

一种基于核距离的核函数度量方法

王裴岩^{1,2} 蔡东风¹

(沈阳航空航天大学知识工程研究中心 沈阳 110136)¹

(南京航空航天大学计算机科学与技术学院 南京 210016)²

摘 要 核方法的效果依赖于所使用的核,因此核的选择和其参数的确定是至关重要的。从特定的数据中学习核需要核度量方法评价核的质量。核排列度量核与学习任务的一致性,因为它具有高效性和有效性,是目前应用最为广泛的核度量方法。然而,有研究表明,核排列仅是最优核函数的充分非必要条件。其主要原因是核排列在特征空间中不具有线性变换不变性。提出了一种新的核度量方法用于核选择,称其为核距离排列。该方法能够克服核排列的局限性,并且同样具有高效性和简单的形式。对比实验表明,该方法能够有效地对核进行度量。

关键词 核方法,核度量,核距离

中图法分类号 TP391 **文献标识码** A

Distance-based Kernel Evaluation Measure

WANG Pei-yan^{1,2} CAI Dong-feng¹

(Knowledge Engineering Research Center, Shenyang Aerospace University, Shenyang 110136, China)¹

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)²

Abstract The success of kernel methods depends on the kernel, thus a choice of a kernel and proper setting of its parameters are of crucial importance. Learning a kernel from the data requires evaluation measures to assess the quality of the kernel. Recently, kernel target alignment (KTA), which measures the degree of agreement between a kernel and a learning task, has been widely used for kernel selection because of its effectiveness and efficiency. However, it is reported that KTA is only a sufficient condition to select a good kernel, but not a necessary condition. The reason is that KTA is not invariant under data translation in the feature space. This paper proposed a new measure for kernel selection named kernel distance target alignment (KDTA). The measure not only overcomes the limitations of KTA but also possesses other properties like simplicity and efficiency. Comparative experiments indicate that the new measure is a good indication of the superiority of a kernel.

Keywords Kernel method, Kernel evaluation measure, Kernel distance

1 引言

核方法(Kernel Method)是目前机器学习领域内的研究焦点之一,已被广泛应用于分类、回归等机器学习任务^[1]。核方法通过一个非线性变换将输入空间映射到高维特征空间,然后在特征空间中使用线性学习算法。非线性变换通过核函数隐式定义,核函数能高效地计算样本在特征空间中的内积,从而不必显式定义和计算从输入空间到特征空间的非线性变换。核函数决定数据在特征空间中的分布,是影响核方法性能的关键因素,因此核函数的选择是核方法研究的核心内容。

核度量(Kernel Evaluation Measure)的研究主要服务于核函数的选择,是核函数选择研究中的一个热点。因此,多种核函数度量标准被相继提出。对于分类问题,结构风险(Structural Risk)^[2]、负对数后验(Negative Log-posterior)^[3]和超核(Hyper kernels)^[4]是较为常用的度量标准。但此类标

准针对特定的核函数,不能给出量化的度量值,并且需要分类器的全部训练过程。交叉验证错误率(Cross-Validation Error)和留一法错误率(Leave-one-out Error)也经常被用作核函数选择的标准,此类方法也需要训练分类器,核函数选择计算开销较大。

目前,一些高效的不依赖于具体机器学习算法的核度量标准被相继提出,如:核排列(Kernel Target Alignment, KTA)^[5]、核极化(Kernel Polarization)^[6]、局部核极化(Local Kernel Polarization)^[7]、特征空间核度量标准(Feature Space Based Kernel Matrix Evaluation Measures)^[8]、 γ ^[9]等。此类度量方法不依赖具体分类器的训练过程,不直接考虑分类器的泛化性能,仅使用训练样本信息,着力衡量训练样本在特征空间中的可分离情况。若不考虑核函数的计算代价,此类方法的计算代价为 $O(n^2)$ (n 为样本数量)。KTA是最早被提出的核度量标准,也是应用最为广泛的度量标准,大量基于KTA

