

一种使用 shapelets 的增量式时间序列分类

丁 剑¹ 王树英²

(北方民族大学计算机科学与工程学院 银川 750021)¹

(北京交通大学计算机与信息技术学院 北京 100044)²

摘 要 根据时间序列数据维度高、实值有序、数据间存在自相关性等特点,对时间序列分类过程进行研究。研究了当前比较流行的时间序列分类方法;从图像处理的角度出发,提出了一种将图片信息转化为时间序列数据的 ITTS 方法。shapelets 作为最能够表示一条时间序列的子序列,随着时间的推移,这个特征序列可能会动态地发生变化。基于这样的思想,提出了一种基于动态发现 shapelets 的增量式时间序列分类算法 IPST。该算法能够较好地动态发现当前最优的 k 个 shapelets,从而提高时间序列分类的准确度。得到的 shapelets 集合还可以与多个传统的分类器结合,从而获得更佳分类效果。

关键词 时间序列,分类,shapelets,图像转化,增量式学习

中图法分类号 TP311 文献标识码 A DOI 10.11896/j.issn.1002-137X.2016.5.048

Incremental Time Series Classification Algorithm Based on Shapelets

DING Jian¹ WANG Shu-ying²

(School of Computer Science and Engineering, Beifang University of Nationalities, Yinchuan 750021, China)¹

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)²

Abstract This paper focused on research of time series classification according to time series features of high dimensionality, ordered real-valued variables, autocorrelation and so on. From the perspective of image processing, this paper proposed a method for ITTS to transform image information to time series. As the best approach to show a subsequence of time series, shapelets would change dynamically as time goes on. Considering this thought, this paper presented time series classification algorithm based on dynamically finding shapelets which is named IPST algorithm. IPST algorithm dynamically discovers current optimal k shapelets well, so as to improve the accuracy of time series classification. The discovered shapelets can be also used with state-of-art classification algorithms, leading to better performance.

Keywords Time series, Classification, Shapelets, Image processing, Incremental learning

1 概述

时间序列是一组序列数据,通常是在相等间隔的时间段内依照给定的采样率对某种潜在过程进行观测的结果^[1]。时间序列分类就是试图给采样结果集中的每一个序列分配标签。时间序列分类问题作为序列分类问题的一个分支^[2],已经在时间序列挖掘领域引起了广泛关注。该问题存在于现实生活中的诸多领域,如在气象中预测天气状况^[3]、健康信息处理中分类心电图或脑电图^[4]、依据用电量来分类家用电器^[5]、依据传感数据来分类行为动作^[6,7]等。

最早的时间序列分类方法^[8]是基于简单趋势的,因此很难解释它的结果。继而研究者又提出了分段表示法,它有一定的抗噪能力,并且权重可以应用于相关的反馈结构中;不足之处是它不能很好地处理异常值。为了克服高维的障碍, Jeng 在 2007 年^[9]试图利用奇异值分解法去选择基本频率数据,然而这就意味着更高的计算花费。在近期的工作中,研究

者比较了 3 个典型的分类器:最近邻、支持向量机和决策森林。尽管高度依赖于已有数据集,但目前看来这 3 种方式都是有效的。带有 DTW 的 1-NN 分类算法^[10]具有更加广泛的应用,尽管它的计算速度受到反复 DTW 计算影响,但是它依然显示了更高的精确度。为了克服这一局限性,研究者提出了基于精确形状平均(Accurate Shape Averaging, ASA)技术的模型构造算法,即每一个训练类仅用一个序列来表示,这样做使得收到的序列仅与每个类的一个均值模型进行比较。除此之外,研究的技术还包括 ARIMA^[11]或 HMM^[12]模型。很多提议方法是基于不同距离度量方式的最近邻(1-NN)算法,如基于 ED 或者 DTW 的 1-NN 等^[13,14]。另外, Buza 等人^[15]还介绍了一种融合不同距离度量方式的 1-NN 分类器。

前面介绍的时间序列分类器在分类时间序列时,倾向于从整个时间序列入手。然而,在许多时间序列数据集中,同一类别的不同时间序列间往往存在着较大的差异性,基于时域相似性的分类器并不能很好地处理这种差异性。而基于形状

