

一种基于 ICA 和 LDA 组合的人脸识别新方法^{*}

郑宇杰¹ 於东军¹ 杨静宇¹ 吴小俊^{2,3} 王卫东¹

(南京理工大学信息学院 南京 210094)¹ (江苏科技大学 镇江 212003)²

(中国科学院机器人学开放研究实验室 沈阳 110015)³

摘要 特征提取是模式识别研究领域的一个热点。本文提出了一种基于独立成分分析和线性鉴别分析的特征提取方法。该方法中引入了零空间的概念,指出了前人算法中的不足之处,并且给出了一个完整的独立成分分析和线性鉴别分析的组合法。在 ORL 和 Yale 人脸数据库上的实验表明了该方法的有效性。

关键词 人脸识别,特征提取,独立成分分析,线性鉴别分析,零空间

A New Method of Fusion of ICA and LDA for Face Recognition

ZHENG Yu-Jie¹ YU Dong-Jun¹ YANG Jing-Yu¹ WU Xiao-Jun^{2,3} WANG Wei-Dong¹

(Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094)¹

(School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang 212003)²

(Robotics Laboratory, Chinese Academy of Science, Shenyang 110015)³

Abstract Feature extraction is one of the hot topics in the field of pattern recognition. Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) are two effective methods of feature extraction. ICA can extract features from high order and LDA can extract features which are useful to classify. However, performance of ICA method combined with LDA is reported as lower than ICA method previously. In this paper, we point out the weakness of the previous methods and a new method of feature extraction based on ICA and LDA with conception of null space is proposed. Experiment results carried on face databases demonstrate the effectiveness of the proposed method.

Keywords Face recognition, Feature extraction, Independent component analysis, Linear discriminant analysis, Null space

1 引言

在模式识别领域,特别是人脸识别、字符识别等领域,特征提取是一个最基本的问题。独立成分分析(ICA)^[1~6]和线性鉴别分析(LDA)^[7,8]是两种经典的特征提取方法,它们被广泛地应用于该领域。独立成分分析方法由于可以从高阶上抽取得到数据的独立特征,因此近年来该方法在人脸识别的特征提取中受到了重视,特别是待识别的人脸处于不同的光照、姿势或表情的情况下。然而,通过 ICA 方法得到的特征信息是更能重构原始数据的信息并不是原始数据的最佳分类信息,因此在模式分类问题上还存在缺陷,而通过 LDA 方法则能得到相应的最佳分类的特征信息。基于这一想法, Bartlett^[4]等提出通过 ICA+LDA 方法得到的特征数据也许能提高分类的性能,但是他们并没有给出有说服力的实验结果。然而跟他们的期望相反的是,在 C. Liu^[9]等人的实验中指出基于 ICA+LDA 方法得到的特征数据的识别性能低于基于 ICA 方法得到的特征数据的识别性能。最近, Juneho Yi^[10]等指出 Liu 方法中采用的特征矢量的维数少于 $(C-1)$, 而最佳的基于 Fisher 准则的 LDA 方法是得到 $(C-1)$ 维的特征矢量(其中的 C 代表类别的数目)。但是 Juneho 的方法中并没有考虑在经过 ICA 变换之后样本的类内散布矩阵存在零空间^[11~13]的问题。本文中,我们将关注在此零空间内的信息,并提出了一种新的基于 ICA 和零空间 LDA 的新方法。

2 基于 ICA 的人脸识别方法

ICA 的最初思想是由 Jutten 和 Herault 提出用来解决鸡尾酒会问题的,目的是想从混合信号中抽取得到相应的独立源信号。它的基本思想如下:

假设得到了 M 个观测信号 $x_i (i=1, 2, \dots, M)$, 每个观测信号都是由 N 个独立的源信号 $s_j (j=1, 2, \dots, N)$ 线性混合而成,即 $x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{iN}s_N, (i=1, 2, \dots, M)$ 。如果用矩阵的形式来表示相应的观测信号和源信号,那么我们可以得到如下的表示形式:

$$X = AS \quad (1)$$

其中 A 为混合矩阵, S 为未知的独立源信号。

ICA 理论的基本出发点是假设各个独立源信号是统计独立的、非高斯的。因此,尽管在这个模型中 A, S 都是未知的,我们仍然可以根据输出信号的独立性原则估计出这些未知的信号。显然,问题的关键是如何度量分离结果的独立性。基于 ICA 理论的基本出发点,信号的独立性度量是通过它的非高斯性度量来实现的,文[6,14]介绍了多种独立性度量原则。本文采用了一种基于负熵的独立性判决准则和分离算法——快速定点独立成分分析 (FASTICA)^[5] 算法。FASTICA 与已经存在的 ICA 算法相比,有很多的优点。如 FASTICA 算法的收敛速度是 3 次的(或至少是 2 次的),而普通 ICA 算法的收敛速度是线性的;FASTICA 算法通过使用一个非线性

