

无线传感器网络下线性支持向量机分布式协同训练方法研究

及歆荣^{①②} 侯翠琴^① 侯义斌^{*①}

^①(北京工业大学嵌入式软件与系统研究所 北京 100124)

^②(河北工程大学信息与电气工程学院 邯郸 056038)

摘要: 针对无线传感器网络中分散在各节点上的训练数据传输到数据融合中心集中训练支持向量机(Support Vector Machine, SVM)时存在的高通信代价和高能量消耗问题, 该文研究了仅依靠相邻节点间的相互协作, 在网内分布式协同训练线性 SVM 的方法。首先, 在各节点分类器决策变量与集中式分类器决策变量相一致的约束下, 对集中式 SVM 训练问题进行等价分解, 然后利用增广拉格朗日乘法, 对分解后的 SVM 问题进行求解和推导, 进而提出基于全局平均一致性的线性 SVM 分布式训练算法(Average Consensus based Distributed Supported Vector Machine, AC-DSVM); 为了降低 AC-DSVM 算法中全局平均一致性的通信开销, 利用相邻节点间的局部平均一致性近似全局平均一致性, 提出基于一次全局平均一致性的线性 SVM 分布式训练算法(Once Average Consensus based Distributed Supported Vector Machine, 1-AC-DSVM)。仿真实验结果表明, 与已有算法相比, AC-DSVM 算法的迭代次数和数据传输量略高, 但其能够完全收敛到集中式训练结果; 1-AC-DSVM 算法具有较好的收敛性, 而且在收敛速度和数据传输量上也表现出显著优势。

关键词: 无线传感器网络; 支持向量机; 分布式学习; 增广拉格朗日乘法; 平均一致性

中图分类号: TP393; TP181

文献标识码: A

文章编号: 1009-5896(2015)03-0708-07

DOI: 10.11999/JEIT140408

Research on the Distributed Training Method for Linear SVM in WSN

Ji Xin-rong^{①②} Hou Cui-qin^① Hou Yi-bin^①

^①(Embedded Software and Systems Institute, Beijing University of Technology, Beijing 100124, China)

^②(School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, China)

Abstract: In Wireless Sensor Network (WSN), transferring all training samples distributed across different nodes to a centralized fusion center for training Support Vector Machine (SVM) significantly increases the communication overhead and energy consumption. Therefore, this paper studies the distributed training approach for linear SVM through the collaboration of neighboring nodes within the networks. First, the centralized linear SVM problem is cast as the solution of coupled decentralized convex optimization sub-problems with consensus constraints on the classifier parameters. Second, the distributed linear SVM problem is solved and derived using the augmented Lagrange multipliers method, and a novel distributed training algorithm, called Average Consensus based Distributed Supported Vector Machine (AC-DSVM), is proposed. To decrease the communication overhead of global average consensus, an improved distributed training algorithm, named Once Average Consensus based Distributed Supported Vector Machine (1-AC-DSVM), is presented, which is only based on once global average consensus. Simulation results show that compared with existing algorithms, AC-DSVM has slightly higher iterations and data traffic, but can converge to the centralized training results; 1-AC-DSVM not only has better convergence, but also has remarkable advantage in convergence speed and data traffic.

Key words: Wireless Sensor Network (WSN); Support Vector Machine (SVM); Distributed learning; Augmented Lagrange multiplier method; Average consensus

1 引言

在无线传感器网络(Wireless Sensor Network,

WSN)的众多应用中, 对监测到的信息进行分类是最重要也是最基础的一项任务^[1,2]。支持向量机(Support Vector Machine, SVM)作为一款性能非常出色的分类器, 因其坚实的理论基础和良好的分类效果, 在 WSN 中得到了越来越广泛的应用^[3-5]。然

2014-03-27 收到, 2014-07-07 改回

国家自然科学基金青年基金(61203377)资助课题

*通信作者: 侯义斌 yhou@bjut.edu.cn

而,在 WSN 中, SVM 的训练数据都分散在各个传感器节点上,仅依靠单个节点上的数据训练效果较差;而通过多跳路由将所有训练数据传输到数据融合中心进行集中式训练,需要占用大量的带宽、消耗大量的能量,这与 WSN 节点上能源替代代价非常高甚至不可替换、带宽资源非常有限相冲突,同时也容易使数据中心周围的节点成为整个系统的瓶颈。针对上述问题,通过相邻节点间的相互协作,在网内分布式协同训练 SVM 的方法研究,近年来引起越来越多研究者的关注。

