基于样本关联感知的无监督深度异常检测模型

席 亮 王瑞东 樊好义 张凤斌
(哈尔滨理工大学计算机科学与技术学院 哈尔滨 150080)

摘 要 异常检测的目标是识别正常模式中的异常模式. 如何充分利用数据的各种特征信息来识别异常是当前异常检测的研究热点之一. 许多数据挖掘及机器学习等方面的智能算法都被用于异常检测规则训练以提高其检测性能. 目前已有模型存在着对复杂数据处理困难、没有充分利用数据样本间关联特征等问题，从而造成异常检测效果不甚理想. 基于此，本文提出一种基于样本关联感知的深度学习模型并用于异常检测. 模型通过对样本的原始特征和样本间的关联关系进行深入分析，利用无向图结构来提取样本间的关联特征，然后基于由特征编码器和图编码器构成的双路自编码器实现对样本的原始特征和关联特征的融合，产生样本在低维特征空间中高质量数据嵌入，然后进行解码重构并计算重构误差和重构特征，最后设计基于高斯混合模型的估计网络，基于重构特征和高质量的数据嵌入估计样本的概率密度，通过给定阈值来进行异常检测. 实验结果表明，本模型的异常检测各项性能指标均比其他基于机器学习和深度学习的异常检测方法提升了 2%左右，参数、消融和噪声实验结果也较其他算法更稳定，可视化实验也能够突出本模型在数据特征提取和充分利用等方面的优势.

关键词 异常检测；图神经网络；高斯混合模型；数据关联
中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2021.02317

Sample-Correlation-Aware Unsupervised Deep Anomaly Detection Model

XI Liang WANG Rui-Dong FAN Hao-Yi ZHANG Feng-Bin
(*School of Computer Science and Technology, Harbin University of Science and Technology, Harbin* 150080)

Abstract The goal of anomaly detection is to identify abnormal patterns within normal patterns. Anomaly detection is applied in different forms in different application scenarios, such as network security, medical image and video monitoring. How to make full use of all kinds of characteristic information of data to identify the anomaly is one of the hot spots of anomaly detection. Many intelligent algorithms and models in data mining, machine learning and deep learning have been used to training anomaly detection rules to improve their detection performances. The training methods of anomaly detection rules are divided into three kinds: supervised, semi-supervised and unsupervised methods, and the third kind is the most popular. Formally, unsupervised deep anomaly detection can be viewed as density estimation from the data distribution. At present, deep anomaly detection models have achieved remarkable results in different application scenarios. But these methods are mainly based on the original features of data for training, and ignore the complex correlation among data samples: there are usually some kinds of correlation among normal samples that abnormal samples do not have. This makes these methods not ideal for anomaly analysis of complex data with characteristics of large data size, high dimension, unbalanced abnormal

收稿日期：2020-02-27；在线发布日期：2020-08-18. 本课题得到黑龙江省自然科学基金项目（No.F2018019）资助. **席 亮(通信作者)**，博士，副教授，主要研究领域为人工智能及应用、网络与信息安全、深度学习等. E-mail: xiliang@hrbust.edu.cn. **王瑞东**，硕士研究生，主要研究领域为网络与信息安全、深度学习等. E-mail: iswangrd@gmail.com. **樊好义（通信作者）**，博士研究生，主要研究领域为人工智能、数据挖掘、深度学习等. E-mail: isfanhy@gmail.com. **张凤斌**，博士，教授，博士生导师，计算机学会(CCF)高级会员（30536S），主要研究领域为人工智能，网络与信息安全.

