粗糙集学习机器泛化性能控制的结构风险最小化方法

刘金福 于达仁
(哈尔滨工业大学能源学院 哈尔滨 150001)

摘 要 对影响粗糙集学习机器泛化性能的因素进行了分析,通过将结构风险最小化原则引入到粗糙集学习中,提出了粗糙集学习的结构风险最小化方法;通过12个UCI数据集上的实验分析,验证了提出方法的有效性。

关键词 粗糙集,泛化性能,结构风险最小化
中图法分类号 TP391 **文献标识码** A

Structural Risk Minimization for Controlling Generalization Performance of Rough Set Learning Machine

LIU Jin-fu YU Da-ren
(School of Energy Science and Engineering, Harbin Institute of Technology, Harbin 150001, China)

Abstract The factors influencing the generalization performance of rough set learning machine were analyzed. Through introducing the principle of structural risk minimization into rough set learning process, structural risk minimization on rough set learning was proposed. Experiments on 12 UCI data sets show that the proposed method is effective for improving the generalization performance of rough set learning machine.

Keywords Rough sets, Generalization performance, Structural risk minimization

粗糙集理论[1]是波兰学者Z. Pawlak于1982年提出的一种新型的用于处理不完备、不精确、不一致信息的数学工具,与Zadeh提出来的词计算理论[2]、张钹和张铃教授提出来的商空间理论[3]合称三大粒度计算理论。在过去的20多年里,该理论取得了长足发展,在特征选择[4]、规则提取[5,6]和分类器构造[7]等方面得到了广泛的应用,已经成为机器学习领域十分活跃的研究分支。然而,作为一个学习机器,其泛化性能的控制问题一直未能得到很好的解决,这在很大程度上影响了粗糙集学习机器的实际应用效果。

本文将从影响粗糙集学习机器泛化性能的因素分析入手,对粗糙集学习机器的泛化性能控制问题展开系统研究,提出问题的解决方法。

1 影响粗糙集学习机器泛化性能的因素

为了从理论上对影响粗糙集学习机器泛化性能的本质因素进行揭示,首先将粗糙集学习问题抽象为一般的机器学习问题,即从给定的待搜索函数集中选择出能够最好地逼近已知数据的函数。由于规则集是粗糙集学习机器实现的分类函数,因此,粗糙集学习机器的待搜索函数就是规则集,待搜索函数集就是所有的待搜索规则集组成的集合。

将粗糙集学习问题抽象为一般的机器学习问题后,就可以借鉴当前机器学习问题的研究成果,从理论上对粗糙集学习机器的泛化性能问题展开分析和研究。由Vladimir N. Vapnik等在20世纪70年代建立的统计学习理论系统地研究了机器学习的问题,特别是有限样本情况下的机器学习问题,这一理论为粗糙集学习机器泛化性能问题的研究提供了直接的理论支持。

统计学习理论指出,一个学习机器的期望风险(泛化能力)与两方面因素有关:一是学习机器在已知数据上的经验风险;二是学习机器的复杂度。对于一个采用完全有界非负损失函数集的学习机器(无界非负和完全有界损失函数集的学习机器有类似结论),期望风险 $R(\alpha)$ 以至少 $1-\eta$ 的概率满足:

$$R(\alpha)\leqslant R_{emp}(\alpha)+\frac{B\xi}{2}\left(1+\sqrt{1+\frac{4R_{emp}(\alpha)}{B\xi}}\right) \tag{1}$$

其中:

1)右边第一项为经验风险,$R_{emp}(\alpha)=\frac{1}{n}\sum_{i=1}^{n}Q(z,\alpha)$,$0\leqslant Q(z,\alpha)\leqslant B$ 为完全有界非负损失函数,$\alpha\in\Lambda$,Λ 为广义参数集合,z 为数据样本对(x,y),n 为数据样本的数目;

2)右边第二项为学习机器的置信范围,与学习机器的复杂度有关。当函数集 $Q(z,\alpha)$,$\alpha\in\Lambda$ 包含无限多个元素且 VC 维 h 有限时,$\xi=4\frac{h(\ln(2n/h+1))-\ln(\eta/4)}{n}$,$VC$ 维 h 代表了学习机器的复杂度。当函数集 $Q(z,\alpha_i)$,$i=1,2,\cdots,i,\cdots,N$ 仅包含 N 个元素时,$\xi=2\frac{\ln N-\ln\eta}{n}$,$N$ 代表了学习机器的复杂度。由于粗糙集学习机器的待搜索规则集集合中的规则集数目有限,因此,粗糙集学习机器置信范围的分析属于这种情况。

