第 27 卷 第 11 期
Vol. 27 No. 11

控 制 与 决 策
Control and Decision

2012 年 11 月
Nov. 2012

文章编号: 1001-0920 (2012) 11-1706-05

基于 Moore-Penrose 逆的加权距离函数研究

黄德才, 陈 欢

(浙江工业大学 计算机科学与技术学院，杭州 310023)

摘 要: “距离”是科学研究与工程技术领域中使用非常广泛的一种度量. 在分析各种距离优、缺点的基础上, 根据马氏距离不受量纲影响, 能描述和处理相关性数据的性能优势, 利用加权 Moore-Penrose (WMP) 广义逆定义了 WMP 马氏距离, 并通过奇异值分解及矩阵的谱分解理论构造其数学形式和计算方法. 理论分析和仿真实验表明, 所提出的方法不仅保持了马氏距离和 MP 马氏距离的优点, 而且克服了它们的缺点, 同时又具有更好的独特性能.

关键词: 相关性数据；马氏距离；Moore-Penrose 广义逆；奇异值分解；谱分解

中图分类号: TP18 **文献标志码:** A

Research of weighted distances based on weighted Moore-Penrose pseudoinverse

HUANG De-cai, *CHEN Huan*

(College of Computer Science & Technology，Zhejiang University of Technology，Hangzhou 310023，China. Correspondent：CHEN Huan，E-mail：godchenhuan@163.com)

Abstract: Distance is a widely used measure in engineering and researching field. The advantages and disadvantages of some distances are analyzed in Euclidean space. Because the Mahalanobis distance is influenced by the dimension and it has great performance of dealing with related data, the weighted Moore-Penrose(WMP) Mahalanobis distance is defined according to WMP pseudoinverse, whose formula is given by singular value decomposition(SVD) and spectral decomposition of matrices. The academic analysis and simulation show that it not only overcomes the disadvantages of non-existence in Mahalanobis distance, but has its own special performances.

Key words: related data；Mahalanobis distance；Moore-Penrose pseudoinverse；SVD；spectral decomposition

1 引 言

随着企业或行业的业务数据不断积累, 形成了海量数据集. 如果单靠人工去整理或理解如此庞大的数据源, 则存在效率和准确性等问题, 因此, 越来越多的企业正通过数据挖掘技术来解决海量数据的整理和知识发现问题, 并为企业决策提供支持. 距离计算方法在数据挖掘中具有极其重要的地位, 因为数据挖掘的本质是从大量数据中发现数据规律以供企业决策所需. 但如何从原始的大量数据中发现规律, 这些规律又如何与决策需求相联系, 这需要用距离 (相似性度量) 来衡量, 例如在聚类分析与分类分析中, 距离常被用作数据实体间相似性的判断准则[1]. 此外, 在神经网络[2]、图像处理[3]、信号处理[4]、生物学[5]等领域中, 距离计算方法也有着广泛的应用.

在各种应用中, 距离的出现形式各不相同, 较常见的距离形式有以下几种.

首先设 $X_1, X_2, \cdots, X_m$ 为 m 个数据个体. 其中: $X_i = (x_{i1}, x_{i2}, \cdots, x_{in})$, $i = 1, 2, \cdots, m$, n 为数据个体 X_i 的属性个数, 则数据总体可表示为 $X = (X_1, X_2, \cdots, X_m)^{\mathrm{T}}$, 即

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}.$$

对于任意两个数据个体 $X_i = (x_{i1}, x_{i2}, \cdots, x_{in})$, $X_j = (x_{j1}, x_{j2}, \cdots, x_{jn})$, 有如下几种距离:

1) 欧氏 (Euclidean) 距离

收稿日期: 2011-05-05；**修回日期:** 2011-07-20.

基金项目: 浙江省重大科技计划项目(2009C11024).