proportion, etc. In view of this, in this paper, a sample-correlation-aware deep learning model is proposed and used for anomaly detection, named sample-correlation-aware unsupervised deep anomaly detection model(SCA-UDLM): First of all, through in-depth analysis of the original sample features and the correlation features among samples, the model uses the *K*-nearest neighbors algorithm to search for similar samples to extract the correlation features among samples and store them in an undigraph structure, where nodes represent the samples and the edges represent the correlation between the two samples. Secondly, the original features and correlation features are fused based on the dual autoencoder composed of feature encoder and graph encoder, and generated high-quality data embedding in the low-dimensional feature space. Thirdly, the decoder decodes the low-dimensional embedding into reconstruction samples with original dimensional, and calculates the reconstruction error and features. Finally, an estimation network based on Gaussian mixture model is designed to estimate the probability density of samples based on the input of the reconstruction features and the low-dimensional embedding which are fused by the additive fusion method, and judge whether these samples are abnormal or not based on the given judgment threshold. A large number of experiments and analyses were made on this model and relevant representative machine learning methods and up-to-date deep learning methods, such as, one class support vector machines (OC-SVM), isolation forests(IF), deep structured energy based models(DSEBM), deep autoencoding GMM(DAGMM), Anomaly generative adversarial network(AnoGAN), adversarially learned anomaly detection(ALAD). The experimental results show that the detection performances of this model are improved by about 2% compared with other related methods, and its experimental results with different parameters, modules and noises are more stable than other methods, which prove the validity of the correlation module. The visualization experimental results can also highlight the advantages of this model in data feature extraction and full utilization.

Keywords anomaly detection; graph neural network; Gaussian mixture model; data correlation

1 引 言

不同的应用领域存在不同的形式的异常，需要进行异常检测来保证系统的正常工作，例如网络安全[1]，医学影像[2]和视频监控[3]等. 异常检测模型（Anomaly Detection Models, ADMs）基于规则库来识别偏离正常行为的异常模式. 如何得到优秀的规则库是异常检测的核心问题. 当今，基于数据挖掘、人工智能、机器学习、深度学习等智能技术的异常检测研究尤为活跃并取得了丰富的研究成果[4].

用于异常检测规则学习的主要方法包括 3 类：有监督、半监督和无监督.（1）有监督方法利用带标签数据进行训练. 标签已指明数据是否异常. 这类方法具有较高的准确率. 然而由于异常特征种类繁多且日新月异，有监督方法通常难以获取全部特征，而且需要积累大量的异常样本来解决数据不平衡问题，而获得足够的异常样本是一个非常大的挑战[4].（2）半监督方法介于有监督和无监督学习之间，基于少量有标记样本和大量无标记样本，通常利用自编码器在无异常的数据上进行半监督训练[5, 6]，利用正常标签比异常标签更容易获取的优势，通过使用单类标签来分离异常值. 半监督方法非常符合现实世界中的实际场景，并能够解决数据不平衡问题. 但是半监督模型对训练数据的质量要求较高，噪声数据会对模型产生较大影响.（3）无监督方法旨在使用未标记数据进行训练，检测没有标记数据的异常值，可充分利用类别未知的训练样本解决模式识别中的各种问题. 这种形式的问题在实际应用中更为常见. 无监督异常检测可以看作是数据分布的密度估计[7]：异常往往位于低概率密度区域. 无监督方法是当前异常检测最为推崇的方法[8].

现有的用于无监督的方法主要包括基于重构的方法、基于聚类的方法、基于支撑域的方法[7].（1）基于重构的方法的代表性方法是利用线性投影的主成分分析（Principal Component Analysis, PCA）[9]及其改进方法[10-12]. 这种方法通过强制降维导致模型对数据的敏感性降低. 另外，深度自编码器（Deep Autoencoder, DAE） 也是目前最为活跃的方法，利用数据的重构误差识别异常[13,14]. 然而，基于重构的方法未考虑压缩数据导致的有效信息的丢失，比

如异常样本可能以正常的方式隐藏在数据中，而这种异常更应该得到关注. 而且这类方法通常无法从数据的低维投影中有效重构出原始数据.（2）基于聚类的方法是另一类流行的异常检测规则学习方法，包括 FCM 算法、多元高斯模型等[15, 16]. 此类方法通常采用先降维后聚类的两步式方法,其降维操作与聚类操作没有直接关系,导致降维后数据无法根据聚类效果进行更新,影响最终检测性能的提升. 针对该问题，许多人将目光转移到高斯混合模型（Gaussian Mixture Model, GMM）提出许多方法，例如深度自编码高斯混合模型（Deep Autoencoding Gaussian Mixture Model, DAGMM）[8]等，通过一个端到端的模型进行训练，后续的聚类分析对数据降维操作进行调整，从而完成数据更新. 然而低质量的训练数据（噪声）会极大影响模型的最终检测效果.（3）基于支撑域的方法也广泛用于异常检测. 这种方法通过学习正常样本的一个有效边界来识别异常，例如单类支持向量机算法（One-Class Support Vector Machine, OC-SVM）[17-19]. 然而，这类方法的效率随着数据维度的增长而降低.

