: 1001-3695(2012)08-2811-06

doi:10.3969/j.issn.1001-3695.2012.08.003

Research of multi-type data fusion in sensor networks

ZHAO Ji-jun, WEI Zhong-cheng, LI Zhi-hua, LIU Hao, LIAN Bin

(School of Information & Electrical Engineering, Hebei University of Engineering, Handan Hebei 056038, China)

Abstract: This paper focused on the new problems of multi-type data fusion. By analyzing the correlation, it explored the correlations of multi-type data which were available for data fusion, proposed the concept of multi-type correlation, and reviewed the current situation of data fusion research that was based on correlation. It also gave the general process of research on multi-type data fusion, and then reviewed related literatures according to the general process. It analysed and compared various kinds of technologies. At last it pointed out the difficulty and sally ports of the problems, and gave the future key points of the research.

Key words: wireless sensor networks (WSN); multi-type data fusion; prediction algorithm; correlation

无线传感器网络 (WSN) 作为一种新型的网络体系, 具有低成本、高精度、易操作等特点, 这使其在科学技术中的地位逐渐凸显出来^[1]。无线传感器网络应用日趋广泛, 受到了研究机构和用户们的推崇; 其发展潜力巨大, 无论研究还是商业领域, 都有着无法估量的发展空间。近几年提出的物联网的概念与无线传感器网络也密不可分, 作为物联网的底层技术支撑, 它承担着数据感知和数据传输的双重压力, 可以说是挑战与机遇并存。一直以来, 无线传感器网络无法逾越的技术屏障就是节点资源有限^[2]。随着技术的不断更新和应用领域的不断推广, 网络中的数据种类逐渐呈现多样化, 这些数据通常来源于不同的节点或同一节点中的多传感器^[3], 不同数据类型属性之间往往存在一定的相关性^[4], 多类型数据的出现给无线传感器网络的数据管理和传输工作带来了更多的挑战。因此, 如何对无线传感器网络中多类型数据进行有效的数据融合, 从而达到能量节省的目的, 是当前无线传感器网络研究所面临的难点问题。

1 多类型数据的相关性特点

1.1 空间—时间相关性

多源性是无线传感器网络监测数据的基本性质^[5]。由于单个传感器节点的监测范围和可靠性是有限的, 为了保障监测

信息的准确性并保持整个网络满足一定的通信连通度, 需要使传感器节点分布达到一定的密度, 有时甚至使多个节点的监测范围互相交叠, 或者多个节点甚至多种不同性能的节点负责监测同一个事件或现象^[3], 这些节点的感知数据在所表述的物理类型属性高度相似甚至相同, 从而使多源数据在物理空间上具有一定程度的空间相关性^[6,7]。同时, 由于物理现象的连续性和无线传感器网络查询操作的限制, 节点连续采样所获得的数据值亦存在一定的相似性, 称之为时间相关性。

鉴于以上两种相关性的存在, 无线传感器网络监测数据存在较大冗余的现象, 必然引起大量节点之间甚至是节点本身的通信干扰, 进而增大数据传输的能耗和时延等^[1,2]。因此, 空间—时间相关性一直是研究数据融合必然考虑的重要因素之一。

1.2 数据类型属性相关性

数据分类现象是随着无线传感器网络规模的扩大和对监测精度的要求出现的^[4]。为了提高监测质量, 从不同参数来反映单个事件或现象的变化或现状, 通常采用多传感器多空间^[8]、多时域、多角度^[9]监测目标信息, 多类型数据必然形成。与此同时, 新型传感器和先进处理技术的涌现, 以及软硬件的改进, 也使得多传感器逐渐普遍, 这些传感器节点自带多种具有不同感知能力的传感器, 能同时感知不同种类的信息, 如土

收稿日期: 2012-02-29; **修回日期:** 2012-03-26 **基金项目:** 河北省自然科学基金资助项目 (F2010001045); 河北省科学技术研究与发展计划资助项目 (10213545); 邯郸市科学技术研究与发展计划资助项目 (1121103137)

作者简介: 赵继军 (1970-), 男, 河北邯郸人, 教授, 硕导, 博士 (后), CCF 会员, 主要研究方向为无线传感器网络、光通信 (zhaojijun@china.com); 魏忠诚 (1987-), 男, 河南商丘人, 硕士, 主要研究方向为无线传感器网络; 李志华 (1978-), 女, 河北邯郸人, 教授, 博士, 主要研究方向为无线传感器网络; 刘昊 (1988-), 男, 湖北荆州人, 硕士研究生, 主要研究方向为无线传感器网络; 连彬 (1988-), 女, 河北邯郸人, 硕士研究生, 主要研究方向为无线传感器网络。

壤^[10]、气体浓度^[11]、温度、湿度、压力、震动、风向、电磁辐射、密度等^[12]。多类型数据现象呈逐渐普遍的发展趋势,但也大大复杂化了数据的管理传输工作。需要指出的是,多类型数据之间往往存在着某种联系,即存在一定程度冗余,这些联系体现在数据类型的属性上,为冗余数据的融合提供了条件。

