

改进的 BT-SVM 应用于电力系统 SSA

朱志慧,李 雷,种冬梅

(南京邮电大学 自动化学院,江苏 南京 210003)

摘 要:随着电力系统的广泛发展,电力系统静态安全评估已变得越来越重要。文中比较了现在几种常用的人工智能方法,选择了支持向量机算法解决这一问题。由于解决大样本问题时,支持向量机所需训练时间显著增加,文中提出了约简样本的方法,并结合适合于电力系统的二叉树结构,提出了一种改进的简化二叉树支持向量机算法。将这种新的支持向量机算法应用于 IEEE57 节点电力系统,结果表明,文中提出的算法取得了比较好的结果,有效可行。

关键词:电力系统;静态安全评估;人工智能;二叉树支持向量机;约简样本

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2012)09-0157-04

Improved Binary Tree Support Vector Machine and Its Application to Power System Static Security Assessment

ZHU Zhi-hui, LI Lei, CHONG Dong-mei

(College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Power system static security assessment is becoming more and more important with the expansion of electrical power system. It compares several common artificial intelligence methods and then selects the support vector machine (SVM) algorithm. Due to the large number of the training set, the training time of SVM increases significantly. To solve this problem, an improved simplifying SVM method is put forward in this article. It decreases the number of the samples and combines the suitable binary tree structure for power system. The proposed simplified SVM algorithm has been applied to IEEE 57-bus power system. The simulation results demonstrate the effectiveness of the proposed algorithm.

Key words: power system; static security assessment; artificial intelligence; binary tree support vector machine (SVM); simplifying sample

0 引 言

电力系统静态安全分析是指判断电力系统在发生预想事故后系统是否会发生过负荷或电压越限的能力。这些年来,随着电力系统规模的不断扩大和发电量的增长,建立可靠的电力安全分析系统,已成为越来越重要的问题。在这种条件下,在线实时控制任何可能危及电力系统安全性和稳定性的干扰的要求也随着时代的发展越来越高。而计算机技术的发展为电力系统的安全性分析管理提供了有利的条件^[1]。

电力系统的安全性分析主要分为静态安全性评估(SSA)和暂态安全性评估(TSA)。文中主要研究的是静态安全分析。自 20 世纪 90 年代以来,研究人员提出了多种方法解决电力系统的可靠性分析。其中蒙特

卡罗模拟系统可靠性评估方法是通过随机生成几个系统状态和评估结构函数或系统函数来实现的^[2];神经网络凭借其良好的分类能力和精度,也是这几年比较流行的方法。

统计学习理论是一个较完善的基于有限样本的理论体系,它是建立在结构风险最小化原则以及 VC 维概念基础上的^[3]。支持向量机(Support Vector Machine, SVM)是在统计学习理论上提出的一种模式识别的新方法,它具有很高的精度和良好的泛化性能。SVM 已经越来越广泛地应用于各个领域,如:人脸识别、字符识别、入侵检测、图像识别、图像分类、信号处理等,包括电力系统的安全分析。

最初的支持向量机都是为两类分类问题而设计的,但多类分类问题在实际的应用中更为普遍,因此需要将支持向量机的优良性能推广到多类分类中去。人们也为此做出了大量研究,提出了许多多类支持向量机算法。文中,电力系统的运行状态可以分为 5 个级别:安全正常状态(A)、不安全正常状态(B)、持久性

收稿日期:2012-01-13;修回日期:2012-04-18

基金项目:国家自然科学基金项目(61070234, 61071167)

作者简介:朱志慧(1987-),女,江苏南通人,硕士研究生,研究方向为数据挖掘与智能计算;李 雷,教授,研究方向为智能信号处理、非线性分析与计算智能。

紧急状态 (C)、稳定性紧急状态 (D) 及恢复状态 (E), 需用到多分类算法。M. Mohammadi 和 G. B. Ghahrebian 提出了一种基于决策树的多分类支持向量机来研究电力系统的在线静态安全问题^[4]。电力系统安全分析需要比较大的训练样本数目, 而此时支持向量机的训练时间和计算机内存将急剧增加。因此文中提出了一种改进二叉树支持向量机, 约简训练样本, 减少样本规模, 以达到更好的使用支持向量机解决电力系统静态安全的目的。

