doi:10.6041/j.issn.1000-1298.2015.S0.027

基于改进 K-means 算法的 WSN 簇头节点数据融合*

高红菊　刘艳哲　陈　莎
(中国农业大学信息与电气工程学院, 北京 100083)

摘要: 无线传感器网络数据融合能够减少节点能耗、延长网络生命周期,近年来受到了广泛关注。已有的应用于农业监测的空间数据融合算法多采用取平均值等方法将一定区域内监测到的数据融合成一个值。而农田环境监测具有监测范围广、监测点多、监测数据量大的特点,监测数据间除了冗余性还具有差异性,因此数据融合应该在消除冗余的同时保留数据的差异。针对农业监测的这一特点,提出在簇头节点应用聚类算法进行空间数据融合,通过聚类减少数据发送量,降低能耗;同时将差异较大的参量聚类到不同类别中以保留数据间的差异。此外,还提出了一种应用于 WSN 簇头节点的自适应改进 K-means 聚类算法,仿真结果表明,所提算法融合后的数据上传量比没有融合减少 41.19%,消除了数据冗余;算法融合前后最大误差低于取平均值法误差的 36%,保留了数据差异性。在没有明确误差要求时,该算法能够在尽量减少数据上传量的同时保持相对误差低于 10%,避免了因聚类个数不当引起的巨大误差。而在有具体误差要求时,该算法融合前后的绝对误差严格低于要求误差。

关键词: 无线传感器网络　改进 K-means 算法　数据差异性　数据融合
中图分类号: TP391　　**文献标识码**: A　　**文章编号**: 1000-1298(2015)S0-0162-06

Fusion of WSN Cluster Head Data Based on Improved K-means Clustering Algorithm

Gao Hongju　Liu Yanzhe　Chen Sha
(*College of Information and Electrical Engineering*, *China Agricultural University*, *Beijing* 100083, *China*)

Abstract: Data fusion for wireless sensor networks (WSN) can reduce the energy consumption of sensor nodes and prolong the network lifetime, so that it has attracted wide spread attention in a variety of applications. The existing algorithms for spatial data fusion that have been used in agricultural monitoring always aggregate the data within a certain area into one value by means of averaging. However, in addition to redundancy resulted from correlation, the sensed data also has variance due to larger monitoring area, more monitoring nodes and larger amount of data in farmland environment. Hence, data fusion in farmland monitoring should retain the differences of data while eliminating the redundancy. The idea that applying data fusion algorithm on WSN cluster head to aggregate spatially correlated data by clustering was proposed. While the parameters whose values are quite different will be clustered into different categories so that differences between the data can be reserved. An improved adaptive K-means clustering algorithm was proposed to be used in cluster head. Simulation results indicate that, the amount of data uploaded with fusion algorithm was decreased by 41.19% compared with that without fusion algorithm, and the maximum error before and after the proposed fusion algorithm is less than 36% of that before and after the averaging fusion method. When there is no clear accuracy requirement, the proposed algorithm can reduce the amount of data uploaded and maintain the relative error less than 10%, avoiding enormous error caused by improper number of clusters. When there are specific accuracy requirements, the relative error produced by the proposed algorithm can meet the error requirements strictly.

Key words: Wireless sensor network　Improved K-means algorithm　Data difference　Data fusion

收稿日期: 2015-10-28　修回日期: 2015-11-09
*国家自然科学基金资助项目(31371531)
作者简介: 高红菊,副教授,主要从事无线传感器网络与农业物联网研究,E-mail: hjgao@cau.edu.cn

引言

随着现代农业的不断发展，用无线传感器网络（Wireless sensor network，WSN）对农业产地环境进行监测已经成为开展精细农业、提高农作物产量的重要方法之一[1-4]。在农业产地环境监测系统中，一定空间范围内会部署多个按周期采样的传感器终端节点，这些节点的空间相对距离较近，其监测值具有一定的空间相关性，导致数据的冗余现象[5]。而传感器网络节点能量有限，减少能耗对延长 WSN 生命周期至关重要[6-7]。在传感器网络节点的工作模块中，无线通信的模块耗能最多[8]，若将包含大量冗余信息的数据全部上传，就会消耗过多能量，所以需要应用数据融合技术来消除冗余，减少数据上传量，延长网络寿命。