综上所述,从数据类型属性相关考虑数据融合是多类型数据融合研究的新方向,也是必然考虑的途径。

本文对类型属性相关性的定义如下:

定义 1 数据类型属性相关性。无线传感器网络不同监测数据类型属性之间存在一定的相似性,这种相似性导致数据之间存在较大冗余,通过数据融合技术选取具有较高质量的数据作为监测数据值,可有效降低这种相关性的冗余度。

1.3 数据类型属性相关性分析

数据类型属性相关性的研究是继时间—空间相关性研究的又一重要途径。首先,数据类型属性相关性不同于空间—时间相关性,是独立于空间—时间相关性以外的一种新的相关性。监测数据类型属性相关性依赖于监测目标的自身类型的固有属性。数据类型属性之间相互依赖、相互转换或互为等价关系^[13]。如何准确挖掘数据类型属性之间的关系,有效评估数据类型属性相关性的程度是研究数据类型属性相关性的重点与难点。其次,数据类型属性相关性具有一定的时—空特征,与时间—空间相关性有紧密的联系。数据类型属性相关性的研究首先要依托于空间—时间相关性。其存在的形式与所表现的特征均体现在空间—时间相关性上,如监测数据类型属性相关的多个传感器在一次采集工作后所要上传的数据在同一物理空间,具有一定的空间相关性竞争;其中两种连续属性物理类型,在时间上依然具有一定的时间相关性。因此结合类型属性相关性与空间—时间相关性是降低多源分类数据类型属性相关性冗余的有效途径,也是必要途径。

2 数据融合研究

2.1 概念与分类

数据融合的目标是通过合并、压缩、过滤、筛选、计算、预测、评估等手段有效降低网络数据传输量,延长网络寿命,提高数据精度和服务质量。目前,针对数据融合的研究主要包括数据融合算法和数据融合位置及模型,数据融合位置研究的任务是将数据融合思想合理体现在各层网络技术,一般位于 MAC 层、路由层^[14]、应用层,以及一些跨层操作的技术;数据融合的模型一般有基于树型、基于多路径和基于簇等几种^[15]。数据融合位置和模型的研究已经逐渐走向成熟,然而其节能作用依然有限,高效的数据融合算法成为当前数据融合研究的瓶颈。

数据融合算法研究是提供数据预处理、筛选、计算、预测和评估等数据处理的方法。文献[5]将数据融合算法分为两大类:数据长度降低的和无数据长度降低的。文献[16]将数据融合分为三大类:渐进数据融合、协商一致数据融合和单跳数据融合。以上数据融合分类方式都是从数据融合操作行为上考虑,只是对现存成果表现形式的一种总结,分类方式不够精确,指导意义不强。更值得指出的是,各种数据融合算法、位置等并没有明确的限制,其出现在各种节能技术时均有交叠,降低了该分类方法的意义。

本文从相关性的角度出发,将无线传感器网络数据融合分

为数据无关性数据融合和数据相关性数据融合。如表 1 所示,不同的类型包含不同的融合方法,各种方法均有优缺点,以下将按照这种分类方式对无线传感器网络数据融合的研究成果进行综述。

表 1 数据融合分类及关键技术比较

融合类型	关键技术	所考虑相关性	优点	缺点
无关性	压缩	无	a) 简单、易实现; b) 融合位置灵活	融合效果较差
	合并、丢弃	无	a) 简单、易实现; b) 融合位置灵活	a) 造成数据缺失; b) 造成误差较大
相关性	抑制发送	空间—时间相关性	a) 融合度高; b) QoS 高; c) 融合位置偏底层	a) 依赖于节点可信度; b) 适用单一数据
	过滤	空间相关性	a) 融合效果较好; b) QoS 较好	a) 依赖于节点可信度; b) 融合偏上层
	评估、预测	属性—时间相关性	a) 融合度高; b) 无缺失值问题	a) 需要输入集(历史数据)、存储空间大; b) 计算复杂度; c) 精度低

2.2 无关性数据融合

传统的数据融合模型一般是与数据无关的数据融合,有专门负责数据融合的节点进行数据融合,是被动的数据融合,此类数据融合一般在数据传输过程中实现^[15]。这种数据融合技术不考虑数据的相关性,负责数据融合的节点收集到数据以后,对所收集的数据信息进行整理,合并或丢弃部分可信度较低的数据,以此来降低网络能耗并提高数据精度。文献[17]设定聚合函数,利用负责数据融合任务的节点对收到的两个或者两个以上的相似数据进行复制压缩,只将众多数据中的一个发往目的节点,以此来降低所需传输的数据量,进而降低网络能耗。文献[14]考虑数据之间的空间相关性,首先利用簇头节点收集各节点冗余信息,然后对收集的信息进行数据融合,降低空间相关的数据冗余,以此减少节点与基站的通信量,进而降低能耗。

数据无关性融合算法一般包含两类压缩技术,一种是在信源编码时,通过压缩降低传输数据的长度,有效降低发送时间;另一种是信源编码时,通过信息整合,将多个信息包压缩至一个数据包中,有效减少数据包数量,以此降低网络能耗。这是一种对信源编码的处理技术,是最早的数据融合技术。