到稿日期:2015-05-06 返修日期:2015-07-15 本文受国家自然科学基金项目(61563001),国家民委科研基金项目(14BFZ008)资助。

丁 剑(1977—),男,硕士,副教授,主要研究方向为计算机应用、数据挖掘, E-mail: ychdj@163.com; 王树英(1988—),女,硕士,主要研究方向为数据挖掘。

相似性的分类算法能够很好地发现区分不同类别的最佳特征^[1]。时间序列 shapelets 是序列中最具辨别性的子序列。Ye^[16,17]最早提出基于 shapelets 的分类算法。该算法采用信息增益度量数据的分裂点,并通过递归搜索最具有辨别性的 shapelets 来构建决策树。Mueen 等人^[7]提出了使用逻辑 shapelets 来构建决策树的思想,并采用加速技术和剪枝策略来提升 Ye 所提算法的可解释性和运行效率。Rakthanmanon^[4]等人针对原有的 shapelets 发现算法在时间效率上的不足,提出了一种基于 SAX^[18]表示的快速 shapelets 发现算法,此算法在提升发现 shapelets 速率的同时保证了一定的分类准确性。

2 预备知识

本节介绍由图像转化为时间序列数据的过程,包括时间序列图像的获取过程、图像预处理过程以及图像转化为时间序列的过程。首先,通过数码相机进行植物叶片或植物整体形状的采集,获得本文所需要的图像。其次,图像预处理^[19]的过程主要是为了提高图像的信噪比,减少图像处理的数据量,并对图形进行边缘检测。最后,将获取到的二值化图像转化为时间序列。

2.1 图像预处理

图形的预处理过程主要包括图像的灰度化、图像的二值化、图形的去噪处理、图形的边缘提取。文中主要采用 Matlab 对图像进行处理。

(1) 图像灰度化

通过采集模块得到的图形数据量非常大,直接对其提取特征参数会影响程序的运行速度,所以首先对其进行灰度变化处理,将其转化为 256 级灰度图像。在 Matlab 中使用 `rgb2gray()` 函数将图像进行灰度化处理,这个函数的输入为 RGB 三维图形,输出为 0 到 255 的 256 级的灰度图像矩阵。另外,图像灰度化处理的函数如式(1)所示,其中, (x, y) 为图像中的坐标点,函数为相应坐标点的值。

$$\text{Gray}(x, y) = a \times R(x, y) + b \times G(x, y) + c \times B(x, y) \quad (1)$$

(2) 图像的去噪处理

图像在采集过程中会受到各种因素的干扰,如光照变化、图像压缩等,这些噪声会使图像的质量下降,给图形的分割、特征提取、模式识别等工作带来诸多不利。因此,在特征提取之前本文对图形进行去噪处理。

传统的去噪方法有邻域平均法、中值滤波法、多图形平均法、高斯平滑滤波、空间域低通滤波和频率域低通滤波等,本文采用中值滤波法进行去噪处理。在 Matlab 中使用 `medfilt2()` 函数对图像进行中值滤波去噪处理。该方法的输入为灰度化图像矩阵,输出为经过滤波处理的灰度化图像矩阵。

(3) 图形二值化

一幅图像包括目标物体、背景和噪声,从多值的数字图像中只取出目标物体,最常用的方法就是设定某一阈值 X ,用 X 将图像的数据分成两大部分:大于 X 的像素群和小于 X 的像素群。很明显,这种二值化方法的关键是 X 值的选取。在 Matlab 中使用 `im2bw(RGB, level)` 函数进行图像二值化处理。这个函数的输入为 0 到 255 的灰度图像矩阵,输出为 0、1 值的二值化矩阵。另外,二值化函数可简记为:

$$\text{当 } G(x, y) < X \text{ 时, } B(x, y) = 1;$$

$$\text{当 } G(x, y) \geq X \text{ 时, } B(x, y) = 0。$$

(4) 图像的边缘提取

在对图像进行去噪和二值化处理之后,还需要对图形内的物体进行边缘提取和轮廓跟踪。这是因为图像的边缘特性是时间序列分类所需要的基本特征。常用的边缘提取方法有 Canny 算子提取方法、Sobel 算子提取方法、Roberts 算子提取方法和 Prewitt 算子提取方法。在 Matlab 中使用 `edge()` 函数进行边缘提取,该方法的输入为需要进行边缘检测的二值化图像矩阵和算子提取方法,输出为经过边缘提取的二值化图像矩阵。另外,本文也使用八邻域法对图形的边缘进行提取,在 Matlab 中使用 `bwlabel()` 函数对矩阵中的连通域进行标号。