由学习机器的期望风险的界可知:当已知数据样本趋向

到稿日期:2009-05-04 返修日期:2009-07-06 本文受国家自然科学基金(No. 60703013),中国博士后科学基金(20080440886)资助。
刘金福(1977—),男,讲师,主要研究方向为自动控制、机器学习、粗糙集等,E-mail:jinfu_liu1977@yahoo.com.cn;于达仁(1966—),教授,博士生导师,主要研究方向为自动控制、机器学习等。

于无穷多时，ξ趋向于零，置信范围趋向于零，学习机器的经验风险代表了期望风险，此时基于经验风险最小化归纳原则构建的学习机器就能够获得满意的泛化性能；当已知数据样本有限时，特别是当 n/h 或 n/N 较小时，学习机器的复杂度则成为影响学习机器泛化性能的一个重要因素，此时要控制学习机器的泛化性能就必须同时控制学习机器的经验风险和复杂度。

在现实问题中，样本数目通常是有限的，因此，对于粗糙集学习机器，要提高其泛化性能，就需要同时对学习机器的经验风险和复杂度进行控制，经验风险和复杂度是影响粗糙集学习机器泛化性能的两个根本因素。

事实上，复杂度控制的思想和理论由来已久。14 世纪英国哲学家 William of Occam 提出的奥卡姆剃刀原理、20 世纪 60 年代提出的解决不适定问题（如密度估计的非参数方法的正则化技术[8]）、20 世纪 70 年代提出的最小描述长度原则[9]等无一不体现着复杂度控制的思想。然而，直到 1992 年 Vapnik 等[10]直接将复杂度控制技术用于学习机器的构造，提出了基于 Δ-间隔分类超平面的支持向量机学习方法，人们才真正意识到学习机器复杂度控制的实际重要性，支持向量机的出现具有重大的里程碑意义。由于支持向量机具有对学习机器复杂度的直接控制机制，其泛化性能在理论上能够得到保障，它也是目前公认的泛化性能最好的学习机器之一。

现在，对学习机器的复杂度实施控制已经成为一种被人们所广为采纳的技术手段，用于改善各种学习机器的泛化性能，典型的例子如决策树学习中的剪枝技术[11]、神经网络学习中神经元数量和连接权值的控制[12]等。实践表明，通过引入复杂度控制技术，学习机器的泛化性能得到了明显的改善。

统计学习理论为人们系统地研究学习机器的泛化性能问题提供了有力的理论基础，它明确指出在有限样本情况下学习机器的复杂度是影响其泛化性能的一个至关重要的因素，大量的事实也证明了复杂度控制对提高学习机器泛化性能的重要性。因此，为了提高粗糙集学习机器的泛化性能，需要对学习机器的复杂度进行有效的控制。

2 粗糙集学习机器缺乏复杂度控制机制

由上一节分析可知，粗糙集学习机器的复杂度由全部待搜索规则集的数目决定，然而，现有的粗糙集学习算法尚未对粗糙集学习机器待搜索规则集集合进行直接的控制，缺乏对复杂度的控制机制，势必将直接影响粗糙集学习机器的泛化性能。接下来，将通过粗糙集学习的属性约简和规则提取两个过程分别对此进行分析。

在目前采用的各种属性约简算法中，基本的做法是保持分类的正域大小 $|POS_B(D)|$ 不变，其中 $POS_B(D)=\bigcup_{X\in U/D}\underline{B}(X)$，$B\subseteq C$ 为条件属性集，D 为决策属性集，U/D 为决策等价类族，$\underline{B}(X)$ 为 X 下近似。由正域的定义可知，正域仅与各决策类的下近似有关，而与组成各决策类下近似的基本信息粒子（等价类）的结构如大小、数目等无关，即与粒化结构无关，因此，属性约简结果与粒化结构无关。举例来说，假设对于两个不同的条件属性子集，它们分别对论域进行粒化后，得到的基本信息粒子以及决策类 X 的下、上近似如图 1 所示，可以看出，尽管两个条件属性子集对论域的粒化结构不同，但由于 X 的下、上近似相同，因此，在属性约简时，这两个条件属性子集不存在任何差异。众所周知，规则提取是基于等价类进行的，因此，图 1 所示的粒化结构的差异将会对提取的规则集产生明显的影响。若不考虑粒化结构的差异，上述属性约简过程将无法对粗糙集学习机器的复杂度进行控制。

