

DOI: 10.11992/tis.201610019

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20170702.0418.002.html>

# 可拓支持向量分类机

陈晓华<sup>1</sup>, 刘大莲<sup>2</sup>, 田英杰<sup>3</sup>, 李兴森<sup>4</sup>

(1. 北京联合大学 教务处, 北京 100101; 2. 北京联合大学 基础部, 北京 100101; 3. 中国科学院 虚拟经济与数据科学研究中心, 北京 100190; 4. 浙江大学 宁波理工学院 管理学院, 浙江 宁波 315100)

**摘要:** 针对分类问题, 基于可拓学的思想, 提出了可拓支持向量分类机算法。与标准的支持向量分类机不同, 可拓支持向量机在进行分类预测的同时, 更侧重于找到那些通过变化特征值而转换类别的样本。文中给出了可拓变量和可拓分类问题的定义, 并构建了求解可拓分类问题的两种可拓支持向量机算法。把可拓学与 SVM 结合是一种新的方向, 文中所提出的算法还有待进一步的理论分析, 将在未来的工作里, 继续探索如何在可拓学的基础上, 构建更加完善的可拓 SVM 方法。

**关键词:** 数据挖掘; 可拓学; 分类; 支持向量机; 最优化; 最优化核函数; 先验知识; 统计学习理论

**中图分类号:** TP181 **文献标志码:** A **文章编号:** 1673-4785(2018)01-0147-05

中文引用格式: 陈晓华, 刘大莲, 田英杰, 等. 可拓支持向量分类机[J]. 智能系统学报, 2018, 13(1): 147-151.

英文引用格式: CHEN Xiaohua, LIU Dalian, TIAN Yingjie, et al. Extension support vector classification machine[J]. CAAI transactions on intelligent systems, 2018, 13(1): 147-151.

## Extension support vector classification machine

CHEN Xiaohua<sup>1</sup>, LIU Dalian<sup>2</sup>, TIAN Yingjie<sup>3</sup>, LI Xingsen<sup>4</sup>

(1. Dean's office, Beijing Union University, Beijing 100101, China; 2. Department of Basic Course Teaching, Beijing Union University, Beijing 100101, China; 3. Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China; 4. School of Management, Ningbo Institute of Technology, Zhejiang University, Zhejiang 315100, China)

**Abstract:** We propose an extension support vector machine (ESVM) to address the classification problem. Unlike the standard support vector machine, ESVM considers samples that can be converted into different labels by changing some feature values. We define the extension variables and extension classification problems and construct the corresponding optimization problem using a heuristic algorithm. In the future, we will improve the proposed method to incorporate the extension theory.

**Keywords:** data mining; extension; classification; support vector machine; optimization; kernel function; prior knowledge; statistical learning theory

随着社会的进步, 科学技术的飞速发展, 当今各种社会生产活动中每时每刻都在产生大量的数据信息, 而随着计算机技术的不断进步, 互联网的出现使得全球数据共享成为现实, 从而全面进入了大数据时代。因此如何将海量数据进行充分有效地挖掘和利用从而指导社会生产是当前数据挖掘领域的核心任务。数据挖掘是 20 世纪 80 年代后期出现的

一类数学、计算机和管理等多学科相结合的交叉科学。其旨在从海量数据中识别、提取有效的、新颖的甚至潜在有用的以及最终可被识别、理解的知识, 从而作为管理决策者进行决策的科学决策依据。数据挖掘作为一个热门研究领域, 涉及的新方法有神经网络、决策树、随机森林、支持向量机、深度学习等。

可拓数据挖掘<sup>[1-6]</sup>是将可拓学的理论和方法<sup>[7-8]</sup>与挖掘数据的方法、技术相结合的一门新技术。它在发现事物的类别转化以及识别潜在的变换知识等

收稿日期: 2016-10-15. 网络出版日期: 2017-07-02.

基金项目: 国家自然科学基金项目 (61472390, 11271361, 71331005);  
北京市自然科学基金项目 (1162005).

通信作者: 刘大莲. E-mail: [ldluck@sina.com](mailto:ldluck@sina.com).

