

# 基于局部重构与全局保持的半监督维数约减算法

韦佳<sup>1</sup> 文贵华<sup>1</sup> 王文丰<sup>2</sup> 王家兵<sup>1</sup>

(华南理工大学计算机科学与工程学院 广州 510006)<sup>1</sup> (南昌工程学院信息工程学院 南昌 330099)<sup>2</sup>

**摘要** 针对基于局部与全局保持的半监督维数约减算法(LGSSDR)对邻域参数选择比较敏感以及对邻域图边权值设定不够准确的问题,提出一种基于局部重构与全局保持的半监督维数约减算法(LRGPSSDR)。该算法通过最小化局部重构误差来确定邻域图的边权值,在保持数据集局部结构的同时能够保持其全局结构。在 Extended YaleB 和 CMU PIE 标准人脸库上的实验结果表明 LRGPSSDR 算法的分类性能要优于其它半监督维数约减算法。

**关键词** 边信息,局部重构,半监督学习,维数约减

**中图法分类号** TP181 **文献标识码** A

## Local Reconstruction and Global Preserving Based Semi-supervised Dimensionality Reduction Algorithm

WEI Jia<sup>1</sup> WEN Gui-hua<sup>1</sup> WANG Wen-feng<sup>2</sup> WANG Jia-bing<sup>1</sup>

(School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China)<sup>1</sup>

(School of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, China)<sup>2</sup>

**Abstract** Considering that Local and Global Preserving Based Semi-Supervised Dimensionality Reduction (LGSSDR) is sensitive to the selection of neighborhood parameter and inaccurate in the setting of the edge weights of neighborhood graph, a new algorithm of Local Reconstruction and Global Preserving Based Semi-Supervised Dimensionality Reduction (LRGPSSDR) was proposed in this paper. The algorithm can set the edge weights of neighborhood graph through minimizing the local reconstruction error and can preserve the global geometric structure of the sampled data set as well as preserving its local geometric structure. The experimental results on Extended YaleB and CMU PIE face database demonstrate that LRGPSSDR is better than other semi-supervised dimensionality reduction algorithms in the performance of classification.

**Keywords** Side-information, Local reconstruction, Semi-supervised learning, Dimensionality reduction

## 1 引言

在很多机器学习和数据挖掘任务中,人们常常面临一个问题:要从大量的未标记数据以及有限的成对约束中学习有用的知识。这种学习方式是半监督学习中的一种,而这些成对约束就称之为边信息。边信息包括两种:一种是正约束,另一种是负约束。正约束表示两个样本属于同一类,但并不知道其确切的类别标签;负约束表示两个样本不属于同一类别。

近来,利用边信息进行特征提取受到了越来越多的关注。Shental 等人提出一种相关成分分析算法(RCA)<sup>[1]</sup>,但该算法只能利用正约束信息,并且忽略了隐藏在大量未标记数据中的潜在信息。Bar-hiller 等人提出一种约束 Fisher 线性判别分析算法(cFLD)<sup>[2]</sup>,但该算法存在着与 RCA 同样的问题。Xing<sup>[3]</sup>等人、Tang<sup>[4]</sup>等人以及 Yeung<sup>[5]</sup>等人提出的算法不仅能够利用正约束信息,而且能够利用负约束信息,但没有利用隐藏在未标记数据中的潜在信息。Wu 等人提出一种迭代自增强相关成分分析算法(ISERCA)<sup>[6]</sup>,该算法能同时利用边信息以及未标记数据中的潜在信息,但是它的缺点是时间复

杂度高并且可能陷入局部极小。Zhang 等人提出一种半监督维数约减算法(SSDR)<sup>[7]</sup>,该算法不仅能够保持边信息的结构,并且能够保持未标记数据所在低维流形的结构,但其缺点是只考虑了低维流形的全局协方差结构而没有考虑其局部结构。

