

数据异常的监测技术综述

吴镜锋 金炜东 唐 鹏

(西南交通大学电气工程学院 成都 610036)

摘要 在目前大数据的环境下,相对于正常数据,异常类数据更难获取,也显得更加重要。异常检测的目的是检测出异于正常主体的活动数据。异常检测适用于机器故障诊断、数据挖掘以及疾病和入侵检测等多个领域。基于目前大量的异常检测方法,主要从异常类数据的有无来阐述,根据这个框架将主要的异常检测方法进行了分类,并评价了这些方法的优劣;最后重点讨论了基于深度学习的大数据异常检测方法,并分别介绍了不同的方法及相关的应用和未来的研究热点。

关键词 异常检测, 监督学习, 无监督学习, 深度学习

中图分类号 TP181 文献标识码 A

Survey on Monitoring Techniques for Data Abnormalities

WU Jing-feng JIN Wei-dong TANG Peng

(School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610036, China)

Abstract In the current environment of large data, anomaly data is more difficult to obtain than normal data, and is more important. The purpose of anomaly detection is to detect activity data from normal subjects. Anomaly detection is widely applied in many fields, such as machine fault detection, data mining and disease detection and intrusion detection. Based on a large number of anomaly detection methods at present, this paper mainly discussed the existence of anomaly data, classified the major anomaly detection methods according to this framework, and put forward the advantages and disadvantages of these methods. Finally, we focused on the large data anomaly detection methods based on the deep learning, and introduced different methods and related applications and future research hotspots respectively.

Keywords Anomaly detection, Supervised learning, Unsupervised learning, Deep learning

异常检测是基于入侵者的活动异常于正常主体的活动这一假设。根据这一理念建立主体正常活动的“活动简档”,将当前主体的活动状况与“活动简档”相比较,当违反其统计规律时,认为该活动可能是“入侵”行为。异常检测的难题在于,如何建立“活动简档”以及如何设计统计算法才能不把正常的操作作为“入侵”或忽略真正的“入侵”行为。

在机器故障诊断、入侵检测以及数据挖掘等领域,一般很容易获得正常类数据,因此这类数据很多;而异常类数据由于获取代价高或获取困难,往往对异常行为了解很少,但异常行为中却蕴含了显著的行为信息,因此,异常检测在这些领域中是非常必要的。异常检测的目的在于检测出不符合期望行为的数据,因此适合应用于故障诊断、入侵检测、欺诈检测以及数据预挖掘等多个领域。

异常检测的对象被称为离群点或野点(outlier)。异常检测也称为偏差检测,因为异常对象的属性值明显偏离期望的或常见的属性值。异常检测也称为例外挖掘(exception mining),因为异常在某种意义上是例外的,主要使用的术语是异常或离群点。

一般来说,对异常检测方法的分类可以根据异常类数据的有无来确定。在异常检测中,往往由于异常类数据的缺乏,

导致不能使用有监督学习的方法来解决,而只能使用正常类这一类数据进行学习,因此通常属于无监督范畴。而要利用有监督学习算法来进行异常检测则有两种思路,一种是人为地产生异常类数据,再采用传统分类的方法对异常类数据进行检测,如SVM(Support Vector Machine)等方法;另一种是利用少量异常类数据,结合一些能够处理两类数据不平衡问题的特殊算法对数据进行异常检测。其异常产生的原因主要如下:1)数据来源于不同的类,某个数据对象可能不同于其他数据对象(即异常),因为它属于一个不同的类型或类。Hawkins的离群点定义如下:离群点是一个观测值,它与其他观测值的差别如此之大,以至于怀疑它是由不同机制产生的。2)自然变异:许多数据集可以用一个统计分布建模,如正态分布建模,其中数据对象的概率随对象到分布中心距离的增加而急剧减少。换言之,大部分数据对象靠近中心,数据对象显著地不同于这个平均对象的似然性很小。3)数据测量和收集误差:数据收集和测量过程中的误差是另一个异常源。剔除这类异常是数据预处理的关注点。

由于异常检测的应用十分广泛^[1-4],本文首先根据对象数据是否包括异常类分为无监督和有监督两类,对主流的基于统计类的基于无监督型的方法进行了总结;而后总结了监督

