

改进PSO-SVM在说话人识别中的应用

李明, 张勇, 李军权, 张亚芬

(兰州理工大学计算机与通信学院 兰州 730050)

【摘要】为了加快粒子群优化算法的收敛速度,增强全局的搜索能力,通过对粒子群优化算法中惯性权重和全局最优值的分析,提出了一种根据迭代次数而自适应变化的惯性权重的粒子群优化方法。改进后的粒子群算法在防止陷入局部最优的能力方面有了明显的增强,同时,给出了应用粒子群优化算法训练支持向量机的方法,并将其应用于说话人识别。实验结果证实了在说话人识别中改进PSO-SVM方法比其他传统方法能获得更好的识别精度和识别速度。

关键词 惯性权重; 粒子群优化算法; 说话人识别; 支持向量机
中图分类号 TP391.4 文献标识码 A

Application of Improved PSO-SVM Approach in Speaker Recognition

LI Ming, ZHANG Yong, LI Jun-quan, ZHANG Ya-fen

(School of Computer and Communication, Lanzhou University of Technology Lanzhou 730050)

Abstract In order to increase the convergent speed and to improve the overall searching ability of the algorithm, a Particle Swarm Optimization (PSO) method is proposed with adaptive inertia weight by the change of the number of iterations based on the analysis of inertia weight global best fitness of the PSO. The improved PSO increases the ability to avoid local optimum. Then a speaker recognition method using this improved algorithm to train Support Vector Machine (SVM) is presented. The experimental results show that the presented SVM method optimized by PSO for speaker recognition can achieve higher recognition accuracy and higher recognition speed.

Key words inertia weight; particle swarm optimization; speaker recognition; support vector machine

语音是人的自然属性之一,每个人的语音都带有强烈的个人色彩,这使得通过分析语音信号来识别说话人成为可能。用语音来鉴别说话人身份有许多独特的优点,如语音是人的固有特征,不会丢失或遗忘;语音信号采集方便,系统设备成本低等。说话人识别技术可用于刑侦破案、机要保密、语音加密口令、指挥系统、电子语音锁、玩具和家用电器等,因此具有广泛的应用前景。

支持向量机(Support Vector Machine, SVM)由于其出色的学习性能,已成为机器学习领域的一个研究热点,并在说话人识别、人脸识别等领域中得到应用^[1]。它可避免以往机器学习中的欠学习问题和过学习问题,在非线性和小样本问题中表现出良好特性。文献[2]首先将SVM用于说话人识别,SVM训练的实质为求解一个带有界约束和线性等式约束的凸二次规划问题,其识别准确率高,但计算量大、运算速度慢。粒子群优化算法^[3](Particle Swarm

Optimization, PSO)最初是模仿鸟群、蜂群、鱼群寻找食物的社会行为来建立的。它具有不容易陷入局部最小,并且算法简单实现容易的优点,所以在寻优过程中计算量大大减少,节省了寻优的时间。由于PSO能求解无约束优化问题,文献[4]提出了一种LPSO算法,可以用于求解含有线性等式约束的优化问题。文献[5]将LPSO算法应用于SVM的训练中,但主要针对的是大样本问题,只是采用LPSO算法求解二次规划。文献[6]提出了用粒子群算法PSO优化参数的PSO-SVM短期交通流预测模型,实现了数据降维并且保持了交通流序列的特征,实验证明了该模型的有效性。本文提出了一种根据迭代次数而自适应变化的惯性权重的粒子群优化方法。

1 支持向量机

支持向量机^[7]是一种基于结构风险最小化的二元分类器,其本质思想就是寻找最优超平面将属于

收稿时间:2007-09-14

基金支持:甘肃省教育厅科研项目(0603-10)

作者简介:李明(1959-),男,教授,主要从事数据挖掘和人工智能方面的研究;张勇(1981-),男,硕士,主要从事说话人识别方面的研究;李军权(1982-),男,硕士,主要从事智能信息处理方面的研究;张亚芬(1984-),女,硕士,主要从事说话人识别方面的研究。

两个类别的样本无误地分开，且使分类间隙最大。所谓的分类间隙是指两类中通过离分类超平面最近的样本且平行于分类超平面的两个超平面间的距离，由支持向量决定的最优分类超平面，是从线性可分情况下的最优分类面发展而来的。SVM只是把数据映射到高维空间，在高维空间进行线性支持向量机。