在 SVM 分布式协同训练的相关研究中,专门针对 WSN 的特性提出训练 SVM 的算法比较少。其中,文献[6,7]针对 WSN 的特性,提出了以在相邻节点间传输支持向量的协作方式,每个节点使用局部训练样本和从邻居节点发送过来的支持向量进行增量训练,将全部节点训练完后得到的支持向量集合作为集中式训练结果的近似。由于该算法对每个节点的增量训练只进行一次,所以训练结果收敛性差,而且当支持向量规模很大时,通信代价仍然会很大。文献[8,9]提出了基于一致性约束的分布式线性 SVM 的训练算法 MoM-DSVM,该算法在相邻节点决策变量相等的约束下,利用并行优化技术,以在相邻节点间传递少量决策结果的方式,经过多次迭代来实现各节点训练结果的一致。由于仅依靠相邻节点间的一致性约束,该算法存在收敛速度慢和收敛性差等缺点。此外,为了解决 SVM 在大规模数据或分布式数据上的训练问题,学者们提出了许多并行或分布式训练 SVM 的算法^[10-12],但这些算法没有考虑 WSN 资源受限的特性,不适于 WSN 应用。

针对已有研究中存在的通信代价高、收敛性差和收敛速度慢等问题,本文致力于研究在不依赖数据融合中心的 WSN 中,仅依靠相邻节点间的相互协作,在各节点分类器决策变量与集中式分类器决策变量相一致的约束下,利用分布式优化方法,对分布式训练线性 SVM 的问题进行求解和推导。基于此,提出了基于平均一致性的分布式线性 SVM 训练算法 AC-DSVM 和改进算法 1-AC-DSVM(once Average Consensus based Distributed SVM)。仿真实验从收敛性、收敛速度和通信代价 3 个方面对提出的两个算法进行实验验证,并与已有算法进行对比分析。

2 分布式线性 SVM 协同训练

2.1 问题描述

考虑一个由 J 个节点构成的 WSN。为了清楚表达 WSN 节点间的相邻关系,使用无向图 $G = \{V, E\}$

对其进行表示,其中, V 表示 WSN 中 J 个节点的集合, E 代表相邻节点间边的集合,节点 J 的所有邻居节点用集合 B_j 表示。此外,假定图 G 是连通的,即图中的任意两个节点之间至少存在一条路径,也就是图中不存在孤立节点。为了方便表述节点上的训练样本,每个节点 $j \in J$ 上的训练样本都用一个训练样本集 $S_j := \{(x_{jn}, y_{jn}) : n = 1, 2, \dots, N_j\}$ 来表示,其中 $x_{jn} \in \mathbf{R}^p$ 是训练样本的特征向量, p 是特征向量维数, $y_{jn} \in \mathbf{Y} := \{-1, 1\}$ 是训练样本对应的分类标签, N_j 是训练样本的数量。本文针对线性 SVM 训练的问题,研究在 WSN 下,仅依赖相邻节点间的相互协作,在网内分布式协同训练 SVM 的方法。

2.2 线性 SVM 训练问题的分解和推导

SVM 的训练问题可以归结为求解一个受约束的凸二次规划问题^[13]。在 WSN 中,训练样本都分散在各个传感器节点上,而且不同传感器节点上的训练样本个数有可能不同。因此,本研究采用了文献[14]中给出的与训练样本数量相关的二次规划问题形式。线性 SVM 问题使用此形式的表示如式(1):

$$\left. \begin{aligned} \min_{\mathbf{W}, b, \xi} \quad & \frac{1}{2} \|\mathbf{W}\|^2 + \frac{C}{N} \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & y_i (\mathbf{W}^T \cdot \mathbf{x}_i - b) \geq 1 - \xi_i \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, N \end{aligned} \right\} \quad (1)$$

在式(1)中, \mathbf{W} 和 b 分别表示要求解的最优分类超平面的权向量和阈值。 \mathbf{x}_i 和 y_i 对应的是第 i 个训练样本的特征向量和类别标签。 ξ_i 和 C 分别表示松弛变量和惩罚因子,是对个别离群点训练样本进行间隔放松和施加惩罚, ξ_i 是一个未知量,而 C 是一个给定值。 N 表示训练样本的个数。

假定分散在各个传感器节点上的训练样本都可以集中起来进行 SVM 训练,此时,线性 SVM 训练问题对应的凸二次规划问题形式如式(2):

$$\left. \begin{aligned} \min_{\mathbf{W}, b, \{\xi_{jn}\}} \quad & \frac{1}{2} \|\mathbf{W}\|^2 + \frac{C}{\sum_{j=1}^J N_j} \sum_{j=1}^J \sum_{n=1}^{N_j} \xi_{jn} \\ \text{s.t.} \quad & y_{jn} (\mathbf{W}^T \cdot \mathbf{x}_{jn} - b) \geq 1 - \xi_{jn} \\ & \xi_{jn} \geq 0, \quad \forall j \in J, n = 1, 2, \dots, N_j \end{aligned} \right\} \quad (2)$$