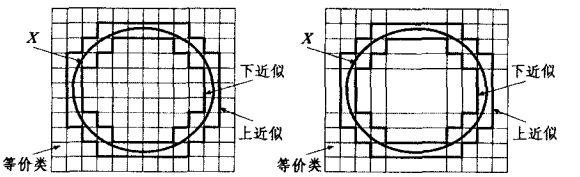

图 1　下、上近似相同而粒化结构不同的情形

在属性约简过程中，最小约简和最小值域空间约简考虑了条件属性数目和属性组合值域空间的大小，是目前仅有的在粗糙集学习算法中对粒化结构实施直接控制的尝试。不可否认，条件属性数目和属性值域空间的大小与提取的规则集之间存在一定的联系，然而它们之间并不存在直接的一一对应关系，这是因为每个等价类并不直接产生规则，最终的规则一般都是由若干个等价类合并后得到的。可见，最小约简和最小值域空间约简对粒化结构的考虑，并不是直接从粗糙集学习机器的规则集出发，也就达不到有效控制粗糙集学习机器复杂度的目的。

完成属性约简后，在约简的条件属性集上能够进行规则提取。目前的规则提取算法大体上可以分为 3 类：最小规则集提取算法、全部规则集提取算法和用户满意规则集提取算法。然而，各种规则提取算法都没有对待搜索规则集集合进行直接的控制，用户满足规则集提取算法采用的阈值条件仅仅是针对规则集中的个别规则进行的间接控制，并不是对待搜索规则集集合进行的控制。

从上述粗糙集学习的属性约简和规则提取过程可以看出，目前的粗糙集学习算法尚缺乏对粗糙集学习机器待搜索规则集集合进行直接控制的机制，即不具备对学习机器的复杂度进行控制的机制，由上一节分析可知，这势必会直接影响粗糙集学习机器的泛化性能。

3 粗糙集学习的结构风险最小化

结构风险最小化归纳原则[13]（SRM 原则）是由 Vapnik 和 Chervonenkis 在 1974 年提出的用于学习过程构建的原则，这一原则使得学习机器的复杂度成为一个直接可控的变量，从而使学习机器的经验风险和复杂度能够被同时控制。

基于 SRM 原则，构建粗糙集学习过程如下：

1）构造粗糙集学习机器待搜索规则集集合的嵌套结构 $S_1\subset S_2\subset\cdots\subset S_i\subset\cdots$，其中 $S_i=\{Q(z,\alpha),\alpha\in\Lambda_i\}$，$Q(z,\alpha)$ 为上述 3 类损失函数之一，则结构元素的复杂度满足 $N_1<N_2<\cdots<N_i<\cdots$，其中 N_i 为 S_i 中规则集的数目，随着结构元素序号的增加，经验风险将减小，复杂度将增加，从而置信范围将增加；

2）从结构中选择一个合适的元素 S^*，使得在 S^* 中最小化经验风险，学习机器的期望风险也最小，则在 S^* 中使经验风险最小的规则集即为粗糙集学习机器的最优规则集，结构元素 S^* 的选择即实现了对粗糙集学习机器复杂度的控制。