在以前的工作中提出了一种基于局部与全局保持的半监督维数约减方法(Local and Global Preserving Based Semi-Supervised Dimensionality Reduction, LGSSDR)<sup>[8]</sup>,它不仅能保持正负约束信息而且能够保持数据集所在低维流形的全局以及局部信息。另外本算法能够计算出变换矩阵并容易地处理未见样本。但是,考虑到 LGSSDR 算法对邻域参数的设定比较敏感,并且对邻域图边权值的设定不能够准确反映样本的局部几何结构,所以本文在 LGSSDR 算法的基础上提出一种基于局部重构误差与全局保持的半监督维数约减方法(Local Reconstruction and Global Preserving Based Semi-Supervised Dimensionality Reduction, LRGPSSDR),该算法不仅具有 LGSSDR 算法的全部优点,而且有其独有的优点,那就是通过最小化局部重建误差的方法来确定邻域图的边权值,对邻域

到稿日期:2010-09-06 返修日期:2010-12-28 本文受国家自然科学基金(60973083),华南理工大学中央高校基本科研业务费专项资金(2009ZM0189,2009ZM0175)资助。

韦佳(1982-),男,博士,讲师,主要研究方向为流形学习、半监督学习,E-mail:csjwei@scut.edu.cn.

参数的设定也不如 LGSSDR 算法敏感。

## 2 问题描述

基于边信息的半监督线性维数约减问题的基本设置如下:给定样本集  $X = \{x_1, x_2, \dots, x_n\} \subset R^D$  以及边信息  $M$  和  $C$ , 其中  $M$  为正约束,  $C$  为负约束。也即如果  $x_i$  与  $x_j$  属于同一类, 那么  $(x_i, x_j) \in M$ ; 如果  $x_i$  与  $x_j$  分别属于两个不同的类, 那么  $(x_i, x_j) \in C$ 。所期望的结果是从上述给定的条件中学习线性变换矩阵  $W = [w_1, w_2, \dots, w_d] \in R^{D \times d}$  ( $d \ll D$ ), 使得原数据经由变换矩阵所得的低维投影为  $Y = \{y_1, y_2, \dots, y_n\} \subset R^d$  (其中  $y_i = W^T x_i$ ), 该低维投影不仅能保持  $M$  和  $C$  中的边信息, 即  $M$  中的点对在低维投影中相互靠近而  $C$  中的点对在低维投影中相互远离, 而且能保持原始数据集的内低维流形结构。

## 3 LGSSDR 算法介绍

LGSSDR 的基本原理是使不属于同一类的数据点离得越远越好(负约束), 而属于同一类的数据点靠得越近越好(正约束)。同时, 为了能够利用大量未标记数据中的潜在信息, 做出如下假设(邻域假设或流形假设): 高维空间中相互靠近的点在低维投影空间中也是相互靠近的。举例来说, 如果高维空间中的两点  $x_i$  和  $x_j$  距离很近, 那么它们的低维投影点  $y_i$  与  $y_j$  之间的距离也应该很近。另一方面, 还希望高维空间中非邻近点的低维投影应该尽量散开(非邻域假设)。因此, LGSSDR 的最佳投影方向可以通过如下准则得到:

$$w^* = \operatorname{argmax}_w \frac{Q_c + \alpha Q_f}{Q_m + \beta Q_n} = \operatorname{argmax}_w \frac{w^T X(L^c + \alpha L^f) X^T w}{w^T X(L^m + \beta L^n) X^T w} \quad (1)$$