式(2)是由 J 组分散的训练样本构成的集中式训练对应的凸二次规划问题形式,和同样训练样本条件下不分组情况下的表示形式式(1)是完全一致的。为了有利于 SVM 训练在各节点上进行,将式(2)变换为等价的易于分解的形式^[15],如式(3):

$$\left. \begin{aligned} \min_{\{\mathbf{W}_j, b_j, \xi_{jn}\}} & \frac{1}{2} \sum_{j=1}^J \|\mathbf{W}_j\|^2 + J \frac{C}{\sum_{j=1}^J N_j} \sum_{j=1}^J \sum_{n=1}^{N_j} \xi_{jn} \\ \text{s.t.} & y_{jn} (\mathbf{W}_j^T \cdot \mathbf{x}_{jn} - b) \geq 1 - \xi_{jn} \\ & \xi_{jn} \geq 0 \\ & \mathbf{W}_j = \mathbf{W}^*, b_j = b^*, \quad \forall j \in J, n=1, 2, \dots, N_j \end{aligned} \right\} (3)$$

其中, \mathbf{W}^* 和 b^* 是式(2)中集中式训练时要求解的权向量和阈值, 而 \mathbf{W}_j 和 b_j 是节点 j 上要求解的权向量和阈值。式(2)和式(3)的等价是通过等式约束 $\mathbf{W}_j = \mathbf{W}^*$, $b_j = b^*$ 来保证的。式(3)的分解形式对应到每个节点 j 上, 即每个节点 j 上求解的二次规划问题形式, 如式(4):

$$\left. \begin{aligned} \min_{\{\mathbf{W}_j, b_j, \xi_{jn}\}} & \frac{1}{2} \|\mathbf{W}_j\|^2 + \frac{C}{N_j} \sum_{n=1}^{N_j} \xi_{jn} \\ \text{s.t.} & y_{jn} (\mathbf{W}_j^T \cdot \mathbf{x}_{jn} - b) \geq 1 - \xi_{jn} \\ & \xi_{jn} \geq 0 \\ & \mathbf{W}_j = \mathbf{W}^*, b_j = b^*, \quad \forall j \in J, n = 1, 2, \dots, N_j \end{aligned} \right\} (4)$$

为求解在每个节点上如式(4)的优化问题, 利用增广拉格朗日方法对等式约束 $\mathbf{W}_j = \mathbf{W}^*$ (因为 b_j 或 b^* 的结果是由 \mathbf{W}_j 和 \mathbf{W}^* 计算得到的) 构建增广拉格朗日函数^[16,17], 如式(5):

$$\left. \begin{aligned} \min L(\mathbf{W}_j, b_j, \xi_{jn}, \mathbf{p}_j) \\ = \min \left\{ \frac{1}{2} \|\mathbf{W}_j\|^2 + \frac{C}{N_j} \sum_{n=1}^{N_j} \xi_{jn} + \mathbf{p}_j(t)^T \right. \\ \left. \cdot (\mathbf{W}_j - \mathbf{W}^*) + \frac{c(t)}{2} \|\mathbf{W}_j - \mathbf{W}^*\|_2^2 \right\} \\ \text{s.t.} & y_{jn} (\mathbf{W}_j^T \cdot \mathbf{x}_{jn} - b) \geq 1 - \xi_{jn} \\ & \xi_{jn} \geq 0, \quad \forall j \in J, n = 1, 2, \dots, N_j \end{aligned} \right\} (5)$$

其中, 由 $\mathbf{W}_j - \mathbf{W}^*$ 构成的二次项是增广拉格朗日方法中用以保证原目标函数的严格凸性^[15]。 $\mathbf{p}_j(t)$ 是等式约束 $\mathbf{W}_j = \mathbf{W}^*$ 上的拉格朗日乘子系数, $c(t)$ 是一个大于零的常数。对式(5)问题的求解, 使用乘子方法^[15,16], 该方法采用乘子 $\mathbf{p}_j(t+1)$ 的迭代形式, 如式(6)所示。其中, $\mathbf{W}^*(t+1)$ 是集中式训练结果的权向量, 是一个公共的耦合变量, 即在每个节点 j 上都存在该变量, 为此, 我们基于当变量 $\mathbf{W}_j(t+1)$ 的值越接近 $\mathbf{W}^*(t+1)$ 时 $\mathbf{p}_j(t+1)$ 的结果应该趋近 0 的事实, 利用 $\mathbf{p}_j(t+1)$ 的迭代形式, 对各节点得到的局部结果 $\mathbf{W}_j(t)$, $\mathbf{p}_j(t)$ 求平均来得到 $\mathbf{W}^*(t+1)$ 的迭代形式, 如式(7)所示。当 $\mathbf{W}^*(t+1)$ 的值确定后, 各个节点就可以对式(5)的优化问题进行求解, 从而求出 $\mathbf{W}_j(t+1)$, 如式(8)所示。这样, 对于式(5)问题的求解可以根据式(6)–式(8)的迭代形式进行求解。

$$\mathbf{p}_j(t+1) = \mathbf{p}_j(t) + c(t) [\mathbf{W}_j(t+1) - \mathbf{W}^*(t+1)] \quad (6)$$

$$\mathbf{W}^*(t+1) = \frac{\sum_{j=1}^J \mathbf{W}_j(t)}{J} - \frac{\sum_{j=1}^J \mathbf{p}_j(t)}{Jc(t)} \quad (7)$$

$$\mathbf{W}_j(t+1) = \arg \min_{\mathbf{W}_j \in P_j} \left\{ \frac{1}{2} \|\mathbf{W}_j\|^2 + \frac{C}{N_j} \sum_{n=1}^{N_j} \xi_{jn} + \mathbf{p}_j(t)^T (\mathbf{W}_j - \mathbf{W}^*) + \frac{c(t)}{2} \|\mathbf{W}_j - \mathbf{W}^*\|_2^2 \right\} \quad (8)$$