^{*} 基金项目:国家自然科学基金资助(编号:60472060),南京理工大学科研发展基金。郑宇杰 博士研究生。

函数直接找到了任何非高斯分布的独立成分以及独立成分能一个个估计等等。

当 ICA 算法^[3]被应用于图像处理时,其模型中的 X 通常是指零均值和白化后的影像,因此首先要对原始训练样本按照式(2)进行零均值化。

$$X = (X - \bar{X}) / \sqrt{D} \quad (2)$$

其中式(2)右边的 X 为训练样本集, \bar{X} 为训练样本集的平均图像, D 为训练样本的标准方差。

零均值化之后,图像的白化过程是根据 PCA 算法得到均值化后影像 X 的特征值矩阵 Λ 和特征向量矩阵 U , 其中 Λ 中的特征值按照从大到小的顺序排列, U 中的特征向量按照对应的特征值的顺序排列。取 Λ 的前 m 个特征值构成矩阵 Λ^m , 则对应的特征向量构成的矩阵为 U^m 。由此,我们可以定义白化矩阵 G 为:

$$G = (\Lambda^m)^{-1/2} U^{mT} \quad (3)$$

因此,零均值后的训练样本集 X 的白化过程为:

$$X' = GX \quad (4)$$

对原始图像数据经过预处理之后,我们可以得到 ICA 的独立成分 S 和混合矩阵 A 为:

$$S = W^T X' = W^T GX, A = (W^T G)^{-1} = G^{-1} W \quad (5)$$

3 线性鉴别分析

线性鉴别分析方法^[8]选择与类内散布矩阵正交的向量作为特征脸空间,从而能压制图像之间与识别信息无关的差异,并且对光照及人脸表情变化都不太敏感。其目的是从高维特征空间中提取出最具有鉴别能力的低维特征,我们希望投影后,在低维空间里不同类别的样本尽可能分得开些,同时希望每个类别内部样本尽量密集,也就是说,样本类间离散度越大越好,而样本类内离散度越小越好。

假设有 c 类已知的模式 $w_1, w_2, \dots, w_c, X = \{x_i\}, i = 1, 2, \dots, n$ 为一个具有 f 维的样本集。任何一个在样本集 X 中的样本 x_i 均属于某个已知类别 w_j , 即 $x_i \in w_j, i = 1, 2, \dots, n, j = 1, 2, \dots, c$ 。

那么样本类间离散度矩阵定义为:

$$s_b = \sum_{i=1}^c p_i (u_i - u)(u_i - u)^T \quad (6)$$

其中, u_i 是 w_i 类的均值, u 是所有样本的均值, p_i 是先验概率。

样本类内离散度矩阵定义为:

$$s_w = \sum_{i=1}^c p_i e_i \quad (7)$$

其中 $e_i = E[(x - u_i)(x - u_i)^T | x \in w_i]$ (8)

我们希望投影后,在低维空间里得到的为最有助于分类的信息。因此,如果类内散布矩阵 s_w 是非奇异矩阵,最优的投影方向 W_{opt} 就是使得样本类间离散度矩阵和样本类内离散度矩阵的行列式比值最大的那些正交特征向量。因此 Fisher 准则函数定义为:

$$J(W_{opt}) = \arg \max_w \frac{W^T s_b W}{W^T s_w W} \quad (9)$$

通过线性代数理论,我们可知 W_{opt} 就是满足如下等式的解:

$$s_b W_i = \lambda_i s_w W_i \quad (10)$$

通常的 Fisher 准则函数通过取对应矩阵 $s_w^{-1} s_b$ 大于零的特征值 λ_i 所对应的特征向量来得到投影空间,此时我们注意到该矩阵最多只有 $c-1$ 个非零特征值。