方面有着广阔的应用前景。

本文吸取了可拓学寻找变换知识的思想,考虑事物的某种性质的变化所带来的问题转化,结合当前已经非常成熟的支持向量机模型,从而提出了可拓支持向量机这一新的可拓数据挖掘新方法。

## 1 支持向量机

支持向量机 (support vector machine, SVM)<sup>[9-18]</sup> 是一种通用的、有效的机器学习算法,主要建立在统计理论 (statistical learning theory, SLT)、最优化理论及核理论的三大理论基础之上。以分类问题为例, SVM 首先把输入映射到某一个高维的特征空间中,利用统计学习理论中的结构风险最小化原则把分类问题转化为一个凸二次规划问题来解决;然后利用最优化理论中的对偶理论,得到该凸规划的对偶问题,使得引入核函数成为可能;最后再利用核理论来处理非线性分类的情况。与数据挖掘中的其他方法相比, SVM 理论相对完备,较好地解决了其他方法中经常出现的一些棘手问题。随着科学技术的进步和完善,支持向量机的研究也更为深入广泛,许多新的理论、算法层出不穷,应用领域更为广阔。

标准的两类分类问题为: 给定训练集

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \subseteq \mathbb{R}^n \quad (1)$$

式中:  $x_i \in \mathbb{R}^n, y_i \in \mathbb{Y} = \{1, -1\}, i = 1, 2, \dots, l$ 。在此基础上, 试寻找定义在  $\mathbb{R}^n$  空间上的一个实值函数  $g(x)$ , 从而得到决策函数:

$$f(x) = \text{sgn}(g(x)) \quad (2)$$

以便推断任意  $x$  对应的输出  $y$ 。若  $g(x)$  是线性函数  $g(x) = (\mathbf{w} \cdot x) + b$ , 则式 (2) 对应着用超平面  $(\mathbf{w} \cdot x) + b = 0$  将空间  $\mathbb{R}^n$  分成两部分, 称为线性分类机。

基于结构风险最小化原则, SVM 构造出的线性分类机的原始最优化问题为

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (3)$$

$$\text{s.t. } y_i(\mathbf{w} \cdot x_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, l \quad (4)$$

$$\xi_i \geq 0, i = 1, 2, \dots, l \quad (5)$$

式中  $\xi = [\xi_1 \dots \xi_l]^T, C \geq 0$  是一个惩罚参数。通过求解上述问题的对偶问题 (见算法 1 中的优化问题), SVM 建立了处理二分类的线性算法 (算法 1)。

### 算法 1 SVM

输入 输入训练集 (见 (1) 式);

输出 分划超平面以及决策函数。

1) 输入训练集 (见式 (1)), 并选择合适的惩罚参数  $C \geq 0$ 。

2) 构造并求解凸二次规划问题:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j - \sum_{j=1}^l \alpha_j$$

$$\text{s.t. } \sum_{j=1}^l y_j \alpha_j = 0$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, l$$

得到解  $\alpha^* = [\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*]^T$ 。

3) 计算最优超平面参数  $(\mathbf{w}^*, b^*)$ :

$$\mathbf{w}^* = \sum_{i=1}^l \alpha_i^* y_i x_i;$$

$$b^* = y_j - (\mathbf{w}^* \cdot x_j), j \in \{j | 0 \leq \alpha_j^* \leq C\}$$

4) 构造分划超平面  $(\mathbf{w}^* \cdot x) + b = 0$ , 从而求出决策函数  $f(x) = \text{sgn}((\mathbf{w}^* \cdot x) + b)$ 。

算法 1 为针对线性分类问题的 SVM 算法, 事实上在很多实际问题中, 分类问题并不能被线性分划。这时需引入从空间  $\mathbb{R}^n$  到 Hilbert 空间的变换  $x = \Phi(x)$ , 将训练集变换到特征空间并在该空间执行算法 1, 相应地,  $(x_i, y_i)$  转化为  $(\Phi(x_i), \Phi(x_j))$ , 而  $(\Phi(x_i), \Phi(x_j))$  的计算可以通过引进核函数  $K(x, x') = (\Phi(x), \Phi(x'))$  代替计算。通常, 常用的核函数有 RBF 核函数、多项式核函数、B 样条核函数等。