其中,

$$\begin{cases} Q_m = \sum_{\substack{(x_i, x_j) \in M \\ \text{where } (x_i, x_j) \in M \text{ or } (x_j, x_i) \in M}} (w^T x_i - w^T x_j)^2 \\ = 2w^T X(D^m - S^m) X^T w = 2w^T X L^m X^T w \\ S_{ij}^m = \begin{cases} 1, & \text{if } (x_i, x_j) \in M \text{ or } (x_j, x_i) \in M \\ 0, & \text{else} \end{cases} \\ D_i^m = \sum_j S_{ij}^m \\ L^m = D^m - S^m \\ Q_c = \sum_{\substack{(x_i, x_j) \in C \\ \text{where } (x_i, x_j) \in C \text{ or } (x_j, x_i) \in C}} (w^T x_i - w^T x_j)^2 \\ = 2w^T X(D^c - S^c) X^T w = 2w^T X L^c X^T w \\ S_{ij}^c = \begin{cases} 1, & \text{if } (x_i, x_j) \in C \text{ or } (x_j, x_i) \in C \\ 0, & \text{else} \end{cases} \\ D_i^c = \sum_j S_{ij}^c \\ L^c = D^c - S^c \end{cases}$$

$$\begin{cases} Q_n = \sum_{ij} (w^T x_i - w^T x_j)^2 S_{ij}^n \\ = 2w^T X(D^n - S^n) X^T w = 2w^T X L^n X^T w \\ S_{ij}^n = \begin{cases} 1, & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{else} \end{cases} \\ N_k(x_j) \text{ 表示点 } x_j \text{ 的 } k \text{ 近邻点的集合} \\ D_i^n = \sum_j S_{ij}^n \\ L^n = D^n - S^n \end{cases}$$

$$\begin{cases} Q_f = \sum_{ij} (w^T x_i - w^T x_j)^2 S_{ij}^f \\ = 2w^T X(D^f - S^f) X^T w = 2w^T X L^f X^T w \\ S_{ij}^f = \begin{cases} 1, & \text{if } x_i \notin N_k(x_j) \text{ and } x_j \notin N_k(x_i) \\ 0, & \text{else} \end{cases} \\ D_i^f = \sum_j S_{ij}^f \\ L^f = D^f - S^f \end{cases}$$

式(1)为一个广义瑞利商问题。如果  $X(L^m + \beta L^n) X^T$  是非奇异的, 那么它的解为下式最大广义特征值所对应的特征向量:

$$X(L^c + \alpha L^f) X^T w = \lambda X(L^m + \beta L^n) X^T w \quad (2)$$

如果  $d > 1$ , 那么取前  $d$  个最大非零广义特征值所对应的特征向量即可组成变换矩阵  $W$ 。

## 4 LRGPSDR 算法

### 4.1 目标函数

在 LLE 算法<sup>[9]</sup>中, Roweis 等假设输入空间的邻域是局部线性的, 这样一来, 输入空间中的每个点都可以由其邻域点的线性组合重构而成。我们在 LRGPSDR 算法中也作出这样的假设, 在这种情况下, 输入空间局部重构误差可以写为:

$$\varepsilon_1(A) = \sum_i \|x_i - \sum_{j: x_j \in N(x_i)} A_{ij} x_j\|^2 \quad (3)$$

式中,  $N(x_i)$  表示  $x_i$  的邻域(本文使用  $k$  邻域)。在满足约束  $\sum_{j: x_j \in N(x_i)} A_{ij} = 1$  (如果  $x_j$  不属于  $x_i$  的邻域, 那么  $A_{ij} = 0$ ) 的条件下, 为了使局部重构误差取得最小值, 可以用 LLE 算法中所用的方法来计算矩阵  $A$ 。

在低维空间中, 为了保持输入空间中的局部邻域结构不变, 根据矩阵  $A$  中的输入空间邻域信息, 低维嵌入空间局部重构误差也必须取最小值:

$$\varepsilon_2(Y) = \sum_i \|y_i - \sum_{j: x_j \in N(x_i)} A_{ij} y_j\|^2 = \operatorname{trace}(YMY^T) \quad (4)$$

式中,  $M = (I - A)^T (I - A)$ ,  $I$  是单位矩阵。当希望结果为线性变换时, 也即  $y_i = W^T x_i$  时, 上式可以写为:

$$\varepsilon_2(Y) = \operatorname{trace}(W^T X M X^T W) \quad (5)$$

如果目标维数为 1, 容易得知  $\varepsilon_2(Y) = w^T X M X^T w$ 。为了方便讨论, 下面仅讨论目标维数为 1 的情况下 LRGPSDR 算法的最优解, 目标维数大于 1 的情况可以很容易推广得到。