WSN 数据融合多用于温室，温室数据的特点是监测区域小，数据变化不大，同一时刻各节点采集到的数据非常相近，所以现有的空间数据融合均是在汇聚节点将终端节点上传的数据融合成一个值，从而消除数据冗余[9-11]。与温室监测相比，农田监测具有监测范围广、监测点多、监测数据量大的特点[12]，因此农田监测系统中节点监测的数据除了具有冗余性还具有差异性。目前还没有研究能够在传感器节点上应用空间数据融合算法，在消除农田监测数据空间冗余性的同时保留数据的空间差异性。针对这一空白，本文提出在传感器网络节点上应用聚类算法实现这一功能。

农田无线传感器网络规模大、节点分布不均，多采用分层路由协议[13-14]，空间相近的数据会先被发送到簇头节点，经过处理后再被发送到汇聚节点。本文在簇头节点应用空间数据融合算法，对同一时刻接收到的数据进行聚类，将参量相近的数据聚成一类发送，从而减少数据发送量，消除数据冗余；与此同时差异大的参量被聚类到不同的类别中，保留数据的差异性。

1　网络平台及算法

1.1　无线传感器网络结构

由于农田区域大、节点分布不均，农田监测一般采用分层拓扑结构的 WSN。如图 1 所示，WSN 包括大量终端节点和 1 个汇聚节点。终端节点根据某种规则形成不同的簇，每个簇由一个簇头和其他成员节点组成[15]，他们被部署在需要监测的地区，通过自组织方式构成网络。成员节点将采集到的农作物产地环境信息发送给簇头，簇头对接收到的数据进行融合，并将融合后的数据沿着其他传感器节点逐跳地进行传输，经过多跳路由到达汇聚节点。最后通过 GPRS 网络等上传到互联网中的服务器。用户可以在平台上对数据进行监测和查询。

考虑到已经存在很多种簇头选择算法[16-17]，本文主要研究簇头确定后，在簇头上应用空间数据融合算法。该算法主要作用是将同一时刻该簇内的其他成员节点所采集的数据进行融合，以保证在保留数据差异性的同时能够消除数据冗余。需要说明的是，由于成员节点发送数据的时间并不完全相同，可以将一定时间内接收到的数据看作是同步的数据。例如，若成员节点每隔 1 h 发送一次数据，则簇头节点可以将 1 h 内的数据进行存储，每隔 1 h 进行一次融合，将融合后的数据发送到汇聚节点。

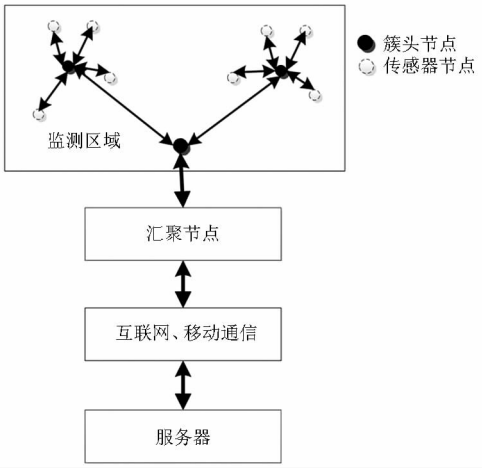

图 1　无线传感器网络拓扑结构

Fig. 1　Topology structure of wireless sensor network

1.2　融合算法

由于 WSN 节点内存较小，计算能力有限，所以选用的算法不宜太过复杂。K-means 算法是一种应用最广泛的基于划分的聚类算法，该算法理论可靠，过程简单，在处理大数据集时，该算法是相对可伸缩和有效的快速收敛算法[18]，因此本文选用 K-means 算法对数据进行融合。

但传统的 K-means 算法需要指定聚类中心个数 k，这就需要提前获得数据的先验知识。而在农田环境监测中先验知识无法获取，因此，能够自适应地确定最佳聚类中心是影响聚类质量的关键因素。本文提出了一种基于密度自适应确定聚类中心的改进 K-means 算法（以下简称改进 K-means 算法），使之能够根据接收到的数据情况以及误差要求自主确定聚类个数并选取合适的聚类中心。该算法能够在没有明确误差要求时自动确定聚类个数，从而在避免误差过大的同时尽量减少聚类个数，减少数据上传量；在有具体误差要求时实现自适应聚类以满足要求。