## 2 可拓支持向量机

在实际问题中, 我们不仅关注对未知样本的类别预测, 更关心已有类别或被预测了类别的样本能否转化到相反类别, 在什么条件下会转化, 所付出的代价有多大。这些在可拓学里被称为变换的知识。挖掘这些变换的知识的一个途径, 就是找到那些可以变换的特征或变量。以疾病的辅助诊断为例, 患者患有疾病的相关特征 (变量) 有很多, 比如性别、年龄、身高、体重、血压 (低压)、胆固醇、血常规等, 这些变量中, 对某个患者来说, 其性别就是不可变换的变量, 而血压则可以通过吃药等治疗手段变高或变低。那么在解决疾病辅助诊断的分类问题时, 我们一方面希望决策函数能准确地预测未知样本的类别, 另一方面希望这个决策函数能找到那些可以互相转换类别的样本, 这样就可以找到更多的变换的知识。首先给出以下定义。

**可拓变量** 若向量  $\mathbf{x} = ([x]_1, \dots, [x]_n)$  中分量  $[x]_j$  的值可以变化, 则称  $[x]_j$  为可拓变量。

**可拓变量集合** 由全体可拓变量构成的集合则称为可拓变量集合, 记为  $\mathbf{E}$ 。

**可拓变量区间** 可拓变量  $[x]_j$  的值所变化的区间  $[a_j, b_j]$  称为可拓变量区间。

**可拓分类问题** 给定训练集

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \subseteq (\mathbb{R}^n \times \mathbb{Y})^l \quad (6)$$

式中:  $x_i \in \mathbb{R}^n, [x_i]_j, j \in \mathbf{E}$ , 可在某个区间内变化,  $y_i \in \mathbb{Y} = \{1, -1\}, i = 1, 2, \dots, l$ 。基于此, 在  $\mathbb{R}^n$  空间上寻找一个实

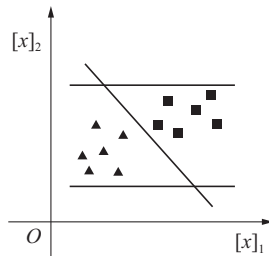
值函数 $g(x)$ , 得到决策函数:

$$f(x) = \text{sgn}(g(x)) \quad (7)$$

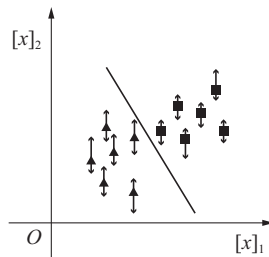
以便推断任意  $x$  对应的输出  $y$ , 并判断其是否可以变换为相反的输出  $-y$ 。

**可拓样本** 若样本点  $x$  可以通过改变某个或某些可拓变量的值而改变类别标号, 则称  $x$  为可拓样本。

下面给出可拓分类问题示意性的例子。图 1 中给出了两类点, 分别用  $\blacktriangle$  和  $\blacksquare$  表示, 分量  $[x]_2$  为可拓变量。图 1(a) 所示的是所有点的可拓变量区间均满足  $[x]_2 \in [a, b]$  的情况, 即所有点被上下两条直线  $[x]_2 = a$  和  $[x]_2 = b$  所控制; 图 1(b) 所示的是每个点的可拓变量区间不同的情况, 即满足  $[x_i]_j \in [a'_j, b'_j]$ , 在图中就是每个点分别被双箭头线段所控制。



(a) 可拓变量区间固定



(b) 可拓变量区间变化

图 1 可拓分类问题

Fig. 1 Extension classification problem

注意, 与标准的两类分类问题不同, 这里除了给出标准的训练集外, 还给出了每个点的可拓变量区间。据此可以构建算法 2: 可拓 SVM1。

#### 算法 2 可拓 SVM1

**输入** 输入训练集 (见式 (1));

**输出** 决策函数。

1) 输入训练集 (见式 (1)), 并选择合适的惩罚参数  $C > 0$ 。

2) 构造并求解凸二次规划问题:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j - \sum_{j=1}^l \alpha_j \\ \text{s.t.} & \sum_{j=1}^l y_j \alpha_j = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, l \end{aligned}$$

得到解  $\alpha^* = [\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*]^T$ 。

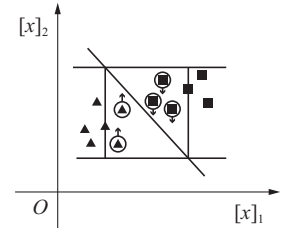
3) 计算最优超平面参数:

$$\begin{aligned} \mathbf{w}^* &= \sum_{i=1}^l \alpha_i^* y_i x_i \\ b^* &= y_j - (\mathbf{w}^* \cdot x_j), j \in \{j | 0 \leq \alpha_j^* \leq C\} \end{aligned}$$

