

doi:10.3969/j.issn.1001-2400.2016.05.013

面向复杂工业大数据的实时特征提取方法

孔宪光, 章 雄, 马洪波, 常建涛, 牛 萌

(西安电子科技大学 工业大数据技术研究中心, 陕西 西安 710071)

摘要: 工业大数据具有大体量、多源性、连续采样和价值密度低等特点, 造成其数据复杂度高、实时性强和异常数据多。而传统的特征提取方法已无法满足复杂工业大数据的实时性要求, 同时工业大数据的处理方法不同于基于互联网的数据流处理方法, 其对精度要求较高。针对该问题, 提出一种鲁棒的增量在线特征提取方法, 即鲁棒增量主成分分析, 采用滑动窗口动态更新数据, 过滤窗口内的异常数据点; 然后对窗口内数据进行增量主成分分析, 从而满足工业大数据处理的精度及实时性要求。实验结果表明, 该方法可有效对数据流进行实时的特征提取, 并达到一定的精度要求。

关键词: 工业大数据; 实时性与鲁棒性; 滑动窗口; 主成分分析; 离群点检测; 特征提取

中图分类号: TP391 **文献标识码:** A **文章编号:** 1001-2400(2016)05-0070-05

Real time feature extraction method for complex industrial big data

KONG Xianguang, ZHANG Xiong, MA Hongbo, CHANG Jiantao, NIU Meng

(Industrial Big Data Technology Research Center, Xidian Univ., Xi'an 710071, China)

Abstract: Industrial big data have the traits of big volume, multi-sources, continuous sampling and low value density, which results in high complexity, real-time and high abnormality. Traditional feature extraction methods cannot meet the real-time requirements of complex industrial big data. In addition, the processing method for industrial big data is different from the internet data stream processing method, which has a higher accuracy requirement. Therefore, this paper proposes a robust incremental on-line feature extraction method as the Robust Incremental Principal Component Analysis. It uses the sliding window to update new coming data dynamically and filter the abnormal data in windows, then the incremental principal component analysis is implemented on data in windows in order to meet the accuracy and real-time requirements of industrial big data processing. Experimental results show that the proposed method can effectively extract the data stream in real time with high accuracy.

Key Words: industrial big data; real-time and robustness; sliding window; principal component analysis; outlier detection; feature extraction

近年来,德国工业 4.0 及中国制造 2025 的提出,使得工业大数据的应用走向了发展的舞台^[1]。工业大数据分析是智能制造的基础,也是支撑未来制造智能化的方向。工业大数据是一个新的概念,泛指工业领域的大数据,包括企业内部制造系统数据和企业外部的大数据,具有大体量、多源性、连续采样、价值密度低、复杂度高和动态性强等特点。工业大数据的这些特点使其不同于互联网等数据流,分析难度与对分析精度的要求相对较高。多源性及复杂性导致了工业大数据的高维性,为准确进行工业大数据分析与预测,需要对工业大数据降维。特征提取是一种常用的特征降维方法,其原理是利用原始的特征空间构建一个新的低维空间,消除冗余特征及不相关特征,有效降低数据的维度。由于工业大数据对处理分析的实时性要求很高,传统的特征提取方法如主

收稿日期:2015-08-13

网络出版时间:2015-12-10

基金项目:中央高校基本科研业务费大数据群资助项目(BDY231423);国家自然科学基金资助项目(51505357);陕西省国际科技合作与交流计划资助项目(2016-KW-048)

作者简介:孔宪光(1974—),男,副教授,博士,E-mail: kongxg@vip.sina.com.