型的异常检测算法的最新进展,重点探讨了基于深度学习的大数据异常检测方法;最后给出了半监督学习在异常检测中的优势和应用前景。

1 无监督异常检测方法

由于通常在异常检测中异常类数据较少甚至没有,因此不能直接借用监督学习方法,而仅能利用已知的正常类样本进行学习和训练,这些方法均属于无监督异常检测方法。根据所用原理,无监督异常检测方法可分为基于密度估计、基于重构和基于支撑域 3 类^[5]。

1.1 基于密度估计的方法

在无监督异常检测方法中,由于只有目标类样本可供学习,因此最简单直接的方式就是通过参数化或非参数化方法来估计训练样本的密度模型,并设置密度阈值,小于该阈值的即被认为异常。

参数方法是首先构建模型,包括一元高斯分布、多元高斯分布、混合高斯模型,然后通过高斯分布的两个参数设定阈值来进行异常检测。而非参数方法通过对训练集上的概率密度进行建模,根据经验风险设置相应的概率密度阈值,分类时将所有概率密度低于该阈值的测试样本判为异常^[6]。另外,对于其他一些本质是利用 KNN 估计样本点的密度的算法,其构建可达性密度和局部异常因子两个标准量,通过判断测试样本的局部异常度来检测是否属于正常^[7]。

1.2 基于重构的方法

基于重构的方法也可以称为基于模型的方法,其基本思想是假设正常样本数据符合某模型,而将不符合该模型的样本数据判为异常。根据模型的维数和形状,可分为点重构、平面重构和曲面重构 3 种方法。点重构的方法是基于样本充分聚集并且可以由一系列的中心点表示,并设置到这些中心的距离阈值进行异常检测。比如 K-均值、K-中心、LVQ 等算法,它们都是以样本点到最近簇类中心的距离作为重构误差来进行异常检测,都采用“两步走”的方法设计异常检测器:首先进行聚类,然后采用各自的算法将所有训练样本到最近簇类中心的距离(根据算法而定)作为阈值^[5]。但是 K-均值和 K-中心算法对于选择簇类中心非常敏感,一般只能得到局部最优解,并且其中 K 值的选取仍然是一个难点^[8-9]。对于要求处理平面簇形的数据时,平面重构的方法能较好地适应这种场景,代表性的方法是 K-平面聚类,它与 K-均值聚类算法原理类似,采用样本点到所属簇类超平面的平均距离作为重构误差阈值,即构建 K 个带状簇类数据描述来实现异常检测。主分量分析(Principal Component Analysis, PCA)是一种线性降维方法,其主要思想是降低数据集的维数,尽可能保证数据集的信息。基于 PCA 的异常检测有以下几个优点^[10]: 1)该方法不做任何分布的假设;2)PCA 可以使数据降维后得到一个简单的分类器而不丢失有价值的信息;3)检测阶段的用时较少,便于实时检测。主曲线是 PCA 中第一主分量的非线性推广,其根据每个点到主曲线的正交距离即重构误差以及所允许的经验误差的约束设定正交距离的阈值,形成以主曲线为中心的柱形数据描述。

1.3 基于支撑域的方法

基于支撑域的方法假设正常数据的描述形状,在符合给定经验误差的基础上,通过最小化体积来保证给定正常样本

的正确接收率,同时最小化异常样本的误接收率,从而使得分类器的分类性能 ROC 曲线向上移动,提高分类器异常检测的性能^[5]。其中主要包括以下几种方法,在异常检测领域,由于一般仅有一类样本,因此原有的 SVM 不再适用。Scholkopf 提出了 one-class SVM 用于异常检测^[11],它假设原点为唯一异常点,寻求一个超平面以最大间隔将目标类样本和原点分开,并形成半个空间的数据描述。由于半个空间的数据描述可能太松,最小包围球(Minimal Enclosing Ball, MEB)的方法是寻找一个球形描述,其包含所有的正常类训练样本,并最小化超球(R, a)的体积,但是其由于是一种硬间隔的方法,因此易受到野点的影响。而支持向量数据描述(Support Vector Data Description, SVDD)为了避免因为野点带来的危害,采用了软间隔的思想,即不严格要求训练样本到球心 a 的平方距离小于等于 R 平方,而是对大于 R 平方的点进行惩罚。图 1 为 SVDD 算法在二维下的原理图,图中横纵坐标分别代表数据中的两种特征。谭真臻等人^[4]利用这种方法实现了对图像中的航空发动机磨损状态的识别,即通过图像处理技术人工提取发动机磨损图片中的特征,利用 SVDD 异常检测算法构建了便于航空发动机状态识别的系统。第三种方法是条带数据描述 Slab SVM,在 one-class SVM 的基础上,又提出了 Slab SVM,其在求解 one-class SVM 的超平面时,在数据两侧各用一个超平面将目标类样本进行约束,从而使目标类样本落在一个条带中。