1 支持向量机及多类分类方法

1.1 支持向量机模型

Vapnik 等人在 1995 年提出基于统计学习理论的一种模式识别新方法——支持向量机 (SVM)^[4]。支持向量机的基本原理要求找到一个最优分类超平面, 它能够满足分类要求及保证分类精度, 并能使两个不同类之间的距离达到最大化。理论上来说, 支持向量机可以用来解决线性问题, 对于非线性问题, Vapnik 等人引入核映射方法, 将其转化为高维空间的线性可分问题^[5]。

对于训练样本集 $\{(x_i, y_i)\}_{i=1}^l \in R^n \times \{+1, -1\}$, $i = 1, 2, \dots, l$, 此时训练支持向量机所得的最优平面为 $(w \cdot x) + b = 0$, $w \in R^N$, $b \in R$, w 为最优平面的法向量, b 为常数项。下面的决策函数便可用来区分位于最优平面两边的两类的任意一点:

$$f(x) = \text{sgn}((w \cdot x) + b) \quad (1)$$

求此平面即求最优化问题:

$$\min \tau(w) = \frac{1}{2} \|w\|^2 \quad (2)$$

约束条件为:

$$y_i((w \cdot x_i) + b) \geq 1, i = 1, 2, \dots, l \quad (3)$$

求解这个约束优化问题, 引入 Lagrange 函数, 再根据最优性条件——KKT 条件, 求解得最优分类决策函数:

$$f(x) = \text{sgn}\left(\sum_{i=1}^l a_i y_i (x \cdot x_i) + b\right) \quad (4)$$

其中 $a_i > 0$ 为 Lagrange 系数。

在非线性问题中, 引入核函数 $K(x, x')$, K 对应某高维特征空间中的内积:

$$K(x, x') = \varphi(x) \cdot \varphi(x') \quad (5)$$

则求解所得的最优分类函数为:

$$f(x) = \text{sgn}\left(\sum_{i=1}^l a_i y_i K(x, x_i) + b\right) \quad (6)$$

1.2 多分类支持向量机

支持向量机最初是用来解决两类分类问题的, 但在实际的应用中更多地接触的是多分类问题, 这就必

须涉及到多分类支持向量机。有两种多分类支持向量机方法: 第一个是组合一系列的二元分类器, 使其实现多分类; 第二个方法是直接在决策函数上进行改进, 处理所有数据, 使其建立 K 分类支持向量机^[6]。基于第二种方法计算量十分大, 通常采用第一种方法, 目前有四种比较常用的组合二元分类器的方法。

1) 一对多方法 (OAA) 是最早实现多分类支持向量机的方法^[7]。它对于 K 类问题构造 K 个二元分类器, 第 i 个分类器中只有第 i 类的训练样本为正, 其余样本均为负。这种方法可能存在一个测试样本可能属于多类或者都不属于任何一类情况, 其泛化性能较差, 训练困难。

2) 一对一方法 (OAO), 该方法共构造 $K(K-1)/2$ 个分类器, 每个分类器使得 K 类中的所有可能的两类训练样本进行训练^[8]。最后采用投票法, 得出 x 属于得票数最多的那个类。这种方法可能存在多个类投票相同的情况。

3) 导向非循环图方法 (DAG), 它的训练阶段和 OAO 方法比较接近, 有 $K(K-1)/2$ 个二元分类器^[9]。但在决策阶段, 它使用了一个具有 $K(K-1)/2$ 个内部节点和 K 个叶子节点的导向非循环图 (DAG)。测试样本从根节点开始, 用二元决策函数的输出值决定其所走的路径: 左侧或者右侧。这种方法的缺点在于根节点的选择直接影响分类结果。

4) 基于二叉树的支持向量机 (BT-SVM) 将所有的类分成两个子类, 每个子类再进行划分, 如此循环, 最终 K 类样本需构造 $K-1$ 个二元支持向量机^[10]。BT-SVM 的优点在于不存在不可分区域, 但是对其二叉树的结构要求也非常的高, 因此确定一个好的二叉树结构至关重要。