网络出版地址: <http://www.cnki.net/kcms/detail/61.1076.TN.20151210.1529.026.html>

成分分析(Principal Component Analysis, PCA)^[2]、线性判别分析(Linear Discriminant Analysis, LDA)和偏最小二乘法(Partial Least Squares, PLS)^[3]等均不能满足工业大数据的实时性及要求。

数据流实时性问题的通常解决方法是以传统特征提取为基础,对其进行快速更新:①在主成分分析方面,文献[4]提出了一种分块增量核主成分分析算法用来降低核主成分分析在特征分解上的计算代价,该算法基于累积率增加特征空间,可不保留历史数据而更新特征空间,并针对每一个数据块只计算1次特征值来求解特征空间。文献[5]提出了一种可以增量的更新数据特征的主成分分析法(Candid Covariance-free Incremental PCA, CCIPCA),该算法无需计算协方差矩阵,不仅提高了计算速度,而且面对高维数据时可快速收敛;②在线性判别分析方面,文献[6]提出了一种构造方法,用于更新判别特征空间,包含新类的数据以一种随机块的方式加入到初始判别特征空间中。文献[7]基于最小二乘线性判别式分析,提出了一种新的增量线性判别式分析算法(Least Square-Incremental Linear Discriminant Analysis, LS-ILDA),当新数据到达时,算法增量更新线性判别式分析算法(Linear Discriminant Analysis, LDA)的最小二乘解,具有低时耗性和高准确性。文献[8]提出了一种完备奇异值分解的增量线性判别分析方法,并用于面部识别中。文献[9]提出一种快速增量线性判别式分析算法,可很好地处理新样本的更新问题,具备较低的空间和时间复杂度;③在偏最小二乘法方面,文献[10]提出了一种增量偏最小二乘法用于处理大规模数据流。

上述增量特征提取方法很好地解决了数据流的实时性问题,却存在一定的精度问题。针对工业大数据应用对数据分析结果的高精度要求,笔者在主成分分析的基础上结合异常点检测,提出了一种鲁棒增量主成分分析(Robust Incremental Principal Component Analysis, RIPCA)方法,该方法在对数据流进行增量特征提取之前,对数据进行异常点的检测及过滤。文中创新点在于提出了一种实时的鲁棒增量主成分分析方法,它不仅可实时对数据进行降维,而且可提高降维的精度。

1 主成分分析及可增量更新数据特征的主成分分析法的原理及鲁棒性

特征提取是模式识别及机器学习中很重要的一个部分。目前,线性特征提取方法已被广泛用于文本分类和面部识别中^[11]。在工业大数据领域,线性特征提取主要应用于故障检测、故障诊断及质量控制中。其中,PCA与LDA是两种最广泛的线性子空间学习方法,而线性判别分析主要关注点在小样本问题^[12],即当处理的数据集比较大时,类内散度矩阵会出现奇异值。因此,文中主要探讨主成分分析。

1.1 主成分分析

主成分分析(PCA)是一种最流行的特征提取方法,也是一种无监督的技术。PCA通常是找到一系列可最大化所有数据方差的正交基向量。假设数据样本点 $x(1), x(2), \dots, x(n)$ 是一个 p 维中心化向量。PCA的目的是就找到一个子空间,子空间中的基准向量对应着最大方差的方向, $C = (1/n) \sum_{i=1}^n x(i)x(i)^T$,为样本的协方差矩阵。目标函数可定义为

$$J(W) = \text{tr} W^T C W, \quad (1)$$

其中, W 为最佳投影矩阵。主成分分析的目标便为在解空间 $H^{p \times k} = \{W \in \mathbb{R}^{p \times k}, W^T W = I\}$ 中最大化目标函数 $J(W)$ 。已经证明 W 的列向量为协方差矩阵 C 的 k 个主特征向量。PCA算法将原始数据投影到一个 k 维($k < p$)子空间中。新的低维特征向量可以表示为 $W^T x$ 。

PCA算法虽然可有效降低数据的维度,但仍存在以下几个缺点:

- (1) PCA对数据点中的异常值十分敏感,存在严重的鲁棒性问题。
- (2) PCA中的特征值分解过程计算复杂度,时间消耗大,不利于进行在线分析。
- (3) PCA是一种批处理算法,数据维度过高时,计算复杂度会过高,不能满足很多实际需求的需求。