当类内散布矩阵奇异的时候, Fisher 准则函数也可以采用如下的形式:

$$J(W_{opt}) = \arg \max_w \frac{W^T s_b W}{W^T s_i W} \quad (11)$$

其中的 s_i 为样本的总体散布矩阵。

4 我们的方法

为了在 ICA 方法中同时取得有效的分类信息,引入 LDA 方法不失为一个合理的想法,因此 Juneho 等提出了一个 ICA + LDA 的算法。其思想是在经过 ICA 变换的空间内采用 LDA 算法抽取得到相应的分类特征信息,同时抽取得到的最终特征矢量维数为 $(C-1)$, 其中 C 为总的样本类别数,这个是在 Fisher 鉴别准则下所能抽取得到的最大的特征维数。但是该方法并没有考虑经过 ICA 变换的空间中会出现新的类内散布矩阵奇异的情况,此时仅利用基于 Fisher 鉴别准则的线性分析方法是否会丢弃相应的分类信息呢? 那么,如何有效地得到更多鉴别信息,以提高该方法的识别率呢? 在本文中,我们改进了 Juneho 的方法,得出了一个新的基于 ICA 和零空间概念的 LDA 组合的新方法。

文[8]中指出,假设原始图像数据空间为 R^n , 那么所有的最优鉴别矢量都能从低维空间 $R^m (m = rank(S_i))$ 中得到而不丢失任何的信息。根据 ICA 理论,原始图像也能从高维的 R^n 空间转换到低维的 R^m 空间,转换的公式如下:

$$X = ICY \quad (12)$$

其中的 $IC = (IC_1, IC_2, \dots, IC_m)$ 代表从原始图像中得到的独立成分。

通过线性变换之后,我们可以得到在 R^m 空间的 Fisher 鉴别准则:

$$J(X) = \frac{Y^T (IC^T S_b IC) Y}{Y^T (IC^T S_i IC) Y} = \frac{Y^T \tilde{S}_b Y}{Y^T \tilde{S}_i Y} = \tilde{J}(Y) \quad (13)$$

其中的 $\tilde{S}_b = IC^T S_b IC$ 和 $\tilde{S}_i = IC^T S_i IC$ 分别被看作是在 R^m 空间的类间散布矩阵和总体散布矩阵。因此,我们很容易得到:

定理 1 假设 Y_1, Y_2, \dots, Y_d 是通过 ICA 变换空间的 Fisher 鉴别函数 $\tilde{J}(Y)$ 得到的最优鉴别矢量,那么 $X_1 = ICY_1, X_2 = ICY_2, \dots, X_d = ICY_d$ 就是基于 $J(x)$ 得到的最优鉴别矢量。

因此,总的线性变换可以如下定义:

$$Z = W^T X \quad (14)$$

其中

$$W^T = (X_1, X_2, \dots, X_d)^T = (ICY_1, ICY_2, \dots, ICY_d)^T = (Y_1, Y_2, \dots, Y_d)^T IC^T \quad (15)$$

因此我们现在的问题是如何取得在经过 ICA 变换的子空间 R^m 中的最优鉴别矢量 Y_1, Y_2, \dots, Y_d 。因为在经 ICA 降维的子空间 R^m 中,仍然会存在 $m > rank(\tilde{S}_w)$ 的问题,所以对此子空间做特征提取的时候就应当考虑存在于 \tilde{S}_w 对应的零空间中的特征信息。

假设 $\gamma_1, \gamma_2, \dots, \gamma_m$ 是降维之后空间类内散布矩阵 \tilde{S}_w 的 m 个特征向量,并且其对应的特征值按照从大到小的顺序排列 $\lambda_1 > \lambda_2 > \dots > \lambda_m$ 。如果存在零空间,我们假定其中的前 m_1 个特征值是正值。因此,对应的特征矢量可以分为 2 部分:特征值为 0 的值对应的特征向量 $P_1 = (\gamma_{m_1+1}, \gamma_{m_1+2}, \dots, \gamma_m)$ 和特征值不为 0 的值对应的特征向量 $P_2 = (\gamma_1, \gamma_2, \dots, \gamma_{m_1})$ 。此时利用 LDA 做特征抽取的时候必须同时考虑在这两个不同

的子空间的信息。基于以上关于零空间的分析,我们可以得到以下完整的算法描述:

步骤 1:对原始图像做 ICA 变换,将原始图像变换到 m ($m = \text{rank}(s_i)$) 维空间。计算得到 R^m 空间类内散布矩阵 \tilde{S}_w 的正交特征向量 $\gamma_1, \gamma_2, \dots, \gamma_m$ 。所有的特征向量按照对应特征值从大到小的顺序排列,其中前 $m-1$ 个特征向量对应大于 0 的特征值。