2.2 图像转化为时间序列

图像转化为时间序列的过程主要包括质心的求取以及时间序列转化的过程。

(1) 图像质心的求取

图像质心的求取即求取出形状的中心点位置。质心求取的目的是计算质心到边缘点的距离,而后用得到的距离作为时间序列的属性值。由上面的八邻域法获取到的边缘标号矩阵,本文简单地把横坐标和纵坐标的均值作为该图像区域的质心。在 Matlab 中,利用坐标均值计算得到质心点并且进行了标注。

(2) 图像转化为时间序列

由上述步骤得到图像的边缘和质心之后,首先设置要取得的采样点的个数 $nsamp$ 。在采样之前,本文首先确定采样的起始点以及采取下一个样本点的方向;然后在图像的边缘点上均匀地取 $nsamp$ 个点,分别计算 $nsamp$ 个样本点与质心之间的距离,并将其作为时间序列的属性值。这样,每一幅图像都将被转化成一条维度为 n 的时间序列,并且将与之对应的类标进行标注。以此类推,最终得到转化为时间序列的训练集合测试集数据。

以上就是一幅图像信息转化为时间序列的全部过程,这个过程称为 ITTS 方法,具体的算法流程包括 5 步:

Step 1 图像预处理阶段,即图像灰度化、去噪处理、二值化、边缘提取;

Step 2 求取质心点;

Step 3 获取边缘点;

Step 4 求得质心到边缘点的距离;

Step 5 形成时间序列。

3 IPST 算法

在现实生活中可以发现,辨别一种类别有时不能单纯地依靠一种特征,而是需要通过几种特征来辨别。由此,一个时间序列数据集中也可能有多个具有辨别性的特征。基于此目的,将 shapelet 的发现过程与分类过程分离。本文目标是找到 k 个具有辨别性的子序列,即 k 个 shapelets,用 k 个 shapelets 的特征与分类器结合进行分类。使用 shapelet 进行分类有很多优点。首先,shapelet 是最具有辨别性的子序列,将使得分类具有可解释性。其次,shapelet 分类器比其他分类器更简洁,分类速度就会很快;最后,shapelet 只是进行阶段独立的基于形状的相似性探索,从而它的分类不需要进行整个序列的扫描。

3.1 Shapelets 转化算法 PST

为了解决递归处理的过程,本文提出了一种 shapelets 转化方法即 PST,用于存储 k 个 shapelets。相应的实现过程共分为 3 个阶段,过程如算法 1 Algorithm PST 所示。首先,单次扫描数据集,产生 k 个最好的 shapelets。需要注意的是,尽管 k 是需要设置的参数值,但是 k 也存在需要设定的最大上限值,并且对每一个 shapelets 的性能没有影响。其次,采用交叉验证的方式评估最终的转化数据集中使用的 shapelets 的合适数目。最后,计算 shapelets 与数据集中时间序列的距离,就得到了新的 shapelets 转化数据集。

算法 1 Algorithm PST

输入:数据集 D , shapelet 候选集的最大值 $MAXLEN$,最小值 $MINLEN$

输出:shapelets 转化后的二维表

//生成 k 个最好的 shapelets

GenerateKshapelets(D , $MAXLEN$, $MINLEN$) {

1. While(D 中的每一条时间序列) {

2. //生成候选集

$W = \text{GenerateCandidates}()$;

3. While(W 中的每一条候选子序列 S) {

4. //计算 S 到 W 中序列的距离

$\text{Distance} = \text{findDistances}(S, W)$;

5. //评价候选集

$\text{Quality} = \text{assessCandidate}(S, \text{Distance})$;

6. shapelets.add(S , Quality);

}

}

7. //将产生 shapelets 排序

$\text{sortByQuality}(\text{shapelets})$;

8. //移除相似的 shapelets

$\text{removeSelfSimilar}(\text{shapelets})$;

9. 得到最好的 k 个 shapelets,这里采用的评价方式为 F 分布

$k\text{Shapelets} = \text{merge}(k, k\text{Shapelets}, \text{shapelets})$;