1.2 可增量更新数据特征主成分分析算法

针对PCA算法不能实时处理数据的问题,文献[5]以PCA为基础,提出了CCIPCA算法。该算法避免了PCA中的协方差矩阵的特征值分解的同时,对新到达的数据进行增量计算,不需要重新扫描所有的数据,因而减少了计算的复杂度,可用于在线数据流降维的处理。

<http://www.xdxb.net>

假设数据流按样本向量 $u(1), u(2), \dots$, 收集, 向量可能无限大. 每个 $u(n), n=1, 2, \dots$, 是一个 d 维向量, 其中, d 可能为 5 000, 甚至更大. 不失一般性, 假设 $u(n)$ 的均值为 0. $A = \{u(n)u^T(n)\}$, 是一个 $d \times d$ 维的协方差矩阵, 采用增量更新的方式计算协方差矩阵, 即

$$v(n) = \frac{1}{n} \sum_{i=1}^n u(i)u^T(i) \frac{v(i-1)}{\|v(i-1)\|}. \quad (2)$$

令 $v(0) = v(1)$, 即数据分布的第 1 个方向. 对于增量估计, 式(2)可写成一种递归的形式为

$$v(n) = \frac{n-1}{n} v(n-1) + \frac{1}{n} u(n)u^T(n) \frac{v(n-1)}{\|v(n-1)\|}, \quad (3)$$

其中, $v = \lambda x$, 为样本协方差矩阵, 特征向量和特征值可分别计算 $x = v/\|v\|$ 和 $\lambda = \|v\|$ 来得到. 上述方法得到的是第 1 阶向量, 第 2 阶向量可通过第 1 阶向量的投影得到, 即

$$u_2(n) = u_1(n) - u_1^T(n) \frac{v_1(n-1)}{\|v_1(n-1)\|} \frac{v_1(n-1)}{\|v_1(n-1)\|}, \quad (4)$$

其中, $u_1(n) = u(n)$. 在完备空间中, $u_2(n)$ 被用作下一个迭代的输入.

CCIPCA 算法避免了协方差矩阵的特征值分解, 极大减少了计算复杂度; 同时, 该算法具备快速收敛的特征, 因而可对数据流进行实时降维. 但同主成分分析一样, 该算法对数据流中的异常点十分敏感, 导致降维精度受异常点影响较大, 而实际工业大数据存在许多的异常点, 因而该算法不适用于处理实际工业大数据问题. 同时, 考虑工业大数据的实时性要求, 需要准确处理动态更新的实时数据流. 因此, 文中提出一种基于滑动窗口的鲁棒增量主成分分析方法.

2 鲁棒增量主成分分析算法

为提高增量主成分分析算法的鲁棒性, 必须去除数据流中离群点对算法造成的影响. 利用滑动窗口获取近期数据, 并对其离群点检测, 去除离群点. 同时, 可去除过期数据对算法造成的影响.

2.1 利用反 k 近邻法判别数据流离群点

数据流异常点检测是数据流挖掘领域中的一个非常重要的工作. 基于反 k 近邻的数据流离群点算法利用滑动窗口更新当前数据, 实时检测数据流的离群点^[13].

假设数据流表示为 $\langle u_1, t_1 \rangle, \langle u_2, t_2 \rangle, \dots$, 其中, $\langle u, t \rangle$ 是数据项的表示, u 为数据流上的样本数据点, t 为该样本点进入或流出滑动窗口的时间. 假设滑动窗口的大小为 w , 在任意时间点 t_n , 滑动窗口模型的查询范围为 $\{u_{\max(0, t_n-w+1)}, \dots, u_n\}$. 时间点 $\max(0, t_n-w+1)$ 之前的数据为过期数据, 可忽略不计. 文献[14]关于离群点定性定义如下.