步骤 2:得到特征值 0 所对应的特征向量 $P_1 = (\gamma_{m+1}, \gamma_{m+2}, \dots, \gamma_m)$ 。那么我们可以得到新的类间散布矩阵 $\tilde{S}_b = P_1^T \tilde{S}_b P_1$ 。由于此时位于 \tilde{S}_w 的零空间中,计算得到 \tilde{S}_b 的正交特征向量 Z_1, Z_2, \dots, Z_l (通常 $l=C-1, C$ 为样本数目)。因此我们可以得到当相应的类内散布矩阵奇异时的最优鉴别矢量集 $Y_j = P_1 Z_j (j=1, 2, \dots, l)$ 。

步骤 3:得到特征值不为 0 所对应的特征向量 $P_2 = (\gamma_1, \gamma_2, \dots, \gamma_m)$ 。那么我们可以得到在新的子空间内对应的类间散布矩阵 $\tilde{S}_b = P_2^T \tilde{S}_b P_2$ 和总体散布矩阵 $\tilde{S}_t = P_2^T \tilde{S}_t P_2$ 。因此

可以根据 Fisher 鉴别准则得到余下的 $(d-l)$ 个最优鉴别矢量 $Z_j (j=l+1, l+2, \dots, d)$ 。由此,我们可以得到当相应的类内散布矩阵非奇异时的最优鉴别矢量集 $Y_j = P_2 Z_j (j=l+1, l+2, \dots, d)$ 。

步骤 4:通过 $Y_j = P_1 Z_j$ 和 $Y_j = P_2 Z_j$ 得到所有在经过 ICA 变换之后子空间内的最优鉴别矢量集 $\text{disc-set} = (Y_1, Y_2, \dots, Y_l, Y_{l+1}, \dots, Y_d)$ 。

步骤 5:得到测试样本在最优鉴别矢量集上的投影数据进行分类。

5 实验结果

本文采用 ORL 和 Yale 人脸数据库验证该算法。ORL 人脸数据库包含了 40 个人在不同表情、不同视角下的 400 幅照片,每个人均有 10 幅照片。图 1 显示了 ORL 人脸数据库的部分人脸图像。在我们的实验中,ORL 人脸数据库的原始图像都被预先调整到 28×23 。

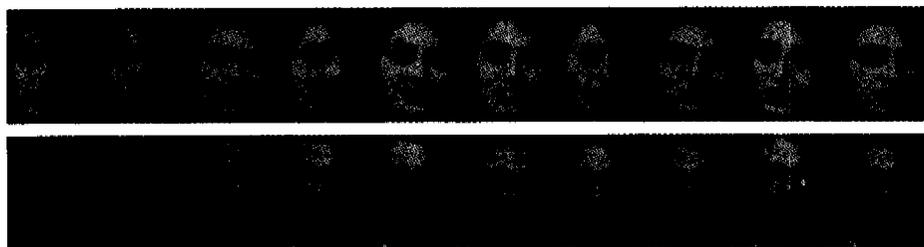


图 1 ORL 人脸数据库的部分人脸

Yale 人脸数据库包含了 15 个人的 165 幅照片。每个人由 11 幅照片所构成,这些照片在不同的表情和光照等条件下:是否带眼镜,高兴,惊讶,伤心等不同情形下拍摄。图 2

显示了该人脸库中的部分人脸图像。该人脸库中的原始图像也预先调整到 40×31 。

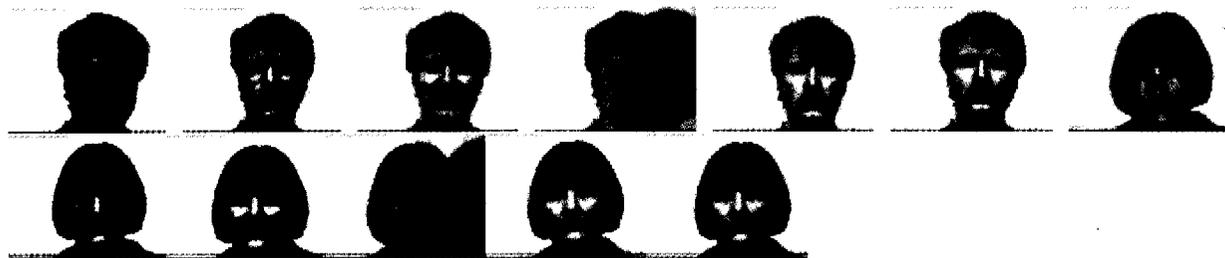


图 2 Yale 人脸数据库的部分人脸

在实验中,我们比较了传统的人脸识别方法和我们方法的识别性能。对于每一组实验结果,训练样本均从人脸数据库中随机产生,每次随机抽取每类人脸的 $k(k=4, 5, 6)$ 个样本组成训练样本集,并将全部的人脸数据作为测试数据。在固定训练样本数目的情况下,对于每次实验均随机地产生 10