与数据无关的数据融合能够提高数据精度和鲁棒性,减少数据冗余,节省能耗,降低网络负载保存网络能量。但其缺点是不能尽早发现相关性数据,在进行数据融合之前节点仍有较多的冗余数据传输,而这些传输也是不必要的。另外,一旦簇头节点受到恶意攻击,基站或者 sink 节点便不能保证收到正确的融合数据。尽管如此,它仍是数据相关性融合研究的基础。

2.3 相关性数据融合

利用相关性进行数据融合是一种有效的节能方式,融合效果较为明显。这些相关性包括时间相关性、空间相关性以及本文重点研究的数据类型属性相关性,针对不同的相关性采用不同的数据融合方法。当前多数算法均为与数据相关的数据融合算法,此类数据融合逐渐出现在底层网络,如路由层、MAC 层等,融合效果更好,降低网络能耗的空间更大。

基于空间相关性的数据融合方法主要包括过滤、筛选、抑制发送等方法;基于时间相关性的数据融合方法包括预测、筛选、抑制发送等。

a) 抑制发送。文献[18~20]均是考虑空间相关性并采用

抑制数据发送的数据融合。根据监测数据质量在空间上的差异,抑制低监测质量数据的发送,能有效降低通信量,提高服务质量。文献[18]提出空间相关性的评估模型,有效利用了空间相关性。文献[19]提出了一种基于服务质量和空间相关性的无线传感器网络 MAC 协议 QS-Sift。该协议赋予事件发生周围各节点不同的优先级,当某节点感知到一个感兴趣的事件发生时,它是否有必要发送监测数据完全取决于其监测数据的质量(SNR),通过这种机制抑制数据质量较差的节点向 sink 节点发送数据,既减少了需要发送的数据量,又保证了高层数据质量的重建工作。文献[20]提出一种基于空间相关性的事件驱动型 MAC 协议,该协议根据节点监测到的事件的强度来赋予节点发送数据的权值,抑制权值过低的节点发送数据,以此来减少信道竞争的冲突,同时也提高了事件监测的精度,减少数据的发送量,降低了网络能耗,达到了数据融合的效果。文献[21]在评估时间、空间相关性的程度的基础之上提出了一种自适应 MAC 机制,以适应网络中时间、空间相关性可能发生的变化,能够有效地最小化网络能耗和延迟,但未能提出如何降低具有时空相关性的数据的冗余的有效方案。

基于空间相关性的抑制类数据融合成果一般依托于 MAC 协议的研究,目前逐渐向高质量数据的筛选及低质量数据抑制策略发展。这种策略的优点在于能够避免过多冗余数据的收集,尤其是在感知层就能降低数据的传输量。但其只适用于一种数据的传输,其提高精度的方法完全依赖于节点在空间上的可信度。可通过对数据可靠性的评估并结合筛选技术改进这一缺陷。

b) 筛选和过滤。筛选技术一般采用代理节点选择机制,选择部分节点以替代具有空间相关性的节点集进行数据上传,有效减少参与通信的节点,降低监测数据冗余度,提高 MAC 层协议性能,从而降低网络能耗;同时也一定程度上解决了拓扑控制的研究目标,即监测与通信需求之间的矛盾。

过滤技术是考虑空间相关性数据融合的一个重要组成部分。文献[22]利用最优的融合簇状态估计的 Krein 空间卡尔曼滤波方法,得到信息形式的鲁棒卡尔曼滤波。簇头节点通过所处簇的观测模型,利用信息形式的鲁棒卡尔曼滤波实现离散形式的卡尔曼滤波。簇头节点将状态估计和可逆的误差协方差矩阵传送到中心基站,中心基站融合簇状态估计产生全局状况估计。文献[23]通过建立模型将错误数据分成临时错误数据和永久错误数据,首先通过时间相关性将临时错误数据滤除,再通过空间相关性将永久错误数据滤除。

筛选和过滤技术一般是从节点的可靠性来等价衡量监测数据的可信度,主要研究目标是精确评估数据的可靠性。由于不考虑监测数据本身的可靠性,筛选和过滤没能最精确化网络监测值,只是对数据作简单处理,未能充分挖掘相关性存在的价值。同样,筛选和过滤技术可结合评估实现数据的进一步评估。

c) 预测。文献[24]综述了多传感器信息融合的研究进展,其中预测算法是最适合于多类型数据融合。另外,鉴于监测数据的时间连续性,预测被视为弥补当前数据融合缺陷的关键性技术。当前阻碍预测技术发展的主要瓶颈在于预测算法。

综上所述,各种数据融合均有优缺点,因此没有任何一种技术能满足数据融合的全部需要,针对不同的需求通过不同技术的结合是实现高效率数据融合有效且可行的方法。

3 数据融合预测技术研究

多类型数据融合考虑利用数据类型属性相关性、时间相关性。本章探讨多类型数据融合的预测技术,并给出多类型数据融合研究的一般流程,包括如图 1 所示的三个重要环节:a) 数据采集;b) 相关性分析;c) 数据预测。