图 2 给出了粗糙集学习的结构风险最小化原理图，SRM 原则给出了一种使粗糙集学习机器在经验风险和复杂度之间进行折衷的方法，其对结构元素的选择即实现了对学习机器

复杂度的直接控制。

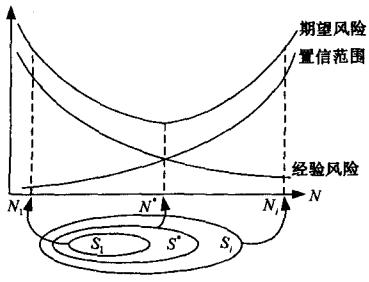

图 2　粗糙集学习的结构风险最小化原理图

一个规则集实际上实现了一种从前件集合到后件集合的映射关系，因此，对于给定的前件和后件集合，全部规则集集合即为从前件集合到后件集合的所有映射关系组成的集合。假设给定前件集合的元素数目为 N_C，后件集合的元素数目为 N_D，则从前件集合到后件集合的所有映射关系的数目为 $N_D^{N_C}$，因此，全部规则集的数目为 $N_D^{N_C}$；当将前件集合中元素的数目减少为 N_B 时，全部规则集的数目将减少为 $N_D^{N_B}$，若此时的规则集集合记为 S_B，原规则集集合记为 S_C，则有 $S_B \subset S_C$。从上述分析可以看出，对于给定的前件和后件集合，通过逐步减少前件集合中元素的数目，能够得到规则集集合之间的嵌套结构。由于前件集合中元素的数目即为每个规则集中规则的数目，因此，要得到粗糙集学习机器待搜索规则集集合的嵌套结构，只需逐步减少所提取规则集中规则的数目，规则集中规则的数目可以用来代表粗糙集学习机器的复杂度。

基于 SRM 原则，通过逐步减少所提取规则集中规则的数目，构造粗糙集学习机器待搜索规则集集合的嵌套结构，能够使粗糙集学习机器的复杂度在粗糙集学习的属性约简和规则提取过程中成为一个直接可控量，从而能够对粗糙集学习机器的复杂度进行直接有效的控制，这将从理论上为粗糙集学习机器泛化性能的提高提供保证。

4　结构风险最小化算法

4.1　基于遗传多目标优化的结构风险最小化算法

根据 SRM 原则，为了最小化粗糙集方法的结构风险，需要同时最小化经验风险和置信范围，然而这两个项是相互矛盾的，因此，这是一个多目标优化问题。

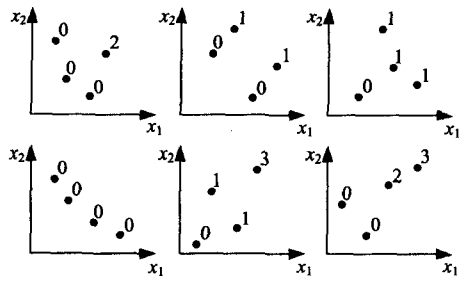

图 3　支配排序方法示例

在多目标优化问题中，需要考虑的优化目标不是单一的，并且各目标之间是相互矛盾的，因此，不存在唯一的全局最优解，而是存在多个最优解的集合，集合中的元素就全体目标而言是不可比较的，一般称为 pareto 最优解集[14]。支配排序法[15]能够被用来求取 pareto 最优解集，该方法首先对目标向量进行 pareto 评分；然后，基于目标向量的 pareto 评分，得到目标向量的支配排序；最后，根据目标向量的支配排序，得到 pareto 最优解集。图 3 给出了利用支配排序法求取 pareto 最优解集的例子，图中坐标轴 x_1 和 x_2 分别代表两个待优化目标，数字代表目标向量的支配排序值，排序值越小，所得到的解越优，排序值 0 为 pareto 最优解。

基于支配排序法，结合遗传算法，可以构建基于遗传多目标优化的结构风险最小化算法如下：首先，针对遗传算法产生的一系列候选条件属性子集，计算复杂度和经验风险；然后，针对每一复杂度，统计其对应的最小经验风险，并与历史计算结果比较，更新每一复杂度所对应的最小经验风险；最后，利用支配排序法，计算每个条件属性子集的支配排序值，将其作为适应度，从而可以利用遗传算法对这一目标优化问题进行进化计算，得到具有最小支配排序值的解的集合，即为 pareto 最优解集，进而选择其中一个合适的 pareto 最优解使粗糙集学习的结构风险最小化。

4.2　启发式结构风险最小化算法

对于多目标优化问题，除了求取 pareto 最优解集，另一种处理方法是对各目标进行加权平均，从而得到一个加权意义下的综合目标，然后利用综合目标将多目标优化问题转化为单目标优化问题。