到稿日期:2013-05-20 返修日期:2013-07-23 本文受国家自然科学基金(61073123),沈航校青年教师自选课题(201106Y)资助。

王裴岩(1983—),男,博士生,主要研究方向为机器学习、信息抽取, E-mail: wangpy_kerc@163.com; 蔡东风(1958—),男,博士,教授,主要研究方向为机器学习、自然语言处理、信息检索。

的工作相继被提出^[12-15]。然而,文献[8]证明了 KTA 是最优核函数的充分非必要条件,即好的核函数也可能获得较低的 KTA 值,其原因是 KTA 在特征空间中不具有线性变换不变性,即:在特征空间中进行变换 $\varphi(x) \rightarrow \varphi(x) + \delta$,那么 KTA 值将会是 δ 的函数。

本文提出一种新的基于核距离^[10]的核函数度量标准,称其为核距离排列(Kernel Distance Target Alignment, KD-TA)。该方法不依赖于分类器的训练过程,计算高效,形式简单,与 KTA 相比,对于特征空间中的线性变换具有不变性。本文第 2 节介绍 KTA;第 3 节提出 KD-TA 并于第 4 节对其进行分析;第 5 节采用核函数选择实验验证所提方法的有效性;最后是结论和对未来工作的展望。

2 核函数与核排列

2.1 核函数形式描述

本文涉及的核函数形式描述如下:

$$k(\mathbf{x}_1, \mathbf{x}_2) = \langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle, \phi: X \rightarrow H \quad (1)$$

X 为输入空间, $X \subset \mathbb{R}^n$, H 为特征空间。给定样本集 $D = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, 其中 $\mathbf{x}_i \in X, y_i = \{+1, -1\}$, 并且有 n_+ 个样本属于“+1”类,有 n_- 个样本属于“-1”类。

2.2 核排列

KTA 由 Cristianini 等人^[5]提出,对于二分类问题,KTA 计算核矩阵与理想目标矩阵的对齐程度。定义核函数 k 的核矩阵为:

$$[K]_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j) \quad (2)$$

定义理想目标矩阵为:

$$[Y]_{i,j} = y_i \cdot y_j = \begin{cases} +1, & y_i = y_j \\ -1, & y_i \neq y_j \end{cases} \quad (3)$$

定义 KTA 为:

$$A(K, Y) = \frac{\langle K, Y \rangle_F}{\sqrt{\langle K, K \rangle_F} \sqrt{\langle Y, Y \rangle_F}} \quad (4)$$

其中 $\langle \cdot, \cdot \rangle_F$ 为 Frobenius 内积。因此,KTA 为归一化后的核矩阵与理想目标矩阵间的 Frobenius 内积。KTA 取值区间为 $[-1, +1]$,值越高证明核矩阵与理想目标矩阵对齐程度越高,核函数越好。KTA 具有如下性质^[11]:

1. KTA 仅使用训练样本集,计算时间复杂度为 $O(n^2)$;

2. KTA 具有简单的形式,可被用作核参数优化过程的目标函数;

3. 在训练集上估计的 KTA 是高度集中于它的期望值,这意味着如果训练集上估计的 KTA 值较大,那么它在测试集上的值也较大;

4. 若 KTA 的期望值较大,那么在特征空间中的线性分类器的泛化能力较强,则分类器的泛化误差率上界越小。

由于具有上述性质,KTA 被广泛应用于核参数优化^[12,13]、多核学习^[2]、特征选择^[14,15]等方面。然而,Nguyen 和 Ho 证明 KTA 仅是最优核函数的充分非必要条件,其原因是 KTA 依赖于样本在特征空间中的绝对位置,不具有线性变换不变性^[8]。Nguyen 和 Ho 采用实例证明了 KTA 可能错误地估计最优核函数与最差核函数,并且对于一般情况,其值受限于特定上界。