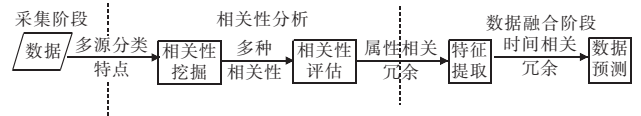


图 1 多类型数据融合流程

3.1 数据采集

影响网络设计、运行的最原始因素在于网络信源的特征^[25],数据采集方法会决定信源特征的表现形式。以往信源模式受监测目标特征影响,网络技术的研究只为满足目标监测的需要,未能对信源加以操作。

近年有少量文献针对信源的复杂性进行有效简化,使得其表现形式更理想化,有效降低了 WSN 能量消耗。针对信源的两操作是:a) 依据空间相关性,抑制或选择部分节点上传数据;b) 依据时间相关性,控制节点上传数据周期(控制采样周期或设置缓冲区控制通信周期)。

3.2 相关性分析

针对相关性的挖掘早在无线传感器网络出现之前就已经受到很多重视,如多媒体、图像处理、模式识别、机器人、电子工艺技术等众多领域均有涉及。无线传感器网络中相关性的研究来源于电子工艺技术^[26~28]。文献[26]提出了基于主成分分析(principal component analysis, PCA)的统计时序分析算法,用于预测考虑内模参数变量空间相关性影响时电路延迟的概率分布情况。文献[27]为提取芯片设计中工艺参数和空间相关性的统计特征,通过解决带约束的非线性最小二乘问题建立空间相关性提取函数,采用改进的替代投影算法抽象出空间相关性矩阵。文献[28]指出具有空间相关性的现象通常由高斯随机过程建模,求解其均值函数和协方差函数并在高斯随机过程被指定在一个线性模型中的条件下提出了客观贝叶斯分析模型。以上相关性挖掘成果只适用于电子工艺技术,能够准确地评估出真实环境的相关性情况,但其均未考虑算法的复杂度和网络环境下的可操作性,因此不能适应计算、存储和能量资源有限的无线传感器网络。

文献[29]是文献[27]针对相关性研究在时域上的推广。在噪声数据中提取时间相关性函数参数,给出两个时刻采集数据的协方差,为了精确求得时间相关性函数和空间相关性变量,该文献将时间变量提取问题转换成一个带约束的非线性最小二乘问题的优化问题。

文献[30]提出熵函数评估方法,两属性 A 和 B 之间的相关性程度可通过式(1)进行测量。

$$U(A, B) = 2 \times \frac{H(A) + H(B) - H(A, B)}{H(A) + H(B)} \quad (1)$$

$$H(X) = - \sum_{i=1}^n p(a_i) \log_2 p(a_i) \quad (2)$$

$$H(A, B) = - \sum_{i=1}^n \sum_{j=1}^m p(a_i, b_j) \log_2 p(a_i, b_j) \quad (3)$$

其中: $H(X)$ 是熵函数,由每个属性的概率分布密度决定; $H(A, B)$ 是 A 和 B 的联合熵函数,由 A 和 B 合并计算得出。

文献[4]提出了一种用于挖掘具有时间相关的分类数据之间相关性的算法。该算法首先用高斯图模型将数据的时间、空间和属性三维关系降低为二维矩阵,然后通过求解高斯图模型的结构挖掘分类数据之间的相关性。

文献[31]针对稠密型无线传感器网络簇头节点读取数据高斯场时的相关性的问题,对物理世界的两种相关性模式进行了研究。该文献首先通过量化理论和信息率失真理论得到服从高斯分布相关性数据的联合熵,然后研究物理世界中可能存在的相关性模型,以此最终确定联合熵与传感器节点物理位置的关系,并由协方差矩阵表示联合熵函数。文献[32,33]均采用协方差矩阵表示数据类型属性间的关系。

文献[34]利用多项式回归算法评估三维无线传感器网络中的空间相关性。该技术首先建立具有两种类型节点的二叉树:树节点和感知节点。感知节点感知物理属性并且将它们的坐标和感应值上传至离它最近的树节点。树节点在接收值上适应多项式函数并将回归系数上传至父亲树节点。该过程由叶子节点开始,至根节点结束。过程结束时,根节点获取整个传感器网络的多项式函数,如空间坐标函数值。当 sink 节点查询根节点时,根节点可以使用多项式函数来计算任意坐标边界节点的属性值。

相关性分析是多类型数据融合研究的基础和关键,相关性的挖掘与评估的技术瓶颈是如何解决评估精度和算法复杂度之间的矛盾问题。

3.3 数据预测

预测技术一般针对具有时间相关性的数据信息,采用预测算法实现对采集数据走势的实时预测,若预测值与监测值相似,则可使用预测值替代监测值,节省了信息传输过程。该技术的主要研究目标是降低节点到 sink 节点的数据通信,数据预测能降低网络的总体能耗。本节综述了当前比较常见的预测技术,并将技术分为回归算法、智能算法和预测模型三种类型,各种技术的优缺点比较如表 2 所示。

