

# 一种改进的贝叶斯算法在垃圾邮件过滤中的研究<sup>\*</sup>

马小龙

(甘肃民族师范学院 计算机科学系,甘肃 合作 747000)

**摘要:** 研究了改进的基于 SVM-EM 算法融合的朴素贝叶斯文本分类算法以及在垃圾邮件过滤中的应用。针对朴素贝叶斯算法无法处理基于特征组合产生的变化结果,以及过分依赖于样本空间的分布和内在不稳定性的缺陷,造成了算法时间复杂度的增加。为了解决上述问题,提出了一种改进的基于 SVM-EM 算法的朴素贝叶斯算法,提出的方法充分结合了朴素贝叶斯算法简单高效、EM 算法对缺失属性的填补、支持向量机三种算法的优点,首先利用非线性变换和结构风险最小化原则将流量分类转换为二次寻优问题,然后要求 EM 算法对朴素贝叶斯算法要求条件独立性假设进行填补,最后利用朴素贝叶斯算法过滤邮件,提高分类准确性和稳定性。仿真实验结果表明,与传统的邮件过滤算法相比,该方法能够快速得到最优分类特征子集,大大提高了垃圾邮件过滤的准确率和稳定性。

**关键词:** 文本分类;垃圾邮件;朴素贝叶斯;支持向量机;EM

**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1001-3695(2012)03-1091-04

**doi:**10.3969/j.issn.1001-3695.2012.03.080

## Research of spam-filtering based on optimized naive Bayesian algorithm

MA Xiao-long

(Dept. of Computer, Gansu Normal University for Nationalities, Hezuo Gansu 747000, China)

**Abstract:** This paper discussed improvement of naive Bayesian text classification algorithms based on the SVM-EM algorithms and applications in spam filtering. Naive Bayes algorithm cannot handle the results based on the feature-based combination changes feature-based, and dependent on the distribution of sample space and the inherent instability of the defect, causing the algorithm complexity increases. To solve the above problems, this paper proposed an improved algorithm based on SVM-EM naive Bayes algorithm, which was combined with naive Bayes algorithm's simple and efficient, the advantages of filling the missing property of EM, the advantages of support vector machines (SVM) algorithms, first made nonlinear transformation and structural risk minimization flow into the second classification optimization problem, and then asked the EM algorithm to fill the requirements of the conditional independence assumptions for Bayesian algorithm. Finally, using Bayesian algorithms to improve the mail filtering classification accuracy and stability. Simulation results show that the proposed method can quickly obtain the optimal feature subset classification, greatly improve the spam filtering accuracy and stability compared to traditional methods of mail filtering algorithm.

**Key words:** text classification; spam e-mail; naive Bayesian; SVM; EM

## 0 引言

随着 Internet 网络的迅猛发展, e-mail 已成为人们信息交流和交换的一种重要通信方式。E-mail 使用简单、传播迅速,给人们带来极大便利的同时,大量的商业广告、色情、反动垃圾邮件和病毒邮件的泛滥浪费了网络带宽和存储空间,逐渐成为 Internet 上的一个严重的安全问题。邮件系统的安全性和可靠性就成为了研究人员关注的焦点,对垃圾邮件的智能分析识别和过滤技术需求极其强烈。

目前常用的垃圾邮件过滤方法有:a) 基于规则的过滤方法,其原理是通过与预设规则比较来判定是否为垃圾邮件,而这些规则通常是静态设置的,在一定程度上满足基本需求,但缺少可信度的知识学习策略,在规律不明显的应用领域中过滤效果较差;b) 黑白名单过滤方法,其原理为将邮件发送方的邮箱或 IP 设为黑名单,以此作为过滤依据,当对方是采用 IP 代理、动态 IP、邮件地址欺骗、伪造或隐藏信源地址等方式发送邮件的话,垃圾邮件无法过滤;c) 基于智能统计和学习的方法,基于统计的方法有 KNN、SVM、Bayes、Rocchio 等分类算法,