}

// 数据转化过程

DataTransformation($k\text{Shapelets}$, D) {

1. While(D 中每一条序列 T) {

2. While($k\text{Shapelets}$ 中每一条 shapelet) {

3. $\text{Dist} = \text{subsequenceDist}(\text{shapelet}, T)$;

}

}

4. // 返回 k 个 shapelet 到 D 中 n 个时间序列的距离

 return Dist ;

}

上述过程中涉及到 k 的大小设定问题,如果 k 值设置得过小,可能会造成没有足够的信息进行分类决策;如果 k 值设置得过大,可能在转化后的数据集上造成过度分类问题,甚至减小了重要 shapelet 的影响力。为此,本文采用了两种方式来处理这个问题。一种方式是使用 $m/2$ 个 shapelets 进行数据转化,其中 m 为数据集中一条数据的长度;另一种方式是使用五重交叉验证自动选择 shapelets 的数目。

根据 PST 算法所示的产生 shapelets 的过程,以 k 个 shapelets 为特征,以 k 个 shapelets 到每条时间序列的距离为属性值,产生一个新的分类数据集,如表 1 所列。用新产生的数据集进行分类时,本文希望在保持原有方法可解释性的同

时,能够减小时间消耗,提高分类精确度。

表 1 由 k 个 shapelets 转化后的新的数据集

	shapelet ₁	shapelet ₂	...	shapelet _k
T_1	X_{11}	X_{12}	...	X_{1k}
T_2	X_{21}	X_{22}	...	X_{2k}
...	...	...	...	...
T_n	X_{n1}	X_{n2}	...	X_{nk}

本文转化这个数据集的目的主要是使产生 shapelet 的过程与构建分类器的过程分离,使得转化后的 shapelets 数据集能够与其他分类器进行结合。正如表 1 所列,这是一个典型的二维表数据集,能够与贝叶斯网络、C4.5、1-NN、SVM 等分类器结合进行时序分类。

3.2 增量式 shapelets 发现算法 IPST

本文提出了一种基于增量式的时间序列分类算法 IPST,实现了动态产生 shapelets 的过程,预计这样的动态生成过程能够使时间序列分类的精确更好。

如上节所述,经过 shapelets 转化后,能够产生一个如表 1 所列的二维数据集。特征提取后的数据集就可以作为一般的数据挖掘任务进行处理,并且使用常见的数据挖掘算法产生研究人员所需要的学习过程。然而,在日常生活中有很多信息都不是静态不变的,它会随着时间的推移持续发生变化。对于这种动态变化的信息,如何进行时间序列分类算法处理是本文研究的重点内容。

Shapelets 在现实生活中可以作为一种模式,正如叶子的茎、股票的变化规律等。那么可以看出,shapelets 随着实例的增加是在动态变化的。本文从动态发现 shapelets 的过程出发,进行时间序列分类的增量式研究。详细算法过程如算法 2 Algorithm ISPT 所示,在这里提出了一种滑动窗口机制,每次处理的时间序列数据为窗口的大小。

算法 2 Algorithm ISPT(动态发现 shapelets 的过程)

输入:时间序列数据

输出: k 个 shapelets

1. 初始化。根据 shapelets 的 k 值初始化滑动窗口,记录 k 个 shapelets 信息的结构。

2. 判断滑动窗口中的实例个数。小于 k 值时,继续滑动;等于 k 值时,运行 GenerateKshapelets(D , $MAXLEN$, $MINLEN$) 方法找到 k 个最好的 shapelets,并将其离散化,并与常见分类模型结合生成分类器。

3. 继续滑动窗口到下一条实例,判断实例分类结果是否正确。如果正确,继续滑动到下一条;如果不正确,对当前实例生成候选集 D ,更新 GenerateKshapelets(D , $MAXLEN$, $MINLEN$) 方法。

4. 产生分类精确度最好的 k 个 shapelets。

具体而言,第 1 步设置滑动窗口的大小为 k ,是为了与发现 shapelets 的数目相统一。本文使用了方差分布中的 F 分布评价产生的 shapelets,主要原因是:

(1) IPST 算法试图生成 shapelets 的集合,并不是要分裂这个数据集;

(2) 算法关注的是类值分布的变化情况;