组不同的训练样本集,以统计相应算法的识别性能。其中的 ICA 算法采用 FASTICA 算法,最终分类器采用 $K(K=3)$ 近邻分类器。在 ORL 人脸数据库和 Yale 人脸数据库上的平均识别率和相应的标准方差如表 1 所示。

表 1 传统人脸识别方法和我们的方法在 ORL 和 Yale 人脸数据库上的识别率

Face database	Training samples of per-class	ICA method	Fisherface	Eigenface	Proposed method
ORL	6	92.78±0.95	94.05±0.67	89.45±1.32	95.73±0.95
	5	90.78±0.80	92.03±0.75	88.45±1.19	93.95±0.87
	4	87.60±1.53	89.68±1.11	85.48±1.29	91.13±1.68
Yale	6	91.33±1.18	93.12±1.14	91.09±1.03	96.79±1.40
	5	91.27±1.86	93.42±1.45	91.40±1.27	96.91±1.70
	4	86.12±2.26	94.36±3.41	88.97±1.93	94.36±3.41

从表 1 的识别性能比较上可以看出我们提出的方法明显

地优于传统的人脸识别方法。基于 ICA+CLDA 的人脸特征

提取方法,从高阶上抽取了人脸的代数特征,使得抽取得到的各个特征在统计上是独立的,改进了 PCA 算法中只得到不相关特征的弱点。CLDA 方法的加入,使得抽取得到的特征保留了有助于分类的信息,并且本文方法中又利用零空间的信息,得到了更多的分类信息,消除了由于光照、表情等对于人脸分类带来的影响。图 3 和图 4 分别显示不同的人脸识别方法在 ORL 和 Yale 人脸数据库中 10 次不同实验下的识别性能比较,其中的训练样本集由每类人脸的 6 个样本随机组成。从识别性能比较图中可以看出,该方法的识别率始终优于传统的人脸识别方法。由此说明在 ICA 方法中加入了能得到分类信息的 LDA 方法之后,人脸的识别性能有了明显的提高,并且该方法的鲁棒性较好。

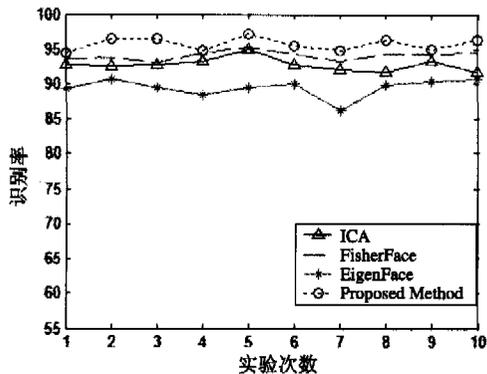


图 3 ORL 人脸数据中不同训练样本集下的识别率比较

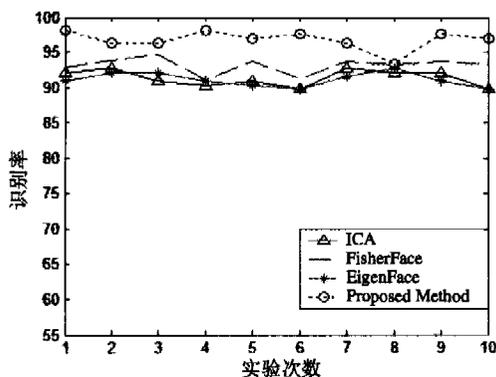


图 4 Yale 人脸数据中不同训练样本集下的识别率比较

结论 本文在 ICA 的基础上,提出了一种基于零空间概

念的 LDA 方法与 ICA 方法相结合的算法。通过 ICA 算法可得到统计独立的特征,为了从这个特征向量中得到更多的有助于分类的信息,结合 LDA 方法以得到保留了更多分类信息的特征数据,并且引入零空间的概念,抽取得到经过 ICA 变换之后的所有的分类信息。从而从数据的高级特性中得到了更多有助于分类的特征信息。在 ORL 和 Yale 人脸数据库的实验表明 ICA+LDA 的方法识别率优于 ICA 方法的识别率。

