

一种面向无线传感器网络的数据汇聚模型

吕林涛¹, 段娟云¹, 李翠²

(1. 西安理工大学计算机学院, 西安 710048; 2. 西安财经学院信息学院, 西安 710061)

摘要: 针对无线传感器网络能量和计算能力严重受限等问题, 通过对数据汇聚技术的研究, 该文提出一种面向无线传感器网络的数据汇聚模型及实现算法。该模型将数据挖掘算法集成在每个传感器网络节点上, 原始数据通过数据挖掘算法处理后, 降低数据维数、消除数据冗余、减少网络通信量、延长网络寿命。仿真结果表明, 该模型具有一定的理论和实用价值。

关键词: 无线传感器网络; 数据汇聚; 数据挖掘

Data Aggregation Model for Wireless Sensor Network

LV Lin-tao¹, DUAN Juan-yun¹, LI Cui²

(1. College of Computer, Xi'an University of Technology, Xi'an 710048;

2. School of Information, Xi'an University of Finance and Economics, Xi'an 710061)

【Abstract】 Aiming at the severe energy and computing resource constraints of Wireless Sensor Network(WSN), based on the analysis and study of data aggregation technology, a data aggregation model for WSN is proposed. Constructed data mining algorithm can be integrated in each sensor network node. Data dimension is reduced and data redundancy is eliminated after the raw-data is processed by data mining algorithm, and the communication traffic is decreased and the life of WSN is extended. Simulation results prove to be highly effective, so it is very valuable in practice.

【Key words】 Wireless Sensor Network(WSN); data aggregation; data mining

1 概述

微电子技术、计算机技术和无线通信技术推动了低功耗、多功能传感器的快速发展, 使其在微小体积内能够集成信息采集、数据处理和无线通信等多种功能。这些微型传感器通过自组织方式就构成了无线传感器网络^[1-2]。

无线传感器网络(Wireless Sensor Network, WSN)的能量、计算能力、存储能力、通信能力受到严重限制。文献[3]提到: 数据传输消耗了总能量的70%, 必须合理利用能量, 努力延长 WSN 工作寿命。而原始监测数据中有大量冗余信息, 通过数据汇聚来减少数据传输量是减少能耗最重要的技术之一。

近年来, 无线传感器网络中基于数据挖掘技术进行数据汇聚已成为计算机工业界与学术界的研究热点。本文基于数据挖掘技术, 提出了一种有效的面向无线传感器网络的数据汇聚模型并加以实现。

2 面向无线传感器网络的数据汇聚模型

目前, 无线传感器网络的数据汇聚主要途径有:

- (1) 以数据为中心的汇聚路由;
- (2) 聚集函数及其安全计算;
- (3) 数据挖掘技术。

途径(1)~途径(2)虽然也可以达到减少数据通信量, 延长传感器节点寿命的目的, 但其在数据汇聚过程中丢失了大量的原始数据结构, 掩饰了令人感兴趣的局部变化。为了获得数据不同粒度的表示, 采用数据挖掘技术进行数据汇聚较为理想。

无线传感器节点能量极其有限, 每一个传感器将其采集

到的数据传送给一个中心服务器, 再在服务器上进行数据挖掘存在一个显著缺点就是能量消耗大、带宽占用也较大, 并且无法灵活地处理大量传感器的情况。目前传感器网络中的数据挖掘研究主要集中在分布式数据挖掘, 要求每一个传感器在本地完成一些计算, 然后将部分结果传递给其他节点, 最后得到无线传感器网络的全局数据汇聚结果。

应用分布式计算模式解决传感器节点数据挖掘。面向无线传感器网络的数据汇聚模型见图1。

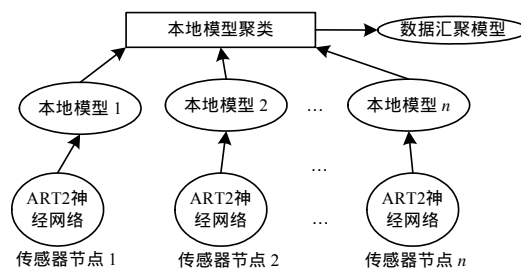


图1 面向无线传感器网络的数据汇聚模型

模型的构造过程如下:

- (1) 基于粗糙集和神经网络构造本地数据挖掘算法;
- (2) 将构造好的算法集成在传感器节点上, 传感器输入数据通过本地算法处理后得到本地数据模型;