上述方法在异常检测中都取得了显著效果. 但这些方法主要基于数据的原始特征，忽略了数据样本之间的复杂关联性，即正常样本之间通常具有异常样本不具备的某种关联性. 这些方法对有数据量大、维度过高、异常比例不均衡等特点的复杂数据，如网络流量、文本分析、医学药品检测等进行异常分析效果不甚理想. 因此，本文针对具有以上特点的复杂数据进行了关联性研究后发现，相似样本间的关联性可以在特征学习期间通过利用来自邻居的代表性特征生成用于异常检测的高质量样本嵌入. 然而，基于样本间的关联性建模方法与那些仅需要捕获非线性结构的常规特征学习模型有很大不同. 因此，如何进行样本间的关联性分析并有效地与样本的原始特征结合用于异常检测的集成特征学习目前尚未得到解决. 基于此，本文提出了一种基于样本关联感知的无监督深度异常检测模型（Sample-Correlation-Aware Unsupervised Deep Anomaly Detection Model, SCA-UDADM）. 它同时考虑数据原始特征和数据样本间的复杂关联性,并采用端到端的方式进行训练,然后利用高质量的样本嵌入进行异常检测，从而达到更好的检测效果. 算法与传统方法的区别如图 1 所示. 具体而言，首先以图结构对数据样本进行关联性建模，其中节点表示样本，边缘表示特征空间中两样本间的关联性；然后，设计基于样本关联感知的无监督深度学习模型（Sample-Correlation-Aware Unsupervised Deep Learning Model, SCA-UDLM），采用由特征编码器和图编码器组成的双路编码器将样本的原始特征和样本间的关联特征共同编码到低维潜在空间中,并设计解码器对编码后的数据进行数据重构. 最后，基于高斯混合模型实现一个单独的估计网络来估计数据的概率密度完成异常检测任务.

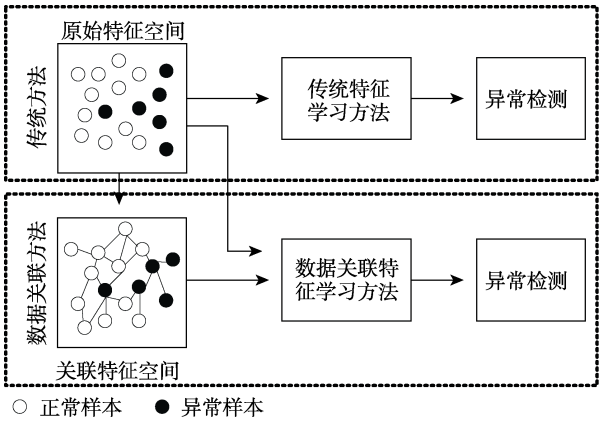

图 1　用于异常检测的关联感知特征学习

本文的主要贡献如下：

（1）本文重点研究异常检测中以前一直被忽视的数据样本之间的关联性特征分析. 利用图结构对数据的关联性建模，利用图神经网络（Graph neural network, GNN）对数据的关联性特征进行提取，用于异常检测任务.

（2）为了充分利用各种数据特征完成更精确的异常检测，本文提出双路自编码器来学习样本的原始特征和关联特征. 利用特征编码器学习原始特征，图编码器来学习关联特征，最终融合编码器学习到的特征来进行异常检测任务.

论文剩余部分的结构安排如下：第 2 部分介绍本文模型涉及的基于无监督深度学习的异常检测和图神经网络相关内容. 第 3 部分详述本文模型的实现过程，并进行相关方法的对比分析. 第 4 部分是实验结果与分析. 最后进行结论性总结和展望.