常见的基于二叉树的支持向量机结构有两种: 偏二叉树和完全二叉树^[11]。半对半 (HAH) 算法是基于完全二叉树结构的, 在文献 [12] 中发现与其它算法相比, HAH 在测试速度、分类器模型的大小及准确率上均有优势。由于电力系统静态安全分析的样本分为 5 种状态, 且状态 A 的样本数目最大, 5 种样本在特征空间中的分布不均衡, 各个聚类中心彼此之间也不等距, 文中采用了一种介于偏二叉树和完全二叉树之间的方法, 即一种改进的半对半 (HAH) 算法。其结构见图 1。

2 改进的简化支持向量机

二叉树支持向量机不存在不可分区域, 但是它存在错误累积的问题, 若在根节点划分出现错误, 会把错误一直延续下去, 后续节点分类错误就会越来越大。因此, 二叉树的结构非常重要, 在二叉树的每一个节点应该使最容易分开的类尽可能的分开。而且, 电力系

统具有大量的训练样本数目,而遇到大样本问题时,支持向量机所需的训练时间和计算机所需内存都急剧增加。由帕累托的 80/20 法则,在任何特定的群体中,重要的因子占少数,而不重要的因子通常占多数^[13]。支持向量机中能成为支持向量的样本通常在边界上,基于电力系统的大样本特点,约简边界样本能起到简化支持向量机的作用^[14]。基于以上因素,文中提出了改进的二叉树支持向量机算法。

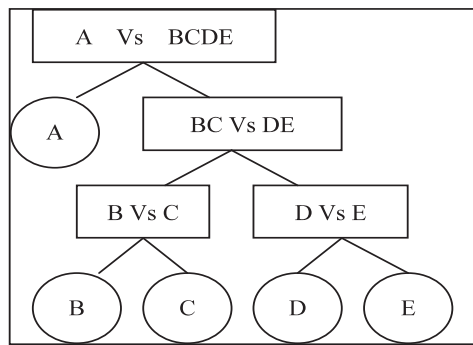


图 1 近似完全二叉树分类结构

2.1 约简样本方法

在两类分类问题中,若有训练样本: $(x_1, \dots, x_m, x_{m+1}, \dots, x_n)$, 其中 n 为训练样本总数, x_1 到 x_m 为 A 类样本, x_{m+1} 到 x_n 为 B 类样本。若两类为非线性情况,使用核函数原元空间的向量映射到某高维特征空间中,此时训练样本变为: $(\varphi(x_1), \dots, \varphi(x_m), \varphi(x_{m+1}), \dots, \varphi(x_n))$

求 A、B 两类的类中心:

$$C_A^\varphi = \frac{1}{m} \sum_{i=1}^m \varphi(x_i) \quad (7)$$

$$C_B^\varphi = \frac{1}{n-m} \sum_{i=1}^{n-m} \varphi(x_i) \quad (8)$$

两类类中心距离为:

$$D^\varphi = C_A^\varphi - C_B^\varphi \quad (9)$$

两类的中点向量为:

$$M^\varphi = \frac{1}{2} D^\varphi \quad (10)$$

(1) 取阈值 $(\frac{1}{3} D^\varphi)$, 计算所有样本 $\varphi(x_i)$ 与中点向量 M^φ 的距离, 所有距离小于阈值的样本为边界样本。

(2) 对于非边界样本, 先以两类中心为中心样本, 计算所有非边界样本与所有的中心样本的最小距离, 若大于指定数值 $(\frac{2}{3} D^\varphi)$, 则该样本也作为中心样本, 否则约简该样本。根据此要求依次检查所有的非边界样本, 使其成为中心样本或被约简。

最后去掉类中心样本, 使其余所有的中心样本和边界样本一起构成训练样本。

2.2 等效距离

设两个样本 x_a 、 x_b 分别属于类 A 和类 B, 非线性情况下, 两个样本之间的欧氏距离可以表示为:

$$d(x_a, x_b) = \|\varphi(x_a) - \varphi(x_b)\|_2 = \sqrt{K(x_a, x_a) - 2K(x_a, x_b) + K(x_b, x_b)} \quad (11)$$