作者简介: 黄德才(1958−), 男, 教授, 博士生导师, 从事数据挖掘、网格调度、供应链管理等研究；陈欢(1987−), 男, 硕士生, 从事数据挖掘、云计算的研究.

$$d(X_i, X_j) = \|X_i - X_j\| = \Big[\sum_{i=1}^{n}(x_{ik} - x_{jk})^2\Big]^{1/2};$$

2) 绝对值距离 (Manhattan 距离)

$$d(X_i, X_j) = \sum_{k=1}^{n}|x_{ik} - x_{jk}|;$$

3) 切比雪夫距离 (Chebyshev 距离)

$$d(X_i, X_j) = \max_k |x_{ik} - x_{jk}|;$$

4) 明可夫斯基距离 (Minkowski 距离)[6]

$$d(X_i, X_j) = \Big[\sum_{k=1}^{n}|x_{ik} - x_{jk}|^p\Big]^{1/P}.$$

由分析可知, 距离 1) ∼ 距离 3) 是距离 4) 中 p 取特殊值或趋于无穷时的某种情况, 而且只有欧氏距离才具有平移不变性. 当量纲一定时, 两个数据 X_i 与 X_j 越相似, 他们的上述 4 种距离计算值越小, 反之亦然. 显然, 量纲在计算类比过程中起着重要的作用. 为了避免量纲带来的影响, 人们通常采用标准化的手段[10]进行处理, 如对标准化后的数据再求欧氏距离, 亦称为方差加权距离. 但是考虑到标准化的实质是缩小数据属性的差异性, 它必然造成数据均值及方差这两大重要信息的丢失, 存在一定的缺陷. 另外, 上述 4 种距离没有考虑数据属性间的相关性以及总体的变异对"距离"远近的影响. 一个变异大的总体可能与更多样品近些, 这些都使得上述距离的应用受到了较大的限制.

5) Camberra 距离[7]

$$d(X_i, X_j) = \sum_{k=1}^{n}\frac{|x_{ik} - x_{jk}|}{x_{ik} + x_{jk}},$$

$$x_{ik}, x_{jk} \geqslant 0,\ x_{ik} + x_{jk} \neq 0.$$

Camberra 距离消除了量纲的影响, 但仍不能反映属性间的相关性, 而且只适用于正实数域, 这完全不能满足大多数科学研究领域的实际需求.

6) 马氏 (Mahalanobis) 距离

$$d(X_i, X_j) = [(X_i - X_j)S^{-1}(X_i - X_j)^{\mathrm{T}}]^{1/2},$$

其中 S 为数据总体 X 的协方差矩阵.

马氏距离对于一切非奇异线性变换都是不变的, 即它不受量纲的影响. 另外, 由于协方差阵的引进, 它可以忽略冗余数据, 并充分考虑了数据属性的相关性. 但由于协方差矩阵是一种客观的计算度量, 它所反映的数据相关性并不一定与人们主观需求的相关性一致, 当其与所需求的相关性不一致时会使数据挖掘的结果更加恶劣; 而且它在带来较大的时间复杂度的同时要求协方差矩阵可逆, 从而使得马氏距离不一定存在, 这限制了马氏距离的应用范围. 因为部分情况下协方差矩阵的逆矩阵是不存在的, 如数据总体 $T = (t_1, t_2)^{\mathrm{T}}$, $t_1 = t_2 = (1, 1)$, 其协方差阵为 $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, 不可逆, 故导致马氏距离不存在.

除了上述距离外, Harmming 距离[8]、Hausdroff 距离等也在各自领域中有着广泛的应用, 但也仅限于个别领域.

综上所述, 马氏距离的起点相对较高, 它是从统计的角度观察问题, 在处理相关性及量纲数据上具有较大的准确性. 在实际数据挖掘或其他研究过程中, 需要处理的数据源一般均有量纲并且属性间具有相关性, 因此, 马氏距离被广泛用于各领域的各个方面, 如生物学[9]、计算机学[10]、离群点的检测、大数据集的取样等[11]. 但是协方差矩阵不仅带来了较大的时间复杂度, 而且要求它可逆, 这使得马氏距离并不一定存在, 从而严重限制了马氏距离的应用. 作者在文献 [12] 中已针对这个问题提出了适用于任何数据集的 MP 马氏距离, 但是它并没有解决协方差矩阵带来的相关性可能与科研所需要的信息相悖的问题.