基金项目: 陕西省自然科学基金资助项目(07JK339)

作者简介: 吕林涛(1954 -), 男, 教授, 主研方向: 计算机网络, 网络信息安全, 数据挖掘; 段娟云、李翠, 硕士

收稿日期: 2007-10-15 **E-mail:** andy5127@163.com

(3)通过分布式聚类得到数据汇聚模型。

3 面向无线传感器网络的数据汇聚模型实现

本文提出了一种基于粗糙集与神经网络技术实现 WSN 的数据汇聚模型。实现过程描述如下：

- (1)获取数据库中的原始数据，利用粗糙集离散化算法得到初始决策表；
- (2)利用粗糙集属性约简算法对初始决策表进行属性约简，得到最简决策表；
- (3)用最简决策表训练 ART2 神经网络；
- (4)将构造好的 ART2 神经网络集成在传感器网络节点上。通过分布式数据聚类获得传感器网络的数据汇聚模型。面向无线传感器网络的数据汇聚算法流程如图 2 所示。

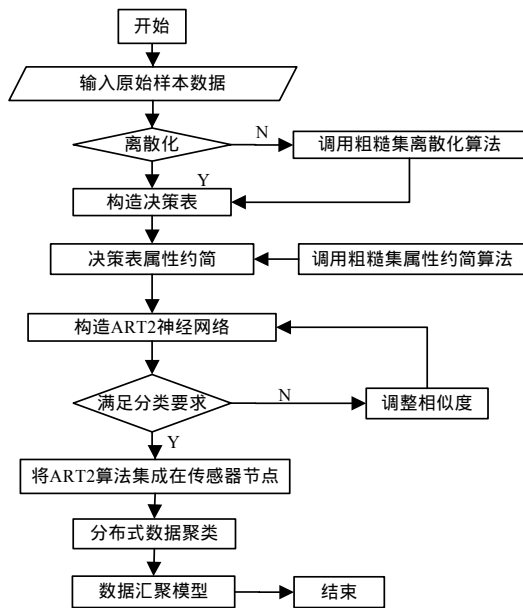


图 2 算法流程

3.1 粗糙集离散化算法

粗糙集理论研究的元素只能是离散值对象，但传感器数据中存在值为实数值的连续数据，需要进行离散化。本文采用离散化算法——无监督等距离划分算法。该算法根据给定参数，在不考虑每个断点段中属性值个数的多少情况下，将属性值域划分为等距离的断点段。基于无监督等距离划分的高散化算法描述如下：

输入 由传感器原始数据得到条件属性集 $C = \{c_1, c_2, \dots, c_n\}$ 。

输出 决策表 $T = \langle C, D \rangle$ ，其中， $D = \{d_1, d_2, \dots, d_n\}$ 为决策属性集； $C = \{c_1, c_2, \dots, c_n\}$ 为离散化后的条件属性集。

(1)计算每个条件属性值域的断点个数，并按断点个数对条件属性 c_i 进行降序排序。

(2) $d=2$ ；/*步长初始化*/

(3) $\text{int } i, j$ ；

(4) $n = \text{content}(C)$ ；/*计算条件属性个数 n */

(5) $\text{while}(i < n+1)\{$

If $i \neq 1$ then $d = C_{\text{imax}}/m$ ；/* C_{imax} 表示 c_i 值域中的最大值*/

$C_{\text{imax}} = \max(c_i)$ ； $C_{\text{imin}} = \min(c_i)$ ；/* C_{imin} 表示 c_i 值域中的最小值*/

$m = C_{\text{imax}}/d$ ；/* m 为新断点集的个数*/

$c_{ij} = C_{\text{imin}} + d \times (j-1)$ ；

/* c_{ij} 为离散化后条件属性 c_i 的第 j 个断点的值*/

$i++$ ；

}end while

(6)离散化结束；

(7)输出决策表 T 。

3.2 粗糙集属性约简算法

对决策表 T 进行属性约简，在保持决策表 T 决策属性和条件属性之间的依赖关系不发生变化的前提下，去除决策表中的冗余属性。约简决策表 T 的过程就是抽取分类规则的过程。属性约简是粗糙集的核心技术之一，目前有多种算法如：基于加权平均属性重要度的属性约简方法，基于二进制分辨的粗糙集属性约简算法等。经研究分析，本文采用基于二进制分辨的粗糙集属性约简算法。