1.1 线性支持向量机

给定数据集 $\{x_i, y_i\}, i=1, 2, \dots, N, y_i \in \{-1, +1\}, x_i \in \mathcal{R}^d$ ，假如这两类样本是线性可分的，则SVM学习的结果是找到一个最优超平面能够将这两类样本分开且分类间隙最大。离超平面最近的样本为支持向量，从某种程度上说，是样本中少量的支持向量决定了最优超平面。支持向量机原理如图1所示，图中“○”与“△”符号分别代表两个待分类的样本。

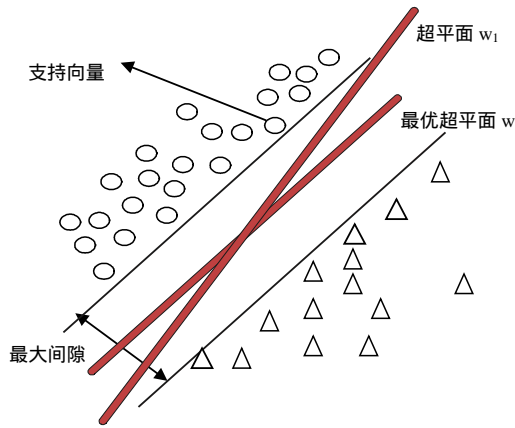


图1 支持向量机原理图

对于给定的数据集 $\{x_i, y_i\}, i=1, 2, \dots, N, y_i \in \{-1, +1\}, x_i \in \mathcal{R}^d$ ，可用超平面：

$$w \cdot x + b = 0 \tag{1}$$

进行分离。参数 w, b 约束于：

$$\min |w \cdot x + b| = 1 \tag{2}$$

超平面约束于：

$$y_i (\langle w \cdot x_i \rangle + b) = 1 \quad i=1, 2, \dots, N \tag{3}$$

在式(3)约束下，最小化 $\frac{1}{2} \|w\|^2$ ，引入拉格朗日泛函形式：

$$l(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N a_i [y_i (\langle w \cdot x_i \rangle + b) - 1] \tag{4}$$

式中 a_i 是拉格朗日因子。对 w, b 最小化 $l(w, b, a)$ ，有满足最优解的条件：

$$\frac{\partial l(w, b, a)}{\partial b} = 0 \Rightarrow \sum_{i=1}^N a_i^* y_i = 0 \tag{5}$$

$$\frac{\partial l(w, b, a)}{\partial w} = 0 \Rightarrow w^* = \sum_{i=1}^N a_i^* x_i y_i \tag{6}$$

将式(5)、(6)带入式(4)，得到对偶最优优化问题：

$$\begin{aligned} \max_a W(a) &= \max_a \{ \min_{w, b} l(w, b, a) \} = \\ & \max_a \left\{ \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j y_i y_j \langle x_i \cdot x_j \rangle \right\} \tag{7} \\ \text{s.t.} \quad & \sum_{i=1}^N a_i y_i = 0 \quad 0 \leq a_i \leq C, i=1, 2, \dots, n \end{aligned}$$

根据式(5)和 $a_i > 0$ ，求得解 a_i^* 。 a_i 是二次规划(QP)优化问题所求解的拉格朗日因子，每一个训练样本对应一个 a_i ，其解中不为零的 a_i 所对应的样本就是支持向量； C 为惩罚参数。QP问题是一个凸最优化问题^[8]，所以SVM的解是唯一的，也是全局最优的。设 a_i^* 为二次规划所得的解，则得到最优分类函数为：

$$f(x) = \text{sign}(\langle w^T \cdot x \rangle + b) = \text{sign} \left(\sum_{i=1}^{N_v} a_i^* y_i \langle x_i \cdot x \rangle + b \right) \tag{8}$$

式中 N_v 为支持向量数。对于线性不可分的样本引入松弛变量，详情参考文献[9]。

1.2 非线性支持向量机

首先把原始数据经过非线性映射 $\phi: \mathcal{R}^d \rightarrow F$ ，映射到高维空间，在高维空间进行线性分类计算，求得解，得到分类函数为：

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^{N_v} a_i^* y_i \phi(x_i)^T \phi(x) + b \right\}$$

式中 N_v 为支持向量数。

引入核概念，在说话人识别中一般采用多项式核函数 $K(x, x_i) = [(x \cdot x_i) + 1]^q$ ，则决策函数定义为：

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^{N_v} a_i^* y_i k(x_i \cdot x) + b \right\} \tag{9}$$