其中最为经典的就是朴素贝叶斯过滤算法,它是一种简单而高效的基于概率统计的分类算法,在邮件过滤技术中有广泛的应用,能适应垃圾邮件变化。朴素贝叶斯过滤算法中的所有属性都参与分类,但是在实际邮件过滤中,统计数据都有不同程度的缺失,而且条件独立性的假设极大影响了分类性能和准确度。

EM 算法同样有坚定的理论基础兼具稳定的特点,被广泛地应用于缺损数据、截尾数据、成群数据等不完全数据的处理。该算法不仅可以对朴素贝叶斯算法要求条件独立性假设进行填补,还可以对类别属性进行填补。本文针对传统方法和朴素贝叶斯在垃圾邮件过滤应用中的局限性,对垃圾邮件的特点进行了系统分析和研究,提出了一种基于 SVM-EM 算法的改进朴素贝叶斯复合智能分类算法。该算法充分融合朴素贝叶斯简单高效、EM 算法对不完全数据处理的优点,通过对邮件过滤关键字的缺失属性的估计值作为 EM 算法初始值,并计算极大似然估计完成缺失属性的填补,获取适合的最大 EM 收敛值和加速收敛,然后利用朴素贝叶斯分类算法对完整数据集进行分类,提高垃圾邮件过滤的精确度和性能。

最后仿真实验结果表明:该复合算法十分有效,不仅能够快速得到最优分类特征子集,而且大大提高了其学习和分类的效率 and 准确率,降低了过滤器的错误率,具有较高的实用价值。笔者采用 Spam Assassin 邮件集和 Yahoo 邮箱中多年的大量垃圾邮件和正常邮件样本的标准数据集,对算法进行了大量测试,训练集达到 5 000 之后,准确率达到了 98.4%。

## 1 朴素贝叶斯过滤技术

### 1.1 朴素贝叶斯基本描述

朴素贝叶斯过滤技术以贝叶斯定理为基础,通过数据的先验概率,利用贝叶斯式计算出其后验概率,并选择具有最大后验概率的类作为该对象所属的类。

在给定训练数据  $T$  时,确定假设空间  $\lambda$  中的最佳假设。贝叶斯算法提供了从先验概率  $P(\lambda)$  以及  $P(T)$  和  $P(T|\lambda)$  计算后验概率  $P(\lambda|T)$  的方法。其公式为

$$P(\lambda|T) = \frac{P(T|\lambda)P(\lambda)}{P(T)}$$

其中: $P(\lambda)$  表示  $\lambda$  的先验概率; $P(T)$  表示待观察的训练数据  $T$  的先验概率; $P(T|\lambda)$  表示给定  $\lambda$  时观察数据  $T$  对应的概率; $P(\lambda|T)$  为  $\lambda$  的后验概率,即给定训练数据  $T$  时  $\lambda$  成立的概率。假设集合  $\lambda$  并寻找观察数据  $T$  在其中的概率最大的假设  $h \in \lambda$ ,称为极大后验(MAP)假设。利用贝叶斯式计算每个待选假设的后验概率,以此来确定 MAP 假设。

$$h_{MAP} \equiv \operatorname{argmax}_{h \in H} P(\lambda|T) = \operatorname{argmax}_{h \in H} \frac{P(T|\lambda)P(\lambda)}{P(T)} = \operatorname{argmax}_{h \in H} P(T|\lambda)P(\lambda)$$