其中, 式(8)中 $\mathbf{W}_j \in P_j$, P_j 表示式(5)中的约束条件; 各个节点上对 $\mathbf{W}_j(t+1)$ 的求解, 利用拉格朗日对偶理论^[13], 构建拉格朗日函数, 如式(9):

$$\begin{aligned} L(\mathbf{W}_j, b_j, \xi_{jn}, \lambda_{jn}, \mu_{jn}) \\ = \frac{1}{2} \|\mathbf{W}_j\|^2 + \frac{C}{N_j} \sum_{n=1}^{N_j} \xi_{jn} + \mathbf{p}_j(t)^T (\mathbf{W}_j - \mathbf{W}^*) \\ + \frac{c(t)}{2} \|\mathbf{W}_j - \mathbf{W}^*\|_2^2 \\ - \sum_{n=1}^{N_j} \lambda_{jn} [y_{jn} (\mathbf{W}_j^T \cdot \mathbf{x}_{jn} - b) - 1 + \xi_{jn}] \\ - \sum_{n=1}^{N_j} \mu_{jn} \xi_{jn}, \quad \forall j \in J, n = 1, 2, \dots, N_j \end{aligned} \quad (9)$$

其中, λ_{jn} 和 μ_{jn} 为拉格朗日乘子, 对偶问题形式如式(10)所示。

$$\left. \begin{aligned} \max L(\mathbf{W}_j, b_j, \xi_{jn}, \lambda_{jn}, \mu_{jn}) \\ \text{s.t.} & \frac{\partial L}{\partial \mathbf{W}_j} = 0 \\ & \frac{\partial L}{\partial b_j} = 0 \\ & \frac{\partial L}{\partial \xi_{jn}} = 0 \\ & \lambda_{jn} \geq 0 \\ & \mu_{jn} \geq 0, \quad \forall j \in J, n = 1, 2, \dots, N_j \end{aligned} \right\} (10)$$

其中, 约束 $\frac{\partial L}{\partial \mathbf{W}_j} = 0$ 可以推导出 $\mathbf{W}_j(t+1)$ 的迭代形式, 如式(11); 而约束 $\frac{\partial L}{\partial b_j} = 0$ 和 $\frac{\partial L}{\partial \xi_{jn}} = 0$ 可以分别推导出式(12)和式(13)中的结果。

$$\mathbf{W}_j(t+1) = \mathbf{u}_j^{-1} [\mathbf{p}_j(t) + c(t)\mathbf{W}^*(t+1) + \mathbf{A}_j^T \boldsymbol{\lambda}_j] \quad (11)$$

$$\sum_{n=1}^{N_j} \lambda_{jn} y_{jn} = 0, \quad \forall j \in J, n = 1, 2, \dots, N_j \quad (12)$$

$$0 \leq \lambda_{jn} \leq \frac{C}{N_j}, \quad \forall j \in J, n = 1, 2, \dots, N_j \quad (13)$$

在式(11)中, $\mathbf{A}_j = \mathbf{Y}_j \mathbf{X}_j$, \mathbf{Y}_j 是节点 j 上的所有训练样本的标签列向量, 而 \mathbf{X}_j 是对应的所有训练样本的特征向量构成的特征矩阵; $\mathbf{u}_j = (1 + c(t))\mathbf{I}_{p+1}$, \mathbf{I}_{p+1}

是 $p+1$ 维的单位矩阵。 λ_j 是由该节点上所有训练样本的 λ_{jn} 构成的列向量, 是未知的。将式(11), 式(12), 式(13)的结果代入式(10)的目标函数中, 可以得到如式(14)所示的目标函数:

$$\begin{aligned} L(\mathbf{W}_j, b_j, \xi_{jn}, \lambda_{jn}, \mu_{jn}) \\ = \left[\mathbf{1}_j - \mathbf{A}_j \mathbf{u}_j^{-1} (\mathbf{p}_j(t) + c(t) \mathbf{W}^*(t+1)) \right]^T \\ \cdot \lambda_j - \frac{1}{2} \lambda_j^T \mathbf{A}_j \mathbf{u}_j^{-1} \mathbf{A}_j^T \lambda_j, \\ \forall j \in J, n = 1, 2, \dots, N_j \end{aligned} \quad (14)$$