2 粒子群优化算法^[2]

粒子群算法是继遗传算法、蚁群算法后的又一种新的群体智能算法^[10]，目前已成为进化算法的一个重要分支，并在神经网络训练、TSP问题、非线性规划、多目标优化等一些领域得到了成功应用。在PSO算法中，有 N 个粒子组成的粒子群，其中第 i 个粒子的位置为 x_i ，可根据它来计算相应的适应值，根据适应值可衡量粒子 i 的优劣。每个粒子的历史最优位置 p_{id} 以及整个粒子群的历史最优位置 p_{gd} 在迭代计算过程中将被记录，粒子通过跟踪两个历史最优位置来更新自己的速度和位置，搜索解空间中的最优解。粒子群算法的数学描述如下：

$$\begin{aligned} v_{id} &= \omega v_{id} + c_1 \text{rand}_1() (p_{id} - x_{id}) + \\ & c_2 \text{rand}_2() (p_{gd} - x_{id}) \end{aligned} \tag{10}$$

$$\begin{cases} v_{id} = v_{max} & v_{id} > v_{max} \\ v_{id} = -v_{max} & v_{id} < -v_{max} \end{cases} \quad (11)$$

$$x_{id} = x_{id} + v_{id} \quad (12)$$

式中 ω 为惯性权值; c_1 和 c_2 都为正常数, 称为加速系数; $\text{rand}_1()$ 和 $\text{rand}_2()$ 是两个随机数发生器; p_{id} 是当前粒子的历史最优位置; p_{gd} 是整个粒子群的历史最优位置。搜索时, 粒子的位置被最大位置、最小位置所限制, 如果粒子的位置超出所给的范围, 则粒子的位置将被限制为最大位置 x_{max} 或最小位置 x_{min} 。同样, 粒子的速度也被最大速度 v_{max} 和最小速度 v_{min} 所限制。粒子的最大速度 v_{max} 决定了解空间的搜索精度, 对 v_{max} 进行动态调整可以使算法具有较好的自适应寻优效果。

3 改进 PSO 算法训练支持向量机用于说话人识别

SVM 训练算法是为了求解不为零、非负的训练样本点 Lagrange 乘子, 即支持向量点, 从而求出权值 ω 和 b^* , 确定分类平面。所以训练支持向量机最核心的问题是解决二次规划(QP)问题, 求解 Lagrange 乘子。SVM 用于模式识别将要解决一个二次优化问题(QP), 即式(7)。二次规划(QP)算法是解决该问题最好的方法。虽然目前有一些现有的 QP 算法和软件包, 但是由于算法的限制, 往往只能解决较小数据集的问题, 对解决具有大量训练样本的问题, 就显得无能为力。在说话人识别中, 语音样本帧的量通常都比较大, 因此解决大样本的 SVM 训练问题就成为一个重要的问题。粒子群优化算法是一种新型的具有高效率全局搜索能力的进化算法。PSO 算法调整参数少, 容易实现, 收敛能力强, 实现时间短。将 PSO 算法应用于训练支持向量机, 可以很好地解决二次规划问题。

3.1 改进 PSO 算法

根据文献[11]中的定理1证明 PSO 进化过程与粒子速度无关, 定理1的重要性在于说明 PSO 算法可以没有粒子速度的概念, 避免了人为确定参数 $[-v_{max}, v_{max}]$ 而影响粒子的收敛速度和收敛精度。不含速度项的粒子群优化方程可以简化为:

$$x_{id}^{t+1} = \omega x_{id}^t + c_1 r_1 (p_{id} - x_{id}^t) + c_2 r_2 (p_{gd} - x_{id}^t) \quad (13)$$

式中 右边的第一项为“历史(History)”部分, 表示过去对现在的影响, 通过 ω 调节影响程度; 第二项为“认知(Cognition)”部分, 表示粒子对自身的思考; 第三项为“社会(Social)”部分, 表示与邻居粒子的比较和模仿, 实现粒子间的信息共享与合作。这样