表 2 预测技术对比

技术分类	预测技术	优点	缺点
回归算法	简单线性回归	算法简单、复杂度低	a) 预测精度低; b) 需频繁更新预测模型
	多元线性回归	预测精度高	算法复杂度高
	基因表达式编程	简单、线性、易于操作	a) 收敛性差、效率低; b) 难以跳出局部最优解
人工智能	神经网络	a) 解决非线性问题; b) 灵活性强	a) 结构设计复杂; b) 难以确定隐藏层神经元数量
	群智能算法	解决非线性问题;	a) 复杂度高;b) 收敛性差; c) 易陷入局部最优解
	Kalman 滤波器	收敛速度快	a) 要求精确的模型和噪声统计; b) 不易预测非线性变化; c) 需存储较多参数
预测模型	蒙特卡洛过滤器	能够预测非线性变化	需存储较多参数
	高斯-马尔可夫移动模型	实时性强	只适用于类似目标移动数据预测

1) 回归算法 数据预测算法的目标是提高预测精度,降低预测算法复杂度。ADAGA-P^[35]将数据预测引入至查询处理过程,将预测算法植入到传感器节点,通过线性回归方程在数据汇聚过程中完成数据预测任务。Sink 节点计算简单线性回归方程的系数,该系数代表一系列节点采样值。Sink 节点将线性回归系数发送给各传感器节点,各节点比较所收集数据与预

测数据的差值,判定差值是否超限,若超限则将数据传送到 sink 节点,sink 节点用收到的数据值重新计算回归方程系数,因此预测精度可以不断调整。线性回归算法的优点在于算法简单,复杂度低,缺点是预测精度较差,需频繁更新预测模型,造成一定能耗。文献[36]通过分析感知数据的相关性,将感知数据作为多元线性回归模型和评估方法的输入集,降低了预测输入集的大小,提高了评估精度,利用多元线性回归数据预测算法预测缺失的感知数据。与简单线性回归相比,计算多元线性回归函数系数更易于提高 sink 节点的预测精度^[37],但其计算复杂度相应有所增加。

2) 人工智能 基因表达式编程^[38,39]是由葡萄牙学者 Ferreira 于 2001 年正式发表的科研成果。该成果是在遗传算法和遗传程序设计的基础上发展起来的基于基因组和表现型组的新型遗传算法,它综合了遗传算法和遗传程序设计具有染色体简单、线性、易于进行遗传操作等优点。GEP-MSDA^[33]考虑多源数据环境,引入基因表达式编程技术,根据传感器节点采集的各种历史数据建立模型来预测未来数据,从而有效地减少网络中的数据传输量,节省节点能量,达到延长网络生命周期的目的,与传统的预测算法相比更有效、更灵活,但难以摆脱基因表达式变成收敛性差、效率低、难以跳出局部最优解等缺陷。

人工神经网络算法是模拟人思维的第二种方式。思维学普遍认为,人类大脑的思维分为抽象(逻辑)思维、形象(直观)思维和灵感(顿悟)思维三种基本方式。这是一个非线性动力学系统,其特色在于信息的分布式存储和并行协同处理。虽然单个神经元的结构极其简单,功能有限,但大量神经元构成的网络系统所能实现的行为却是极其丰富多彩的。1986 年,Rumelhart 等人提出了多层前馈神经网络的学习算法,即 BP 神经网络算法。BP 网络的工作原理对预测算法及模型的研究有很好的启示,但 BP 网络结构设计是非常困难和复杂的问题,特别是确定隐藏层神经元的准确数量,这直接决定了整个网络的泛化能力。目前已有大量文献针对 BP 网络进行改进。

粒子群优化算法^[40,41]在 1995 年由 Kennedy 和 Eberhart 提出,它是一个基于群体的优化算法,由随机解决方案群进行初始化,在工程和计算机科学中作为一种问题解决方法得到了广泛应用。文献[32]提出了一个通过粒子群算法改进的基于 BP 神经网络的数据预测机制。源节点和 sink 节点具有相同的数据预测模型,同时预测数据,当源节点监测到的数据和预测的数据之间的差值不超过设定门限值时,源节点不发送数据,sink 节点取预测值,否则传送数据,并更新预测模型。为使数据预测输入集简单可靠,文献[32]还提出了基于粒子群算法的特征选择算法。文献[42]提出了一种基于时间相关性预测机制的数据融合技术,并将该技术应用于工厂监控,利用粒子群算法改进了 BP 神经网络预测算法的性能,有效降低了具有时间相关性融入的数据传输所带来的能量消耗。

文献[43]提出了自适应人工鱼群算法,并用自适应人工鱼群算法训练 BP 神经网络取得了良好的效果。人工鱼群算法^[44]是一种根据鱼群的行为特点,并应用动物自治体的模型,是一种自下而上设计的新型的寻优策略。经验证,这是一种有效的寻优模式。