3 核距离排列

本节将介绍提出的核度量方法,使用核距离矩阵代替

KTA 中的核矩阵,计算核距离矩阵与理想目标矩阵的对齐程度。本文将该方法称为核距离排列(Kernel Distance Target Alignment, KD-TA)。其中核距离的定义为:

$$\| \phi(\mathbf{x}_1) - \phi(\mathbf{x}_2) \|^2 = k(\mathbf{x}_1, \mathbf{x}_1) + k(\mathbf{x}_2, \mathbf{x}_2) - 2k(\mathbf{x}_1, \mathbf{x}_2) \quad (5)$$

对应于核函数的核距离矩阵为:

$$[D]_{i,j} = \| \phi(\mathbf{x}_i) - \phi(\mathbf{x}_j) \|^2 = k(\mathbf{x}_i, \mathbf{x}_i) + k(\mathbf{x}_j, \mathbf{x}_j) - 2k(\mathbf{x}_i, \mathbf{x}_j) \quad (6)$$

核函数可被视为样本在特征空间中相似性或接近程度的度量,其值越高,则两样本越相似或越接近。与之相反,距离经常被视为非相似性度量,即距离越大,则两样本在特征空间中的相似性越低^[17]。因此,对于 KD-TA,其理想目标矩阵定义为:

$$[Y]_{i,j} = \begin{cases} 0, & y_i = y_j \\ 1, & y_i \neq y_j \end{cases} \quad (7)$$

核距离矩阵 D 的核距离排列定义为:

$$A(D, Y) = \frac{\langle D, Y \rangle_F}{\sqrt{\langle D, D \rangle_F} \sqrt{\langle Y, Y \rangle_F}} \quad (8)$$

KD-TA 的取值范围为 $[-1, +1]$,其值越高,则两矩阵的对齐程度越高,那么样本在核函数 k 所构建的特征空间中可分性越高。计算 KD-TA 的时间复杂度为 $O(3n^2)$,需要计算 $k(\mathbf{x}_i, \mathbf{x}_j)$, $k(\mathbf{x}_i, \mathbf{x}_i)$ 和 $k(\mathbf{x}_j, \mathbf{x}_j)$ 3 个核函数。KD-TA 比 KTA 计算复杂度要高,但仍然比结构风险和交叉验证的方法效率高。

4 对核距离排列的分析

KD-TA 的提出考虑了两点因素,首先,核距离比核更适于度量样本在特征空间中的相似程度;其次,核距离在特征空间中具有线性变换不变性。本节将详细分析上述两点因素。

4.1 相似性度量

核经常被视为样本在特征空间中的相似程度或接近程度的度量^[6,7,16]。相似性度量被定义为一种样本集的笛卡尔积到实数的映射函数($S: X \times X \rightarrow \mathbb{R}$),并且满足非负性、对称性和最大性^[17]。

非负性:

$$\forall \mathbf{x}_i, \mathbf{x}_j \in X, S(\mathbf{x}_i, \mathbf{x}_j) \geq 0;$$

对称性:

$$\forall \mathbf{x}_i, \mathbf{x}_j \in X, S(\mathbf{x}_i, \mathbf{x}_j) = S(\mathbf{x}_j, \mathbf{x}_i);$$

最大性:

$$\forall \mathbf{x}_i, \mathbf{x}_j \in X, S(\mathbf{x}_i, \mathbf{x}_i) \geq S(\mathbf{x}_i, \mathbf{x}_j)。$$

然而,并不是所有的核都满足上述 3 点性质。如: sigmoid(tanh)核不满足非负性,线性核和多项式核不满足最大性。因此,将核视为样本在特征空间中的相似性度量并不总是合理的。与之相反,核距离满足非负性、对称性以及最大性相对的最小性:

$$\forall \mathbf{x}_i, \mathbf{x}_j \in X, S(\mathbf{x}_i, \mathbf{x}_i) \leq S(\mathbf{x}_i, \mathbf{x}_j);$$