为了确定新构造的距离函数更适用于数据挖掘等领域, 本文针对马氏距离的上述缺陷, 基于 MP 马氏距离提出一种处理相关性数据性能更优越的距离函数 —— WMP 马氏距离, 同时利用矩阵的谱分解理论, 围绕权矩阵的确定进行了研究, 进而根据奇异值分解理论构造并证明了其数学形式. 仿真实验进一步验证了 WMP 马氏距离的优越性能. 从而实现了对于任何数据集都存在 MP 马氏距离, 而且相关性、准确性及稳定性均更强的 WMP 马氏距离的构造.

2　基于 Moore-Penrose 逆的距离函数

2.1　基于 Moore-Penrose 逆的 MP 马氏距离

定义 1　对于任意矩阵 A, 必存在唯一矩阵 B, 其阶数与 A^{T} 相同, 使得 B 同时满足如下方程:

$$ABA = A,\ (AB)^{\mathrm{T}} = AB,$$
$$BAB = B,\ (BA)^{\mathrm{T}} = BA. \tag{1}$$

其中 B 为 A 的 Moore-Penrose 广义逆矩阵, 记为 A^+. 式 (1) 称为 Penrose 方程组[13].

定义 2　设 X_i 与 X_j 是来自数据总体 X 的两个数据, S 为总体 X 的协方差矩阵, 则有距离函数

$$d_{mp}(X_i, X_j) = [(X_i - X_j)S^+(X_i - X_j)^{\mathrm{T}}]^{1/2}, \tag{2}$$

其中 S^+ 为协方差矩阵 S 的 Moore-Penrose 逆. 式 (2) 称为 MP 马氏距离[12].

MP 马氏距离的实质是通过奇异值分解, 用协方差阵的广义逆阵代替逆矩阵, 避免了马氏距离不存在的情况, 它继承了马氏距离的相关性, 同时解决了协方差阵不可逆的问题. 由于它满足距离的 3 条基本性质, 说明它是一个距离.

2.2 基于加权 Moore-Penrose 逆的 WMP 马氏距离

虽然MP马氏距离解决了马氏距离不存在的问题, 但是仍没给出当协方差矩阵反映的数据信息与人们需求的信息不一致时的解决方案. 考虑如下加权 Penrose 方程组:

$$\begin{aligned}&ABA=A,\ (MAB)^{\mathrm{T}}=MAB,\\&BAB=B,\ (NBA)^{\mathrm{T}}=NBA.\end{aligned}\tag{3}$$

如果作变换 $\tilde{A}=M^{1/2}AN^{-1/2}$, $\tilde{B}=N^{1/2}\times BM^{-1/2}$, M 和 N 分别为 m, n 阶 Hermite 正定矩阵, 则可以得到 2.1 节中的 Penrose 方程组, 即

$$\begin{aligned}&\tilde{A}\tilde{B}\tilde{A}=\tilde{A},\ (\tilde{A}\tilde{B})^{\mathrm{T}}=\tilde{A}\tilde{B},\\&\tilde{B}\tilde{A}\tilde{B}=\tilde{B},\ (\tilde{B}\tilde{A})^{\mathrm{T}}=\tilde{B}\tilde{A}.\end{aligned}\tag{4}$$

根据实对称矩阵的谱分解理论, 有 $M=\sum\limits_{i=1}^{m}\lambda_ie_ie_i^{\mathrm{T}}$, λ_i 为 M 的特征根, e_i 为标准化正交特征向量, 则有

$$M^{1/2}=\sum_{i=1}^{m}\sqrt{\lambda_i}e_ie_i^{\mathrm{T}},$$

$$M^{-1/2}=(M^{1/2})^{-1}=\sum_{i=1}^{m}\frac{1}{\sqrt{\lambda_i}}e_ie_i^{\mathrm{T}}.$$

同理, 若 β_i 为 N 的特征根, v_i 为对应的标准化特征向量, 则有

$$N^{1/2}=\sum_{i=1}^{n}\sqrt{\beta_i}v_iv_i^{\mathrm{T}},$$

$$N^{-1/2}=(N^{1/2})^{-1}=\sum_{i=1}^{n}\frac{1}{\sqrt{\beta_i}}v_iv_i^{\mathrm{T}}.$$

定义 3 对于任意矩阵 $A\in R^{m\times n}$, M 和 N 分别是 m 与 n 阶 Hermite 正定阵, 若矩阵 B 满足如下加权 Penrose 方程组:

$$\begin{aligned}&ABA=A,\ (MAB)^{\mathrm{T}}=MAB,\\&BAB=B,\ (NBA)^{\mathrm{T}}=NBA,\end{aligned}\tag{5}$$

则称 B 为 A 的 (正定) 加权 Moore-Penrose 逆, 并且它是唯一的, 记作 A^{+}_{MN}.