基于二进制分辨的粗糙集属性约简算法描述如下：

输入 决策表 $T = \langle U, C \cup \{d\}, V, f \rangle$ ，其中， $U = \{u_1, u_2, \dots, u_n\}$ ， $C = \{c_1, c_2, \dots, c_m\}$ 。

输出 约简后的决策表 Redu 。

(1)初始化 $\text{Redu} = C$ ，建立决策表 T 的二进制分辨矩阵 M ；

(2)删除 M 中所有单元均为“0”的行；

/*删除不一致实例对*/

(3) $\text{While}(M \neq \emptyset)\{$

计算 M 中每个选择 $c_i \in C$ 列对应的“1”单元个数最大值，设为 $\text{Number}(c_i)$ ；令 $\text{Number}(c_i)$ 为其中的最大值；

$\text{Redu} = \text{Redu} \cup \{c_i\}$ ；

从 M 中删除 c_i 列中单元值为“1”的所有行；

从 M 中删除 c_i 列；}endwhile

(4)假设 $\text{Redu} = \{r_1, r_2, \dots, r_k\}$ 中包含了 k 个属性，它们以被选入 Redu 的先后顺序排列， r_k 为最后一个被选入的属性；

(5)建立决策表 $\text{TR} = \langle U, \text{Redu} \cup \{d\}, V, f \rangle$ 的二进制分辨矩阵 MR ；

(6)删除 MR 中所有单元均为“0”的行；

(7)for $i=2$ to k {

删去 MR 中的 r_i 列；

If(MR 中没有出现单元值全为“0”的行)

$\text{Redu} = \text{Redu} - \{r_i\}$ ；

Else

将 r_i 列放回 MR 中；

Endif }endfor

步骤(1)~步骤(3)选择整个条件属性 C 的一个子集，它能够区分所有 C 能区分的实例。也就是说，它具有和整个条件属性相同的分类能力。这种关系可以使用粗集理论的属性依赖度来表示。步骤(4)~步骤(7)则删除这个子集中冗余的属性，以使约简达到最小。

3.3 构造 ART2 神经网络

图 3 给出了 ART2 网络一个处理单元的结构。

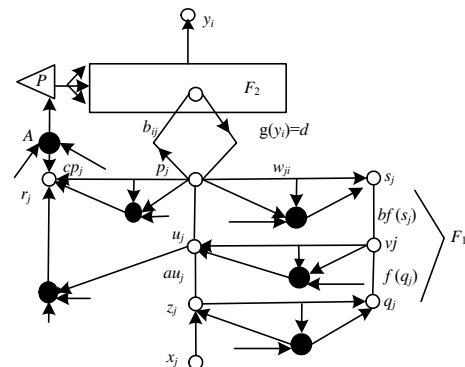


图 3 ART2 神经网络结构

本文采用 ART2(Adaptive Resonance Theory)实现无线传

传感器网络数据汇聚。ART2 分为 2 层 :比较层 F_1 和识别层 F_2 , F_1 和 F_2 构成系统的短期记忆 (Short Term Memory, STM) ,它们之间的连接通路构成系统的长期记忆(Long Term Memory, LTM)。 F_1 层分为上、中、下 3 层完成对输入矢量的去噪、归一化处理以及将处理结果与某一类模式中进行线性组合。 F_2 层完成神经元之间的竞争学习。

ART2 神经网络算法描述如下 :

输入 输入矢量 $X=[x_0, x_1, \dots, x_{n-1}]^T$,则在 F_1 中有相应的 N 个处理单元。

输出 输出矢量 $\{Y\}$ 是一个 M 维实数向量 , $\{Y\}=[y_0, y_1, \dots, y_{M-1}]^T$,在 F_2 中有 M 个神经元 , 分别对应待识别的样本模式类别。

(1)初始化

$W_{ij}(0)=1/(n+1)$; /* W_{ij} 为 F_1 到 F_2 的权系数 */

$W_{ji}=1$; /* W_{ji} 为 F_1 到 F_2 的权系数 */

$\theta_j=1$; /* 警戒值 */

int $L=M$;

(2)经过 F_1 的计算得到输出向量 S 。

(3)进行匹配度计算(K_1)

for F_2 每个神经元 j {

$t_j=\sum_i W_{ij} \cdot S_i$; /* t_j 为 j 的激活值 ; S_i 为比较层神经元的输出 */

}end for

(4)选择最优匹配神经元 c

If 激活量 $t_c=\max(t_j)$ then 输出 y_j ;