2　相关工作

2.1　基于无监督深度学习的异常检测

如前文所言，基于无监督深度学习的异常检测仅根据数据的潜在特性进行训练并用于异常检测可解决带标签数据难以获取的问题[20]. 学习样本数据的概率密度分布是所有无监督深度异常检测的核心：假设正常事件发生的概率高于异常数据的发生的概率. 目前最流行的概率密度估计方法是 GMM，

它将一个整体的对象分为多个基于高斯分布的子对象，从而学习到更加准确的样本分布[16]. 除此之外，无监督学习方法，例如玻尔兹曼机、深度玻尔兹曼机、深度信念网络（Deep Belief Network, DBN）[21]、广义去噪自编码器[22]、循环神经网络（Recurrent Neural Network, RNN）[22]以及长短期记忆网络（Long Short Term Memory, LSTM）[23]都被用于异常检测.

近年来，无监督深度异常检测模型取得了许多优秀的研究成果，整体效果要优于基于传统方法的异常检测模型. 其中基于 DAE 的异常检测模型效果最为显著. 文献[7]采用 DAE 和 GMM 来进行异常检测：DAE 对数据进行特征提取和降维操作，利用基于 GMM 的估计网络对样本的概率密度进行估计. 作者为了解决降维操作导致的有效信息丢失问题，将重构误差作为数据特征的一部分与降维后的数据一起作为估计网络的输入，更有利于模型准确估计样本的概率密度. 然而，此模型并没有对数据隐含的关联关系进行挖掘，还具有改进的空间. 文献[24]提出了深度结构能量模型(Deep Structured Energy Based Model, DSEBM)，采用深度能量结构直接对数据分布建模，集成不同类型的数据（如静态数据、顺序数据和空间数据）并与正则化的 DAE 连接完成复杂的数据采样和模型训练. 能量模型能够降低训练中数据信息的损失，因此本文基于能量来设计了模型的目标函数.

另外，近年来基于生成对抗网络（Generative Adversarial Network, GAN）的异常检测也取得了很多优秀成果. 文献[25]首次提出了基于 GAN 的异常检测算法. 由于单纯地利用 GAN 进行异常检测会增加检测时间，作者在训练中基于训练 DAE 的方式学习数据的潜在分布来解决 GAN 训练时间长的问题，但是模型训练仍然比较复杂. 文献[26]提出了一种基于双向 GAN 的异常检测方法，使用基于反向学习特征的重构误差来确定样本是否异常，同时确保训练过程中初始样本空间和潜在空间的一致性，使 GAN 稳定训练，显著提高了异常检测性能. 然而，基于 GAN 的相关模型虽然学习样本分布的能力较强，但也并未充分挖掘数据间关联性特征.

基于以上相关研究现状分析可以看出，目前无监督深度异常检测方法具有传统异常检测模型无法比拟的优势. 因此，本文将以基于无监督深度学习模型为研究对象，并对数据间的关联性进行建模，深入挖掘数据样本间的关联性从而更高效精确地进行模型训练以提高检测性能.

2.2 图神经网络

近年来，图表被广泛用于建立成对关系的模型，如文献分析、社交网络和生物信息网络等[27]. GNN 作为一种基于图结构的深度学习模型，在各个应用中表现出了卓越的性能. 文献[28]利用 GNN 基于图结构对节点的原始特征进行学习从而重构特征空间. 文献[29-31]引入卷积运算提出了图卷积神经网络（Graph Convolutional Network, GCN），如图 2 所示，优化了基于图结构的节点特征学习. GCN 的最新研究成果是图注意力网络（Graph ATtention, GAT）[32]，其核心是注意力机制，可高效处理可变大小的数据，将其应用于同类型的节点连接中，并驱使模型将注意力集中在数据的关键部分，从而得到更精确的训练结果，已被广泛应用于各种应用场景，例如文本分析[33]，知识图[34]和图像处理[35]等.