设 $X=(x_1,x_2,\cdots,x_m)$ 为 m 个待优化目标所组成的目标向量，$W=(w_1,w_2,\cdots w_m)$ 为目标向量 X 的加权向量，则在加权意义下的综合目标能被定义为：

$$O(X,W)=\sum_{i=1}^{m} w_i x_i \tag{2}$$

对于给定的决策表，为了使粗糙集方法的结构风险最小化，可以将经验风险和复杂度转化为一个综合指标，重新定义属性重要度。对于给定的决策表 $IS=\langle U,A=C\cup D,V,f\rangle$，设 $B\subseteq C, a\in C-B$，定义在条件属性集 B 的基础上粗糙集方法对应的经验风险为 $1-\gamma_B(D)$，$\gamma_B(D)$ 为近似质量，$\gamma_B(D)=|POS_B(D)|/|U|$；所搜索函数集的复杂度为 $h(B)$，则 a 在当前条件属性集 B 的基础上相对于决策属性 D 的重要度能够被定义为：

$$SIG_{stru}(a,B,D)=(\gamma_{B\cup\{a\}}(D)-\gamma_B(D))-w\left(\frac{h(B\cup\{a\})-h(B)}{n}\right) \tag{3}$$

基于新的属性重要度评价指标，通过递归地选取具有最大重要度的属性，当剩余任意属性的重要度 $SIG_{stru}(a,B,D)\leqslant 0$ 时算法结束，能够得到一个新的启发式属性约简算法。通过选择一个合适的 w，能够使粗糙集方法的结构风险最小化。

5　实验分析

通过将机器学习领域中广泛采用的控制机器学习方法泛化性能的基本理论——结构风险最小化原则引入到粗糙集方法中，提出了粗糙集学习的结构风险控制方法，设计了基于遗传多目标优化和启发式的结构风险最小化算法。这部分将利用 12 个 UCI 数据集[16]，开展实验分析，对提出方法的有效性进行评价。

图 4 给出了 soybean 数据集上粗糙集方法获得的分类精度 Acc、近似质量 Qua、值域空间大小 N_v、属性数目 N_a、规则支尺度 $Supp_r$ 和规则长度 Len_r 随规则数目 N_r 的变化规律。可以看出，当规则数目为一个合适值时，粗糙集方法的分类精度达到最大，泛化性能最好；而当规则数目继续增加时，虽然

近似质量增加，但分类精度却下降，泛化性能降低。可见，为了提高粗糙集学习机器的泛化性能，需要对粗糙集学习机器的复杂度进行控制。

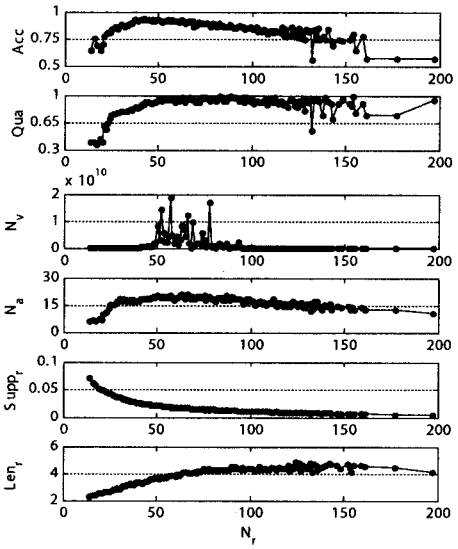

图 4　粗糙集方法各项性能指标随规则数目的变化

对于常规的粗糙集方法，通过引入属性约简的停止阈值 ϵ 也能起到控制粗糙集方法结构风险、提高粗糙集方法泛化性能的目的，由于该方法具有简单有效的特点，目前已获得了较为广泛的应用。为了验证本文提出的结构风险最小化算法的有效性，接下来将开展粗糙集学习的各种结构风险最小化算法的比较研究。

表 1 给出了常规粗糙集学习算法以及本文提出的基于遗传多目标优化和启发式结构风险最小化算法取得的分类精度。表中所有实验结果均为采用十字交叉验证法得到的平均结果，其中，hRS 代表近似质量阈值 ϵ 取 0.95 时的常规粗糙集方法，hSRM_R 代表 w 取 0.5 时的启发式结构风险最小化粗糙集方法，SRM_R 代表基于遗传多目标优化的结构风险最小化粗糙集方法。从这些实验结果，可以看出：

1）SRM_R 和 hSRM_R 获得的分类精度优于常规粗糙集方法 hRS，这说明结构风险最小化能够提高粗糙集学习机器的泛化性能。

2）hSRM_R 获得的分类精度接近于 SRM_R，这说明启发式结构风险最小化算法也能够获得较为满意的性能，为了提高算法的计算效率，可以采用启发式结构风险最小化算法来提高粗糙集学习机器的泛化性能。