这样, 式(10)中的对偶问题可以简化为:

$$\left. \begin{aligned} \min \quad & \frac{1}{2} \lambda_j^T \mathbf{A}_j \mathbf{u}_j^{-1} \mathbf{A}_j^T \lambda_j \\ & - \left[\mathbf{1}_j - \mathbf{A}_j \mathbf{u}_j^{-1} (\mathbf{p}_j(t) + c(t) \mathbf{W}^*(t+1)) \right]^T \lambda_j \\ \text{s.t.} \quad & \sum_{n=1}^{N_j} \lambda_{jn} y_{jn} = 0 \\ & 0 \leq \lambda_{jn} \leq \frac{C}{N_j}, \quad \forall j \in J, n = 1, 2, \dots, N_j \end{aligned} \right\} \quad (15)$$

在式(15)的优化问题中, 因只含有一个列变量 λ_j , 可以利用传统的二次规划问题进行求解。当求解得到 λ_j 后, 利用式(11), 就可以求出每个节点上的 $\mathbf{W}_j(t+1)$ 。至此, WSN 中线性 SVM 分布式协同训练问题可以按照式(7)、式(15)、式(11)、式(6)的顺序进行迭代求解。

2.3 分布式线性 SVM 协同训练算法

上述线性 SVM 训练的推导形式式(7)中, 公共变量 $\mathbf{W}^*(t+1)$ 需要计算所有节点得到的局部训练结果 $\mathbf{W}_j(t)$ 和 $\mathbf{p}_j(t)$ 的平均值, 即各节点进行一次局部训练后就需要对训练结果 $\mathbf{W}_j(t)$ 和 $\mathbf{p}_j(t)$ 进行一次全局平均(简称“平均一致性”)。因每次迭代都是基于平均一致性来进行训练, 所以可以保证各节点能够收敛到比较优的训练结果, 而且也可以加快各节点的局部收敛速度。为此, 提出了基于平均一致性的分布式线性 SVM 训练算法(Average Consensus based Distributed Supported Vector Machine, AC-DSVM), 具体描述如表 1 所示。

在 AC-DSVM 算法中, 每迭代一次就要进行一次平均一致性的计算。然而, 在 WSN 无数据融合中心或特殊节点的情况下, 只依靠相邻节点间的协作来实现全局平均一致性, 会大大增加网络的通信开销。为了使 AC-DSVM 算法的思想能在无数据融合中心的 WSN 中有效应用, 降低节点间通信代价, 对其进行了改进。将 AC-DSVM 算法中的平均一致性思想应用在相邻节点间, 即在邻居节点间进行训练结果的局部平均一致性, 以此作为对全局平均一

表 1 AC-DSVM 算法

算法 1	AC-DSVM 算法
步骤 1	初始化 $k = 0$, 初始化 $\mathbf{W}_j[0], \mathbf{p}_j[0], \forall j, \mathbf{W}^*[1]$ 为 0 向量;
步骤 2	设 $k = 1$;
步骤 3	每个节点利用式(15)、式(11)、式(6), 计算 $\mathbf{W}_j[k], b_j[k], \mathbf{p}_j[k], \forall j$;
步骤 4	运行全局平均一致性算法, 利用式(7), 计算 $\mathbf{W}^*[k+1]$;
步骤 5	修改 $k = k + 1$; 如果满足收敛规则停止执行, 否则, 转到步骤 3 执行。

致性的近似, 进行各节点的局部训练, 直到各个节点的训练结果局部收敛。然后通过一次全局的平均一致性计算, 得到最优训练结果的近似解。基于此想法, 提出了基于一次平均一致性的分布式线性 SVM 训练算法(Once Average Consensus based Distributed Supported Vector Machine, 1-AC-DSVM), 具体描述如表 2 所示。

表 2 1-AC-DSVM 算法

算法 2	1-AC-DSVM 算法
步骤 1	初始化 $k = 0, \mathbf{W}_j[0], \mathbf{p}_j[0], \forall j, \mathbf{W}^*[1]$, 为 0 向量;
步骤 2	每个节点利用式(15)、式(11)、式(6), 计算 $\mathbf{W}_j[1], b_j[1], \mathbf{p}_j[1], \forall j$;
步骤 3	各节点将其结果广播给其邻居节点;
步骤 4	各节点接受邻居节点发送过来的数据, 和本地结果利用式(7), 进行局部平均一致性操作;
步骤 5	以步骤 4 中得到的平均一致性结果进行下次迭代操作, 重复步骤 2~步骤 4, 直到局部收敛;
步骤 6	对各节点收敛到的局部结果 \mathbf{W}_j^* 和相应的 b_j^* , 进行一次全局平均一致性的计算。

在上述两个算法中, 都要求全局平均一致性。但是, 在 WSN 中仅通过相邻节点间的相互协作对各节点上数据求平均一致性, 至今还没有成熟统一的方法^[17,18]。为此, 本研究使用最简单的洪泛方法^[17]来实现全局平均一致性, 其思想是各节点将自己的训练结果和其邻居节点传递过来的最新训练结果保存到本地, 并将新接收到的训练结果转发给其它邻居节点, 在经过与网络结构相关的固定迭代次数后, 各节点可以得到网络中其它所有节点的数据, 从而求得全局平均一致。