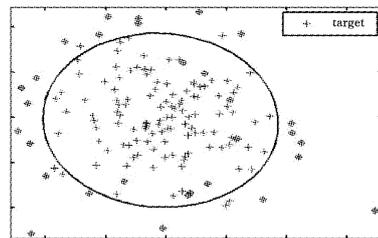


图 1 SVDD 二维示意图

2 有监督异常检测方法

上述方法均是基于无监督学习框架,都是利用已有目标类样本获得样本的数据描述。对监督学习方法的研究已经非常成熟,不少方法都是利用两类分类方法进行异常检测^[5]。其中主要包括两种技术路线:1)人工生成异常类样本,并使用现有的两类分类方法进行异常检测;2)利用现存的少量的异常类样本构成极端不平衡问题后,设计一些能够处理不平衡的算法进行异常检测。

2.1 人工生成异常样本的异常检测方法

其中包括以下 3 种方法:目标类样本的原点生成,均匀分布的异常样本生成,以及边界异常样本的生成。其中均匀分布是将异常检测问题看作一个二类分类问题,其采用目标类样本空间上的均匀分布生成异常类样本,然后用这些异常类样本和已有目标类样本构成 SVM 进行异常检测^[12]。但异常类样本服从目标类样本空间上的均匀分布的假设非常困难且苛刻,未必符合真正的异常类样本分布。边界异常样本的生成通过增加负类样本的方法扩展了无监督异常检测方法中仅有的目标类样本集 X ,进而采用标准的 ν -SVM 训练分类器进行异常检测。这种方法可有效地避免均匀分布方法带来的问题。

2.2 利用现有异常样本的异常检测方法

有时异常检测任务中会存在极少数的异常类样本,此时鉴于现有的两类方法对样本类别分布不平衡的适应性不佳,提出了一些能够较好地适应这种情形的异常检测方法。1)带负类样本的支持向量数据描述(SVDD with negative examples, NSVDD),其中假设负类样本尽量在超球外,正类样本在球体内,训练方式与 SVDD 相同,但其正负类样本间隔为 0,推广能力有限。2)最小包围最大排斥机(Minimum Enclosing Maximum Excluding Machines, MEMEM),它是对 NSVDD 方法的改进,目标是 minimized 正类样本的超球体以及最大化正负样本的间隔,这类方法具有更好的适应性。

3 基于深度学习的大数据异常检测方法

在某些领域中,传统目标检测和异常检测方法难以应对海量的数据^[13-14]。由于传统特征提取困难且不准确,因此需要一种能够自动提取海量特征的方法;通过对大量数据本身的学习,获取其中最有效的特征表征,并通过建立相对复杂的网络结构,充分挖掘数据之间的关联,建立强大的分类器。深度学习就为自动提取目标特征提供了一个有效的框架^[15]。在异常检测中,目标的特征提取是目标检测的关键。在目前大数据的背景下,深度学习即通过构建具有多个隐层的机器学习模型,从大量的训练数据中获得更准确、有用的特征,从而提高分类或异常检测的效率。深度学习可以看成是一种特征学习,动机在于建立和模拟人脑的神经网络,它是模仿人脑机制来进行分析学习的一种方法,在图像数据中学习最能表达目标的特征。本文针对深度学习中的方法,分为基于有监督学习和无监督学习两种原理进行介绍。

3.1 有监督深度学习的异常检测方法

在异常检测中,一个不可避免的问题就是特征提取。由于传统方法的局限性,在大数据形式下,特征提取往往会耗费大量的人力和物力,且很难提取到完整和实用的特征信息;而人工神经网络由于其特殊的结构和方式,能够广泛并准确地提取到数据的特征,这对于实际应用中的异常检测无疑有很大的提高。在基于有监督的深度学习方法中,最具代表性的就是卷积神经网络。