证明 由 MP 逆矩阵的唯一性[12]易知, $\tilde{B}$ 是 $\tilde{A}$ 的唯一 MP 逆. 又因 M、N 确定后, $\tilde{A}$ 和 $\tilde{B}$ 分别与 A 和 B 一一对应, 所以, 唯一的 A 确定唯一的 B, 即 A^{+}_{MN} 唯一. □

根据上述理论可知, $\tilde{B}$ 是 $\tilde{A}$ 的 Moore-Penrose 广义逆矩阵, 而 $\tilde{A}=M^{1/2}AN^{-1/2}$, $\tilde{B}=N^{1/2}BM^{-1/2}$, 所以有 $N^{1/2}BM^{-1/2}=(M^{1/2}AN^{-1/2})^{+}$, 即

$$A^{+}_{MN}=B=N^{-1/2}(M^{1/2}AN^{-1/2})^{+}M^{1/2}.$$

加权 Moore-Penrose 逆保留了逆的一般性质[12], 如 $(A^{+}_{MN})^{+}_{MN}=A$; $A^{+}_{MN}A=I$ 等.

定理 1 (加权 Moore-Penrose 逆的构造) 若 A 是 $m\times n$ 矩阵, M 和 N 分别为 m 和 n 阶 Hermite 正定阵, 令 $\tilde{A}=M^{1/2}AN^{-1/2}$, 且其奇异值分解形式为 $\tilde{A}=UHV^{\mathrm{T}}$, $H=\begin{bmatrix}\Sigma&0\\0&0\end{bmatrix}$, $\Sigma=\mathrm{diag}(a_1,a_2,\cdots,a_r)$, $a_i>0$, r 是矩阵 $\tilde{A}$ 的秩, U 和 V 为正交阵, 则 A 的加 M、N 权逆

$$A^{+}_{MN}=N^{-1/2}VTU^{\mathrm{T}}M^{1/2},\tag{6}$$

其中 $T=\begin{bmatrix}\Sigma^{-1}&0\\0&0\end{bmatrix}$.

证明 在文献 [12] 中作者已经证明了构造的 A^{+} 满足 Penrose 方程组, 所以有

$$\tilde{B}=\tilde{A}^{+},$$

$$N^{1/2}BM^{-1/2}=(M^{1/2}AN^{-1/2})^{+}=VTU^{\mathrm{T}},$$

$$A^{+}_{MN}=B=N^{-1/2}VTU^{\mathrm{T}}M^{1/2}.$$ □

根据上述理论, 可以得出 n 阶协方差矩阵 S 的加 M、N 权逆为

$$S^{+}_{MN}=B=N^{-1/2}(M^{1/2}SN^{-1/2})^{+}M^{1/2}.\tag{7}$$

其中

$$M^{1/2}=\sum_{i=1}^{m}\sqrt{\alpha_i}e_ie_i^{\mathrm{T}},$$

$$M^{-1/2}=(M^{1/2})^{-1}=\sum_{i=1}^{m}\frac{1}{\sqrt{\alpha_i}}e_ie_i^{\mathrm{T}},$$

$$N^{1/2}=\sum_{i=1}^{n}\sqrt{\beta_i}v_iv_i^{\mathrm{T}},$$

$$N^{-1/2}=(N^{1/2})^{-1}=\sum_{i=1}^{n}\frac{1}{\sqrt{\beta_i}}v_iv_i^{\mathrm{T}},$$

$$(M^{1/2}SN^{-1/2})^{+}=VTU^{\mathrm{T}}.$$

V,T,U 分别通过下述方式求解:

令 $\tilde{S}=M^{1/2}SN^{-1/2}$, 若其奇异值分解形式为 $\tilde{S}=UHV^{\mathrm{T}}$, $H=\begin{bmatrix}\Sigma&0\\0&0\end{bmatrix}$, $\Sigma=\mathrm{diag}(a_1,a_2,\cdots,a_r)$, $a_i>0$, r 是矩阵 $\tilde{S}$ 的秩, U 和 V 为正交阵, 则 $T=\begin{bmatrix}\Sigma^{-1}&0\\0&0\end{bmatrix}$.