定义 1 离群点是一个与其他点差别很大的观测值, 以至于怀疑它是由不同的机制产生的.

定义 2 k -distance(p) 为对任意自然数 k , 定义 p 的 k 距离为 p 和某个对象 o 之间的距离 $d(p, o)$, 且 o 满足:

至少存在 k 个对象 $o' \in D/\{p\}$, 使 $d(p, o') \leq d(p, o)$;

至少存在 $k-1$ 个对象 $o' \in D/\{p\}$, 使 $d(p, o') < d(p, o)$.

定义 3 p 的 k 距离邻域 $N_k(p)$ 为给定 p 的 k -distance(p), p 的 k 距离邻域包含所有与 p 距离不超过 k -distance(p) 的对象.

定义 4 p 的反 k 近邻 $N_{RN_k}(p)$ 为把 p 当作 k 近邻的对象的集合. $|N_{RN_k}(p)|$ 表示集合的大小, 即反 k 近邻个数, 且 $N_{RN_k}(p) = \{q \mid q \in D, p \in N_k(q)\}$.

定义 5 基于反 k 近邻的离群点为当前窗口反 k 近邻最小的 m 个对象作为本次查询基于反 k 近邻的离群点.

若 p 的反 k 近邻 $|N_{RN_k}(p)|$ 少, 则认为 p 的偏离程度大, 即离群度大. 因此, 基于 p 的反 k 近邻个数可以判定 p 的离群度.

基于上述定义对 CCIPCA 算法进行修正, 提出新的鲁棒增量分析算法.

<http://www.xdxb.net>

2.2 鲁棒增量主成分分析算法介绍

原始 CCIPCA 算法在所有的数据集上对协方差矩阵进行增量更新,然而设备状态评估更多考虑的是其历史数据,对当前数据考虑不够,不足以很好地表征当前设备状态.因此,文中以周期更新的滑动窗口对 CCIPCA 算法进行修改,更多地考虑近期获取的数据,并对获取的数据进行离群点处理.

假设滑动窗口的大小为 w ,基本窗口大小为 b ,当前滑动窗口内的时间序列数据为 u_1, u_2, \dots, u_w ,对应的协方差矩阵为 v_m^{old} ,当新的基本窗口内 b 个数据到达时,滑动内数据更新为 $x_{b+1}, \dots, x_w, x_{w+1}, \dots, x_{w+b}$,对应的协方差矩阵可表示为

$$v_m^{\text{new}} = v_m^{\text{old}} + \frac{1}{w} \left(\sum_{i=1}^b u(w+i) u^T(w+i) \frac{v(w+i-1)}{\|v(w+i-1)\|} - \sum_{i=1}^b u(i) u^T(i) \frac{v(i-1)}{\|v(i-1)\|} \right) \quad (5)$$

由式(5)可以看出,滑动窗口内更新后的数据的协方差矩阵等于上一个窗口内数据的协方差矩阵加上一个增量,可避免重复扫描数据带来的时间损耗,满足在线实时分析的需求.

基于上述滑动窗口增量更新,结合反 k 近邻离群点检测,提出鲁棒的增量主成分分析算法描述如下:

(1) 初始化滑动窗口宽度 w ,用户指定窗口内离群点数 m ,整数 k ,主特征向量数 p .

(2) 扫描一遍当前窗口内数据,更新当前窗口受影响的 k 近邻表与反 k 近邻表,删除 $|N_{\text{RN}_k}(p)|$ 最小的 m 个对象.

(3) 对删除 $|N_{\text{RN}_k}(p)|$ 最小的 m 个对象后的滑动窗口内数据利用式(5)计算样本协方差矩阵,并计算相应的特征向量及特征值.

(4) 利用式(4)计算下一阶特征向量,直到特征向量数为 p .