图 2 为卷积神经网络的结果图,它是一个多层的神经网络,每层由多个二维平面组成,而每个平面由多个独立的神经元组成。

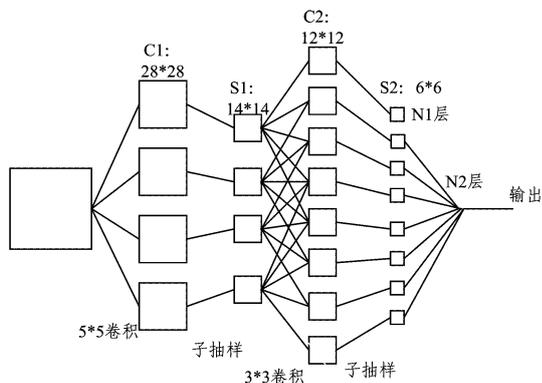


图 2 卷积神经网络结构图

输入图与 3 个可训练的滤波器和可加偏置进行卷积,卷积后在 C1 层产生特征映射图,再进行像素的加权求和等,通

过函数得到 S 层的特征映射图,最终像素连接成向量输入到传统的神经网络得到输出。其中 C 层为特征提取层, S 层为特征映射层。卷积神经网络的训练算法和传统 BP 算法类似,主要包括两个阶段:1)向前传播阶段,从样本集中取一个样本(M, Np),将 M 输入网络并计算相应的实际输出 Op; 2)向后传播阶段,计算实际输出 Op 和理想输出 Np 的差,并按照最小化误差的方法反向传播调整权重矩阵。

卷积神经网络是一种特殊的深层神经网络模型。卷积神经网络对图像进行处理时有以下几个优点^[16]:1)输入图像和网络的拓扑结构能很好地吻合;2)特征提取和模式分类同时进行;3)权重共享可以减少网络的训练参数,使网络结构的适应性更强。目前通过卷积神经网络进行目标检测和视频目标识别的应用非常广泛。郭美银等人利用卷积神经网络对分割后的视频图像进行特征提取,采用多数表决规则实现了对失真视频图像的分类,具有较强的鲁棒性和实用性^[17]。李岳云等人利用深度卷积神经网络对 MSAR 数据库中的图片进行特征学习后,将得到的置信度融入随机场,实现了显著性的检测^[18]。类似地,芮挺等人利用深度卷积神经网络实现了对行人的异常检测。对于利用卷积神经网络等深度学习进行目标异常检测,主要是利用卷积神经网络进行特征提取来避免人工提取特征的缺陷,比如在时间上耗费较多,提取的特征信息不全面等。另外,结合异常检测算法进行异常检测也产生了一些新的方法,比如 CNN+SVDD、CNN+K-均值等。

3.2 无监督深度学习的异常检测方法

机器学习方法可以分为生成方法和判别方法,所对应学到的模型分别是生成式模型和判别式模型,其中生成式模型在无监督深度学习方面占据主要位置。无监督生成模型通过学习真实数据的本质特征,刻画出样本数据的分布特征,生成与训练样本相似的新数据。由于模型的参数远少于训练数据的量,因此模型能够发现并有效地内化数据的本质,并生成这些数据。生成式模型可以用于在没有目标类标签信息的情况下捕捉观测到可见数据的高阶相关性。在传统的深度学习框架中,深度生成模型可以从这些网络中采样来有效生成样本。

3.2.1 生成式对抗网络(GAN)

生成式对抗网络(Generative Adversarial Networks, GAN)是近两年流行的生成式模型,它是由 GoodFellow Ian^[20]于 2014 年基于“零和博弈”的思想提出的。生成式对抗网络包含一个生成模型 G 和一个判别模型 D,生成模型捕捉样本数据的分布,判别模型用于判断输出样本是真实样本的概率。GAN 相当于是一个二分类器,其结构如图 3 所示。