LRGPSDR 算法的目标向量可以定义为:

$$w^* = \operatorname{argmax}_w \frac{Q_c + \alpha Q_f}{Q_m + \beta \varepsilon_2} = \operatorname{argmax}_w \frac{w^T X(L^c + \alpha L^f) X^T w}{w^T X(L^m + \beta M) X^T w} \quad (6)$$

式中,  $Q_c, Q_m, Q_f$  的定义与 LGSSDR 中的一样。它的解为下式最大广义特征值所对应的特征向量:

$$X(L^c + \alpha L^f) X^T w = \lambda X(L^m + \beta M) X^T w \quad (7)$$

易知, 如果  $d > 1$ , 那么取前  $d$  个最大非零广义特征值所对应的特征向量即可组成变换矩阵  $W$ 。

### 4.2 算法描述

输入: 样本集  $X = \{x_1, x_2, \dots, x_n\} \subset R^D$ , 正约束  $M$ , 负约束  $C$ , 邻域参数  $k, \alpha, \beta$ ;

输出: 变换矩阵  $W \in R^{D \times d}$  ( $d \ll D$ )。

第一步 构造  $L^c, L^m$  和  $L^f$ , 并根据邻域参数  $k$ , 利用局部重构误差方法构造  $M$ 。

第二步 构造维数约减的优化目标, 最优投影向量  $w^*$  如式(6)所示。

第三步 计算线性变换矩阵。令线性变换矩阵  $W =$

$[w_1, w_2, \dots, w_d](d \ll D)$ , 那么  $w_1, w_2, \dots, w_d$  就是式(7)的最大的  $d$  个广义特征值所对应的广义特征向量。

## 5 实验及分析

为了验证 LRGPSSDR 算法的有效性, 本节通过几个实验展示它与其他算法相比较的结果。比较指标是降维后的低维投影在分类问题上效果的好与差(假设在训练时只知道边信息, 而分类时知道所有训练样本的标签, 所使用的分类方法为最近邻分类法)。本实验将在 Extended YaleB<sup>[10]</sup> 和 CMU PIE<sup>[11]</sup> 人脸数据集上对 LRGPSSDR, LGSSDR, SSSR 以及基线方法(直接在原始输入空间上使用最近邻分类法)进行比较。在下面的实验中, 边信息都是通过从训练样本中随机选取样本点对来获取。如果某一样本点对的两个样本属于同一类, 则把该点对放入正约束中, 反之则放入负约束中。在计算之前先用 PCA 算法对训练样本进行预处理(保留其中 98% 的主成分)。所有算法的实验结果都是 100 次不同边信息情况下的平均值。

Extended YaleB 人脸数据集是由 38 人、每人 64 幅在不同光照下的 256 级灰度图像组成的, 其中每张图像都被裁剪为  $32 \times 32$  大小(见图 1)。从中随机选择每人的 50 个样本作为训练集, 剩余的 14 个样本作为测试集。



图 1 Extended YaleB 人脸数据集中的一些实例

CMU 的 PIE 人脸数据集共包括 68 个人的 41368 张脸部图片, 这些脸部图片是在不同的姿态、光照和表情的条件下采集而来的。其中每张图像都被裁剪为  $32 \times 32$  大小, 每张照片都是 256 级的灰度图片(如图 2 所示)。本实验随机选择其中 24 个人、每人 170 张照片作为实验用例, 然后随机选取其中的 50 张图片作为训练集, 剩余的 120 张图片作为测试集。