在优化的过程中不必考虑速度这一项, 可以减少计算量, 加快寻优速度。

PSO 在优化前期收敛速度很快, 但在优化后期收敛速度变得缓慢, 因而导致收敛精度低。这主要是粒子群难以摆脱局部极值的原因。很多学者提出了许多改进方法, 如杂交 PSO、变异 PSO、自适应 PSO、重新初始化粒子群、结合模拟退火等局部寻优策略等^[12]。但是, 这些策略主要用于调整参数 ω 、 r_1 、 r_2 、 a 和变量 x_{id} 。这些改进策略都不同程度地提高了收敛速度和精度。由于较大的 ω 可以增强 PSO 的全局搜索能力, 而较小的 ω 则使算法有较强的局部搜索能力。根据 ω 的变化特点, 本文改进了 PSO 算法中的 ω 权重法, ω 仍然随迭代次数线性递增, 当迭代次数到达某个阈值时, 便重置 ω 的初值。这样 ω 随迭代次数变化而变化, 有助于算法摆脱局部极值, 增强 PSO 的全局搜索能力, 找出最优解。将 PSO 算法应用于训练支持向量机, 可以很好解决二次规划问题, 提高支持向量机的训练速度和分类精度, 继而提高说话人的识别精度和识别速度。具体设置如下:

$$\begin{cases} \omega = \omega_{max} & \text{iter} < \text{iter}_{\text{阈值}} \\ \omega = \frac{\text{iter} - \text{iter}_{\text{阈值}}}{\text{iter}_{\text{总}} - \text{iter}_{\text{阈值}}} (\omega_{max} - \omega_{min}) + \omega_{min} & \text{iter} \geq \text{iter}_{\text{阈值}} \end{cases} \quad (14)$$

式中 iter 为迭代次数; $\text{iter}_{\text{总}}$ 为设置的总的迭代次数; $\text{iter}_{\text{阈值}}$ 为迭代次数的阈值; ω_{max} 是惯性权重的最大值; ω_{min} 是惯性权重最小值。当 ω 重置的时候, 非全局最优的粒子也按一定概率随机的重新初始化, 但仍保留“生前”的记忆, 即它还记得全局最优和个体最优。这样使得搜索的范围更加扩大。粒子优化算法步骤如下:

- (1) 初始化粒子群 (a_1, a_2, \dots, a_m) , 确定群体规模 m , 给定算法的最大、最小权重因子 ω_{max} 、 ω_{min} 值, 设定算法总的迭代次数 $\text{iter}_{\text{总}}$ 和迭代次数的阈值 $\text{iter}_{\text{阈值}}$ 。
- (2) 将每个粒子的个体极值 p_{id} 设置为当前位置, 利用适应函数即式(15)计算每个粒子的适应度, 取适应度最好的粒子所对应的个体极值作为最初的全局极值 p_{gd} 。
- (3) 按照式(13)、(14)进行迭代计算, 更新粒子的位置。
- (4) 由式(15)评价每个粒子的适应值。
- (5) 将每个粒子的适应值与其 p_{id} 对应的适应值比较, 若优, 更新 p_{id} , 否则保留原值。
- (6) 将更新后的每个粒子的 p_{id} 与全局极值 p_{gd} 比较, 若优, 更新 p_{gd} , 否则保留原值。
- (7) 判断是否满足终止条件, 若达到最大迭代次数或所得解不再变化就终止迭代, 否则返回到步骤(3)。

3.2 改进PSO训练支持向量机

由SVM的数学模型可知,需要求解的是支持向量系数 a_i ($i=1,2,\dots,m$), 自适应函数为:

$$f(p) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j K(x_i, x_j) \quad (15)$$

根据PSO算法和SVM的要求,粒子的初始位置应满足条件 $\sum_{i=1}^n a_i y_i = 0$ 和 $0 < a_i < C$ ($i=1,2,\dots,n$)。

PSO在优化前期收敛速度很快,但在优化后期收敛速度变得缓慢,导致收敛精度低,使搜索空间受到极大地限制,直接影响到SVM的训练速度和性能,这是粒子群难以摆脱局部极值的主要原因。本文改进权重 ω 来增强PSO的全局搜索能力,找出最优的支持向量系数,从而提高SVM的训练速度和性能。采用改进PSO训练SVM的过程如下:

(1) 初始化PSO的参数,并按照本文的方法对粒子群进行初始化,设置当前迭代数 $iter=1$ 。(2) 对每个粒子执行下述操作:按式(15)计算各粒子的适应度,更新相应的 p_{id} 和 p_{gd} ;按式(14)计算当前惯性权重 ω ;用式(13)更新粒子的位置矢量。(3) $iter=iter+1$,如果 $iter > iter_{总}$,则 p_{gd} 即为SVM的训练结果,否则返回步骤(2)重复上述过程。(4) 根据更新各粒子的适应值,计算支持向量系数 a_i ($i=1,2,\dots,m$)。(5) 根据式(9)计算决策函数。