参考文献

- 1 Juten C, Herault J. Blind separation of source[J],part 1; An adaptive algorithm base on neuromimetic architecture. Signal Processing, 1991, 24(1): 1~10
- 2 Comon P. Independent component analysis—A new concept? [J]. Signal Processing, 1994, 36(3): 287~314
- 3 Bartlett M S, Movellan J R, Sejnowski T J. Face Recognition by Independent Component Analysis[J]. IEEE Transaction on Neural Network, 2002, 13(6): 1450~1464
- 4 Stewart B M, Sejnowski T J. Independent components of face images; A representation for face recognition[A]. In: Proc. of the 4th Annual Joint Symposium on Neural Computation, Pasadena, CA, May 1997
- 5 Hyvarinen A. Fast and robust fixed-point algorithm for independent component analysis [J]. IEEE Transaction on Neural Network, 1999, 10(3): 626~634
- 6 杨竹青,李勇,胡德文.独立成分分析方法综述[J].自动化学报, 2002, 28(5): 762~772
- 7 Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs Fisherfaces; Recognition Using Class Specific Linear Projection [J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 1997, 19(7): 711~720
- 8 Yang J, Yang J Y. Why can LDA be performed in PCA transformed space? [J]. Pattern Recognition, 2003, 36(2): 563~566
- 9 Liu C J, Wechsler H. Comparative assessment of independents component analysis(ICA) for face recognition[A]. In: Second Int. Conf. Audio and Video-based Biometric person Authentication, 1999. 22~24
- 10 Yi J, Kim J, Choi J, Han J, Lee E. Face Recognition Based on ICA Combined with FLD[A]. Biometric Authentication, 2002. 10~18
- 11 Yu H, Weng J. Using Discriminant Eigenfeatures for Image Retrieval[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 1996, 18(8): 831~836
- 12 Chen L F, Liao H Y M, Lin J C, Ko M T, Yj G J. A New LDA-based Face Recognition System Which Can Solve the Small Sample Size Problem[J]. Pattern Recognition, 2000, 33(10): 1713~1726
- 13 Yang J, Frangi A F, Yang J Y, Zhang D, Jin Z. KPCA plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature extraction and Recognition. IEEE Trans, Pattern Analysis and Machine Intelligence, 2005, 27(2): 230~244
- 14 Hyvarinen A, Oja E. Independent component analysis; algorithms and applications [J]. IEEE Trans on Neural Network, 2000, 13(4-5): 411~430

(上接第 189 页)

与 Apriori 算法相比, BOM 算法利用位矩阵可以更方便地来处理数据,矩阵的编码方式也被用来判断一个待生成的 k-项集是不是频繁项集,大大地提高了挖掘关联规则的效率,并且基于矩阵的算法也便于在实际应用中被实现和理解。

结论 为了挖掘大型事务数据库中数据间所有的关联规则,文章提出了一个新的算法 BOM 算法,并且与经典的挖掘关联规则算法 Apriori 算法做了分析对比。BOM 算法具有良好的性能,它不需要多次扫描事务数据库;不需要在内存中存储大量的候选项集,相反,对于一个新的待生成的 k-项集,会立即被判断是否是一个频繁项集,如果不是,则不生成该 k-项集。因此,与 Apriori 算法相比,对于大型交易事务数据库,算法节约了大量的时间和空间,具有良好的性能。

在以后的研究中,将在以下几个方面开展:对如何高效地获得频繁项集的大小 k 进行研究寻找一种更好的数据结构来

替代矩阵,进一步降低对内存的需求,提高算法的性能。频繁项集出现的位置并没有给予太多考虑,但是对于一些应用,这将是很有用的,这也是进一步研究的课题。

参考文献

- 1 Agrawal R, Srikant R. Mining sequential patterns: [IBM Research Report]. 1995
- 2 Agrawal R, Imielinski T, Swami A. Mining association rules between sets in large databases. In: Proc. the ACM SIGMOD Conf. Management of Data, May 1993. 207~216
- 3 Agrawal R, Srikant R. Fast algorithm for mining association rules; [IBM Research Reprt]. 1994
- 4 Agrawal R, Mannila H, Toivonen H, et al. Fast Discovery of Association Rules. In Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996. 306~328
- 5 Huang Liusheng, Chen Huaping, Wang Xun, Cheng Guoliang, A Fast Algorithm for Mining Association Rules. In J. Comput. Sci. & Technol., 2000, 15(6): 619~624
- 6 Jiawei Han, Micheline Kamber. Data Mining ; Concepts and Techniques[C]. Mongan Kaufmann publishers, 2000. 225~278