$P(T)$  是不依赖于  $\lambda$  的常量。在某些情况下,可假定  $\lambda$  中每个假设有相同的先验概率(即对  $H$  中任意  $\lambda_i$  和  $\lambda_j$ ,  $P(\lambda_i) = P(\lambda_j)$ )。进一步简化,只需考虑  $P(T|\lambda)$  来寻找极大可能假设。 $P(T|\lambda)$  常称为给定  $\lambda$  时数据  $T$  的似然度,而使  $P(T|\lambda)$  最大的假设被称为极大似然  $h_{ML}$ 。

$$h_{ML} \equiv \operatorname{argmax}_{h \in H} P(T|\lambda)$$

朴素贝叶斯分类器应用的学习任务中,每个实例  $x$  可由属性值的合取描述,而目标函数  $f(x)$  从某有限集合  $V$  中取值。贝叶斯方法的新实例分类目标是在给定描述实例的属性值  $\{a_1, a_2, \dots, a_n\}$  下,得到最可能的目标值  $V_{MAP}$ 。

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n)$$

基于贝叶斯公式重写为

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} \frac{P(a_1, a_2, \dots, a_n | v_j)P(v_j)}{P(a_1, a_2, \dots, a_n)} = \operatorname{argmax}_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j)P(v_j)$$

朴素贝叶斯 NB 分类器算法基于一个简单的假定:估算目标值时属性  $\{a_1, a_2, \dots, a_n\}$  之间条件是相互独立的,则观察到  $a_1, a_2, \dots, a_n$  的联合概率正好是对每个单独属性的概率乘积:

$$P(a_1, a_2, \dots, a_n | v_j) = \prod_i P(a_i | v_j)$$

将其代入上式中,可得到 NB 分类器所使用的方法:

$$v_{NB} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

当所需的条件独立性能够被满足时,朴素贝叶斯分类器输出的 ( $v_{NB}$ ) 等于 MAP 分类。

### 1.2 朴素贝叶斯邮件过滤流程和步骤

朴素贝叶斯分类算法利用贝叶斯定理的优势,在垃圾邮件过滤应用中有广泛应用,是过滤垃圾邮件最为精确的技术。在智能垃圾邮件过滤技术中,通过贝叶斯过滤器的自我学习智能技术,能有效保护电子邮件的正常通信,过滤垃圾邮件的骚扰。朴素贝叶斯分类分为三个阶段:

a)准备工作阶段。收集大量正常和垃圾邮件作为样本,确定特征属性,并对每个特征属性进行适当划分,然后提取邮件样本中主题和信体中的字符串,建立对应的数据库分类,输出特征属性和训练样本。

b)分类器训练阶段。计算每个类别在训练样本中的出现频率及每个特征属性划分对每个类别的条件概率估计。创建贝叶斯概率库,统计出每个字串在垃圾邮件中出现的概率以及在正常邮件中出现的概率,然后根据公式计算出邮件中含某字串则为垃圾邮件的概率。

c)应用阶段。使用训练好的 Bayes 分类器对待分类项进行分类,其输入是分类器和待分类项,输出是待分类项与类别的映射关系。通过对邮件样本的训练, Bayes 分类器可以自动获取垃圾邮件的特征,并根据特征的变化,对邮件进行有效过滤。

## 2 基于改进的朴素贝叶斯过滤算法

朴素贝叶斯算法是一种简单而高效的分类算法,但是它的条件独立性假设极大影响了分类性能,在此提出了一种基于支持向量机(SVM)、EM 算法融合的改进朴素贝叶斯分类算法,并通过实验结果验证了该算法的有效性。

SVM-EM-NB 算法通过贝叶斯统计方法对邮件进行分类,挖掘有效的数据信息,并结合支持 SVM 技术预测垃圾邮件概率,以及贝叶斯前验分布和后验分布用来估计 SVM 中的参数。通过 NB 和 SVM-EM-NB 的预测结果表明,该文提出的 SVM-NB 方法提高了垃圾邮件的过滤精度,是一种有效的方法。

### 2.1 支持向量机训练

支持向量机是目前一种新兴的技术,在垃圾邮件过滤方面越来越得到重视。SVM 的提出有很深的理论背景,其训练的本质是解决一个二次规划问题,得到的是全局最优解。

SVM 训练的基本思想概括为两点:a)针对线性可分情况进行分析,通过使用核函数与非线性转换算法和将低维输入空间线性不可分的样本变化为高维特征空间并使其线性可分,从而使得高维特征空间可以采用线性算法对样本的非线性特征进行线性分析;b)基于结构风险最小化理论在特征空间中构建最优分割超平面,使得学习器得到全局最优化。