2.3 时间与空间复杂度分析

当前窗口大小计为 b ,当前窗口内一个 k 近邻对就着一个反 k 近邻,协方差更新与数据量大小线性相关.因此,算法的空间开销为 $O(N(2d+2k))$.

文献[11]证明反 k 近邻的时间复杂度为 $O(b)$,RIPCA 算法的复杂度与数据量和特征数量线性相关,因此,该算法的时间复杂度为 $O(b)$.

3 仿真实验与结果分析

文中对 RIPCA 算法与 CCIPCA 算法在处理速度与处理精度上分别进行仿真实验与比较,测试数据集为某半导体制造数据,该数据集包含了 1567 行记录,每条记录共 590 个特征,对数据集中的缺失值,用均值进行填充后,利用时间戳,将其仿真成一段实时的数据流,并将其应用于算法比较中.表 1 取其前 10 个记录,前 10 个特征.

表 1 部分实验数据

记录	特征 1	特征 2	特征 3	特征 4	特征 5	特征 6	特征 7	特征 8	特征 9	特征 10
1	3 030.93	2 564.00	2 187.733 30	1 411.126 5	1.360 2	100	97.613 3	0.124 2	1.500 5	0.016 2
2	3 095.78	2 465.14	2 230.422 20	1 463.660 6	0.829 4	100	102.343 3	0.124 7	1.496 6	-0.000 5
3	2 932.16	2 559.94	2 186.411 10	1 698.017 2	1.510 2	100	95.487 8	0.124 1	1.443 6	0.004 1
4	3 032.24	2 502.87	2 233.366 70	1 326.520 0	1.533 4	100	100.396 7	0.123 5	1.530 1	-0.003 1
5	2 946.25	2 432.84	2 233.366 70	1 326.520 0	1.533 4	100	100.396 7	0.123 5	1.528 7	0.016 7
6	330.27	2 430.12	2 230.422 22	1 463.660 6	0.829 4	100	102.343 3	0.124 7	1.581 6	-0.027 0
7	3 058.88	2 690.15	2 248.900 00	1 004.469 2	0.784 4	100	106.240 0	0.118 5	1.515 3	0.015 7
8	2 967.68	2 600.47	2 248.900 00	1 004.469 2	0.788 4	100	106.240 0	0.118 5	1.535 8	0.011 1
9	3 061.11	2 428.37	2 248.900 00	1 004.469 2	0.788 4	100	106.240 0	0.118 5	1.538 1	0.015 9
10	2 994.05	2 548.21	2 195.122 20	1 046.146 8	1.320 4	100	103.340 0	0.122 3	1.514 4	-0.019 0

文中算法利用 python 实现,窗口大小设为 500, $k=15$.对两个算法分别重复 10 次取平均值,作为最后的运行结果,首先,对比各算法处理实时数据流的速度.分别将数据集降维成 5~20 个主成分,每个实验重复 10 次,取其平均值作为各算法的处理时间,实验流程如图 1 所示,实验结果如图 2 所示.

由图 2 可以得出,将数据降成维度 15 以内时,CCIPCA 算法的处理速度要高于 RIPCA 算法的.将数据降成

<http://www.xdxb.net>

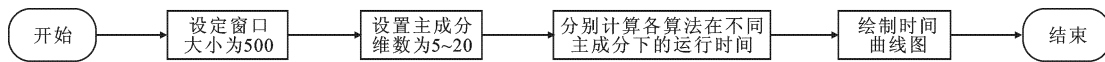


图 1 算法运行时间对比实验流程示意图

维度 15 后,鲁棒增量主成分分析的处理速度变快. 两者的处理速度在 10 ms 左右,基本满足数据流的处理要求.

在数据集中添加 10% 的异常点以验证算法的鲁棒性,对数据分别进行降维处理,算法流程如图 3 所示.

取 $k=5$, 窗口宽度(即窗口中样本点个数)对异常点的检测率的影响如图 4 所示. 可以看出,随着窗口宽度的增加,算法中异常检测部分对异常点的检测效果不断提升,因此,可有效剔除异常点对主成分分析造成的影响.