并且满足三角不等式:

$$\forall \mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k \in X, S(\mathbf{x}_i, \mathbf{x}_j) \leq S(\mathbf{x}_i, \mathbf{x}_k) + S(\mathbf{x}_j, \mathbf{x}_k)。$$

由此可见,核距离比核更适于反映样本在特征空间的相对关系。然而,需要特别指出的是当核是非正定核时(如: sigmoid 核),式(5)为负值,并且核距离不再满足上述性质,因此本文提出的方法不适于度量非正定核。

4.2 不变性

核是样本在特征空间中的内积,显然,距离具有线性变换不变性,而内积不具有该性质。本文采用文献[8]中使用的实例来证明 KDTA 的线性变换不变性。

最好情况:核将所有属于类“+1”的样本映射到特征空间中的点 φ_+ ,将所有属于类“-1”的样本映射到特征空间中的点 φ_- 。假设 $\|\varphi_+ - \varphi_+\|^2 = \|\varphi_- - \varphi_-\|^2 = 0$, $\|\varphi_+ - \varphi_-\|^2 = \alpha$ 且 $\alpha > 0$; $\langle \varphi_+, \varphi_+ \rangle = \langle \varphi_-, \varphi_- \rangle = 1$, $\langle \varphi_+, \varphi_- \rangle = \beta$ 且 $-1 \leq \beta < 1$ 。对于任何 α 和 β 值,度量方法都应将该核度量为最优核。KDTA 的值为:

$$A(D, Y) = \frac{2 \cdot n_+ \cdot n_- \cdot \alpha}{\sqrt{2 \cdot n_+ \cdot n_- \cdot \alpha^2} \cdot \sqrt{2 \cdot n_+ \cdot n_-}} = 1 \quad (9)$$

由此可见,KDTA 的值为 1(最高值),并且与 α 无关。

在此种情况下 KTA 的值为:

$$A(K, Y) = \frac{n_+^2 + n_-^2 - 2n_+ n_- \beta}{\sqrt{n_+^2 + n_-^2 + 2n_+ n_- \beta^2} \cdot n} \quad (10)$$

式(10)的值完全依赖于 β ,当 β 为 -1 时, $A(K, Y)$ 为 1,当 β 的值趋近于 1 时, $A(K, Y)$ 趋近于 $(n_+ - n_-)^2 / n^2$ 。在此种情况下, $A(K, Y)$ 的取值区间为 $((n_+ - n_-)^2 / n^2, 1]$ 。

最坏情况:核将每类一半的样本映射到特征空间中的点 φ_+ ,将另一半的点映射到特征空间的点 φ_- 。假设 $\|\varphi_+ - \varphi_+\|^2 = \|\varphi_- - \varphi_-\|^2 = 0$, $\|\varphi_+ - \varphi_-\|^2 = \alpha$ 且 $\alpha > 0$; $\langle \varphi_+, \varphi_+ \rangle = \langle \varphi_-, \varphi_- \rangle = 1$, $\langle \varphi_+, \varphi_- \rangle = \beta$ 且 $-1 \leq \beta < 1$ 。对于任何 α 和 β 值,度量方法对该核的度量值应较低。若在特征空间中使用线性分类器,那么分类准确率应为 1/2(错误地分类了一半的样本)。KDTA 的值为:

$$A(D, Y) = \frac{n_+ \cdot n_- \cdot \alpha}{\sqrt{\frac{n^2 \cdot \alpha^2}{2}} \cdot \sqrt{2 \cdot n_+ \cdot n_-}} = \frac{\sqrt{n_+ \cdot n_-}}{n} \quad (11)$$