SVM 分类器的优点在于通用性较好,可以提高泛化性能、解决非线性问题,且分类精度高、分类速度与训练样本个数无关,和朴素贝叶斯邮件过滤算法融合将大大提高查准率和查全率。

设分类方程为  $x \cdot w + b = 0$ ,对它进行归一化,使得对线性可分的样本集  $(x_i, y_i), i = 1, \dots, n, x \in R^d, y \in \{+1, -1\}$ , 满足约束条件  $y_i [(w \cdot x_i) + b] - 1 \geq 0, i = 1, \dots, n$ 。

利用 Lagrange 优化方法将上述最优分类面问题转换为对偶形式,即在约束条件

$$\sum_{i=1}^n y_i \alpha_i = 0 \text{ 和 } \alpha_i \geq 0, i = 1, \dots, n$$

下对  $\alpha_i$  求解下列函数的最大值:

$$\text{Max } Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

其中,  $\alpha_i$  为原问题中与每个约束条件对应的 Lagrange 乘子。这样就转换为一个不等式约束下二次函数寻优的问题,即存在唯一解。对应的样本是支持向量,采用线性分类解上述问题后得到的最优分类函数为

$$f(x) = \text{sgn} \{ (w \cdot x) + b \} = \text{sgn} \left[ \sum_{i=1}^n \alpha_i y_i (x_i \cdot x) + b \right]$$

其中,  $b$  表示的是分类阈值,可以用任一个支持向量求得。

对非线性问题,通过非线性映射  $\Phi$  把输入空间的样本转换为某个高维特征空间。核函数  $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ , 这样,在高维空间进行内积运算时,无须了解变换  $\Phi$  的形式。这样不仅能实现非线性变换后的线性分类,还没有增加时间复杂度,此时对偶形式变为  $\text{Max } Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j)$ , 而相应的分类函数变为

$$f(x) = \text{sgn} \left( \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right)$$

SVM 将大的分类工作量在输入空间而不是在高维特征空间中完成,解决算法可能导致的维数灾难问题,快速解决二次规划的问题,具有较高的训练准确度。

### 2.2 EM 算法和朴素贝叶斯分类

EM 期望最大化算法 (expectation-maximization) 是一种迭代方法,主要用来计算后验分布的极大似然估计和用于不完整数据的填补应用。其基本步骤为:a) 在已知变量和当前参数的情况下估计缺失属性的期望,即 E 步;b) 重估参数的期望,保证参数值是在 E 步中补充完整数据的极大似然值,即 M 步最大化问题,最后就可以保证能得到局部收敛最大值。最优分类面如图 1 所示。

定义两个样本空间  $X$  和  $Y$ , 其中  $X$  是完整数据空间,  $Y$  是观察数据 (即 incomplete data), 令  $Z$  表示添加数据那么  $X = (Y, Z)$ ;  $\varphi$  为参数集合;  $g(y|\varphi)$  表示观察后验概率密度函数;  $f(x|\varphi)$  表示添加数据  $Z$  后得到的后验密度函数;  $k(x|y, \varphi)$  表示给定数据  $\varphi$  和观察数据  $y$  下  $x$  的条件密度函数, 则  $k(x|y, \varphi) = f(x|\varphi)/g(y|\varphi)$ 。定义似然函数  $L(\varphi) = \log(g(y|\varphi))$  得出  $L(\varphi) = \log(f(x|\varphi)) - \log k(x|y, \varphi)$ 。定义函数:  $Q(\varphi'|\varphi) = E(\log(f(x|\varphi')|y, \varphi))$  和  $H(\varphi'|\varphi) = E(\log(k(x|y, \varphi')|y, \varphi))$ , 最后得到  $Q(\varphi'|\varphi) = L(\varphi') + H(\varphi'|\varphi)$ 。