(3) 这里还考虑了处理多类标的分类问题。

本文提出的基于 shapelets 的增量式时序分类方法具有一些优缺点。它采用动态生成 shapelets 的方法,更利于找到最优的便于分类的子序列;且 shapelets 本身就有一定的可解释,但是这个可解释的过程并不能直观地感受到。

4 实验分析

4.1 实验数据

本文所涉及到的时间序列分类算法在 Weka¹⁾ 平台上使用 Java 语言实现。为了评价 shapelet 转化的性能,本文选择了 UCR²⁾ 时间序列数据库中 14 个数据集,它们分别是 Adiac、Beef、ChlorineConcentration、Coffee、DiatomSizeReduction、ECG FiveDays、FaceFour、GunPoint、ItalyPowerDemand、Lightning-7、MedicalImages、SonyAIBORobotSurface、Symbols、SyntheticControl、Trace。本文之所以选择这些特殊数据集,主要是因为这些数据集分为测试集和训练集。因此,所有的 shapelet 选择、模型选择和分类器构建过程都是在训练集上进行的,测试集仅仅用于评价最终训练好的分类器。

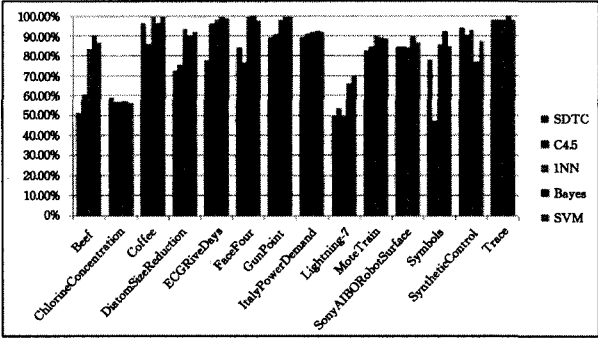


图 1 在 shapelet 转化后的数据集上经典分类器的精确度

4.2 实验分析

首先比较基于 shapelets 的各经典分类器的性能。从图 1 和表 2 可以看出,在 shapelet 转化后的数据集上构建的分类器的精确度与在分类器过程中产生 shapelet 的 SDTC 分类器的精确度非常相近,甚至可能更好。使用贝叶斯分类器进行时间序列的分类具有较好的分类精确度。

表 2 结合 shapelets 与数据挖掘分类算法得到的正确率

m/2 shapelets	SDTC	C4.5	1NN	Bayes	SVM
Beef	51.32%	60.31%	83.33%	90.12%	86.67%
ChlorineConcentration	58.80%	56.48%	56.93%	57.08%	56.15%
Coffee	96.43%	85.71%	100.00%	96.43%	100.00%
DiatomSizeReduction	72.22%	75.16%	93.46%	90.02%	92.16%
ECGRiveDays	77.47%	96.17%	98.37%	99.54%	87.95%
FaceFour	84.09%	76.14%	100.00%	100.00%	95.73%
GunPoint	89.33%	90.67%	98.00%	99.33%	98.35%
ItalyPowerDemand	89.21%	90.96%	92.13%	92.42%	92.13%
Lightning-7	49.32%	53.42%	49.32%	65.75%	69.86%
MoteTrain	82.51%	84.42%	90.34%	89.06%	88.66%
SonyAIBORobotSurface	84.53%	84.53%	84.03%	89.68%	86.69%
Symbols	77.99%	47.14%	85.63%	92.26%	84.62%
SyntheticControl	94.33%	90.33%	93.00%	76.67%	87.33%
Trace	98.00%	98.00%	98.00%	100.00%	98.00%

接着比较基于 shapelets 的增量式时间序列分类算法 IPST 与 PST 算法的性能。由表 3 和图 2 可以看出,基于 shapelets 的增量式时间序列分类算法 IPST 的分类精确度比基于 shapelets 的时间序列分类算法 PST 的要高,说明在动态生成 shapelets 的过程中,实时对时间序列数据集的子序列进行更新,能够提高时间序列的分类精确度。

表 3 PST 算法和 IPST 算法的精确度比较

数据集	IPST	PST
CBF	95.36%	86.67%
ChlorineConcentration	85.56%	56.15%
Coffee	95.26%	92.16%
Cricket	91.21%	87.95%
DiatomSizeReduction	94.63%	95.73%
FaceAll	99.54%	98.35%
Gun_Point	95.21%	92.13%
ItalyPowerDemand	85.23%	69.86%
MALLAT	92.14%	88.66%
MedicalImages	93.25%	86.69%
SonyAIBORobotSurface	89.35%	84.62%
Symbols	92.31%	87.33%