图 2 CMU PIE 人脸数据集中的一些实例

为了研究边信息数量(Number of Constraints, NOC)对 SSSR, LGSSDR 和 LRGPSSDR 算法的影响, 在 Extended YaleB 数据集和 CMU PIE 数据集上分别计算当目标维数  $d=30$  和  $d=100$  时这两种算法在不同边信息数量情况下的分类精度, 实验结果如图 3 和图 4 所示(其中 LGSSDR 和 LRGPSSDR 算法的参数均为  $k=5, \alpha=0.1, \beta=0.1$ )。从结果中可以看出, 当目标维数  $d=30$  时, 在 Extended YaleB 数据集上基线方法的分类效果最好, SSSR 算法的分类效果最差, 这是由于目标维数降得过低, 导致其它维数约减算法不如基线方法。但是相比于 SSSR 和 LGSSDR 算法, LRGPSSDR 算法的分类效果还是最好的, 这说明 LRGPSSDR 算法在目标维数较低的情况下也能得到较好的结果。在 CMU PIE 数据集上, LRGPSSDR 算法取得了最好的分类效果, 而 LGSSDR 算法的分类效果却是最差的, 这是由于 LGSSDR 算法对邻域参数比较敏感。这充分说明了 LRGPSSDR 的优点, 即对邻域图边权值的设定更为精确, 对邻域参数选择不敏感。而且相比于 SSSR 算法, LRGPSSDR 算法在边信息数量较少的情况下也能得到很好的分类效果。当目标维数  $d=100$  时, 在 Ex-

tended YaleB 数据集上, SSSR 算法的分类效果依然不如基线方法好, 这是因为 SSSR 算法只保持了全局结构, 而没有保持局部结构, 可见仅保持全局结构是远远不够的。LGSSDR 算法由于既保持了全局结构又保持了局部结构, 因此得到的结果要优于基线方法和 SSSR 算法。LRGPSSDR 算法得到的结果是最好的, 因为它不仅具有 LGSSDR 算法的优点, 而且对邻域图边权值的设定更精确, 更能反映局部结构的实际情况。CMU PIE 数据集上的情况与 Extended YaleB 数据集基本相同, 唯一不同之处在于在该数据集上 SSSR 算法的分类效果要好于基线方法。

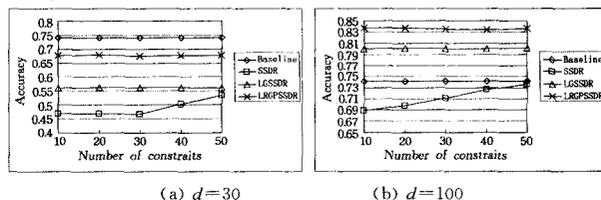


图 3 各种算法在不同边信息数量情况下在 Extended YaleB 数据集上的分类精度

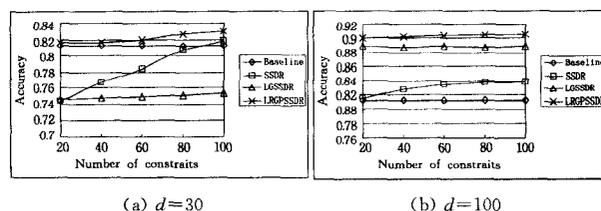


图 4 各种算法在不同边信息数量情况下在 CMU PIE 数据集上的分类精度

为了研究邻域参数  $k$  对 LGSSDR 和 LRGPSSDR 算法的影响, 在 Extended YaleB 数据集和 CMU PIE 数据集上分别计算当目标维数  $d=30$  和  $d=100$  时这两种算法在不同邻域大小情况下的分类精度, 实验结果如图 5 和图 6 所示(NOC=20)。从实验结果中可以看出, LRGPSSDR 算法在任何邻域参数  $k$  下的结果都要好于 LGSSDR 算法的结果(除 Extended YaleB 数据集在  $d=100, k=2$  时的情况之外), 并且其稳定性要好于 LGSSDR 算法。也就是说, LRGPSSDR 算法最好的结果与最坏的结果之间的差距要小于 LGSSDR 算法最好的结果与最坏的结果之间的差距, 这说明本算法对邻域参数的设置不如 LGSSDR 算法敏感, 给实际应用带来了很大的方便。