群智能算法展现出一定的灵活性和适应性,目前已有大量群智能算法被提出。群智能算法主要启发于不同物种的不同寻优行为,对不同算法的改进也主要通过改进适应度函数和寻

优行为来提高算法收敛性、稳定性和跳出局部最优解的能力。然而,群智能算法收敛性差、易于陷入局部最优解等缺陷依然没有解决。事实上,生物群行为并不能完全反映现实事物特征,依据群智能思想而又不完全依赖才是解决复杂问题的出路。

3) 预测模型 文献[45]提出一种基于预测的数据聚集协议,该协议设计了一个双队列机制,用于传感器节点和 sink 节点的连续预测数据序列的同步,避免了连续预测的累计误差。协议提出了三个基于预测的数据聚集方法,即灰度模型数据聚集(GMDA)、卡尔曼滤波(Kalman-filter)数据聚集(KFDA)、联合灰度模型—卡尔曼滤波数据聚集(CoGKDA)。通过集成灰度模型快速建模的优点和卡尔曼滤波计算数据序列噪声上的优势,CoGKDA 展现出了预测精度高、通信开销低、计算复杂度低等优点。卡尔曼滤波器在解决非线性变化的预测存在较大困难,这一点连续的蒙特卡洛过滤器^[46]表现较好,但所有的过滤器都需要存储大量的参数,这对存储资源有限的无线传感器网络来说无疑也是致命的。

高斯—马尔可夫移动模型是一种用于预测目标轨迹最好的移动模型之一,因为它能及时捕获目标运动速度的相关性。因此,该模型适用于一些类似移动轨迹变化的数据预测;另外它还能够用不同的高斯—马尔可夫参数复制其他常见的移动模型。GMPE_MLH^[47]用最大似然估计的方法替代传统的自相关技术或者递归最小二乘估计技术评估高斯—马尔可夫参数,避免了后者需要大量历史信息的缺陷。同时,GMPE_MLH 模型产生的高斯—马尔可夫参数的估计值和真实值之间的差异是微不足道的,并且能够通过自相关技术或递归最小二乘估计提供移动目标位置的预测值与高斯—马尔可夫参数估计的比较。

文献[48]提出了一种新颖的选择性数据预测模型,认为数据预测模型并不总是节省能量的,并且过于复杂的数据预测算法对于 WSN 的硬件来说是很大的挑战。文中通过公式计算并证明了在何种情况下采用数据预测模型可以降低能耗,并给出了明确的判决公式以决定是否使用数据预测模型。这是首次有学者研究数据预测模型本身对全网能耗的影响。文献[49]构建并提出了一个基于随机融合进行预测能力估计的方法,该方法融合了非稳定性行为和随机通信特征的共同影响,所产生的随机融合方程评估出具有非理想化信道和自组织连接通信网络的分布式估计能力。

文献[50]提出基于预测模型的过滤机制,解决由传输时间冗余数据而造成的能量浪费。该机制由三部分组成:a) 预测模块用于最小化时域变化规则;b) 自学习模块用于更新模型;c) 驱动模块用于数据过滤操作的控制。在同样 QoS 保障下降低了以往以错误驱动规则和门限分布规则为过滤机制中间件时的传输负载。由于随机变量过程被看成灰度变量过程,灰度理论有很强的随机和突变处理能力。支持向量机解决传统学习算法的局部最小点、维度灾难、小样本学习、过度拟合等问题,并且有很强的泛化能力。在特征信息有限的情况下,它能够进一步探索隐含的分类信息,并且预测未来特征信息以提高系统性能。GMSVM^[51]将灰色模型的预测理论引入支持向量机,实现动态实时多步预测,并且用黎曼几何分析和实验时间序列提高 RBF 核函数的计算速度。该方法具有较高的预测精度和灵活性。

综上所述,预测技术在多类型数据融合中取得了很大效果,但没有一种预测技术能满足数据融合的全部需要,因此,依赖于数据类型属性—时间相关性的数据预测技术具有很大的研究空间。针对不同的网络,应采用不同技术相互弥补,才能实现更好的效果。

4 结束语

本文讨论了当前数据融合研究所面临的数据间相关性冗余较为复杂、融合难度大等问题;综述了数据融合研究现状;提出数据类型属性相关性的概念,这种相关性独立于时间—空间相关性之外,又具有一定的时空特征。本文指出,数据预测是解决多类型数据融合的有效方式,并给出了多类型属性相关数据融合研究的一般流程并在各部分指出了技术研究的难点、突破口以及未来的研究重点。

参考文献:

- [1] YICK J, MUKHERJEE B, GHOSAL D. Wireless sensor network survey [J]. *Computer Networks*, 2008, 52(12): 2292-2330.
- [2] ANASTASI G, CONTI M, FRANCESCO M D, et al. Energy conservation in wireless sensor networks: a survey [J]. *Ad hoc Networks*, 2009, 7(3): 537-568.
- [3] 向敏, 石为人. 基于数据关联性的无线传感器网络簇内数据管理算法[J]. *自动化学报*, 2010, 36(9): 1343-1350.
- [4] CAO Dong, QIAO Xiu-quan, JUDITH G, et al. Mining data correlation from multi-faceted sensor data in Internet of things [J]. *China Communications*, 2011, 8(1): 132-138.
- [5] FASOLO E, ROSSI M, WIDMER J, et al. In-network aggregation techniques for wireless sensor networks: a survey [J]. *IEEE Wireless Communications*, 2007, 14(2): 70-87.
- [6] XU Ying-qi, LEE W. Exploring spatial correlation for link quality estimation in wireless sensor networks [C]//Proc of the 4th Annual IEEE International Conference on Pervasive Computing and Communications. 2006: 200-211.
- [7] VURAN M B, AKAN O B. Spatio-temporal characteristics of point and field sources in wireless sensor networks [C]//Proc of IEEE International Conference on Communications. 2006: 234-239.
- [8] JAGGI N, KAR K. Multi-sensor event detection under temporal correlations with renewable energy sources [C]//Proc of the 7th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks Program. 2009: 1-9.
- [9] VINYALS M, RODRIGUEZ-AGUILAR J A, CERQUIDES J. A survey on sensor networks from a multiagent perspective [J]. *The Computer Journal*, 2011, 54(3): 455-470.
- [10] ARDALAN A A, JAFARI M. Multi-sensor approach to settlement analysis of earth dams [J]. *Computational Geosciences*, 2012, 16(1): 123-138.
- [11] KUMAR M R R, DEVASTHALE A, LEVY G, et al. A multi-sensor climatological view of double ITCZs over the Indian Ocean [J]. *International Journal of Remote Sensing*, 2012, 33(9): 2925-2936.
- [12] BHATTACHARYA S, PAL K, PAL S K. Multi-sensor based prediction of metal deposition in pulsed gas metal arc welding using various soft computing models [J]. *Applied Soft Computing*, 2011, 12(1): 498-505.
- [13] GREEN O, BARTZANAS T, LØKKE M, et al. Spatial and temporal variation of temperature and oxygen concentration inside silage stacks [J]. *Biosystems Engineering*, 2011, 111(2): 155-165.