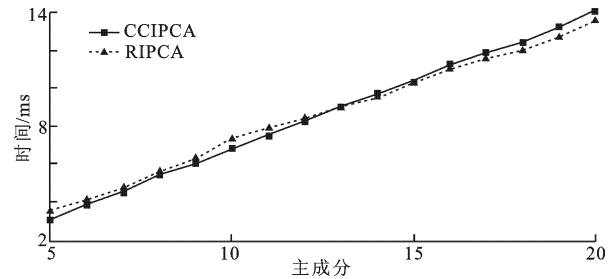


图 2 算法运行时间对比示意图

累积贡献率 $G(m) = \sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i$, 其中, λ_i 为第 i 个

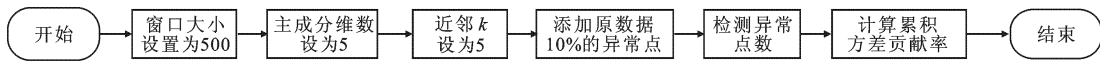


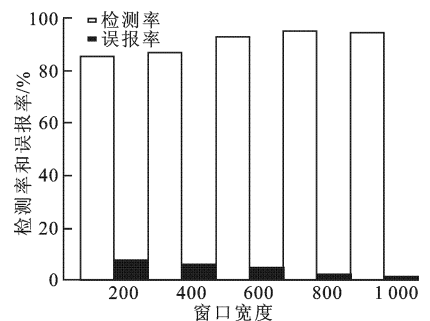
图 3 鲁棒性对比分析流程示意图

特征值, m 表示取前 m 个主成分, p 为所有主成分的个数. 累积贡献率能反映原来变量的信息, 其值越大表示主成分越能反映原变量的信息. 取 $k=5$, RIPCA 算法与 CCIPCA 算法在噪声点下的累积方差对比如表 2 所示.

表 2 累积方差对比

算法	CCIPCA	RIPCA
累积方差百分比/%	81.7	93.2

可见, 对数据进行异常处理后, 在提取相同的主成分前提下, 鲁棒增量主成分分析 RIPCA 算法可获得更高的累积方差贡献率, 更能反映原变量的信息, 因而, 提取的主成分更能代表原数据集特征.

图 4 $k=5$ 时, 异常检测结果示意图

4 结 束 语

简要介绍了一种常用线性特征提取方法即主成分分析, 指出了该方法的计算复杂度与鲁棒性问题. CCIPCA 算法可解决数据流的实时性问题, 但存在一定的噪声敏感性, 不适用于实际数据流的处理. 因此, 文中以 CCIPCA 算法为基础, 采用周期更新的滑动窗口, 对当前窗口数据实现增量更新方式, 并对当前窗口数据进行离群点检测去除离群点, 减少离群点对增量主成分分析的影响. 实验结果表明, 该算法在满足数据流实时性的同时, 可有效对数据进行降维处理, 并保证一定的精度.

参考文献:

- [1] 彭训文. 工业 4.0: 从自动生产到智能制造[J]. 大飞机, 2014(7): 40-42.
PENG Xunwen. Industrie 4.0: From Automatic Production to Intelligent Manufacturing[J]. Jetliner, 2014(7): 40-42.
- [2] 王晓慧. 线性判别分析与主成分分析及其相关研究评述[J]. 中山大学研究生学刊(自然科学、医学版), 2007, 28(4): 50-61.
WANG Xiaohui. A Summary of LDA, PCA and Relative Work[J]. Journal of the Graduates Sun Yat-Sen University (Natural Sciences, Medicine), 2007, 28(4): 50-61.

(下转第 152 页)

<http://www.xdxn.net>

Model [J]. Journal of Xidian University, 2013, 40(4): 90-95.