1.2.1 传统 K-means 算法

K-means 算法是数据挖掘十大经典算法之一，是一种非常经典的基于划分的聚类算法。该算法的基本思想是：以空间中 k 个点为初始中心对已有数据进行聚类，把数据归类到与初始中心最相似的类别中[19]。然后通过迭代的方法，不断更新各聚类中心的值，直至相邻 2 次的聚类中心没有变化为止。

假设要把数据分为 k 个类别，算法的实现基本步骤如下：

(1)选择 k 个类别的初始聚类中心。

(2)在每一次的迭代中，求任意一个样本到各初始聚类中心的距离，将该样本归类到与之相距最短的中心所在的类。

(3)利用均值等方法更新该类的中心值。

(4)对于所有的 n 个聚类中心，如果利用步骤(2)、(3)的迭代法更新后，值保持不变，则迭代结束，否则继续迭代。

在传统 K-means 聚类算法中，聚类中心个数 k 需要预先给定，这就使得 k 有一定的主观性及盲目性。而 k 的选择对聚类结果影响很大，聚类结果会随着 k 的改变而改变，不合适的 k 会影响聚类质量，或使算法陷入局部最优。如果 k 偏小，会造成硬性划分，影响聚类质量；如果 k 偏大，则需要上传的数据量增大，增加了簇头节点的能源消耗。所以选择一个合适的 k 至关重要。

在大田监测中，情况复杂多变，且簇头是自适应选择的，这使得簇头每一次接收到的数据个数以及数据来源不定，导致每一次需要融合的数据情况有很大差异，如果用固定的聚类中心 k，则会影响聚类时间和质量。因此，本文提出自适应改进K-means 算法，该算法可以根据每一次接收到的数据情况自适应地确定聚类中心个数，从而得到更好的聚类结果。

1.2.2 改进 K-means 算法

基于密度自适应确定聚类中心的改进 K-means 算法思想为：根据每一次接收到的参量值的最大值、最小值及数据个数，来确定计算密度的半径 r 和每个聚类中心间的最小距离 d，从而实现自适应聚类。需要强调的是，为了保证每一次的聚类中心个数在合理范围内，合理地确定 d 和 r 至关重要。本文考虑了没有明确误差要求和有具体误差要求 2 种情况。

1.2.2.1 没有明确的融合误差要求

当没有明确的误差要求时，需要根据接收到的数据情况自主确定聚类中心，使得在避免误差过大的同时能够尽量减少聚类个数。对于聚类中心个数 k，很多研究[20]都使用经验规则 $k\leqslant\sqrt{n}$，其中 n 为聚类数据个数，文献[21]也证明了该规则的合理性。本文提出基于密度自适应确定聚类中心的改进 K-means 算法，该算法确定的 d 可以在不同的情况下保证 k 在指定范围内。

在本算法中，d 的计算公式为

$$d=\frac{p_{\max}-p_{\min}}{\sqrt{n}} \tag{1}$$

式中 $p_{\max}$——参量的最大值

$p_{\min}$——参量的最小值

也就是说，聚类中心之间的距离 d_i 满足

$$d_i\leqslant\frac{p_{\max}-p_{\min}}{\sqrt{n}} \tag{2}$$

而聚类中心的个数 k 满足

$$k=\frac{p_{\max}-p_{\min}}{d_i} \tag{3}$$

由式(2)、(3)可知 $k\leqslant\sqrt{n}$，从而保证了每一次的聚类中心个数在合理范围内。

为了减小时间复杂度，本算法的 r 没有再用算法确定，而是根据基于密度的算法的常规经验，直接采用

$$r=\frac{d}{2}=\frac{p_{\max}-p_{\min}}{2\sqrt{n}} \tag{4}$$

1.2.2.2 要求数据的融合值与实际值的差小于 e

在本算法中，只要使 $d=r=e/2$ 即可，该算法会根据接收到的数据情况自主确定聚类个数。

1.2.3 本文算法实现步骤

算法输入为同一时刻簇头节点接收到的 n 个节点传来的节点地址及参量，算法输出为聚类得到的 k 个聚类中心及属于该聚类中心的节点地址。

算法实现基本步骤如下：

(1)将每一次接收到的数据放入结构体数组 Recievedata 中，每个结构体变量包括节点地址 Recievedata. address、节点参量 Recievedata. parameter。找出结构体数组中参量的最大值 $p_{\max}$、参量最小值 $p_{\min}$、以及数据个数 n 或者融合误差 $e/2$。