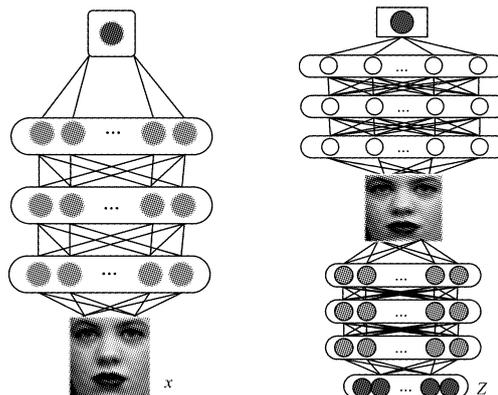


图 3 GAN 的训练和识别

模型的优化过程是一个“二元极小极大博弈”问题,训练时固定一个模型,更新另外一个模型的参数,交替迭代,使得对方的错误最大化,最后 G 能估测出样本数据的分布。相比于其他网络,生成式对抗网络的优势在于不需要构建一个成本函数,它利用深度学习自身的缺陷(普遍网络对于对抗样本具有很高的脆弱性)实现了模型内部的监督学习。对于异常检测,生成式对抗网络通过两种样本之间的对抗学习能够产生更好的样本,提高了网络的稳定性,结合异常检测算法能够达到更好的效果。

随着研究的深入,生成式对抗网络又得到了一些改进,比如在 G 和 D 的建模中加入条件变量 Y 用于指导数据的生成,这些条件变量可以是类别标签,或者修复图像、视频等数据,这就是条件生成式网络(CGAN)。生成式模型的功能之一便是异常检测,生成式模型可以从关联输入移动到输出之外,可以实现半监督分类、数据操作、填空、图像修复、去噪、one-shot 生成等多种应用,最直接的体现之一是 Emily Denton 和 Rob Fergus 等人在文献[21]中制作的合成图像可以非常接近真实图像。

3.2.2 单次学习(One-shot Learning)

在目标识别中,由于庞大的数据需要耗费大量的时间、人力和物力,因此要准确地识别物体而又要高效地利用训练样本是目前机器学习的研究热点。目前,谷歌的 DeepMind 找到了一种新的方法,称为“单次学习”(One-shot Learning),他们通过对深度学习的算法进行一些调整,使之通过一个或几个例子就可以识别出图片中的物体。其中一种方法是通过贝叶斯方法来构建最大似然(ML)和最大后验(MAP)的模型,其利用之前学习类别的先验知识,对象类别由概率模型表示,先验知识表示为模型参数的概率密度函数,对象的后验模型类别通过观测值更新先验获得[22]。图 4 给出文献[22]中的部分实验过程。

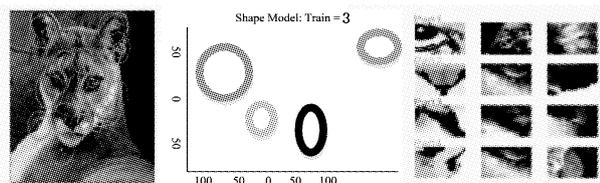


图 4 单次学习实验过程[22]

最后当用该方法来检测含多类目标样本的数据库时,贝叶斯方法产生信息模型,在训练样本很少的情况下,可以达到很好的效果。同样地,单次学习的方法可以用于异常检测,其利用相同类别的先验知识作为概率模型,并设定相应的阈值,当用其检测新的样本时,可以判断是否异常。

3.3 半监督深度学习的异常检测方法

半监督学习是模式识别和机器学习领域研究的重点问题,是监督学习和无监督学习相结合的一种学习方法。它主要考虑如何利用少量的标注样本和大量的未标注样本进行训练和分类。目前,在半监督学习中有两个基本的假设,分别是聚类假设和流形假设。聚类假设是指处在相同聚类(cluster)中的示例有较大的可能拥有相同的标记。根据该假设,决策边界就应该尽量通过数据较为稀疏的地方,从而避免把稠密的聚类中的数据点分到决策边界两侧。在这一假设下,大量

未标记示例的作用就是帮助探明示例空间中数据分布的稠密和稀疏区域,从而指导学习算法对利用有标记示例学习到的决策边界进行调整,使其尽量通过数据分布的稀疏区域。聚类假设简单、直观,常以不同的方式直接用于各种半监督学习算法的设计中。T. Joachims 在文献[23]中提出了 TSVM 算法,在训练过程中,该算法不断修改 SVM 的划分超平面,并交换超平面两侧某些未标记示例的可能标记,使得 SVM 在所有训练数据(包括有标记和未标记示例)上最大化间隔(margin),从而得到一个既通过数据相对稀疏的区域又尽可能正确划分有标记示例的超平面。图 5 为直推式 SVM 和 TSVM 的比较简图。