- [4] 苏哲, 许录平, 王婷, 等. 一种新的脉冲星累积脉冲轮廓时间延迟测量算法[J]. 宇航学报, 2011, 32(6): 1256-1261.
SU Zhe, XU Luping, WANG Ting, et al. A New Time Delay Measurement Algorithm for Pulsar Accumulated Pulse Profile [J]. Journal of Astronautics, 2011, 32(6): 1256-1261.
- [5] EMADZADEH A A, GOLSHAN A R, SPEYER J L. Consistent Estimation of Pulse Delay for X-ray Pulsar Based Relative Navigation [C]//Joint 48th IEEE Conference on Decision and Control, and 28th Chinese Control Conference. Piscataway: IEEE, 2009: 1488-1493.
- [6] EMADZADEH A A, SPEYER J L. On Modeling and Pulse Estimation of X-ray Pulsars [J]. IEEE Transactions on Signal Processing, 2010, 58(9): 4484-4495.
- [7] RINAURO S, COLONNESE S, SCARANO G. Fast Near-maximum Likelihood Phase Estimation of X-ray Pulsars [J]. Signal Processing, 2013, 93(1): 326-331.

(编辑: 齐淑娟)

(上接第 74 页)

- [3] ROSIPAL R, KRÄMER N. Overview and Recent Advances in Partial Least Squares [C]//Lecture Notes in Computer Science: 3940. Berlin: Springer, 2006: 34-51.
- [4] JOSEPH A A, TOKUMOTO T, AZAWA S. Online Feature Extraction Based on Accelerated Kernel Principal Component Analysis for Data Stream [J]. Evolving Systems, 2015: 1-13.
- [5] WENG J Y, ZHANG Y L, HWANG W S. Candid Covariance-free Incremental Principal Component Analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8): 1034-1040.
- [6] PANG S N, OZAWA S, KASABOV N. Incremental Linear Discriminant Analysis for Classification of Data Streams [J]. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 2005, 30(5): 905-914.
- [7] LIU L P, JIANG Y, ZHOU Z H. Least Square Incremental Linear Discriminant Analysis [C]//IEEE International Conference on Data Mining. Piscataway: IEEE, 2009: 298-306.
- [8] LU G F, ZOU J, WANG Y. Incremental Learning of Complete Linear Discriminant Analysis for Face Recognition [J]. Knowledge-Based Systems, 2012, 31: 19-27.
- [9] CHU D, LIAO L Z, et al. Incremental Linear Discriminant Analysis: A Fast Algorithm and Comparisons [J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(11): 2716-2735.
- [10] ZENG X Q, LI Z G. Incremental Partial Least Squares Analysis of Big Streaming Data [J]. Pattern Recognition, 2014, 47(11): 3726-3735.
- [11] 温浩, 卢朝阳, 高鑫学. 融合小波变换和张量 PCA 的人脸识别算法 [J]. 西安电子科技大学学报, 2009, 36(4): 602-607.
WEN Hao, LU Zhaoyang, GAO Xinxue. Face Recognition Algorithm Based on Wavelet Preprocessing and Tensor PCA [J]. Journal of Xidian University, 2009, 36(4): 602-607.
- [12] SHARMA A K, PALIWAL K. Linear Discriminant Analysis for the Small Sample Size Problem: an Overview [J]. International Journal of Machine Learning and Cybernetics, 2013, 6(3): 443-454.
- [13] 张忠平, 梁永欣. 基于反 k 近邻的流数据离群点挖掘算法 [J]. 计算机工程, 2009, 35(12): 11-13.
ZHANG Zhongping, LIANG Yongxin. Stream Data Outlier Mining Algorithm Based on Reverse k Nearest Neighbors [J]. Computer Engineering, 2009, 35(12): 11-13.
- [14] HAWKINS D. Identification of Outliers [M]. London: Chapman and Hal, 1980: 1-45.

(编辑: 齐淑娟)