记  $\varphi^i$  为第  $i+1$  次迭代开始时参数的估计值, 则第  $i+1$  次迭代的两步为

a) 计算每个隐藏变量  $z_{ij}$  的期望值  $E[z_{ij}]$ , 假定当前假设  $h = \langle \mu_1, \mu_2 \rangle$  成立。

b) 计算一个新的极大似然假设  $h' = \langle \mu_1', \mu_2' \rangle$ , 假定由每个隐藏变量  $z_{ij}$  所取的值为 a) 中得到的期望值  $E[z_{ij}]$ , 将假设  $h = \langle \mu_1, \mu_2 \rangle$  替换为新的假设  $h' = \langle \mu_1', \mu_2' \rangle$ , 然后循环计算每个  $z_{ij}$  的期望值。此  $E[z_{ij}]$  正是实例  $x_i$  由第  $j$  个正态分布生成的概率:

$$E[z_{ij}] = \frac{p(x = x_i | \mu = \mu_j)}{\sum_{n=1}^2 p(x = x_i | \mu = \mu_n)} =$$

$$\frac{e^{-\frac{1}{2\sigma^2}(x_i - \mu_j)^2}}{\sum_{n=1}^2 e^{-\frac{1}{2\sigma^2}(x_i - \mu_n)^2}}$$

因此步骤 a) 可由将当前值  $\langle \mu_1, \mu_2 \rangle$  和已知的  $x_i$  代入到上式中实现。在步骤 b), 使用 a) 中得到的  $E[z_{ij}]$  来导出一新的极大似然假设  $h' = \langle \mu_1', \mu_2' \rangle$ 。如后面将讨论到的, 这时的极大似然假设为

$$\mu_j \leftarrow \frac{\sum_{i=1}^m E[z_{ij}] x_i}{\sum_{i=1}^m E[z_{ij}]}$$

新的表达式只是对  $\mu_j$  的加权样本均值, 每个实例的权重为由其第  $j$  个正态分布产生的期望值。

### 2.3 SVM-EM 的朴素贝叶斯算法

用优化训练的 SVM 训练邮件集, 解决一个二次规划问题, 使学习器得到一个全局最优化解, 然后把数据集分成完整集和缺失集, 计算缺失属性的数据项与完整属性数据项的相关度, 取相关度最大的数据项对应的属性作为缺失属性的一个估计值。此估计值作为 EM 算法的初始值, 然后执行 EM 算法的两步, 完成极大似然估计, 用最后估计的值来完成缺失属性的填补, 最后用朴素贝叶斯分类算法对完整数据集进行分类。

输入:  $T = \{X_1, X_2, \dots, X_n\}$ , 其中,  $X_1, X_2, \dots, X_n$  为原始属性集,  $\lambda = \{t_1, t_2, \dots, t_m\}$  为类别属性。

输出: 样本  $X$  的类别。

算法主要步骤如下:

a) 把数据集  $T$  分为两个数据子集  $T_i$  和  $T_j$ 。  $T_i$  中的记录全部为完整记录, 任何属性不含缺失值;  $T_j$  中的记录为不完整记录, 即属性中含有一个及以上的缺失值。