表 1　结构风险最小化粗糙集方法获得的分类精度

数据集	hRS	hSRM_R	SRM_R
hepatitis	0.8650	0.8775	0.9163
Iono	0.8888	0.9201	0.9429
Horse	0.9700	0.9700	0.9725
Votes	0.9609	0.9679	0.9680
credit	0.8232	0.8246	0.8341
Zoo	0.9509	0.9418	0.9518
lymphography	0.7852	0.8043	0.8252
Wine	0.9333	0.9549	0.9386
Flags	0.5787	0.6389	0.6555
Autos	0.7686	0.7645	0.7917
images	0.8333	0.8714	0.8714
soybean	0.8945	0.9253	0.9282
平均值	0.8544	0.8718	0.8830

结束语　本文基于统计学习理论，分析了影响粗糙集学习机器泛化性能的因素，提出经验风险和复杂度是影响粗糙集学习机器泛化性能的两个根本因素，为了提高粗糙集学习机器的泛化性能，需要对粗糙集学习机器的复杂度进行控制。

结构风险最小化原则是机器学习领域中广泛用于控制学习方法复杂度的学习原则，为此，本文进一步将结构风险最小化原则引入到粗糙集方法中，提出了粗糙集学习的结构风险最小化方法，并且分别设计了基于遗传多目标优化和启发式的结构风险最小化算法。通过开展 12 个 UCI 数据集上的实验，发现提出的方法能明显提高粗糙集学习机器的泛化性能，验证了所提方法的有效性。

参 考 文 献

[1] Pawlak Z. Rough Sets [J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356

[2] Zadeh L A. Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic[J]. Fuzzy Set and System, 1997, 90(2): 111-127

[3] 张铃，张钹. 模糊商空间理论(模糊粒度计算方法)[J]. 软件学报，2003，14(4)：770-776

[4] Han S Q, Wang J. Reduct and Attribute Order[J]. Journal of Computer Science and Technology, 2004, 19(4): 429-449

[5] Leung Y, Li D Y. Maximal Consistent Block Technique for Rule Acquisition in Incomplete Information Systems[J]. Information Sciences, 2003, 153: 85-106

[6] Zheng Z, Wang G Y. RRIA: A Rough Set and Rule Tree Based Incremental Knowledge Acquisition Algorithm[J]. Fundamenta Informaticae, 2004, 59(2/3): 299-313

[7] Wu C D, Yue Y, Li M X. The Rough Set Theory and Applications[J]. Engineering Computations, 2004, 21(5/6): 488-511

[8] Parzen E. On Estimation of Probability Density Function and Model[J]. Annals of Mathematical Statistics, 1962, 33: 1065-1076

[9] Rissanen J. Modeling by Shortest Data Description[J]. Automatica, 1978, 14: 465-471

[10] Vapnik V N. Principles of Risk Minimization for Learning Theory[C]// J. E. Moody, et al., eds. Advances in Neural Information Processing Systems. Morgan Kaufmann, San Mateo, CA, 1992, 4: 831-838

[11] Bramer M. Using J-pruning to Reduce Overfitting in Classification Trees[J]. Knowledge-based Systems, 2002, 15(5/6): 301-308

[12] Schittenkopf C, Deco G, Brauer W. Two Strategies to Avoid Overfitting in Feedforward Networks[J]. Neural Networks, 1997, 10(3): 505-516

[13] Vapnik V N, Chervonenkis A J. Theory of Pattern Recognition [M]. Nauka, Moscow, 1974: 1-353

[14] Xie T, Chen H W. Evolutionary Algorithms for MultiObjective Optimization and Decision-Making Problems[J]. Engineering Science, 2002, 4(2): 59-67

[15] Foncea C M, Fleming P J. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization[C]// S. Forrest, eds. Proceedings of the 5th International Conference on Genetic Algorithms. San Mateo, California, 1993: 416-423

[16] Blake C, Keogh E, Merz C J. UCI Repository of Machine Learning Databases[D]. Dept. of Information and Computer Science, Univ. of California, Irvine, 1998. http:// www. ics. uci. edu/ ~ mlearn/MLRepository. html