两类样本之间的距离通常用样本中心的欧氏距离 d_{AB} 来表示。但是这种情况不能很好地考虑两类样本的分布特点, 很难正确地反映两类样本的可分性, 尤其对于文中的电力系统静态安全样本分布特点。因此引入等效距离的概念, 两类样本类 A 与类 B 的等效距离表示为:

$$D_{AB} = d_{AB} / (\sigma_A + \sigma_B) \quad (12)$$

其中 σ_A 为特征空间中 A 类样本的标准差, σ_B 为特征空间中 B 类样本的标准差, 等效距离越大的两类越容易被分出。

2.3 文中的算法流程

根据电力系统静态安全评估的五种状态的样本分布特点, 文中提出了一种近似完全二叉树的支持向量机结构, 首先设计 SVM 识别出 A 类, 再区分 (B、C) 类和 (D、E) 类, 最后设计两个 SVM 分别识别 B、C 类, D、E 类。

具体算法步骤如下:

- (1) 采用高斯径向基核函数计算每个类与其它类之间的欧式距离及等效距离;
- (2) 分出与其它类等效距离最大的一类;
- (3) 对于剩下的类, 把两个等效距离最小的类合并成一个类 M, 计算这个类 M 和其它类的等效距离;
- (4) 找出与类 M 等效距离最大的类 N;
- (5) 对类 M 和类 N 分别重复步骤 (3), 即不断找出分别与 M、N 类等效距离最小的类并将其合并, 直至所有的类都被并入 M 类或 N 类中, 所有剩下的类被分为两类: M 类和 N 类;
- (6) 依次重复步骤 (3) 到步骤 (5), 即不断使两个大类一分为二, 知道所有的类都被完全分开为止。

3 实验和分析

3.1 基础知识介绍

文中使用性能指标 PI_{MW} 对电力系统遇到干扰的严重性进行量化并对其安全状态分类:

$$PI_{MW} = \sum_{l=1}^{N_L} \left(\frac{w_l}{2n} \right) * \left(\frac{P_l}{P_l^{\lim}} \right)^{2n} \quad (13)$$

其中: P_l 为支路 l 的功率流; P_l^{\lim} 为支路 l 功率的限值; N_L 为支路的总数目; w_l 为支路 l 功率的权重 (文中设其 = 1); n 为补偿函数的指数^[15]。

文中提出的改进的简化二叉树支持向量机算法模

型的性能由下面的性能参数来决定:

(1) 均方误差 (MSE)

$$MSE = \frac{1}{k} \sum_{i=1}^k |DO_i - AO_i|^2 \tag{14}$$

其中: k 为总的类别数; DO_i 为从离线模拟得到的预期结果; AO_i 为由支持向量机模型得出的实际结果。

(2) 总准确率 (TA)。

$$TA = \frac{\sum_{i=1}^k T_{k_i}}{\sum_{i=1}^k T_{k_i} + \sum_{i=1}^k F_{k_i}} \tag{15}$$

其中: k 为总的类别数; T_{k_i} 为分类正确的样本总数; F_{k_i} 为分类错误的样本总数。

(3) 平衡错误率 (BER)。

$$BER = \frac{1}{k} \left(\sum_{i=1}^k \frac{F_{k_i}}{T_{k_i} + F_{k_i}} \right) \tag{16}$$

3.2 实验结果和分析

文中提出了一种基于二叉树的改进的简化支持向量机算法,并应用于电力系统静态安全评估中,采用 IEEE57 节点电力系统进行实验分析。实验中采用 LIBSVM 软件包^[16],基于 Windows 操作系统 VC++6.0 环境。

系统需定义两类样本数据:训练样本和测试样本。实验时通过改变总负载来考虑不同的操作状态,所有节点的负载在基本状态的 50% 到 200% 之间变化,共生成 1532 个不同的操作点。根据文中提出的约简样本的方法,选择了其中的 1378 个操作点进行实验。在所有的操作点中,文中随机选取 1213 个操作点作为训练数据,剩余的 165 个操作点作为测试数据。

训练时对所有的突发事件进行模拟,并根据公式 (13) 逐一计算其相应的性能指标 PI_{MW} 。每个突发事件下的每个操作点,其安全状态由 PI_{MW} 决定,再根据安全状态确定每个样本属于 A、B、C、D 或 E 中的哪个类别。