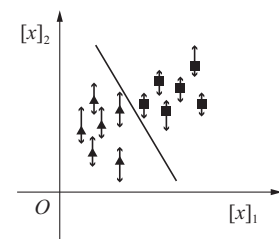
4) 构造分划超平面  $(\mathbf{w}^* \cdot x) + b = 0$ , 从而求出决策函数  $f(x) = \text{sgn}((\mathbf{w}^* \cdot x) + b)$ 。

5) 对输入  $x_k$ , 首先用决策函数  $f(x_k)$  得到其对应的预测类别  $y_k$ , 然后用其可拓变量对应的可拓区间  $\alpha_j^k$  和  $b_j^k$  分别代替  $[x_k]_j$ , 这样对  $|E|$  个可拓变量, 会得到  $2^{|E|}$  个不同的组合值, 相应的, 我们基于  $x_k$  得到了  $2^{|E|}$  个新的输入, 分别用决策函数来判断, 若有一个被判断为  $-y_k$ , 则认为该输入是可变换的。

对图 1(a) 所示的例子, 可以用算法 2 得到图 2(a) 所示的分划线, 因该问题的可拓变量区间为定值, 所以可以直接得到两条垂直的线, 在这两条垂直直线中间的点都是可拓样本 (用圆圈标注), 并用箭头表示该样本在可拓变量的变化方向。对图 1(a) 所示的例子, 用算法 2 得到图 2(b) 所示的分划线, 可以看出没有可拓样本点, 即没有点可以通过可拓变量的区间变化来达到类别标号的改变。



(a) 可拓区间固定的分划线



(b) 可拓区间变化的分划线

图 2 可拓分类 SVM

Fig. 2 Extention classification SVM

进一步, 考虑稍微复杂的情况, 若对上述的可拓分类问题, 增加了先验知识。该先验知识为: 在给定的训练集  $S$  中, 训练点  $x_i$  的可拓变量  $[x_i]_j, i = 1, 2, \dots, l, j \in E$  在可拓区间内变化时, 其  $y_i$  没有变化, 如何寻找决策函数来推断任意  $x$  对应的输出  $y$ , 并判断其是否可以变换为相反的输出  $-y$ 。

可以看出, 这时候训练点的输入就不再是一个点, 而是一个集合。如果限定该集合是一个凸集, 比如为一个球体, 则训练集变为

$$\mathbf{S} = \{(X_1, y_1), (X_2, y_2), \dots, (X_l, y_l)\} \quad (8)$$

式中输入集合  $X_i$  是一个以  $X_i$  为球心、以  $r_i$  为半径的球体,  $y_i \in \mathcal{Y} = \{1, -1\}$ ,  $i = 1, 2, \dots, l$ 。那么该分类问题就转化为鲁棒支持向量机所求解的问题<sup>[15]</sup>, 此时构造的原始最优化问题为

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (9)$$

$$\text{s.t. } y_i(\mathbf{w} \cdot (x_i + x_i u_i) + b) \geq 1 - \xi_i, \quad (10)$$

$$\forall \|u_i\| \leq 1, i = 1, 2, \dots, l$$

$$\xi_i \geq 0, i = 1, 2, \dots, l \quad (11)$$

因为约束式 (10) 含有无穷多个约束, 一般做法是把它转化成一个仅含有有限个约束的问题, 并求解相应的对偶问题<sup>[15]</sup>, 据此构造可拓 SVM2, 见算法 3。

### 算法 3 可拓 SVM2

输入 输入训练集 (见式 (8));

输出 决策函数。

1) 输入训练集 (见式 (8)), 并选择合适的惩罚参数  $C > 0$ 。

2) 构造并求解二阶锥规划:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi, u, v, t} \quad & \frac{1}{2}(u-v) + C \sum_{i=1}^l \xi_i \\ \text{s.t. } \quad & y_i(\mathbf{w} \cdot (x_i + x_i u_i) + b) \geq 1 - \xi_i, i = 1, 2, \dots, l \\ & \xi_i \geq 0, i = 1, 2, \dots, l \\ & u + v = 1 \\ & [u \ v \ t]^T \in \mathbf{L}^3 \\ & [t \ \mathbf{w}]^T \in \mathbf{L}^{n+1} \end{aligned}$$