- [14] HEINZELMAN W R, CHANDRAKASAN A, BALAKRISHNAN H. Energy-efficient communication protocol for wireless microsensor networks[C]//Proc of the 33rd Hawaii International Conference on System Sciences. Washington DC:IEEE Computer Society, 2000:8020.
- [15] MARAIYA K, KANT K, GUPTA N. Study of data fusion in wireless sensor network[C]//Proc of International Conference on Advanced Computing and Communication Technologies. 2011:535-539.
- [16] HUANG Yi. Energy efficient distributed data fusion in multihop wireless sensor networks[D]. Riverside: University of California, 2011.
- [17] KRISHNAMACHARI L, ESTRIN D, WICKER S. The impact of data aggregation in wireless sensor networks[C]//Proc of International Conference on Distributed Computing Systems. 2002:575-578.
- [18] VURAN M, AKYILDIZ I. Spatial correlation-based collaborative medium access control in wireless sensor networks[J]. *IEEE/ACM Trans on Networking*, 2006, 14(2):316-329.
- [19] 赵明, 陈志刚, 葛志辉, 等. QS-Sift: 一种基于服务质量和空间相关性的无线传感器网络 MAC 协议[J]. *小型微型计算机系统*, 2006, 26(11):2007-2011.
- [20] ZHENG Guo-qiang, TANG Sheng-yu. Spatial correlation-based MAC protocol for event-driven wireless sensor networks[J]. *Journal of Networks*, 2011, 6(1):121-128.
- [21] ZHAO Ji-jun, WEI Zhong-cheng, LI Zhi-hua, *et al.* An adaptive MAC protocol based on spatial and temporal correlation in wireless sensor networks[J]. *Journal of Communications*, 2011, 6(3):262-269.
- [22] LEE T H, RA W S, YOON T S, *et al.* Robust Kalman filtering via Krein space estimation[J]. *IEEE Proceedings of Control Theory and Applications*, 2004, 151(1):59-63.
- [23] ZHOU Chun-hua, CHEN Bing, GAO Yong, *et al.* A technique of filtering dirty data based on temporal-spatial correlation in wireless sensor network[J]. *Procedia Environmental Sciences*, 2011, 10:511-516.
- [24] 王耀南, 李树涛. 多传感器信息融合及其应用综述[J]. *控制与决策*, 2001, 16(5):518-522.
- [25] LUO Jun, JIANG Ling-ge, HE Chen, *et al.* Two states CBR modeling of data source in dynamic traffic monitoring sensor networks[J]. *Journal of Shanghai Jiaotong University*, 2007, E-12(05):618-622.
- [26] CHANG H, SAPATNEKAR S S. Statistical timing analysis considering spatial correlations using a single PERT-like traversal[C]//Proc of International Conference on Computer Aided Design. 2003:621-625.
- [27] XIONG Jin-jun, ZOLOTOV V, HE Lei. Robust extraction of spatial correlation[J]. *IEEE Trans on Computer-Aided Design of Integrated Circuits and Systems*, 2007, 26(4):619-631.
- [28] BERGER J O, De OLIVEIRA V, SANSÓ B. Objective bayesian analysis of spatially correlated data[J]. *Journal of the American Statistical Association*, 2001, 96(456):1361-1374.
- [29] IMBRIGLIO L, GRAZIOSI F. Robust extraction of temporal correlation from noisy data[C]//Proc of the 69th IEEE Vehicular Technology Conference. 2009.
- [30] DING C, PENG H. Minimum redundancy feature selection from microarray gene expression data[J]. *Journal of Bioinformatics and Computational Biology*, 2005, 3(2):185-205.
- [31] XU X, WANG W. Spatial correlation in wireless sensor networks[C]//Proc of the 6th IEEE International Conference on Wireless Communications, Networking and Mobile Computing. 2010:3111-3114.
- [32] GUO Wen-zhong, XIONG Nai-xue, ATHANASIOS V V, *et al.* Multi-source temporal data aggregation in wireless sensor networks[J]. *Wireless Personal Communications*, 2011, 56(3):359-370.
- [33] 高洪奎, 郭文忠, 陈国龙, 等. 用 GEP 实现传感器网络多源时域数据融合[J]. *计算机工程与科学*, 2009, 31(9):28-31.
- [34] SHARMA A. Exploiting spatial correlation using tree based polynomial regression in a three dimensional wireless sensor network[D]. Cincinnati: University of Cincinnati, 2007.
- [35] MATOS T B, BRAYNER A, MAIA J E B. Toward in-network data prediction in wireless sensor networks[C]//Proc of ACM Symposium on Applied Computing. 2010:592-596.
- [36] YAN Xiao-zhen, XIE Hong, WANG Tong. A multiple linear regression data predicting method using correlation analysis for wireless sensor networks[C]//Proc of Cross Strait Quad-Regional Radio Science and Wireless Technology Conference. 2011:960-963.
- [37] CARVALHO C, GOMES D G, AGOULMINE N, *et al.* Improving prediction accuracy for WSN data reduction by applying multivariate spatio-temporal correlation[J]. *Sensors*, 2011, 11(11):10010-10037.
- [38] FERREIRA C. Gene expression programming: a new adaptive algorithm for solving problems[J]. *Complex Systems*, 2001, 13(2):87-129.
- [39] FERREIRA C. Gene expression programming in problem solving[C]//Proc of the 6th Online World Conference on Soft Computing in Industrial Applications. 2001.
- [40] KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proc of IEEE International Conference on Neural Networks. 1995:1942-1948.
- [41] EBERHART R, KENNEDY J. A new optimizer using particle swarm theory[C]//Proc of the 6th International Symposium on Micro Machine and Human Science. 1995:39-43.
- [42] SUNG Wen-tsai. Multi-sensors data fusion system for wireless sensors networks of factory monitoring via BPN technology[J]. *Expert Systems with Applications*, 2010, 37(3):2124-2131.
- [43] 师彪, 李郁侠, 于新花, 等. 自适应人工鱼群-BP 神经网络算法在径流预测中的应用[J]. *自然资源学报*, 2009, 24(11):2005-2013.
- [44] 李晓磊, 邵线. 一种基于动物自治体的寻优模式: 鱼群算法[J]. *系统工程理论与实践*, 2002, 22(11):32-38.
- [45] WEI Gui-yi, LING Yun, GUO Bin-feng, *et al.* Prediction-based data aggregation in wireless sensor networks: combining grey model and Kalman filter[J]. *Computer Communications*, 2011, 34(6):793-802.
- [46] YANG Z, WANG X. Joint mobility tracking and handoff in cellular networks via sequential Monte Carlo filtering[J]. *IEEE Trans on Signal Processing*, 2003, 51(1):269-281.
- [47] LIU Bing-hong, CHEN Min-lun, TSAI M J. Message-efficient location prediction for mobile objects in wireless sensor networks using a maximum likelihood technique[J]. *IEEE Trans on Computers*, 2011, 60(6):865-878.
- [48] JIANG Hong-bo, JIN Shu-dong, WANG C. Prediction or not? An energy-efficient framework for clustering-based data collection in wireless sensor networks[J]. *IEEE Trans on Parallel and Distributed Systems*, 2011, 22(6):1064-1071.
- [49] MARTIN T W, CHANG K. A data fusion formulation for decentralized estimation predictions under communications uncertainty[C]//Proc of the 9th International Conference on Information Fusion. 2006:1-7.
- [50] HUANG R, ZHANG Z C, XU G H. Predictive model-aided filtering scheme of data-collection in WSN[J]. *Journal of China Universities of Posts and Telecommunications*, 2011, 18(2):17-24.
- [51] KANG Jian, TANG Li-wei, ZUO Xian-zhang, *et al.* GMSVM-based prediction for temporal data aggregation in sensor networks[C]//Proc of the 5th International Conference on Wireless Communications, Networking and Mobile Computing. 2009:3410-3413.