根据文中所提出的二叉树的结构训练构造的 4 个二元支持向量机,最后识别出每个类的样本个数如表 1 所示:

表 1 IEEE57 节点电力系统的安全状态

学习样本	IEEE 57 节点电力系统
样本总数	1378
安全正常状态 (类 A)	719
不安全正常状态 (类 B)	312
持久性紧急状态 (类 C)	187
稳定性紧急状态 (类 D)	107
恢复状态 (类 E)	53
训练样本总数	1213
测试样本总数	165

分别将文中的方法和传统的 BP 神经网络、一对多 (OAA) 多类分类支持向量机算法以及普通的基于二叉树的多分类支持向量机算法进行比较,结果如表 2 所示:

表 2 电力系统静态安全分析各种分类器的性能比较

分类器	MSE	TA (%)	BER (%)
本文算法	0.13	96.10	0.75
BP 神经网络算法	0.37	91.24	6.01
OAA 算法	0.31	93.47	4.38
BTSVM 算法	0.17	96.14	0.95

结果表明:文中提出的算法在均方误差和平衡错误率上均优于其它算法,精确率和 BTSVM 算法差不多,优于其它两个算法。

4 结束语

文中基于二叉树多类分类支持向量机提出了一种改进的简化支持向量机算法,并将其应用于电力系统静态安全评估中。二叉树支持向量机在训练速度、测试速度及所需支持向量机数量上,其性能均优于一对多、一对一和导向非循环图支持向量机算法。考虑到需应用于样本规模大且样本分布不均衡的电力系统静态安全分析中,文中提出了一种改进的算法,对大样本进行适当的约简,并优化二叉树的分类结构,并将本算法应用于 IEEE57 节点电力系统的静态安全评估中。由实验数据得出文中提出的算法取得有效的结果。将来的工作重点可以考虑结合模糊支持向量机算法进行电力系统静态安全评估研究。

参考文献:

[1] 曾 沅,余贻鑫. 电力系统动态安全域的实用解法[J]. 中国电机工程学报,2003,23(5):24-28.

[2] Thukaram D, Khincha H P, Vijaynarasimha H P. Artificial neural network and support vector machine approach for locating faults in radial distribution systems[J]. IEEE Trans on Power Delivery,2005,20(2):710-721.

[3] 高 隽. 神经网络原理及仿真实例[M]. 北京:机械工业出版社,2007.

[4] Mohammadi M, Gharehpetian G B. Application of multi-class support vector machines for power system on-line static security assessment using DT-based feature and data selection algorithms[J]. Journal of Intelligent & Fuzzy Systems,2009,20(3):133-146.

[5] 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.

[6] 杨 斌,路 游. 基于统计学习理论的支持向量机的分类方法[J]. 计算机技术与发展,2006,16(11):56-58.

质量值 $W(DW)$ 为 0.634,按照“[0,0.5]:一般、[0.5,0.8]:良好、[0.8,1]:优秀”的划分,实验验证的数据集市的设计质量为良好。对于其他数据集市的设计质量,可通过同样的计算方法得出。

表 4 $W_q(t_j,a_i)$ 计算结果

Dimension	
Attribute	$W_d(t_j,a_i)$
POrderID	0.612
DepartmentID	0.673
Measure	
Attribute	$W_m(t_j,a_i)$
OSchedulePHours	0.659
OScheduleCost	0.593

3 结束语

文中提出了一种自动验证数据仓库设计质量的定量分析方法,用数据去衡量数据仓库的设计质量,进一步研究的话,可以用文中提出的方法开发一个自动验证数据仓库设计质量的辅助判断工具。当然,一个数据仓库设计的好坏,往往是一个综合评定的结果。文中得出的结果只不过是定量分析的结果,可以作为判断数据仓库设计质量好坏的一个判断因素,另外,可以结合数据仓库的逻辑结构、设计经验等非定量因素去综合判定数据仓库设计质量的好坏。

参考文献:

[1] Chengalur-Smith I N, Ballou D P, Pazer H L. The impact of data quality information on decision making: an exploratory analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 1999, 11(6): 853-864.