b) 调用 EM 算法, 完成缺失数据填补。

c) 随机选择 4/5 的样本作为训练集, 剩余 1/5 的样本作为测试集, 计算训练集样本的先验概率  $P(\lambda)$ 。

d) 在假设类条件独立的情况下根据贝叶斯公式计算条件概率  $P(X_i | \lambda)$ 。

e) 根据式(3)计算后验概率  $P(\lambda | X_i)$ , 输出类别, 求出分类准确率。

朴素贝叶斯、EM 和 SVM 算法是整个算法的核心部分, 对垃圾邮件、合法邮件、用户发出的邮件和提取的邮件进行学习 and 过滤。

### 3 仿真实验与结果分析

本文所有的实验都是在普通 PC (Intel Core2 2.53 GHz CPU, 2 GB RAM) 上做的实验, 实验软件环境为 Eclipse 3.6 + Java Bayes 算法, 仿真实验的数据来源为: 英文邮件样本采用 Spam Assassin 邮件库, 它包含 1 897 封正常邮件和 4 150 封垃圾邮件; 中文邮件样本采用笔者 Yahoo 邮箱的 443 封正常邮件和 1 341 封垃圾邮件。

实验对上述两个邮件样本进行了综合测试, 基于 Eclipse 3.6 平台分别使用 Java 语言实现的朴素贝叶斯和改进的基于 EM 和朴素贝叶斯算法构建邮件过滤器, 在不同训练集规模、不同特征数量等评估指标下进行邮件训练和邮件判别实验, 评价算法的查全率和误报率, 平均匹配一封大小为 6.0k 的邮件只需要 0.02 s 左右。邮件查全率 = 正确识别邮件数/垃圾邮件总数; 误报率 = 错误识别垃圾邮件为正常邮件数/正常邮件总

数。实验结果数据如表 1 和图 2 所示。

表 1 NB 和 SVM-EM-NB 查全率对比

训练邮件 总数 /封	朴素贝叶斯 查全率 /%	朴素贝叶斯 误报率 /%	改进的 朴素贝叶斯 查全率 /%	改进的朴素贝叶斯 误报率 /%
1000	75.2	9.7	85.5	5.4
2000	79.1	8.8	87.3	5.0
3000	83.4	7.9	91.3	4.6
4000	86.9	7.3	92.9	4.2
5000	89.6	6.8	95.6	4.0
6000	91.9	6.1	96.6	3.6
7831	93.6	5.6	98.4	3.2

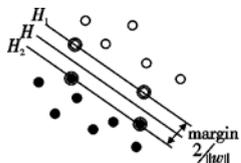


图1 最优分类面

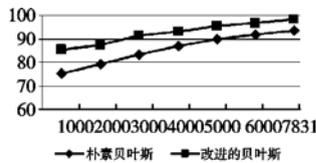


图2 NB和SVM-EM-NB查全率对比

该实验是基于朴素贝叶斯算法和改进的基于 SVM-EM 朴素贝叶斯复合算法进行的。从表 1 可以看出,训练集的大小对查全率及精确度都有较大的影响,一般来说,训练集中邮件样本越多,邮件过滤的性能越好,稳定性越强。

通过实验表明,本文实现的改进朴素贝叶斯复合智能算法,在训练集达到 5 000 之后查全率达到了 95% 以上,远高于朴素贝叶斯算法,这也正说明相对朴素贝叶斯算法的优越性。适当增加训练集样本数,改进的贝叶斯复合智能算法将在查全率和误报率方面有更好的表现。

#### 4 结束语

朴素贝叶斯邮件过滤器简单有效而且可计算性强,可用于更加精确的垃圾邮件过滤器。本文针对传统过滤算法对垃圾

邮件的识别率低的不足,提出了改进的贝叶斯复合智能算法来过滤垃圾邮件。实验结果表明,复合智能算法要优于传统的朴素贝叶斯算法,而且最大程度地保持了算法的自动化和智能化,具有很强的适用性和自我学习性,突破了用朴素贝叶斯方法对大规模数据特征提取时存在的局限性。基于邮件样本集的训练总数的增加会使该算法有更高的精确度,所以笔者后续将对该方面作更深入的研究。