(2)计算出半径 $r=\frac{p_{\max}-p_{\min}}{2\sqrt{n}}$ 或 $r=\frac{e}{2}$。

(3)算出每个参量 Recievedata. parameter 的密度，即 Recievedata. parameter 在半径 r 内包含的参量个数。

(4)将 Recievedata 结构体数组按 Recievedata. parameter 的密度从大到小排列。

(5)把密度最大值作为第 1 个聚类中心放入聚类中心数组 center 中。

(6)判断下一个 Recievedata. parameter 是否在 center 中所有数据的距离 d 外,即参量值是否与所有的聚类中心的距离均大于 d,若是,则将这个参量值放入 center 中;否则,判断下一参量值,直到判断完接收到的所有参量值。

(7)计算每个参量值与 center 中的初始聚类中心之间的距离,并把他们分配到与之最近的聚类中心所在的类;将每一类的参量值取平均值作为融合值。

(8)上传融合值及属于该融合值的终端节点地址。

算法流程图如图 2 所示。

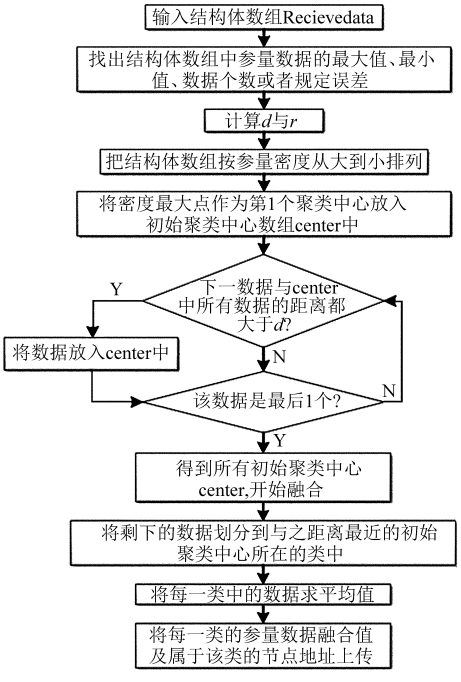

图 2　改进 K-means 算法流程图

Fig. 2　Flowchart of improved K-means algorithm

2　实验结果与分析

为了验证本文算法的有效性,在 Matlab 上对仅求平均值的方法、固定聚类中心个数的 K-means 算法与本文算法分别进行仿真测试,对不同算法的融合结果进行比较。

数据参量采用英特尔伯克利实验室的传感器网络数据中 20 个节点的温度数据。由于该数据集的 5 号、15 号传感器数据较差,所以本实验选择 1 ~ 22 号节点(排除 5 号、15 号)作为融合对象。该实验数据每个节点每 30 s 发送一次数据,而在实际的农田环境监测中,节点一般 1 h 发送一次数据。为了模拟真实情况,筛选出 3 月 1—7 日的每个节点在整点接收到的数据进行融合。也就是每个节点每天选取 24 组数据,7 d 共选取 168 组数据。采用不同算法对每组数据进行聚类。

2.1　没有给定具体误差要求的情况

这种情况下的 k 由每次接收到数据的最大值、最小值及接收到的个数决定。

2.1.1　取平均值法与聚类算法融合前后对比

基于 Matlab R2012 采用取平均值方法与本算法分别对 168 组数据进行融合,得到的融合前后最大温差如图 3 所示。

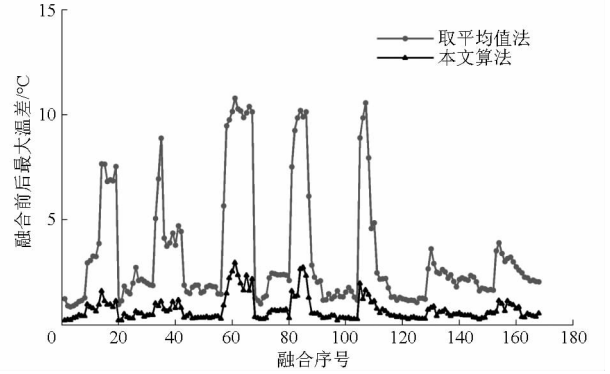

图 3　取平均值法和本文算法融合前后最大温差对比

Fig. 3　Comparison of the average method and the proposed algorithm before and after fusion