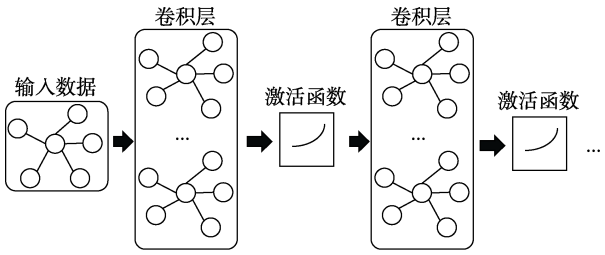

图 2 图卷积神经网络基本模型

目前，基于 GNN 的异常检测也是深度学习重要的应用热点之一. 文献[36]在有噪声标签的视频异常检测中，通过利用 GNN 将视频与标签的关联信息进行学习以此消除噪声标签，并利用动作分类器将视频监控信号从高可信度片段传播到低可信度片段，从而实现对视频中的异常数据进行检测. 文献[37]为了解决属性网络中，因网络稀疏性和数据非线性而导致的模型无法捕获不同信息模式之间的复杂交互问题，提出使用 GCN 来对属性网络的拓扑结构和节点属性进行学习，利用自编码器将 GCN 学习到的结构特征和属性特征一起重建原始数据，利用重构误差来发现属性网络中的异常. 但是该方法无法对非结构化数据进行关联性建模，因此对非结构化数据的异常检测并不理想. 同样地，针对属性网络中网络结构和节点属性之间潜在的跨模态复杂交互问题，文献[38]提出了一种用于属性网络异常检测的深度学习框架(AnomalyDAE)，使用一个由结构自编码器和属性自编码器构成的双路自编码器对数据节点嵌入和属性嵌入进行联合学习，并在结构自编码器中引入图注意力机制，从而实现更有效的网络结构特征捕捉.

由此可见，很多应用背景中的数据特征具备成对的图结构关系，如社交网络中的社区和功能嵌入中的集群等[39]. 然而，还有一些应用场景中的数据特征不具备成对关系，无法直接建立图结构. 这就需要对其进行数据关联分析，从而构造出合适的图结构以便进一步训练. 本文应用背景即是如此.

3　基于数据关联感知的无监督深度异常检测模型

SCA-UDADM是一个用于异常检测的端到端无监督深度学习模型. 模块框架如图 3 所示. SCA-UDADM 主要由图的构建、双路编码器、特征解码器和估计网络 4 个模块组成. 具体而言，数据样本之间的关联性由原始特征空间以图结构生成. 在构建的图中，节点表示样本，边缘表示特征空间中两个样本之间的关联性. 然后使用由特征编码器和图编码器组成的双路编码器将样本的原始特征和关联性特征联合编码到低维潜在空间中，再使用特征解码器进行样本特征重构. 最后，利用基于 GMM 的估计网络进行样本概率密度估计，通过测量样本相对于给定阈值的能量来检测异常.

3.1　相关定义

定义 1.　图的定义为 $\boldsymbol{G}=\{\boldsymbol{V},\boldsymbol{\varepsilon},\boldsymbol{X}\}$，具有 N 个节点和 E 条边. 其中节点集合表示为 $\boldsymbol{V}=\{\boldsymbol{v}_i, i=1,2,...,N\}$，边的集合表示为 $\boldsymbol{\varepsilon}=\{e_i, i=1,2,..,E\}$，其中 $e_i=(v_{i_1}, v_{i_2})$ 表示在节点 $\boldsymbol{v}_{i_1}$ 和 $\boldsymbol{v}_{i_2}$ 之间的一条边. $\boldsymbol{X}\in\mathbb{R}^{N\times F}$ 表示每行代表一个节点的关联特征的特征矩阵，表示特征的维度.

定义 2.　图的邻接矩阵 $\boldsymbol{A}\in\mathbb{R}^{N\times N}$ 表示图的拓扑结构. 如果在节点 $\boldsymbol{v}_i$ 和节点 $\boldsymbol{v}_j$ 之间具有一条边，则邻接矩阵中的元素 a_{ij}=1，否则 a_{ij}=0.

定义 3.　异常检测模型中，给定一组输入数据 $\boldsymbol{S}=\{\boldsymbol{s}_i \mid i=1,...,N\}$，每个样本 $\boldsymbol{s}_i$ 都与 F 维特征 $\boldsymbol{X}_i\in\mathbb{R}^{N\times F}$ 相关联. 模型的目的是学习一个函数 $u(\boldsymbol{X}_i):\mathbb{R}^N\mapsto\mathbb{R}$ 用来计算样本的能量. 最后，基于设定阈值 λ 来判定样本 $\boldsymbol{s}_i$ 是否异常：

$$y_i=\begin{cases}1, & \text{if } u(\boldsymbol{X}_i)\geqslant\lambda\\ 0, & \text{其他}\end{cases} \tag{1}$$

其中，y_i=1 表示异常，y_i=0 表示正常.