3) 计算  $\bar{b}$ , 构造决策函数:

$$f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i (x_i \cdot x) + \sum_{i=1}^l y_i \mathbf{B}_i^T (\bar{\beta}_i^* - \bar{\beta}_i) + \bar{b}\right) \quad (12)$$

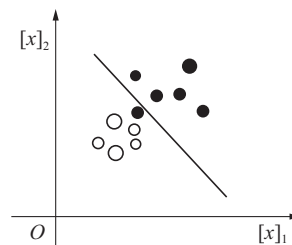
式中  $(\mathbf{B}_i^q)_q = (x \cdot (\Delta_i)_{\cdot, q})$ ,  $q = 1, 2, \dots, n$ 。

4) 对输入  $x_k$ , 首先用决策函数  $f(x_k)$  得到其对应的预测类别  $y_k$ , 然后在其可拓变量对应的可拓区间内随机取值  $m$  次, 这样对  $|\mathbf{E}|$  个可拓变量, 会得到  $m^{|\mathbf{E}|}$  个不同的组合值, 相应的, 基于  $x_k$  得到了  $m^{|\mathbf{E}|}$  个新的输入, 分别用决策函数来判断, 若有一个被判断为  $-y_k$ , 则认为该输入是可变换的。

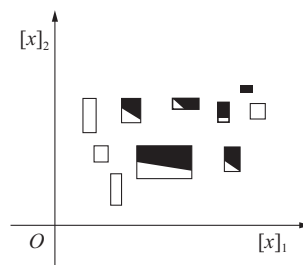
如图 3(a) 所示, 设  $[x]_1$ 、 $[x]_2$  均为可拓变量, 每个点的可拓区间为半径不同的球体, 空心圆是正类, 实心圆是负类。根据算法 3 可得到图中的分划线, 则可知虚线所示的圆的圆心点为正类, 但当该点在圆内变化时, 它的类别会改变。

如果先验知识为: 在给定的训练集  $\mathbf{S}$  中, 训练点  $x_i$  的可拓变量  $[x_i]_j$ ,  $i = 1, 2, \dots, l$ ,  $j \in \mathbf{E}$  在可拓区间内变化时, 其  $y_i$  产生了变化, 如图 3(b) 所示。 $[x]_1$ 、 $[x]_2$  均为可拓变量, 每个点的可拓区间为不同大小的矩形, 白色为正类, 黑色为负类。而黑白相间

的矩形则为点在这些区间内变化时, 类别发生了变化。这时问题就变得复杂, 需要引入可拓集的概念和关联函数来处理, 这是进一步的研究方向。并且还可以将其拓展到应用广泛的多示例、多标签学习等方面<sup>[19-22]</sup>。



(a) 可拓区间为球体



(b) 可拓区间为矩形

图 3 可拓分类 SVM

Fig. 3 Extension classification SVM

## 3 结束语

本文提出了一种新的分类方法——可拓支持向量机, 可拓支持向量机在进行分类预测时较标准的支持向量分类机有所不同, 它更侧重于找到那些通过变化特征值而转换类别的样本。文中给出了可拓变量和可拓分类问题的定义, 并构建了求解可拓分类问题的两种算法。基于此模型, 我们可以在理论和方法上做一系列研究。理论上, 包括对其相应的统计学习理论基础研究, 可拓决策函数集的 VC 维估计, 推广能力的上界估计, 可拓结构风险的实现原理等; 方法上, 包括如何在标准的支持向量机模型中引入关联函数, 构造可拓的核函数, 对其相应的大规模最优化问题的快速求解算法, 模型参数选择问题等。

## 参考文献:

- [1] 蔡文, 杨春燕, 陈文伟, 等. 可拓集与可拓数据挖掘[M]. 北京: 科学出版社, 2008.
- CAI Wen, YANG Chunyan, CHEN Wenwei, et al. Extension set and extension data mining[M]. Beijing: Science Press, 2008.