由图 3 可以看到,取平均值法融合前后的最大温差明显大于本文算法融合前后的最大温差,证明由于大田监测系统中节点间数据有差异,仅仅对数据取平均值则会造成较大误差。经计算,本文算法得到的融合前后的温差至多为取平均值法的 36%,由此可见,本文算法的性能优于取平均值法,可以在对数据进行融合的同时保留数据的差异性,避免产生较大的融合误差。

2.1.2　聚类中心个数分析

用本文提出的基于密度自适应确定聚类中心的改进 K-means 算法对每一时刻的 20 个温度进行融合,每次融合产生的聚类中心个数如表 1 所示。由于数据较多,这里截取前 24 次融合的聚类中心个数结果。

表 1　本文算法聚类中心个数

Tab. 1　Number of cluster centers of proposed algorithm

融合序号	聚类中心个数	融合序号	聚类中心个数	融合序号	聚类中心个数
1	4	9	3	17	3
2	4	10	3	18	3
3	4	11	4	19	3
4	3	12	3	20	4
5	4	13	3	21	4
6	4	14	2	22	4
7	3	15	3	23	4
8	3	16	3	24	3

从表 1 中可看出,本实验的聚类中心个数有 2、3、4 共 3 种。而实验中用到的终端节点共有 20 个,即 $n=20$;而自适应聚类中心算法要求聚类中心个数小于$\sqrt{n}$,因此,理论上得到的聚类中心个数最大为 4。而本实验得到的聚类中心个数与理论值相符,可见基于密度自适应确定聚类中心的改进 K-means 算法会根据接收到数据的不同自动调整聚类中心个数,从而得到最合适的聚类结果。

不进行数据融合时,每次上传全部 20 个节点的温度,168 次共需要上传 3 360 个温度及 3 360 个节点地址;而聚类后每次上传最多 4 个温度,168 次共上传了 592 个温度及 3 360 个节点地址,数据上传量减少了 41.19%,大大消除了数据冗余性,降低了能耗。

2.1.3 聚类中心个数固定的传统 K-means 算法与本文算法融合误差对比

用本文算法与聚类中心个数 k 分别为 4、3、2 的传统 K-means 算法对 168 组数据进行融合,将融合值与实际值进行对比,得到了每次融合前后的最大温差和最大相对误差。最大温差是指每次参与融合的 20 个节点融合前后温度差的最大值。最大相对误差代表最大温差与实际温度的比值。

4 种算法融合前后的最大温差与最大相对误差如图 4、5 所示,为了图片清晰便于分析,只展示了第 100 ~ 168 次聚类结果。

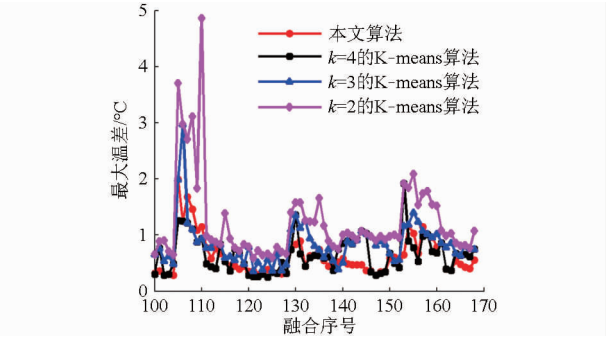

图 4 4 种算法融合前后最大温差

Fig. 4 Maximum temperature difference of four algorithms before and after fusion

为了减少数据上传量,聚类中心的个数应该越少越好,但由图 4 和图 5 可看出,当聚类中心个数选择过小时,在某些时刻会出现巨大误差。例如第 100 ~ 110 次聚类结果中,聚类中心为 2 的 K-means 算法得到的融合前后相对误差高达 20%,这种情况会严重影响监测质量。而当聚类中心个数持续很大时,会导致上传数据量过大,增加能耗。因此根据每次接收到的数据动态确定合适的聚类个数至关重要。