至此可以知道, 只要权矩阵 M、N 确定, 则协方差矩阵 S 的加 M、N 权逆便可计算. 在数据集 X 较小时, M、N 的确定可以由专家组主观意识给出, 但是考虑到实际数据挖掘中的数据集相当庞大, 而 M、N 的阶正好是数据集的属性个数, 因此, 在通常情况下由专家给出权矩阵 M 和 N 是相当困难的. 对此, 本文给出一种客观的、完全依托于数据总体 X 的权矩阵 M、N 计算方法.

对于一个数据集, 协方差矩阵必定是半正定的. 根据对称矩阵的谱分解理论, 协方差矩阵 S 的谱分解形式为

$$S=\lambda_1e_1e_1^{\mathrm{T}}+\lambda_2e_2e_2^{\mathrm{T}}+\cdots+\lambda_ne_ne_n^{\mathrm{T}}.$$

其中: λ_i 为 S 的第 i 个特征值, e_i 为对应的标准化正交特征向量.

从上述分解形式可以看出, 若某一特征值越大, 则说明该特征值对应的特征向量对矩阵 S 的贡献 (影响) 越大.

观察加权 Penrose 方程组的后两个方程, 若将 M、S 绑定, N、S_{MN}^{+} 绑定, 则其形式等价于 Penrose 方程组的后两个方程, 所以 M 与 S 应具有相同的意义, N 与 S_{MN}^{+} 具有相同的意义, 即 N 与 S 应成相反关系. 因此可以构造

$$M=\alpha_1e_1e_1^{\mathrm{T}}+\alpha_2e_2e_2^{\mathrm{T}}+\cdots+\alpha_ne_ne_n^{\mathrm{T}}. \tag{8}$$

其中: $\alpha_i>0$, 为 $\lambda_i+\min\{\lambda_i\}+1$ 标准化后的值, 即

$$\alpha_i=\frac{\lambda_i+|\min\{\lambda_j\}|+1}{\sum\limits_{i=1}^{n}(\lambda_i+|\min\{\lambda_j\}|+1)},$$

λ_i 为 S 的第 i 个特征值, e_i 为对应的标准化特征向量.

$$N=\beta_1v_1v_1^{\mathrm{T}}+\beta_2v_2v_2^{\mathrm{T}}+\cdots+\beta_nv_nv_n^{\mathrm{T}}. \tag{9}$$

其中: $\beta_i>0$, 为 α_i 的倒数经标准化的值, 即

$$\beta_i=(1/\alpha_i)\Big/\sum_{i=1}^{n}\frac{1}{\alpha_i},$$

v_i 为向量 e_i 中各元素倒数后再归一化得到的向量. 例如 $e_i=(e_{1i},e_{2i},\cdots,e_{ni})^{\mathrm{T}}$, 令 $v_i=(v_{1i},v_{2i},\cdots,v_{ni})^{\mathrm{T}}$, $\nu_{ji}=(1/e_{ji})\Big/\sum\limits_{j=1}^{n}\dfrac{1}{e_{ji}}, j=1,2,\cdots,n$.

$\alpha_i>0,\beta_i>0$ 保证了所构造的 M、N 为正定矩阵.

定义 4 (WMP 马氏距离)　若数据 X_i, $X_j\in\{X\}$, S 为数据集 $\{X\}$ 的协方差矩阵, 则有距离函数

$$d_{wmp}(X_i,X_j)=[(X_i-X_j)\mathrm{mol}(S_{MN}^{+})(X_i-X_j)^{\mathrm{T}}]^{1/2}. \tag{10}$$

其中 M、N 为满足式 (8) 和 (9) 的矩阵; S_{MN}^{+} 满足式 (7); $\mathrm{mol}(S_{MN}^{+})$ 表示对矩阵 S_{MN}^{+} 的元素进行取模运算, 即对于 S_{MN}^{+} 中的每一个元素 s_{ij}, 若 s_{ij} 为实数, 则保持不变; 若 s_{ij} 为复数, 则取其模.