4 实验

4.1 语音库及参数设置

语音库由10个人的录音组成,包括5名男性,5名女性,每人语音长度约90 s。本文采用采样频率11 025 Hz,16 b量化,在普通环境下采样。对于文本无关的说话人每个人录制六遍,前三遍用于训练,后三遍用于识别。语音数据首先需进行预处理,包括端点检测、预加重 $1-0.97z^{-1}$,加窗(Hamming窗,帧长20 ms,帧移10 ms),提取12维的MFCC倒谱系数及其一阶、二阶差分系数,形成36维的特征矢量序列。考虑到说话人特性会随时间改变,语音库分两次录制,每次录制三遍,中间间隔一个月左右。

粒子数选择10或20个粒子,最大迭代次数为500,迭代次数阈值选择300,惯性权重最大设置为1.5,惯性权重最小设置为0.7,随机序列 r_1 和 r_2 都设置为1.4,惩罚参数 C 设为100。选择多项式函数作为核函数。采用核主成分分析方法用于说话人特征提取,既可以提取主成分也可以剔除噪声,然后采用改进PSO-SVM方法进行特征的分类识别。

4.2 算法性能分析比较

为了测试本文提出算法的有效性,当粒子数分别取10或20,对本文的改进PSO-SVM算法、PSO-SVM和SVM算法在说话人识别中的性能进行比较。表1为粒子数取10的时候各算法的性能比较,表2为粒子数取20的时候各算法性能比较。本文对各算法的识别率(RR)、识别时间(T)、需要的支持向量的数目(SV)进行了分析、比较。

表1 粒子数取10时各算法性能的比较

算法	RR	T/s	SV
SVM	92.9	3.17	741
PSO-SVM	95.6	1.65	314
改进PSO-SVM	98.6	0.65	193

从表1中可以看出,本方法提高了说话人识别的效率,同时减少了识别的时间,可以很好地满足说话人识别的实用性。这就说明了采用粒子群优化算法对支持向量机参数进行寻优,对训练的SVM具有较高的分类正确率,提高了识别的准确率。改进的粒子群优化算法具有不容易陷入局部最小、算法简单和计算量小的优点,大大减少了寻优的时间,提高了效率。从分类所需要的支持向量数(SV)也可以看出,本方法通过粒子寻优用较少的支持向量数取得了最优的分类平面。由于真正对分类起作用的支持向量数目较少,其计算量也小,从而提高了说话人识别的速度。

表2 粒子数取20时各算法性能的比较

算法	RR	T/s	SV
SVM	92.9	3.17	741
PSO-SVM	97.3	2.26	356
改进PSO-SVM	98.9	1.23	205

从表2中可以看出,粒子数目增加,识别的准确率也增加,其识别时间也必然会延长,粒子寻优找出的有用的支持向量数也越多。从两个表的对比可以看出,虽然粒子数增加识别率略有提高,但是却是以时间为代价换来的,不利于说话人识别的实用性。从大量的实验中发现粒子数在12~15之间取得最好的识别效果,而粒子寻优一般迭代到400左右即可找出最优的分类平面,惯性权重取0.7~1.5之间使粒子具有最强的全局搜索能力。

5 结论

本文通过对粒子群优化算法中惯性权重和全局

最优值的分析,给出了自适应惯性权重的PSO方法,并且将PSO的快速全局优化特点和SVM的非线性拟合特点相结合用于说话人识别。通过实验证明,该方法与其他传统说话人识别处理方法相比较,具有更好的收敛精度,并且对于识别问题具有显著的时间优势。

参 考 文 献

- [1] CAMPBELL W M, CAMPBELL J P, REYNOLDS D A. Support vector machines for speaker and language recognition[J]. *Computer Speech and Language*, 2006, 20: 210-229.
- [2] 曾建潮, 崔志华. 一种保证全局收敛的PSO算法[J]. *计算机研究与发展*, 2004, 41(8): 1333-1338.
- [3] PAQUET U, ENGELBRECHT A P. A new particle swarm optimizer for linearly constrained optimization[C]//Proc of IEEE Congress on Evolutionary Computation. [S.l.]: IEEE, 2003: 227-233.
- [4] PAQUET U, ENGELBRECHT A P. Training support vector machines with particle swarms[C]//Proc of International