由图 4、5 可以看出,本文算法可以根据接收到

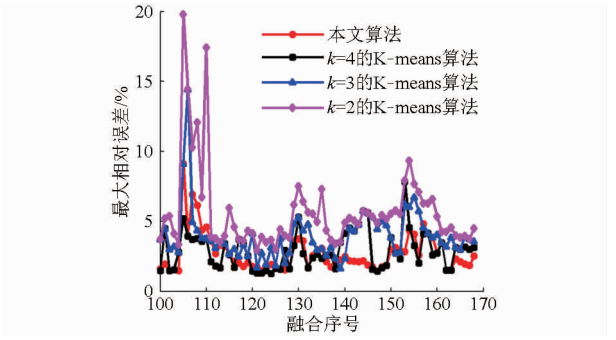

图 5 4 种算法融合前后的最大相对误差

Fig. 5 Maximum relative error of four algorithms before and after fusion

的数据情况自主确定聚类中心个数,从而在尽量减少聚类个数的同时保证融合前后最大相对误差小于 10%。由此可以说明,本文算法可以在保证融合结果理想的情况下减少数据上传量,避免了 K-means 算法因使用固定聚类中心而产生的巨大误差。

2.2 数据的融合值与实际值的差小于 e 的情况

在有些情况下,用户会根据实际情况确定融合误差上限,从而保证融合效果。设要求数据的融合值与实际值的差小于 e,即要求融合前后最大温差小于 e。在这种情况下应用基于密度自适应确定聚类中心的改进 K-means 算法,能够根据误差要求计算半径,从而自主确定聚类中心。

图 6 展示了不同误差要求 e 下的融合前后的最大融合误差。

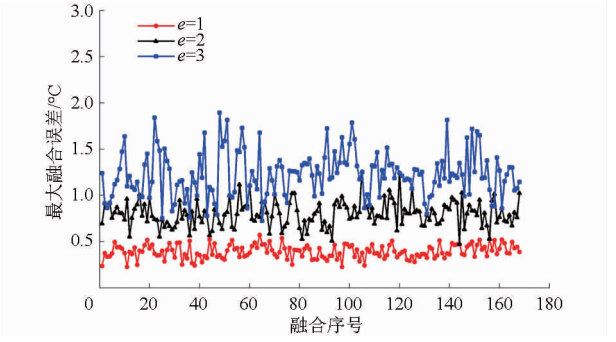

图 6 不同误差要求 e 下的融合误差

Fig. 6 Fusion error with different error requirements e

由图 6 可看出,当误差要求为 e 时,每次聚类得到的融合前后的温差均小于 e。经计算,当 $e=1$ 时,融合误差最大为 0.573 3℃,为要求误差的 57.33%;当 $e=2$ 时,融合误差最大为 1.255 3℃,为要求误差的 62.76%;当 $e=3$ 时,最大误差为 1.895 8℃,为要求误差的 63.19%。由此可见,本文提出的基于密度自适应确定聚类中心的改进 K-means 算法在有明确的误差要求时能够自适应确定聚类中心以满足具体要求。

3　结束语

本文提出的基于密度自适应确定聚类中心的改进 K-means 算法可以根据每次簇头节点接收到的数据情况自适应确定聚类中心，从而在减少数据上传量的同时避免出现过大误差。当没有明确的误差要求时，该算法能够自主确定聚类个数，当有具体的误差要求时，该算法能够根据要求进行聚类以达到聚类误差标准。经仿真结果验证，该算法具有良好的聚类效果，能够在保留数据差异性的同时剔除数据的冗余度，非常适用于农田环境参数监测。

参　考　文　献

1　杨信廷，吴滔，孙传恒，等. 基于 WMSN 的作物环境与长势远程监测系统[J]. 农业机械学报，2013，44(1)：167－173.
Yang Xinting，Wu Tao，Sun Chuanheng，et al. Remote monitoring system of crop environment and growing based on WMSN[J]. Transactions of the Chinese Society for Agricultural Machinery，2013，44(1)：167－173.（in Chinese）

2　Chang Chao，Xian Xiaodong，Hu Ying. Design of precision agriculture remote environment monitoring system based on WSN[J]. Chinese Journal of Sensors & Actuators，2011，24(6)：879－883.

3　葛文杰，赵春江. 农业物联网研究与应用现状及发展对策研究[J]. 农业机械学报，2014，45(7)：222－230.
Ge Wenjie，Zhao Chunjiang. State-of-the-art and developing strategies of agricultural internet of things[J]. Transactions of the Chinese Society for Agricultural Machinery，2014，45(7)：222－230.（in Chinese）

4　Srbinovska M，Gavrovski C，Borozan V，et al. Environmental parameters monitoring in precision agriculture using wireless sensor networks[J]. Journal of Cleaner Production，2015，88：297－307.