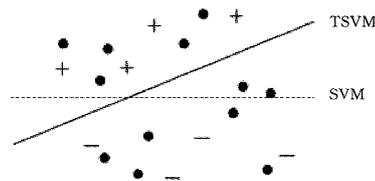


图 5 直推式 SVM[23]

流形假设是指处于一个很小的局部邻域内的示例具有相似的性质,因此,其标记也应该相似。这一假设反映了决策函数的局部平滑性。与聚类假设着眼于整体特性不同,流形假设主要考虑模型的局部特性。在该假设下,大量未标记示例的作用就是让数据空间变得更加稠密,从而有助于更加准确地刻画局部区域的特性,使得决策函数能够更好地进行数据拟合。流形假设也可以容易地直接用于半监督学习算法的设计中。周志华[24]提出基于分歧的半监督学习方法,通过使用多个学习器来对未标记数据进行利用,在学习过程中将未标记数据作为多学习器间信息交互的平台。

在目前大规模数据的形式下,基于半监督学习的行为分析在安全监控、高级人机交互、视频会议、基于行为的视频检索以及医疗诊断等方面有着广泛的应用前景和潜在的经济价值,是当前计算机视觉领域的一个研究热点。行为分析的最终目标是要通过对行为特征数据的分析来获取行为的语义描述与理解,而行为建模和检测是达到上述目标的一个关键步骤。在很多实际应用如(监控系统的人群异常检测)中,通常涉及到对正常状态的建模和待检状态与正常状态的差异检测[25],大部分算法是通过正常数据进行建模。传统的基于监督学习的建模方法虽然能够建立准确的行为模型,但是需要手工标记大量的训练样本,这将造成大量的资源浪费。因此,基于半监督学习的建模成为一种趋势,这类方法可以自动或半自动地建立行为模型,在减轻人的负担的同时,也增强了算法的适用性。

结束语 本文对异常检测方法进行了综合性的阐述。首先介绍了异常检测的定义、背景以及异常检测中“异常点”产生的原因;其次,根据方法对异常类的使用,将其分为无监督和 supervised 两类,无监督的异常检测主要概括了数据处理形式不同的几类方法,而有监督的异常检测方法则从样本的生成方式上进行介绍;然后重点探讨了在大数据形式下的深度学习的方法。由于目前异常检测的方法在原理上已经逐渐成熟,在大规模数据中进行异常检测应着眼于挖掘数据本身的结构信息,而深度学习框架能够较好地适应这一特点。文中

同样以无监督和有监督两种方式探讨了深度学习中的代表性方法以及近年来的研究热点。文章最后总结了半监督学习中异常检测方法的研究热点以及它的应用前景。在目前海量数据的科学背景下,基于深度学习的异常检测方法也将成为研究的焦点。