[2] Aebi D, Perrochon L. Towards improving data quality[C]//Proc of the international conference on information systems and management of data. [s. l.]: [s. n.], 1993: 273-281.

[3] Wang R Y, Strong D M. Beyond accuracy: what data quality

means to data consumers[J]. Journal of Management Information Systems, 1996, 12(4): 5-30.

[4] Karr A F, Sanil A P, Banks D L. Data quality: a statistical perspective[J]. Statistical Methodology, 2006, 3(2): 137-173.

[5] Bouzeghoub M, Kedad Z. A Quality-based Framework for Physical Data Warehouse Design[C]//Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000). Stockholm, Sweden: [s. n.], 2000: 5-6.

[6] Theodoratos D, Bouzeghoub M. Data Currency Quality Factors in Data Warehouse Design[C]//Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'99). Heidelberg, Germany: [s. n.], 1999.

[7] Manfred A, Jeusfeld C Q. Design and Analysis of Quality Information for Data Warehouse[C]//European Commission in ESPRIT Long Term Research Project 22469 DWQ. [s. l.]: [s. n.], 1998.

[8] Vassiliadis P, Bouzeghoub M, Quix C. Towards Quality-oriented Data Warehouse Usage and Evolution[C]//The 11th International Conference on Advanced Information System Engineering. [s. l.]: [s. n.], 2000: 89-115.

[9] Calero C, Piattini M. Towards Data Warehouse Quality Metrics[C]//Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2001). Interlaken, Switzerland: [s. n.], 2001.

[10] Zuse H. A Framework of Software Measurement[M]. Hawthorne, NJ, USA: Walter de Gruyter & Co., 1997.

[11] Wang Y, Wang R Y. Anchoring Data Quality Dimension in Ontological Foundation[J]. Communication of the ACM, 1996, 39(11): 86-95.

[12] Wang R Y, Storey V C, Firth C P. A Framework for Analysis of Data Quality Research[J]. IEEE Transaction on knowledge and data engineering, 1995, 7(4): 623-640.

[13] Pipino L L, Lee Y W, Wang R Y. Data Assessment[J]. Communication of the ACM, 2002, 45(4): 211-218.

(上接第 160 页)

[7] Vapnik V. Statistical Learning Theory[M]. New York: John Wiley & Sons, 1998.

[8] Platt J, Cristianini N, Shawe-Taylor J. Large Margin DAGs for Multiclass Classification[C]//Proceedings of Neural Information Processing Systems. [s. l.]: MIT Press, 1999: 547-553.

[9] 张苗, 张德贤. 多类支持向量机文本分类方法[J]. 计算机技术与发展, 2008, 18(3): 139-141.

[10] Takahashi F, Abe S. Decision Tree-based Multiclass Support Vector Machines[C]//Proceedings of the 9th International Conference on Neural Information Processing. [s. l.]: [s. n.], 2002: 1418-1422.

[11] 胡国胜. 支持向量机及在电力系统中的应用[J]. 高电压技术, 2007, 33(4): 101-105.

[12] Lei H S, Govindaraju V. Half-Against-Half Multi-class Support Vector Machines[C]//Proceedings of Multiple Classifier Systems 2005. USA, California: [s. n.], 2005: 156-164.

[13] 刘健, 刘忠, 熊鹰. 改进的二叉树支持向量机多类分类算法研究[J]. 计算机工程与应用, 2010, 46(33): 117-120.

[14] 刘培胜, 贾银山, 韩云萍. 一种改进的简化支持向量机[J]. 辽宁石油化工大学学报, 2009, 29(1): 76-78.

[15] Srivani J, Swarup K S. Power system static security assessment and evaluation using external system equivalents[J]. Electrical Power and Energy Systems, 2008, 30(2): 83-92.

[16] Chang Chin-Chung, Lin Chih-Jen. LIBSVM: A Library for Support Vector Machines[EB/OL]. 2005. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

改进的 BT-SVM 应用于电力系统 SSA

作者: [朱志慧](#), [李雷](#), [种冬梅](#)
作者单位: [南京邮电大学 自动化学院, 江苏 南京 210003](#)
刊名: [计算机技术与发展](#)
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2012(9)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjtz201209042.aspx