若 $n_+ = n_- = n/2$, $A(D, Y) = 1/2$ 。KDTA 不与 α 相关,但依赖于样本集的分布情况,并且最大值与理论值相符 1/2。

此种情况下 KTA 的值为:

$$A(K, Y) = \frac{(n_+ - n_-)^2 \cdot (1 + \beta) / 2}{n^2 \cdot \sqrt{(1 + \beta^2) / 2}} \quad (12)$$

式(12)的值完全依赖于 β ,当 β 值为 -1 时 KTA 为 0,当 β 值趋近于 1 时, KTA 趋近于 $(n_+ - n_-)^2 / n^2$ (与最好情况相同),在此种情况下, $A(K, Y)$ 的取值区间为 $[0, (n_+ - n_-)^2 / n^2]$ 。

上述实例表明核排列不具有线性变换不变性,依赖于样本在特征空间中的具体位置(与 β 值相关)。核距离排列具有线性变换不变性,不依赖于样本在特征空间中的具体位置(与 α 无关),并且对于最好情况和最坏情况都可以给予准确的度量。

当 $y_i = y_j$ 时,若将理想目标矩阵的值设为 -1, KDTA 的值为(最好情况):

$$A(D, Y) = \frac{2 \cdot n_+ \cdot n_- \cdot \alpha}{\sqrt{2 \cdot n_+ \cdot n_- \cdot \alpha^2} \cdot n} = \frac{\sqrt{2 \cdot n_+ \cdot n_-}}{n} \quad (13)$$

对于最坏情况:

$$A(D, Y) = \frac{-\frac{1}{2} \cdot \alpha \cdot (n_+ - n_-)^2}{\sqrt{\frac{\alpha^2 n^2}{2}} \cdot n} = -\frac{(n_+ - n_-)^2}{\sqrt{2} \cdot n^2} \quad (14)$$

可以看出,无论是最好情况还是最坏情况, KDTA 都完全依赖于样本集的分布。对于最好情况,在任何样本分布情况下, KDTA 都无法达到最优化。对于最坏情况, KDTA 的值为负值,该值差于将所有样本映射为特征空间中的一个点的情况。上述原因是本文将理想目标矩阵值设为 0 而非 -1 的依据。

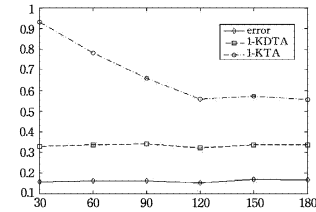
5 实验

本文使用模拟数据和 6 个来自 UCI 的数据集来验证所提出的方法。由于本文提出的方法基于 KTA 的思想,因此仅与 KTA 进行对比实验。在模拟数据集上,将各度量值与交叉验证错误率进行了比较,展示各度量方法对错误率的反映情况,从而展示 KTA 不具有线变换不变性,并且 KDTA 可以克服该限制。在 6 个 UCI 数据集上,进行核选择实验。

分类算法使用 SVM,工具采用 libsvm¹⁾,使用十折交叉验证估计错误率。为了方便展示,展示 1-KTA、1-KDTA 和错误率(error),这些值的区间为 $[0, 1]$ 。

5.1 模拟数据集

模拟数据为二维,使用线性核模拟不同核函数所产生的样本在特征空间中的分布。使用两个 Gaussian 分布来生成数据,类别“+1”的中心在 $\varphi_+ = (1, 0) \in R^2$,类别“-1”的中心在 $\varphi_- = (\cos\beta, \sin\beta) \in R^2$,标准差为 $\text{var}_+ = \text{var}_- = \|\varphi_+ - \varphi_-\| / 2$,每类样本集包含 500 个样例。基于上述条件, β 值不影响样本集的可分性,对于任意 β 值,使用线性核的分类错误率应维持恒定,那么对于不同 β 值线性核的度量值也应维持恒定。本文将 β 值分别赋为 30°、60°、90°、120°、150°和 180°。实验结果如图 1 所示。