- Joint Conference on Neural Networks. [S.l.]: [s.n.], 2003: 1593-1598.
- [5] 曹成涛, 徐建闽. 基于PSO-SVM的短期交通流预测方法[J]. *计算机工程与应用*, 2007, 43(15): 12-14.
- [6] 祁亨年. 支持向量机及其应用研究综述[J]. *计算机工程*, 2004, 30(10): 6-9.
- [7] 孙建成, 张太镒, 刘海员. 基于SVM的多类模拟调制方式识别算法[J]. *电子科技大学学报*, 2006, 35(2): 696-699.
- [8] SU Chao-ton, YANG Chien-hsin. Feature selection for the SVM: an application to hypertension diagnosis[J]. *Expert Systems with Applications*, 2008, 34: 754-763.
- [9] 沈艳, 郭兵, 古天祥. 粒子群优化算法及其与遗传算法的比较[J]. *电子科技大学学报*, 2005, 34(5): 696-699.
- [10] 胡旺, 李志蜀. 一种更简化而高效的粒子群优化算法[J]. *软件学报*, 2007, 18(4): 860-868.
- [11] YIN Peng-yeng. Particle swarm optimization for point pattern matching[J]. *J Vis Commun Image R*, 2006, 17: 143-162.
- [12] HOU Feng-lei, WANG Bing-xi. Text-independent speaker recognition using support vector machines[C]//Proceeding of ICII. [S.l.]: [s.n.], 2001: 402-407.

编辑 漆蓉

(上接第1332页)

4 结 论

物流信息化项目具有投资大、风险高、周期长和其能否成功实施直接关系企业全局的特点,欲将RFID成功融入现代物流信息建设中,任重而道远。从根本上看,RFID能否大规模应用取决于是否解决了RFID应用与现有信息系统之间的应用整合问题,尤其是与企业内部应用系统的整合。本文采用以Web Services架构为基础的SOA模型整合物流企业各信息子系统,打造随需而变的敏捷软件体系架构,并在该信息融合模型基础上充分利用数据挖掘和智能工具为企业建立商业智能,以支持企业的日常运作和长远规划,从而为RFID在物流行业中的推广应用和具体实施提供可参考的解决方案和信息基础设施建设框架。

参 考 文 献

- [1] FINKENZELLER K. RFID handbook: fundamentals and applications in contactless smart cards and identification, second edition[M]. Indianapolis: John Wiley & Sons, 2003: 5-10.
- [2] 谭民, 刘禹, 曾隽芳, 等. RFID技术系统工程及应用指南[M]. 北京: 机械工业出版社, 2007: 278-281.
- [3] HASSAN T, CHATTERJEE S. A taxonomy for RFID[C]//Proceedings of the 39th Hawaii International Conference on

- System Sciences. Hawaii: IEEE Computer Society Press, 2006: 1158-1163.
- [4] RAO K V, NIKITIN P V, LAM S F. Antenna design for UHF RFID tags: a review and a practical application[J]. *IEEE Transaction on Antenna Propagation*, 2005, 53(12): 3870-3876.
- [5] 秦志光, 周世杰, 刘锦德. 智能运输系统中的中间件技术[J]. *电子科技大学学报*, 2002, 31(2): 168-172.
- [6] FLOERKEMEIER C, LAMPE M. RFID middleware design-addressing application requirements and RFID constraints[C]//Proceedings of SOC-EUSAI 2005(Smart Objects Conference). Grenoble: ACM Press, 2005: 265-270.
- [7] KIM Y I, PARK J S, CHEONG T S. Study of RFID middle framework for ubiquitous computing environment[C]//IEEE the 7th International Conference on Advanced Communication Technology. Phoenix Park: IEEE, 2005: 825-830.
- [8] NEWCOMER E, LOMOW G. Understanding SOA with web services[M]. Boston: Addison-Wesley Professional, 2004: 8-23.
- [9] 李斌, 李文锋. NET 2.0在软件再工程中的研究与应用[J]. *四川大学学报*, 2007, 39(增刊): 251-255.
- [10] 王慧斌, 王建颖. 信息系统集成与融合技术及其应用[M]. 北京: 国防工业出版社, 2006: 224-228.
- [11] 陈伟, 丁秋林. 可扩展数据清理软件平台的研究[J]. *电子科技大学学报*, 2006, 35(1): 100-103.
- [12] 焦季成, 刘芳, 刘静, 等. 智能数据挖掘与知识发现[M]. 西安: 西安电子科技大学出版社, 2006: 08-15.

编辑 张俊