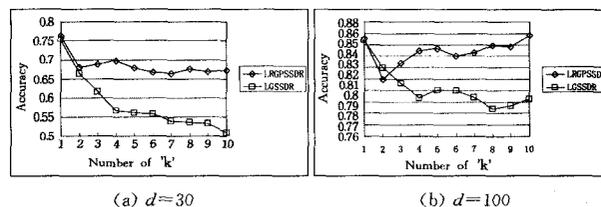


图 5 两种算法在不同邻域大小情况下在 Extended YaleB 数据集上的分类精度

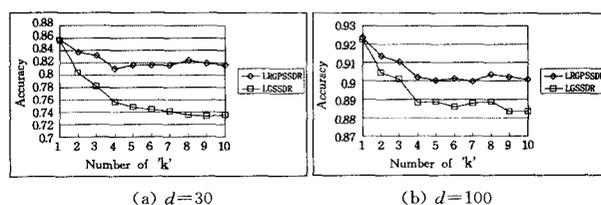


图 6 两种算法在不同邻域大小情况下在 CMU PIE 数据集上的分类精度

**结束语** 由于 LGSSDR 算法对邻域参数选择比较敏感以及对邻域图边权值设定不够准确,在 LGSSDR 算法的基础上提出一种基于局部重构误差与全局保持的半监督维数约减方法。本算法不仅具有 LGSSDR 算法的优点,而且通过最小化局部重构误差的方法对邻域图边权值的设定更为精确,对邻域参数选择也更加容易,为算法的实际应用带来了很大的方便。

但本算法也存在一些不足之处。例如,算法假定样本空间的邻域是局部线性的,而此假定并不是在所有的样本空间中都成立。因此,如何减少此假定对算法的影响,扩大其适用范围,是需要进一步研究的内容。

### 参考文献

- [1] Shental N, Hertz T, Weinshall D, et al. Adjustment Learning and Relevant Component Analysis[C]//LNCS 2353. London: Springer-Verlag,2002:776-790
- [2] Bar-Hillel A, Hertz T, Shental N, et al. Learning a Mahalanobis Metric from Equivalence Constraints[J]. *Journal of Machine Learning Research*,2005,6(6):937-965
- [3] King E P, Ng A Y, Jordan M I, et al. Distance Metric Learning, with Application to Clustering with Side-Information[C]//Advances in Neural Information Processing Systems 15. Cambridge: MIT Press,2003:505-512
- [4] Tang W, Zhong S. Pairwise Constraints-guided Dimensionality

(上接第 181 页)

关系进行了研究,提出了空间线对象连接度的定义,并在此基础上实现了 SLCC 算法。其中间接连接步长(定义 8)和间接连接度(定义 11)在实际应用中要与具体的应用背景相结合,本文未考虑线对象的长度、宽度等因素,也未考虑复杂线对象。SLCC 算法具有一定的应用价值,可以为道路建设的选址、公交线路的设定等提供相关的参考,一般来说良好的道路选址或者公交线路的设定,其道路或线路的连接度都比较高。空间线对象除了自身的属性外,还具有形状、长度、宽度、拓扑关系等空间属性,本文仅仅考虑了线对象相交关系,如何将线对象自身的属性和空间属性综合考虑,还有待深入研究。