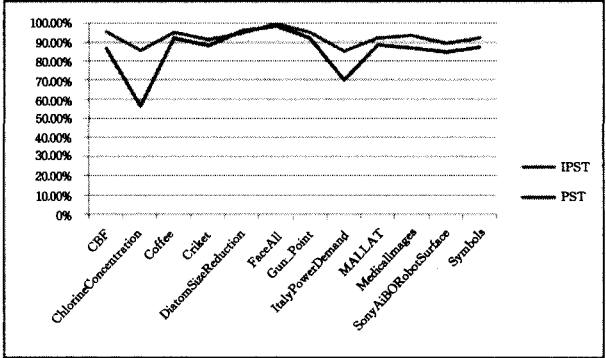


图 2 IPST 算法与 PST 算法比较

结束语 数据挖掘技术的广泛研究,使得复杂数据类型的挖掘得到越来越多研究人员的关注。本文从复杂数据类型中的时间序列数据出发,研究了当前国内外关于时间序列分类的研究现状,提出一种 shapelets 数据预处理转化的 PST 算法,使其能够与其他分类器相结合。根据数据动态变化的特征,提出了基于 shapelets 的增量式时间序列分类算法 IPST。IPST 算法能够动态地发现 shapelets,发现子序列的过程是实时更新的。该方法的缺陷是可解释过程不能直观地感受到,且需要对算法进行进一步的优化,以提高运行效率。

参考文献

[1] Yuan J D, Wang Z H. Review of time series representation and classification technique [J]. Computer Science, 2015, 42(3): 1-7 (in Chinese)
原继东, 王志海. 时间序列的表示与分类算法综述[J]. 计算机科学, 2015, 42(3): 1-7

[2] Xing Z, Pei J, Keogh E J. A brief survey on sequence classification [J]. SIGKDD Explorations, 2010, 12(1): 40-48

[3] McGovern A, Rosendahl D, Brown R, et al. Identifying predictive multi-dimensional time series motifs: an application to severe weather prdition [J]. Data Ming and Knowledge Discovery, 2011, 22: 232-258

[4] Rakthanmanon T, Campana B, Mueen B, et al. Searching and mining trillions of time series subsequences under dynamic time warping[C]//Proceedings of the 18th International Conference on Knowledge Discovery and Data Mining. 2012: 262-270

(下转第 293 页)

¹⁾ <http://www.cs.waikato.ac.nz/ml/weka/>
²⁾ http://www.cs.ucr.edu/~eamonn/time_series_data/

模型的车辆行为识别方法。该方法可以在高速公路监控视频中变道、超车等轨迹数据较少的情况下,通过采用二次谱聚类算法实现对变道、超车、直行等车辆轨迹的有效聚类;此外,该方法还可以将聚类得到的不同类型轨迹数据通过 Baum-Welch 算法来训练与目标轨迹对应的 HMM 模型,并将 HMM 模型的多维概率输出作为随机森林(RF)模型的训练样本输入进行分类识别。实验表明,该 HMM-RF 混合模型方法具有更高的准确率。虽然 HMM-RF 混合模型在一定情况下提高了准确率,但在未来研究工作中可以寻求和引入更多的车辆轨迹特征以进一步提高准确率。另外,谱聚类的计算复杂度为 $O(n^3)$,其中 n 是数据点的个数,其计算过程十分耗时,这限制了它在实际中的应用。我们在未来的研究工作中考虑将 MapReduce 和谱聚类算法结合,以加速谱聚类的计算过程。