3 实验及结果分析

本文以一个由 30 个节点构成的 WSN 网络为例, 使用模拟数据和 UCI 数据集进行了实验。以集中式 C-SVM 训练结果为基准, 对 AC-DSVM 算法,

1-AC-DSVM 算法和文献[8]中的 MoM-DSVM 算法从训练结果收敛性、收敛速度和数据传输量 3 个方面进行了对比实验。

3.1 模拟数据实验

在实验中使用的模拟数据是由 2 维随机向量 (X, Y) 服从 $N(2, 2, 5, 5, 0)$ 和 $N(22, 2, 5, 5, 0)$ 随机产生的两类训练样本数据, 每类训练样本个数均为 600。将生成的训练数据平均分配给 WSN 中的 30 个节点进行训练。同时, 为了研究不同网络结构对算法训练结果的影响, 使用了链状结构和连通度为 0.0896 的网络结构进行了实验。

3.1.1 训练结果收敛性 从图 1 和表 3 显示的训练结果可以看出, AC-DSVM 算法和 1-AC-DSVM 算法得到最优决策分类线能够很好地与集中式 C-SVM 算法的训练结果相吻合, 尤其是 AC-DSVM 算法的训练结果和集中式 C-SVM 算法的训练结果完全吻合; 而 MoM-DSVM 算法得到最优决策分类线与集中式算法 C-SVM 的训练结果吻合得不好, 尤其是阈值 b 差异明显。从表 3 还可以看出, 不同网络结构对算法 AC-DSVM 和 1-AC-DSVM 的训练结果的收敛性影响不明显, 而对 MoM-DSVM 算法的收敛性略有影响。

表 3 4 种训练算法得到的最优分类决策线的权向量和阈值

算法	权向量和阈值	
	链状网络结构	网状网络结构
C-SVM	-0.3816, -0.0091; 0.1858	-0.3816, -0.0091; 0.1858
AC-DSVM	-0.3816, -0.0091; 0.1858	-0.3816, -0.0091; 0.1858
1-AC-DSVM	-0.3816, -0.0091; 0.1856	-0.3816, -0.0091; 0.1857
MoM-DSVM	-0.3817, -0.0093; 0.1936	-0.3816, -0.0092; 0.1957

3.1.2 收敛速度 从图 2 显示的结果可以看出, 两种不同网络结构下, 算法 AC-DSVM, 1-AC-DSVM 和 MoM-DSVM 在网状结构下收敛迭代次数, 分别比链状结构下收敛迭代次数少 41.2%, 25.5% 和 52.6%。表明网络结构对 3 种算法的收敛速度影响很明显, 网状结构比链状结构收敛快; 从图 2 还可以看出, 在链状结构下, AC-DSVM 算法比 MoM-DSVM 的收敛迭代次数多 33.7%; 1-AC-DSVM 算法分别比 AC-DSVM 和 MoM-DSVM 的收敛迭代次数少 89.2% 和 82.1%。在网状结构下, AC-DSVM 算法比 MoM-DSVM 的收敛迭代次数多 65.6%; 1-AC-

DSVM 算法分别比 AC-DSVM 和 MoM-DSVM 的收敛迭代次数少 91.5% 和 88.6%。表明在两种网络结构下, AC-DSVM 算法在收敛速度上明显慢于 MoM-DSVM 算法, 而 1-AC-DSVM 算法在收敛速度上要明显快于算法 AC-DSVM 和 MoM-DSVM。

3.1.3 数据传输量 从图 3 显示的结果可以看出, 两种不同网络结构下, 算法 AC-DSVM 和 1-AC-DSVM 在网状结构下数据传输量, 分别比链状结构下数据传输量多 3.4% 和 2.6%, 而算法 MoM-DSVM 在网状结构下数据传输量比链状结构下数据传输量少 52.6%。表明算法 AC-DSVM 和 1-AC-DSVM 在数据传输量上受网络结构影响不明显, 而算法 MoM-DSVM 在数据传输量上受网络结构影响明显; 从图 3 还可以看出, 在链状结构下, AC-DSVM 算法比 MoM-DSVM 的数据传输量多 70.2%, 算法 1-AC-DSVM 分别比 AC-DSVM 和 MoM-DSVM 的数据传输量少 92% 和 86.4%; 在网状结构下, AC-DSVM 算法比 MoM-DSVM 的数据传输量多 271%, 算法 1-AC-DSVM 分别比 AC-DSVM 和 MoM-DSVM 的数据传输量少 92.1% 和 70.8%。表明在两种网络结构下, AC-DSVM 算法在数据传输量上明显高于 MoM-DSVM 算法, 而 1-AC-DSVM 算法在数据传输量上明显低于算法 AC-DSVM 和 MoM-DSVM。