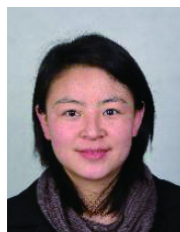


- [2] 杨春燕, 蔡文. 可拓工程[M]. 北京: 科学出版社, 2007.  
YANG Chunyan, CAI Wen. Extension engineering[M]. Beijing: Science Press, 2007.
- [3] CAI Wen. Extension theory and its application[J]. Chinese science bulletin, 1999, 44(17): 1538–1548.
- [4] 李立希, 李铎汶, 杨春燕. 可拓学在数据挖掘中的应用初探[J]. 中国工程科学, 2004, 6(7): 53–59.  
LI Lixi, LI Huawen, YANG Chunyan. Study on the application of extenics in data mining[J]. Engineering science, 2004, 6(7): 53–59.
- [5] 陈文伟, 黄金才. 从数据挖掘到可拓数据挖掘[J]. 智能技术, 2006, 1(2): 50–52.  
CHEN Wenwei, HUANG Jincai. From data mining to extension data mining[J]. Intelligent technology, 2006, 1(2): 50–52.
- [6] 杨春燕, 蔡文. 可拓数据挖掘研究进展[J]. 数学的实践与认识, 2009, 39(4): 136–141.  
YANG Chunyan, CAI Wen. Recent progress in extension data mining[J]. Mathematics in practice and theory, 2009, 39(4): 136–141.
- [7] 蔡文, 杨春燕, 何斌. 可拓逻辑初步[M]. 北京: 科学出版社, 2003.  
CAI Wen, YANG Chunyan, HE Bin. Extension logic[M]. Beijing: Science Press, 2003.
- [8] 蔡文, 杨春燕, 林伟初. 可拓工程方法[M]. 北京: 科学出版社, 1999.  
CAI Wen, YANG Chunyan, LIN Weichu. Extension engineering methods[M]. Beijing: Science Press, 1999.
- [9] CORTES C, VAPNIK V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273–297.
- [10] CRISTIANINI N, SHAW-ETAYLOR J. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge: Cambridge University Press, 2000.
- [11] VAPNIK V N. Statistical learning theory[M]. New York: John Wiley and Sons, 1998.
- [12] NOBLE W S. Support vector machine applications in computational biology[M]//SCHÖELKOPF B, TSUDA K, VERT J P. Kernel Methods in Computational Biology. Cambridge: MIT Press, 2004.
- [13] ADANKON M M, CHERIET M. Model selection for the LS-SVM. application to handwriting recognition[J]. Pattern recognition, 2009, 42(12): 3264–3270.
- [14] TIAN Yingjie, SHI Yong, LIU Xiaohui. Recent advances on support vector machines research[J]. Technological and economic development of economy, 2012, 18(1): 5–33.
- [15] DENG Naiyang, TIAN Yingjie, ZHANG Chunhua. Support vector machines: optimization based theory, algorithms, and extensions[M]. Boca Raton: CRC Press, 2013.
- [16] HUANG Kaizhu, YANG Haiqin, KING I, et al. Learning large margin classifiers locally and globally[C]//Proceeding of the 21st International Conference on Machine Learning. Banff, Alberta, Canada, 2004.
- [17] SHAW-ETAYLOR J, CRISTIANINI N. Kernel methods for pattern analysis[M]. Cambridge: Cambridge University Press, 2004.
- [18] BORGWARDT K M. Kernel methods in bioinformatics [M]//LU H H S, SCHÖELKOPF B, ZHAO Hongyu. Handbook of Statistical Bioinformatics. Berlin, Heidelberg: Springer, 2011: 317–334.
- [19] VAPNIK V, IZMAILOV R. Learning using privileged information: Similarity control and knowledge transfer[J]. Journal of machine learning research, 2015, 16: 2023–2049.
- [20] SUN S. A survey of multi-view machine learning[J]. Neural computing and applications, 23(7): 2031–2038.
- [21] CHEPLYTGINA V, TAX, D M, LOOG M. Multiple instance learning with bag dissimilarities[J]. Pattern recognition, 2015, 48(1): 264–275.
- [22] CARBONNEAU M A, GRANGER E, RAYMOND A J, et al. Robust multiple-instance learning ensembles using random subspace instance selection[J]. Pattern recognition, 2016, 58: 83–99.

#### 作者简介:



陈晓华, 女, 1975 年生, 工程师, 主要研究方向为电力系统及其自动化。



刘大莲, 女, 1978 年生, 副教授, 博士研究生, 主要研究方向为数据挖掘。



田英杰, 男, 1973 年, 研究员, 博士生导师, 主要研究方向为最优化与机器学习。