5　王玲. 无线传感器网络时空相关性数据融合算法研究[D]. 重庆：重庆大学，2014.
Wang Ling. Study of temporal-spatial correlation based data fusion algorithm in wireless sensor networks[D]. Chongqing：Chongqing University，2014.（in Chinese）

6　Liu C，Zhang R. The WSN energy efficient protocol based on adaptive mechanism of sleep[J]. Open Cybernetics & Systemics Journal，2015，9：478－482.

7　Miao Y S，Yuan L，Wu H R，et al. Optimization of energy heterogeneous cluster-head selection in farmland WSN[J]. Applied Mechanics & Materials，2013，441：1010－1015.

8　丁红雨. 无线传感器网络跨层式通信协议的研究[D]. 合肥：合肥工业大学，2011.
Ding Hongyu. The research on wireless sensor network cross-layer commucation protocols[D]. Hefei：Hefei University of Technology，2011.（in Chinese）

9　Wang X H，Xu L H，Wei R H. A new fusion structure model on greenhouse environment data and a new fusion algorithm of sunlight[C]//2014 International Conference on Wireless Communication and Sensor Network，2014：418－424.

10　熊迎军，沈明霞，陆明洲，等. 温室无线传感器网络系统实时数据融合算法[J]. 农业工程学报，2012，28(23)：160－166.
Xiong Yingjun，Shen Mingxia，Lu Mingzhou，et al. Algorithm of real time data fusion for greenhouse WSN system[J]. Transactions of the CSAE，2012，28(23)：160－166.（in Chinese）

11　王纪章，彭玉礼，李萍萍. 基于事件驱动与数据融合的温室 WSN 节能传输模型[J]. 农业机械学报，2013，44(12)：258－261.
Wang Jizhang，Peng Yuli，Li Pingping. Energy transmission model of WSN in greenhouse based on event-driven and data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery，2013，44(12)：258－261.（in Chinese）

12　焦俊，操俊，潘中，等. 基于物联网的农田环境在线监测系统[J]. 农业工程，2014，4(6)：18－23.
Jiao Jun，Cao Jun，Pan Zhong，et al. Farm environmental online monitoring system based on internet of things[J]. Agricultural Engineering，2014，4(6)：18－23.（in Chinese）

13　孙想，吴保国，吴华瑞，等. 能量高效的农田无线传感器网络拓扑关联路由算法[J]. 农业机械学报，2015，46(8)：232－238.
Sun Xiang，Wu Baoguo，Wu Huarui，et al. A topology based energy efficient routing algorithm in farmland wireless sensor network[J]. Transactions of the Chinese Society for Agricultural Machinery，2015，46(8)：232－238.（in Chinese）

14　Karim L，Nasser N，Salti T E. Efficient zone-based routing protocol of sensor network in agriculture monitoring systems[C]// International Conference on Communications & Information Technology，2011：167－170.

15　Hesham Abusaimeh，Yang Shuang-Hua. Dynamic cluster head for lifetime efficiency in WSN[J]. International Journal of Automation and Computing，2009，6(1)：48－54.

16　Sun B，Gui C，Jia Y，et al. Mobility entropy-based cluster head selection algorithm for Ad Hoc networks[J]. Energy Procedia，2011，6(3)：49－55.

17　Liang Q. Clusterhead election for mobile Ad Hoc wireless network[C]//14th IEEE Proceedings on Personal，Indoor and Mobile Radio Communications，PIMRC 2003，2003：1623－1628.

18　高国琴，李明. 基于 K-means 算法的温室移动机器人导航路径识别[J]. 农业工程学报，2014，30(7)：25－33.
Gao Guoqin，Li Ming. Navigating path recognition for greenhouse mobile robot based on K-means algorithm[J]. Transactions of the CSAE，2014，30(7)：25－33.（in Chinese）

19　Kanungo T，Mount D M，Netanyahu N S，et al. An efficient K-means clustering algorithm：analysis and implementation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2002，24(7)：881－892.

20　Rezaee M R，Lelieveldt B P F，Reiber J H C. A new cluster validity index for the fuzzy c-mean[J]. Pattern Recognition Letters，1998，19(3－4)：237－246.

21　于剑，程乾生. 模糊聚类方法中的最佳聚类数的搜索范围[J]. 中国科学，2002，32(2)：274－280.