$y_j=f(t_j)=\begin{cases} 1 & t_j > \theta_j \\ 0 & t_j \leq \theta_j \end{cases}$

/* θ_j 是神经元 j 的阈值 , y_j 是神经元 j 的输出 */

(5)比较和试验警戒值

If $R=|S|/|X| > \rho$; /* S 为 F_1 层的输出向量; X 为输入向量 */

then goto K_3 ;

(6)最优匹配无效及其处理(K_2)

If ($L=$)

/* 识别层中神经元所取得的相似率(R)都小于警戒值(ρ) */

then 网络自动增加一类模式用来表示新的模式类别;

else { $L=L-1$;

系统发出 Reset 信号 ,置神经元 c 为零 ,不允许其再参与竞争 ,

goto K_1 ;}end if

(上接第 99 页)

参考文献

- [1] Clark D. The Design Philosophy of the DARPA Internet Protocols[C]//Proceedings of the ACM Symposium on Communications Architectures and Protocols. Stanford, CA, USA: ACM Press, 1988.
- [2] Johnson D, Perkins C, Arkko J. Mobility Support in IPv6[EB/OL]. (2004-06-10). <http://www.ietf.org/rfc/rfc3775>.
- [3] Moskowitz R, Nikander P. Host Identity Protocol Architecture[EB/OL]. (2004-01-11). <http://www.ietf.org/internet-drafts/draft-ietf-hip-arch-02.txt>.

(7)ART2 网络通过自学习对警戒值 ρ 进行调整以满足分类要求 (K_3)。

4 结束语

仿真实验采用文献[4]的实验数据作为原始数据集,对原始数据集经过离散化后得到 480 个样本数据。将 480 个样本数据集分为 300 个训练数据集和 180 个验证数据集。利用原始训练数据集和经过粗糙集处理过的训练数据集分别训练文献[5]提出的 ART1 神经网络和本文提出的 ART2 神经网络。实验结果表明,经过粗糙集处理过的训练数据集个数为 212 个,比原始的要小,训练网络的时间也相应地缩短。

网络训练结束后,为了对比这两种网络的精度,利用验证数据集对这两个网络进行分类试验,从测试结果分析得出,利用 ART1 神经网络直接进行分类的整体精度为 82.36%,粗糙集 ART2 神经网络的整体精度为 90.12%。由此可以看出利用粗糙集对原始数据处理后,不但增加了网络的训练时间,而且提高了 ART2 神经网络的精度。

利用本模型进行数据汇聚,无线传感器网络数据通信中所传输的不是高维、冗余的原始数据,而是数据模型,具有良好的应用前景。

参考文献

- [1] 李建中, 李金宝, 石胜飞. 传感器网络及其数据管理的概念、问题与进展[J]. 软件学报, 2003, 14(10): 1717-1727.
- [2] 屠燕春, 郭爱煌. 无线传感器网络的路由算法与仿真实现[J]. 计算机工程, 2006, 32(22): 124-126.
- [3] 张建明, 宋迎清, 周四望, 等. 无线传感器网络中数据汇聚技术的研究[J]. 计算机应用, 2006, 26(6): 1273-1278.
- [4] 雷霖, 陈锋, 代传龙, 等. 基于 Rough Sets 的传感器异常数据处理[J]. 电子科技大学学报, 2006, 35(4): 678-681.
- [5] Ulavok A, Davcev D. Data Mining in Wireless Sensor Networks Based on Artificial Neural Networks Algorithms[C]//Proc. of the 1st International Workshop on Data Mining in Sensor Networks of SDM. [S. l.]: SIAM Press, 2005: 10-16.

- [4] Eriksson J. PeerNet: Pushing Peer-to-Peer down the Stack[C]//Proc. of the Int'l Workshop on Peer-to-peer Systems. Berkeley, USA: [s. n.], 2003.
- [5] Balakrishnan H, Lakshminarayanan K, Ratnasamy S. A Layered Naming Architecture for the Internet[C]//Proc. of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. Portland, Oregon, USA: ACM Press, 2004.
- [6] Yumiba H, Imai K, Yabusaki M. IP-based IMT Network Platform[J]. IEEE Pers. Commun., 2001, 8(5): 18-23.