WMP 马氏距离的实质是通过奇异值分解和矩阵的谱分解理论, 用协方差阵的加权Moore-Penrose 逆代替逆矩阵, 它避免了马氏距离不存在的情况, 解决了协方差矩阵不可逆的问题, 而且不受量纲的影响; 同时, 权矩阵 M、N 的精妙构造, 使得 WMP 马氏距离解决了马氏距离中协方差矩阵不能反映数据正确信息的问题, 在处理相关性数据的性能上得到了很大的提高. 当协方差阵正确反映了数据信息时, M、N 强化了信息的体现; 当协方差阵错误反映数据信息时, 权矩阵 M、N 弱化并纠正了错误信息. 由于本文构造的 M、N 均为正定实对称矩阵, 它显然与 MP 马氏距离一样符合距离定义的 3 条性质: 对称性、非负性和三角不等式性.

3　仿真实验

本文对公用测试数据库 uci 中的 wine 数据集进行仿真. 该数据集具有 14 个属性, 第 1 个属性为类属性, 表明各数据属于哪一类, 共有 3 个类. 为使仿真结果更具说服力, 将上述距离运用于经典的 K-Means 聚类算法 (初始类中心从 3 个固有类中随机取). 根据 K-Means 算法对 wine 数据集中后 13 个属性数据仿真, 将得出的聚类结果与原始类结果进行比较, 重复 10 次实验后, 利用误判率来确定新距离的效果, 见表 1.

表 1　基于各种距离的 K-Means 聚类结果比较

K-Means	平均误判率	误判率标准差
基于欧氏距离	0.339 185	0.114 306
基于马氏距离	0.199 117	0.206 680
基于 MP 马氏距离	0.199 117	0.206 680
基于 WMP 马氏距离	0.068 820	0.014 323

仿真结果验证了本文理论的正确性. 从误判率均值分析可知, 马氏距离、MP 马氏距离和 WMP 马氏距离在性能上都优于欧氏距离; 由于 wine 数据集的协方差矩阵可逆, 得到的基于马氏距离的 K-Means 聚类结果与基于 MP 马氏距离的 K-Means 聚类结果是相同的, 只有当协方差阵不可逆时, MP 马氏距离才会体现其优势; 基于 WMP 马氏距离的 K-Means 聚类效果则明显优于基于其他距离的 K-Means 聚类.

误判率标准差则说明了 WMP 马氏距离在稳定性能上也明显优于欧氏距离和马氏距离, 体现的数据相关性更加可靠; 而马氏距离在稳定性上则稍显逊色. 究其原因, 主要是 WMP 马氏距离解决了协方差矩阵可能错误地反映数据间相关性的问题, 即反映的数据相关性与人们研究主题所需要的相关性不一致. 另外, 将上述仿真方法运用于其他公用测试数据库时, 也得出了同样的结果, 这表明 WMP 马氏距离除了具有准确性和稳定性的优点之外, 还具有普适性.

4　结　论

“距离”是科研及工程领域中使用非常多的度量. 本文研究了数据挖掘领域中各种距离的利弊, 提出了一种新的相似性计算方法, 并通过理论与仿真表明了该距离性能的优越性, 它不仅可以取代马氏距离在处理相关性数据上的作用, 而且还具有独特的性能. 考虑到实际应用中的数据大多具有相关性, 因此 WMP 距离的研究具有很大意义. 但是, 马氏距离的一个较大的缺点——高计算复杂度并没有在本文中得到改进. 因此如何减小计算复杂度, 使本文方法得到

更广泛的应用是未来需要研究的课题.

参考文献(References)

[1] Hwang W J, Wen K W. Fast kNN classification algorithm based on partial distance search[J]. Electronics Letters, 1998, 34(21): 2062-2063.

[2] Kohonen T. Self-organized formation of topologically correct feature maps[J]. Biological Cybernetics, 1998, 43(1): 59-69.

[3] 余莉, 王润生. 基于多尺度变形模板的目标检测与识别[J]. 计算机研究与发展, 2002, 39(10): 1325-1330.
(Yu L, Wang R S. Object detection and recognition based on multiscale deformable template[J]. J of Computer Research and Development, 2002, 39(10): 1325-1330.)