参 考 文 献

- [1] Atev S, Miller G, Papanikolopoulos N P. Clustering of vehicle trajectories [J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(3): 647-657
- [2] Morris B T, Trivedi M M. Understanding vehicular traffic behavior from video: a survey of unsupervised approaches [J]. Journal of Electronic Imaging, 2013, 22(4): 6931-6946
- [3] Morris B T, Trivedi M M. Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2287-2301
- [4] Morris B, Trivedi M. Learning trajectory patterns by clustering: Experimental studies and comparative evaluation [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2009 (CVPR 2009). IEEE, 2009: 312-319
- [5] Morris B, Trivedi M. An adaptive scene description for activity analysis in surveillance video [C] // 19th International Conference on Pattern Recognition, 2008 (ICPR 2008). IEEE, 2008: 1-4
- [6] Morris B T, Trivedi M M. Learning and classification of trajectories in dynamic scenes: A general framework for live video analysis [C] // IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance, 2008 (AVSS'08). IEEE, 2008: 154-161
- [7] Hu W, Xiao X, Fu Z, et al. A system for learning statistical motion patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(9): 1450-1464
- [8] Piciarelli C, Micheloni C, Foresti G L. Trajectory-based anomalous event detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(11): 1544-1554
- [9] Breiman L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32
- [10] Zelnik-Manor L, Perona P. Self-tuning spectral clustering [C] // Advances in Neural Information Processing Systems, 2004: 1601-1608
- [11] Chen J, Wang R, Liu L, et al. Clustering of trajectories based on Hausdorff distance [C] // 2011 International Conference on Electronics, Communications and Control (ICECC). IEEE, 2011: 1940-1944
- [12] Hervieu A, Bouthemy P, Le Cadre J P. A HMM-based method for recognizing dynamic video contents from trajectories [C] // IEEE International Conference on Image Processing, 2007 (ICIP 2007). IEEE, 2007, 4: 533-536
- [13] Wang Y, Yang C, Wu X, et al. Kinect based dynamic hand gesture recognition algorithm research [C] // 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). IEEE, 2012: 274-279
- [12] Azzouzi M, Nabney I T. Analyzing time series structure with Hidden Markov Models [C] // Proceedings of the International Conference on Neural Networks and Signal Processing. 1998: 402-40
- [13] Ding H, Trajcevski G, Scheuermann P, et al. Querying and mining of time series data: experimental comparison of representations and distance measures [J]. Proceedings of the International Conference on Very Large Data Bases, 2008, 1(2): 1542-1552
- [14] Keogh E, Kasetty S. On the need for time series data mining benchmarks: a survey and empirical demonstration [J]. Data Mining and Knowledge Discovery, 2003, 7(4): 349-371
- [15] Buza K. Fusion methods for time series classification [D]. Germany: University of Hildesheim, 2011
- [16] Ye L, Keogh E. Time series shapelets: a new primitive for data mining [C] // Proceedings of the 15th International Conference on Knowledge Discovery and Data Mining. 2009: 947-956
- [17] Ye L, Keogh E. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification [J]. Data Mining and Knowledge Discovery, 2011, 22(1/2): 149-182
- [18] Lin J, Keogh E J, Wei L, et al. Experiencing SAX: a novel symbolic representation of time series [J]. Data Mining Knowledge Discovery, 2007, 15(2): 107-144
- [19] Keogh E, Wei L, Xi X, et al. LB-Keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures [C] // Proceedings of the 32nd International Conference on Very Large Data Bases. 2006: 882-893
- [5] Lines J, Bagnall A, Caiger-Smith P, et al. Classification of household devices by electricity usage profiles [C] // Proceedings of the 12th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2011). 2011: 403-412
- [6] Hartmann B, Link N. Gesture recognition with inertial sensors and optimized dtw prototypes [C] // Proceedings of IEEE International Conference on Systems Man and Cybernetics (SMC). 2010: 2102-2109
- [7] Mueen A, Keogh E, Young N. Logical-shapelets: an expressive primitive for time series classification [C] // Proceedings of the 17th International Conference on Knowledge Discovery and Data Mining. 2011: 1154-1162
- [8] Bakshi B R, Stephanopoulos G. Induction of real-time patterns from operating data for diagnosis and supervisory control [J]. Computers & Chemical Engineering, 1994, 18(4): 303-332
- [9] Jeng S L, Huang Y T. Time series classification based on spectral analysis [J]. Communications in Statistics-Simulation and Computation, 2007, 37(1): 132-142
- [10] Chen Y, Hu B, Keogh E, et al. DTW-D: time series semi-supervised learning from a single example [C] // Proceedings of the 19th International Conference on Knowledge Discovery and Data Mining. 2013: 383-391
- [11] Kalpakis K, Gada D, Andputtagunta V. Distance measures for effective clustering of ARIMA time series [C] // Proceedings of the IEEE International Conference on Data Mining. 2001: 273-280

(上接第 260 页)