3.2 UCI 数据集实验

为了验证算法 AC-DSVM 和 1-AC-DSVM 对实际训练样本数据的有效性, 使用 UCI 数据库中的 Wine 数据集进行实验。因 Wine 数据集的样本数据是 13 维的高维数据, 训练得到的最优分类超平面不易平面表现, 所以本实验只对 13 维权向量的分量及阈值的收敛过程和收敛结果进行显示。本文选取了 13 维权向量的分量 w_1, w_3, w_5 和阈值 b 在算法 AC-DSVM, 1-AC-DSVM, C-SVM 和 MoM-DSVM 下的收敛过程和收敛结果进行显示, 如图 4(a)-4(d)。从图 4 中各子图的显示结果可以看出, 算法 AC-DSVM 和 1-AC-DSVM 可以近似收敛到集中式的训练结果, 算法 AC-DSVM 收敛过程中呈现的阶梯状收敛趋势的水平方向和算法 1-AC-DSVM 收敛过程中最后的直接跳变反映了平均一致性算法的运行; 从图 4 中各子图的显示结果还可以看出, 与算法 MoM-DSVM 相比, 算法 AC-DSVM 的收敛速度稍慢, 而算法 1-AC-DSVM 明显快于算法 AC-DSVM 和 MoM-DSVM。

4 结束语

本文在各节点局部分类器决策变量与集中式决

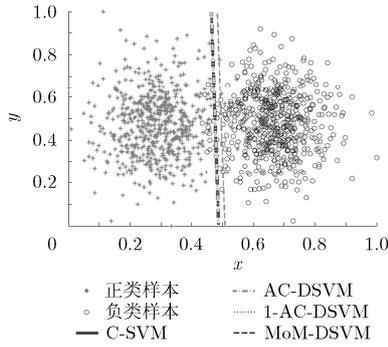


图1 两类训练样本和4种训练算法得到的最优分类决策线

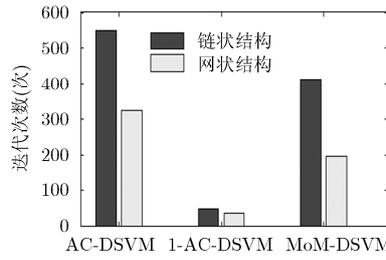


图2 两种网络结构下3种算法的迭代次数

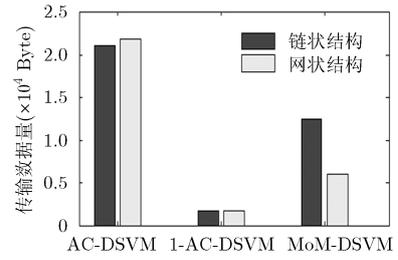
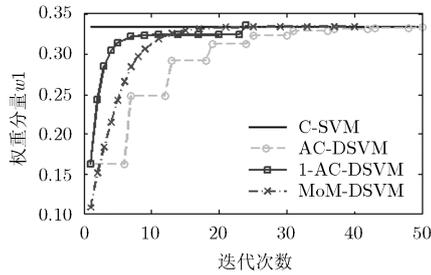
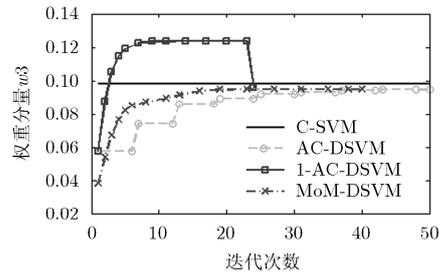


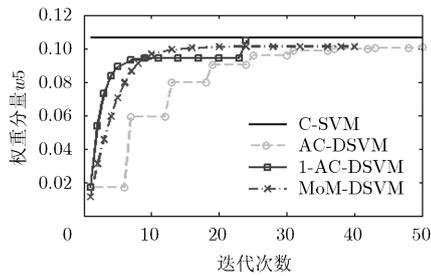
图3 两种网络结构下3种算法的数据传输量



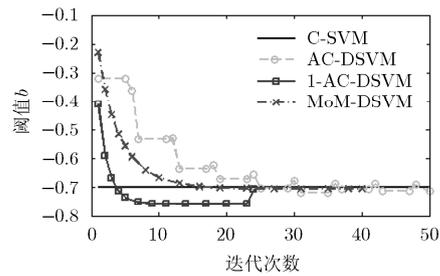
(a) 权重分量 w_1 的收敛过程及结果



(b) 权重分量 w_3 的收敛过程及结果



(c) 权重分量 w_5 的收敛过程及结果



(d) 阈值 b 的收敛过程及结果

图 4 Wine 数据集下 4 种算法的收敛过程和收敛结果

策变量相一致的约束下，利用分布式优化方法，研究了仅依靠相邻节点间的相互协作，在网内分布式协同训练线性 SVM 的方法，提出了基于平均一致性的分布式线性 SVM 训练算法 AC-DSVM 和改进算法 1-AC-DSVM。仿真实验结果表明：与算法 MoM-DSVM 相比，AC-DSVM 算法虽然在收敛速度和数据传输量方面较差，但其可以准确地收敛到集中式训练结果；1-AC-DSVM 算法具有较好收敛性，而且在收敛速度和数据传输量上也表现出了显著优势。此外，不同网络结构对两个算法在收敛性和数据传输量上的影响不明显，对收敛速度影响较明显。1-AC-DSVM 算法因其良好的性能为 WSN 下 SVM 分布式协同训练提供了一种非常有效的方法。由于 SVM 的优势在于解决非线性问题，因此笔者进一步的研究工作是将提出的算法扩展到非线性问题的求解上。