[4] Basseville M. Distance measure for signal-processing and pattern-recognition[J]. Signal Processing, 1989, 18(4): 349-369.

[5] Li L P, Weinberg C R, Darden T A, et al. Gene selection for sample classification based on gene expression data: Study of sensitivity to choice of parameters of the GA/KNN method[J]. Boinformatics, 2001, 17(12): 1131-1142.

[6] Kamimura R, Uchida O. Greedy network-growing by Minkowski distance functions[C]. 2004 IEEE Int Joint Conf on Neural Networks. Budapest, 2004, (4): 2837-2842.

[7] 叶斌, 胡修林, 张蕴玉, 等. 基于3D Zernike矩的三维地形匹配算法及性能分析[J]. 宇航学报, 2007, 28(5): 1241-1245.
(Ye B, Hu X L, Zhang Y Y, et al. 3D terra in matching algorithm and performance analysis based on 3D Zernike moments[J]. J of Astronautics, 2007, 28(5): 1241-1245.)

[8] He Y, Zhang B W, Yao E Y. Weighted inverse minimum spanning tree problems under hamming distance[J]. J of Combinatorial Optimization, 2005, 9(1): 91-100.

[9] Chou K C, Zhang C T. Prediction of protein structural classes[J]. Critical Reviews in Biochemistry and Molecular Biology, 1995, 30(4): 275-349.

[10] Ruiz A, Lopez-de-Teruel P E. Nonlinear kernel-based statistical pattern analysis[J]. IEEE Trans on Neural Networks, 2001, 12(1): 16-32.

[11] Shenk J S, Westerhaus M O. Population definition, sample selection, and calibration procedures for near-infrared reflectance spectroscopy[J]. Crop Science, 1991, 31(2): 469-474.

[12] 陈欢, 黄德才. 基于广义马氏距离的缺损数据补值算法[J]. 计算机科学, 2011, 38(5): 149-153.
(Chen H, Huang D C. Missing data imputation based on generalized mahalanobis distance[J]. Computer Science, 2011, 38(5): 149-153.)

[13] 陈永林. 广义逆矩阵的理论与方法[M]. 南京: 南京师范大学出版社, 2005: 36-40.

(上接第1705页)

[8] Mcbain J, Timusk M. Feature extraction for novelty detection as applied to fault detection in machinery[J]. Pattern Recognition Letter, 2011, 32(7): 1054-1061.

[9] Sbarbaro D, Johansen T A. Analysis of artificial neural networks for pattern-based adaptive control[J]. IEEE Trans on Neural Networks, 2006, 17(5): 1184-1193.

[10] Zhou Xiaojie, Chai Tianyou. Pattern-based hybrid intelligent control for rotary kiln process[C]. The 16th IEEE Int Conf on Control Applications. Singapore, 2007: 30-35.

[11] 王聪, 陈填锐, 刘腾飞. 确定学习与基于数据的建模及控制[J]. 自动化学报, 2009, 35(6): 693-706.
(Wang C, Chen T R, Liu T F. Deterministic learning and data-based modeling and control[J]. Acta Automatica Sinica, 2009, 35(6): 693-706.)

[12] 徐正光. 智能自动化的模式识别方法及其工程实现[D]. 北京: 北京科技大学, 2001.
(Xu Z G. Pattern recognition method of intelligent automation and its implementation in engineering[D]. Beijing: University of Science and Technology Beijing, 2001.)

[13] Xu Z G, Sun C P. Moving pattern-based approach to modeling of a class of complex production processes[C]. Proc of 20th IEEE Int Conf on Fuzzy System. Taipei, 2011: 2282-2287.

[14] Moore R E. Interval analysis[M]. Englewood Cliffs: Prentice-Hall, 1966.

[15] Sun C P, Xu Z G. Extended T-S fuzzy model based on interval arithmetic and its application to interval nonlinear regression analysis[C]. IEEE Int Conf on Fuzzy Systems. Jeju Island, 2009: 1773-1778.

[16] Wang L, Langari R. Complex systems modeling via fuzzy logic[J]. IEEE Trans on System, Man, and Cybernetics, 1996, 26(1): 100-106.