横轴为 β 值,纵轴为分类错误率和各度量方法的度量值

图 1 不同 β 值的实验结果

从图 1 可见,对于不同的 β 值,交叉验证的错误率维持恒定,1-KDTA 也同样维持稳定,并且与交叉验证错误率的变化基本相符。然而,1-KTA 的变动较大,可见 KTA 更依赖于样本在特征空间中的绝对位置。由此, KTA 对于特征空间中的线性变换不具有不变性,而 KDTA 具有该性质。

5.2 UCI 数据集

本文使用来自 UCI 的 6 个数据集进行核函数选择实验,各数据集信息如表 1 所列。对于多分类问题使用“one-vs-others”策略转化为多个二分类问题。线性核(Lin)、多项式核(Poly)和 RBF 核(RBF)作为选择对象,各核函数的参数为 libsvm 的默认值。以交叉验证的错误率作为参照,比较各度量方法对错误率的反映情况。以 Kendall 排序相关系数衡量各度量方法核排序序列与交叉验证错误率核排序序列的一致性,结果如表 2 所列。表 3 列出了每种核度量方法和交叉验证错误率选择的最优核函数列,同时给出了每种度量方法参

¹⁾ <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

照交叉验证错误率的选择错误数。表 4 列出了在每种度量方法中最优核函数排列的位序。

表 1 各数据集信息

数据集名称	样本数	类别数	特征数
Breast Cancer Wisconsin	699	2	10
Pima Indians Diabetes	769	2	9
Ionosphere	351	2	34
Ecoli	366	8	8
Glass	214	7	9
Yeast	1484	10	9

表 2 各核函数与交叉验证的核排序序列 Kendall 相关系数

	1-KDTA	1-KTA
Breast Cancer Wisconsin	-0.3333	0.3333
Pima Indians Diabetes	0.3333	-0.3333
Ionosphere	0.3333	0.3333
Ecoli	1	-0.3333
Glass	-0.3333	-0.3333
Yeast	1	-1
AVG	0.3333	-0.2222

表 3 核函数选择实验结果

	1-KDTA	1-KTA	error
Breast Cancer Wisconsin	Poly	Lin	Lin
Pima Indians Diabetes	RBF	Lin	RBF
Ionosphere	RBF	RBF	RBF
Ecoli	RBF	RBF	Lin
Glass	RBF	Lin	Poly
Yeast	RBF	Lin	RBF
选择错误数	3	4	-

表 4 最优核函数排列位序

	1-KDTA	1-KTA
Breast Cancer Wisconsin	2	1
Pima Indians Diabetes	1	3
Ionosphere	1	1
Ecoli	2	3
Glass	3	2
Yeast	1	3
AVG	1.67	2.17

由表 2 可知, KDTA 的核排序序列与错误率的排序序列一致性更高, 特别是在 Ecoli 和 Yeast 数据上, KDTA 与错误率完全一致, 在两个数据集(Glass 和 Breast Cancer Wisconsin)上为负值(部分序列为错误率排序序列逆序)。KTA 在 4 个数据集上获得了负值, 尤其在 Yeast 上, 其排序序列与错误率完全相反。从表 3 可见, KDTA 的最优核函数错误次数为 3, 而 KTA 为 4。表 4 中, KDTA 的最优核函数排列位序为 1.67, KTA 为 2.17。采用置信度水平为 95% 的 t 检验证明, KDTA 的最优核函数排列位序小于 KTA。

综上, 本文分别使用了模拟数据与 UCI 数据集进行实验, 展示了 KTA 的局限性(不具有线性变换不变性), 并且 KDTA 可以克服该局限, 并且 KDTA 比 KTA 更能反映分类错误率。