#### 参考文献:

- [1] HUANG J Z, NG M K, RONG Hong-qiang, et al. Automated variable weighting in K-means type clustering [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 27(5): 657-668.
- [2] 刘斌,黄铁军,程军,等.一种新的基于统计的自动文本分类方法[J].中文信息学报,2008,16(6):18-24.
- [3] 朱明,王俊普.一种最优特征集的选择算法[J].计算机研究与发展,2006,35(9):803-805.
- [4] 胡佳妮,徐蔚然,郭军,等.中文文本分类中的特征选择算法研究[J].光通信研究,200,5(3):44-46.
- [5] 潘有能.一个自动分词分类系统的实现[J].情报学报,2002,21(1):38-41.
- [6] 邓乃扬,田英杰.数据挖掘中的新方法:支持向量机[M].北京:科学出版社,2004.
- [7] 巩知乐,张德贤,胡明明.一种改进的支持向量机的文本分类算法[J].计算机仿真,2009,26(7):165-168.
- [8] 解冲锋,李星.基于序列的文本自动分类算法[J].软件学报,2011,13(4):783-788.
- [9] 韩家沛,孟小峰,王静,等. Web 挖掘研究[J].计算机研究与发展,2010,38(4):405-411.
- [10] 杜树新,吴铁军.模式识别中的支持向量机方法[J].浙江大学学报:工学版,2008,37(5):521-527.
- [11] 罗印升,李人厚,张雷,等.人工免疫算法在函数优化中的应用[J].西安交通大学学报,2009,7(8):840-843.

(上接第 1090 页)

- [2] ZHAO M, SERVETTO S D. An analysis of the maximum likelihood estimator for localization problems [C]//Proc of IEEE ICBN. 2005: 982-990.
- [3] BAHL P, PADMANABHAN V N. RADAR: an in-building RF-based user location and tracking system [C]//Proc of IEEE INFOCOM. New York: IEEE Computer and Communications Society, 2000: 775-784.
- [4] 周贤伟.无线传感器网络与安全[M].北京:国防工业出版社,2007.
- [5] CAPKUN S, HUBAUX J P. Secure positioning of wireless devices with application to sensor networks [C]//Proc of IEEE INFOCOM. 2005.
- [6] 许力.无线传感器网络的安全和优化[M].北京:电子工业出版社,2010.
- [7] LIU Dong-gang, NING Peng, DU W. Attack-resistant location estimation in sensor networks [C]//Proc of IEEE IPSN. Piscataway: IEEE Press, 2005: 99-106.
- [8] LAZOS L, POOVENDRAN R, CAPKUN S. ROPE: robust position estimation in wireless sensor networks [C]//Proc of IEEE IPSN. [S. l.]: IEEE Press, 2005: 324-331.

- [9] CHEN H, LOU W, MA J, et al. TSCD: a novel secure localization approach for wireless sensor networks [C]//Proc of the 2nd International Conference on Sensor Technologies and Applications. Washington DC: IEEE Computer Society, 2008: 661-666.
- [10] WANG W, BHARGAVA B. Visualization of wormhole in sensor networks [C]//Proc of ACM WiSe. [S. l.]: ACM Press, 2004: 51-60.
- [11] WANG W, LU A. Interactive wormhole detection and evaluation [J]. Information Visualization, 2007, 1(6): 13-17.
- [12] MAHESHWARI R, GAO J, DAS S R. Detecting wormhole attacks in wireless networks using connectivity information [C]//Proc of IEEE INFOCOM. [S. l.]: IEEE Press, 2007: 107-115.
- [13] LAZOS L, POOVENDRAN R, CAPKUN S. SeRLoc: robust localization for wireless sensor networks [J]. ACM Trans on Sensor Networks, 2005, 1(1): 73-100.
- [14] CHEN Hong-long, LOU Wei, WANG Zhi. A secure localization approach against wormhole attacks using distance consistency [C]//Journal on Wireless Communications and Networking Volume. New York: Hindawi Publishing Corporation, 2010: 1-12.
- [15] 郭浩,赵晓辉,张鹏.基于概率定位的稀疏无线传感器网络[J].空军工程大学学报:自然科学版,2008,9(1):26-29.