### 参考文献

- [1] Andrew A, Thomas C, Korniss G. Ecological invasion: spatial clustering and the critical radius[J]. *Evolutionary Ecology Research*,2007(9):375-394
- [2] Wan L, Li Y, Liu W, et al. Application and study of Spatial Cluster and Customer Partitioning[C]//Proc. of the Fourth International Conference on Machine Learning and Cybernetics. Guangzhou,2005:1701-1706
- [3] Tanser F, Bärnighausen T, Cooke G S, et al. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic [J]. *Int. Journal of Epidemiology*,2009(4):1008-1016
- [4] Lloyd S P. Least Squares Quantization in PCM [J]. *IEEE Trans. Information Theory*,1982(28):128-137
- [5] Ng A R, Han J. Efficient and effective clustering method for spatial data mining [C]//Proc. of the 1994 Intl. Conf. Very Large Databases. Santiago, Chile,1994:144-155
- [6] Zhang T, Ramakrishnan R, Livny M. BIRCH: an efficient data clustering method for very large databases [C]//Proc. of the

Reduction[C]//Proc of the 2006 SIAM DM Workshop on Feature Selection for Data Mining. 2006:59-66

- [5] Yeung D Y, Chang H. Extending the Relevant Component Analysis Algorithm for Metric Learning Using Both Positive and Negative Equivalence Constraints [J]. *Pattern Recognition*,2006,39(5):1007-1010
- [6] Wu F, Zhou Y L, Zhang C S. Relevant Linear Feature Extraction Using Side-information and Unlabeled Data [C]//Proc of the 17th International Conference on Pattern Recognition. Washington: IEEE Computer Society,2004:582-585
- [7] Zhang D Q, Zhou Z H, Chen S C. Semi-supervised Dimensionality Reduction [C]//Proc of the 7th SIAM International Conference on Data Mining. 2007:629-634
- [8] 韦佳,彭宏. 基于局部与全局保持的半监督维数约减方法[J]. *软件学报*,2008,19(11):2833-2842
- [9] Roweis S T, Saul L K. Nonlinear Dimensionality Reduction by Locally Linear Embedding [J]. *Science*,2000,290(22):2323-2327
- [10] Georghiades A S, Belhumeur P N, Kriegman D J. From Few to Many: Illumination Cone Models for Face Recognition Under Variable Lighting and Pose [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,2001,23(6):643-660
- [11] Sim T, Barker S, Bsat M. The CMU Pose, Illumination, and Expression Database [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,2003,25(12):1615-1618

1996 Intl. Conf. Management of Data. Montreal, Canada, 1996:103-114

- [7] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases [C]//Proc. of the 1996 Intl. Conf. Knowledge Discovery and Data Mining. Portland, OR, 1996:226-231
- [8] Wang W, Yang J, Muntz R. STING: A statistical information grid approach to spatial data mining [C]//Proc. of the 1997 Intl. Conf. Very Large Data Bases. Athens, Greece, 1997:186-195
- [9] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm [J]. *J. Royal Statistical Society*,1977(39):1-38
- [10] Ji G, Miao J, Bao P. A Spatial Clustering Algorithm Based on Spatial Topological Relations for GML Data [C]//Proc. of Intl. Conf. on Artificial Intelligence and Computational Intelligence. Shanghai China: IEEE Computer Society,2009:298-301
- [11] Ji G, Miao J, Yang M. A Novel Spatial Clustering Algorithm Based on Spatial Adjacent Relations for GML Data [C]//Proc. of Intl. Workshop on Education Technology and Computer Science. Wuhan, China: IEEE Computer Society,2009:278-281
- [12] Ji G, Zhang L. A Spatial Polygon Objects Clustering Algorithm Based on Topological Relations for GML Data [C]//Proc. of Intl. Conference on Information Engineering and Computer Science. Wuhan, China: IEEE Publishing,2009:363-366
- [13] Yang N, Ji G. A Spatial Lines Clustering Algorithm Based on Adjacent Relations for GML Data [C]//Proc. of Intl. Conf. on Information Engineering and Computer Science. Wuhan, China: IEEE Publishing,2009:3593-3596
- [14] 李永红,华一新. 一种快速判断线段相交的方法 [J]. *测绘通报*,2003(7):30-31
- [15] 杨娜,吉桂林. 一种基于相交关系的 GML 空间聚类算法 [J]. *广西师范大学学报:自然科学版*,2009(3):113-117