结束语 本文提出了一种新的基于核距离的核度量方法, 称之为核距离排列。与 KTA 相比, 使用核距离代替了核函数, 度量核距离矩阵与最优目标矩阵的对齐程度。该方法的提出考虑了两点因素, 首先, 核距离较核更适于度量样本间的相似或接近程度; 其次, 核距离在特征空间中具有线性变换不变性。本文采用模拟数据和 UCI 数据, 通过核选择实验验证了所提方法的有效性。

本文没有考虑不均衡数据和多模态数据的情况, 并且没有给出 KDTA 的泛化错误率的理论保证。再者, 本文将 KTA 的核矩阵替换为核距离矩阵, 那么是否可以将其他度量

方法的核矩阵替换为核距离矩阵, 这将是一个很有趣的研究点。在 Breast Cancer Wisconsin 数据集上 KTA 的效果好于 KDTA, 这表明可能存在本文未考虑的其他因素影响度量效果, 继续改进 KDTA 也是未来的主要工作。

参考文献

- [1] Schölkopf B, Smola A. Learning with Kernels [M]. MIT Press, Cambridge, Massachusetts, 2002
- [2] Gönen M, Alpaydin E. Multiple kernel learning algorithms [J]. Journal of Machine Learning Research, 2011, 12: 2211-2268
- [3] Girolami M, Rogers S. Hierarchic bayesian models for kernel learning [C]//Proceedings of the 22nd international conference on machine learning. Bonn, Germany, Springer Verlag, August 2005: 241-248
- [4] Cheng S-O, Smola A J, Williamson R C. Learning the kernel with hyperkernels [J]. Journal of Machine Learning Research, 2005, 6: 1043-1071
- [5] Cristianini N, Shawe-Taylor J, Elisseeff A, et al. On kernel-target alignment [J]. Advances in Neural Information Processing Systems, 2001, 14: 367-373
- [6] Baram Y. Learning by kernel polarization [J]. Neural Computation, 2005, 17: 1264-1275
- [7] Wang Ting-hua, Tian Sheng-feng, Huang Hou-kuan, et al. Learning by local kernel polariz [J]. Neurocomputing, 2009, 72: 3077-3084
- [8] Nguyen C H, Ho Tu-bao. An efficient kernel matrix evaluation measure [J]. Pattern Recognition, 2008, 41: 3366-3372
- [9] Chudzian P. Evaluation measures for kernel optimization [J]. Pattern Recognition Letters, 2012, 33: 1108-1116
- [10] Schölkopf B. The kernel trick for distance [J]. Advances in Neural Information Processing Systems, 2001, 13: 301-307
- [11] Wang Ting-hua, Zhao Dong-yan, Tian Sheng-feng. An overview of kernel alignment and its applications [J]. Artificial Intelligence Review, November 2012
- [12] Camargo J E, Gonzalez F A. A multi-class kernel alignment method for image collection summarization [C]//Proceedings of the 14th Iberoamerican Conference on Pattern Recognition: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Guadalajara, Mexico, Springer Verlag, November 2009: 545-552
- [13] Igel C, Glasmachers T, Mersch B, et al. Gradientbased optimization of kernel-target alignment for sequence kernels applied to bacterial gene start detections [J]. IEEE Transactions on Computational Biology and Bioinformatics, 2007, 4(2): 216-226
- [14] Wong W W L, Burkowski F J. Using kernel alignment to select features of molecular descriptors in a qsar study [J]. IEEE Transactions on Computational Biology and Bioinformatics, 2011, 8(5): 1373-1384
- [15] Ramona M, Richard G, David B. Multiclass feature selection with kernel gram matrix based criteria [J]. IEEE transactions on neural networks and learning systems, 2012, 23(10): 1611-1623
- [16] Hofmann T, Scholkopf B, Smola A J. Kernel methods in machine learning [J]. The Annals of Statistics, 2008, 36(3): 1171-1220
- [17] Lesot M-J, Rifqi M. Similarity measures for binary and numerical data: a survey [J]. International Journal of Knowledge Engineering and Soft Data Paradigms, 2009, 1(1): 63-84