参考文献

- [1] 赖英旭. 基于时间序列分析的工业控制以太网流量异常检测[J]. 北京工业大学学报, 2015(2): 200-206.
- [2] 叶青, 张丽丽. 复杂数据的异常检测算法[J]. 信息技术, 2015(5): 176-179.
- [3] 费欢, 李光辉. 基于 K-means 聚类的 WSN 异常数据监测算法[J]. 计算机工程, 2015(7): 124-128.
- [4] 谭真臻, 陈果, 陈立波. 基于图像分析和野点检测的航空发动机磨损状态识别[J]. 中国机械工程, 2010(7): 827-831.
- [5] 陈斌, 陈松灿, 潘志松, 等. 异常检测综述[J]. 山东大学学报, 2009(6): 19-23.
- [6] YEUNG D, CHOW C. Parzen-window network intrusion detectors[C]// Proceeding of the 16th International Conference on Pattern Recognition (ICPR'02). Washington, DC, USA: IEEE Computer Society, 2002.
- [7] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF: identifying density-based local outliers[J]. SIGMOD Record, 2000, 29(2): 93-104.
- [8] 王军敏, 李艳. 基于 K-均值算法的数据聚类 and 图像分割研究[J]. 平顶山学院学报, 2014, 29(2): 43-45.
- [9] 潘楚, 罗可. 基于改进计算的 K-medoids 聚类算法[J]. 计算机应用, 2014, 34(7): 1997-2000.
- [10] 王坤, 郭云飞. 基于 PCA 的无监督异常检测方法研究[J]. 郑州大学学报, 2004, 36(4): 39-42.
- [11] SCHOLKOPF B, PLAT J, SHAWE-TAYLOR J, et al. Estimating the support of high-dimensional distribution [J]. Neural Computation, 2001, 13(7): 1443-1471.
- [12] STEINWART I, HUSH D, SCOVEL C. A classification framework for anomaly detection[J]. Journal of Machine Learning Research, 2005, 6: 211-232.
- [13] 邓红莉, 杨韬. 一种基于深度学习的异常检测方法[J]. 信息通信, 2015(3): 3-4.
- [14] 高常鑫, 桑农. 基于深度学习的高分辨率遥感影像目标检测[J]. 测绘通报, 2014(S0): 108-111.
- [15] HINTON, RUSLAN R. Reducing the Dimensionality of Data with Neural Network[J]. Science, 2006(313): 504-507.
- [16] 肖柏旭, 张丽静. 基于分流抑制机制的卷积神经网络人脸检测法[J]. 计算机应用, 2006, 26(S1): 46-48.
- [17] 邹美银, 陈黎, 田菁. 基于卷积神经网络的视频图像失真检测及分类[J]. 计算机应用研究, 2016, 33(9): 2827-2830.
- [18] 李岳云, 许悦雷, 马时平. 深度卷积神经网络的显著性检测[J]. 中国图像图形学报, 2016, 21(1): 53-59.
- [19] 芮挺, 费建超, 周游. 基于深度卷积神经网络的行人检测[J]. 计算机工程与应用, 2016, 52(13): 162-166.
- [20] IAN G, POUGET ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. 2014: 2672-2680.
- [21] DENTON E, CHINTALA S, SZLAM A, et al. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks[C]// Advances in Neural Information Processing Systems 28(NIPS 2015). 2015.
- [22] LI F F, ROB F. One-shot Learning of Object Categories[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016: 594-611.
- [23] JOACHIMS T. Transductive Inference for Text Classification using Support Vector Machines [C] // Sixteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc, 1999: 200-209.
- [24] 周志华. 基于分歧的半监督学习[J]. 自动化学报, 2013, 39(11): 1871-1878.
- [25] 杨琳, 苗振江. 一种人群异常行为检测系统的设计与实现[J]. 铁路计算机应用, 2010, 19(7): 37-41.
- [40] HEIZER A, BARZOHAR M, MALAH D. Stable Fitting of 2d Curves and 3d Surfaces by Implicit Polynomials [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2004, 26(10): 1283-1294.
- [41] ZOLLHÖFER M, DAI A. Shading-based Refinement on Volumetric Signed Distance Functions [J]. ACM Transactions on Graphics (TOG), 2015, 34(4): 96.
- [42] ZHENG B, TAKAMATSU J, IKEUCHI K. An Adaptive and Stable Method for Fitting Implicit Polynomial Curves and Surfaces [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010, 32(3): 561-567.
- [43] WU G. An Adaptive Shape Description Method Based on Implicit Polynomial Curves [J]. Acta Electronica Sinica, 2014, 42(3): 505-511.
- [44] MANSON J, PETROVA G, SCHAEFER S. Streaming Surface Reconstruction Using Wavelets [J]. Eurographics Symposium on Geometry Processing, 2008, 27(5): 1411-1420.
- [45] SENER S, UNEL M. Affine Invariant Fitting of Algebraic Curves Using Fourier Descriptors [J]. Pattern Analysis and Applications, 2005, 8(1): 72-83.
- [46] YALCIN H, UNEL M, WOLOVICH W. Implicitization of Parametric Curves by Matrix Annihilation [J]. International Journal of Computer Vision, 2003, 54(1-3): 105-115.
- [47] 陈发来. 曲面隐式化新进展 [J]. 中国科学技术大学学报, 2014, 44(5): 345-361.
- [48] 李彩云, 朱春钢, 王仁宏. 参数曲线的分段近似隐式化 [J]. 高校应用数学学报, 2010, 25(2): 202-210.
- [49] 陈发来. 有理曲线的近似隐式化表示 [J]. 计算机学报, 1998, 21(9): 855-859.

(上接第 23 页)