参考文献

- [1] 吕方旭, 张金成, 郭相科, 等. 基于 WSN 的战场声目标多特征联合智能分类识别[J]. 科学技术与工程, 2013, 13(35): 10713-10721.
Lü Fang-xu, Zhang Jin-cheng, Guo Xiang-ke, et al.. The acoustic target in battlefield intelligent classification and identification with multi-features in WSN[J]. *Science Technology and Engineering*, 2013, 13(35): 10713-10721.
- [2] Taghvaeeyan S and Rajamani R. Portable roadside sensors for vehicle counting, classification, and speed measurement[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 15(1): 73-83.
- [3] Shahid N, Naqvi I H, and Bin Qaisar S B. Quarter-sphere SVM: attribute and spatio-temporal correlations based outlier & event detection in wireless sensor networks[C]. *IEEE Wireless Communication and Networking Conference: Mobile and Wireless Networks*, Paris, 2012: 2048-2053.
- [4] 刘倩, 崔晨, 周杭霞. 改进型 SVM 多类分类算法在无线传感

- 器网络中的应用[J]. 中国计量学院学报, 2013, 24(3): 298-303.
- Liu Qian, Cui Chen, and Zhou Hang-xia. Application of a modified SVM multi-class classification algorithm in wireless sensor networks[J]. *Journal of China University of Metrology*, 2013, 24(3): 298-303.
- [5] Raj A B, Ramesh M V, Kulkarni R V, *et al.* Security enhancement in wireless sensor networks using machine learning[C]. IEEE 14th International Conference on High Performance Computing and Communications, Liverpool, 2012: 1264-1269.
- [6] Flouri K, Beferull-Lozano B, and Tsakalides P. Optimal gossip algorithm for distributed consensus SVM training in wireless sensor networks[C]. Proceedings of the 16th International Conference on Digital Signal Processing (DSP '09), Heraklion, 2009: 886-891.
- [7] Flouri K, Beferull-Lozano B, and Tsakalides P. Training a support-vector machine-based classifier in distributed sensor networks[C]. Proceedings of 14th European Signal Processing Conference, Florence, 2006: 1-5.
- [8] Forero P A, Cano A, and Giannakis G B. Consensus-based distributed linear support vector machines[C]. Proceedings of the International Conference on Information Processing in Sensor Networks, Stockholm, 2010: 35-46.
- [9] Forero P A, Cano A, and Giannakis G B. Consensus-based distributed support vector machines[J]. *Journal of Machine Learning Research*, 2010, 11(May): 1663-1707.
- [10] Lu Y, Roychowdhury V, and Vandenberghe L. Distributed parallel support vector machines in strongly connected networks[J]. *IEEE Transactions on Neural Networks*, 2008, 19(7): 1167-1178.
- [11] Wang Dong-li, Li Jian-xun, and Zhou Yan. Support vector machine for distributed classification: a dynamic consensus approach[C]. IEEE/SP 15th Workshop on Statistical Signal Processing, Wales, 2009: 753-756.
- [12] Kim Woo-jin, Yoo Jae-hyun, and Kim H J. Multi-target tracking using distributed SVM training over wireless sensor networks[C]. IEEE International Conference on Robotics and Automation, Saint Paul, 2012: 2439-2444.
- [13] Shawe T J and Cristianini N. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods [M]. New York: Cambridge University Press, 2000: 121-130.
- [14] Scholkopf B and Smola A J. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond [M]. New York: The MIT Press, 2002: 204-205.
- [15] Bertsekas D P and Tsitsiklis J N. Parallel and Distributed Computation: Numerical Methods[M]. Massachusetts: Athena Scientific, 1997: 243-247.
- [16] Boyd S, Parikh N, and Chu E. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. *Foundations and Trends in Machine Learning*, 2011, 3(1): 1-122.
- [17] Sardellitti S, Giona M, and Barbarossa S. Fast distributed average consensus algorithms based on advection-diffusion processes[J]. *IEEE Transactions on Signal Processing*, 2010, 58(2): 826-842.
- [18] Sardellitti S, Barbarossa S, and Swami A. Average consensus with minimum energy consumption: optimal topology and power allocation[C]. Proceedings of the European Signal Processing Conference (EUSIPCO 2010), Aalborg, 2010: 189-193.
- 及歆荣: 女, 1978 年生, 博士生, 研究方向为无线传感器网络、分布式机器学习.
- 侯翠琴: 女, 1983 年生, 博士, 研究方向为无线传感器网络、分布式机器学习.
- 侯义斌: 男, 1952 年生, 教授, 博士生导师, 主要研究方向为物联网接入技术、新型人机交互技术、嵌入式软件与系统.