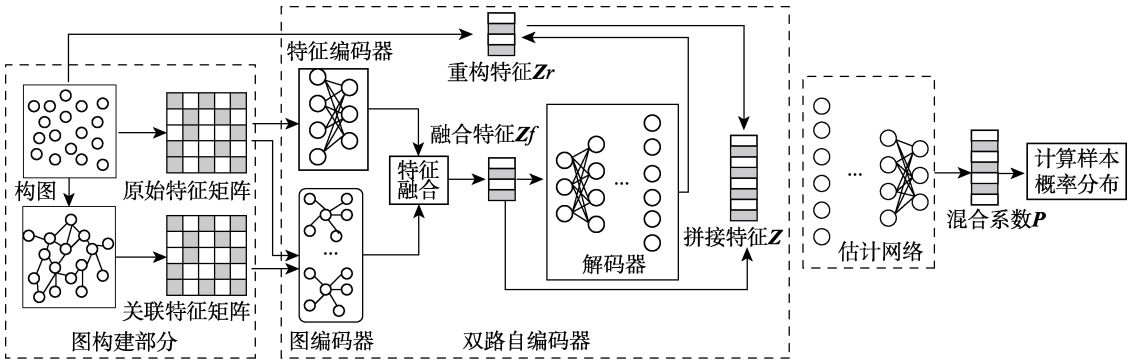

图 3　模型框架

3.2　数据关联模型构建

为了探索数据样本之间的关联性，模型显式地构造了一个图结构来关联来自特征空间的相似样本：给定一组输入样本 $\boldsymbol{S}=\{\boldsymbol{s}_i \mid i=1,...,N\}$，利用 K-NN(K-Nearest Neighbors)算法来确定样本 $\boldsymbol{s}_i$ 特征空间中的 K 个近邻 $\boldsymbol{N}_i=\{\boldsymbol{s}_{ik} \mid k=1,...,K\}$. 然后在 $\boldsymbol{s}_i$ 和它的邻居 $\boldsymbol{s}_{ik}$ 之间分配一个无向边. 最后构造无向图 $\boldsymbol{G}=\{\boldsymbol{V},\boldsymbol{\varepsilon},\boldsymbol{X}\}$，其中 $\boldsymbol{V}=\{\boldsymbol{v}_i=\boldsymbol{x}_i, i=1,2,...,N\}$ 为节点集，$\boldsymbol{\varepsilon}=\{e_{ik}=(\boldsymbol{v}_i,\boldsymbol{v}_{ik}) \mid \boldsymbol{v}_{ik}\in\boldsymbol{N}_i\}$ 为边缘集，$\boldsymbol{X}\in\mathbb{R}^{N\times F}$ 为节点特征矩阵，如图 4 所示. 基于所构造的图，显式地捕获样本间的关联性，然后利用样本间的信息和模型的传播机制进行特征学习.

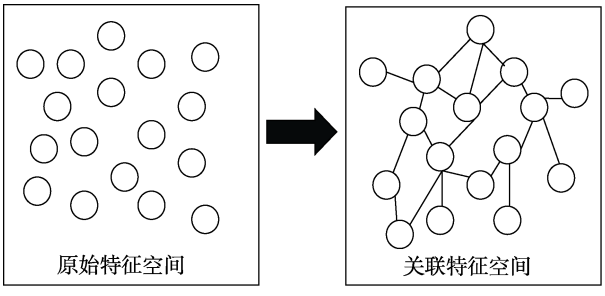

图 4　数据关联模型构建示意图

算法 1.　数据关联模型构建

输入：输入样本 $\boldsymbol{S}=\{\boldsymbol{s}_i \mid i=1,...,N\}$；近邻的数量 k；集合大小为 N

输出：无向图 $\boldsymbol{G}$

函数：*knn*：*K*-NN 算法；*dis*：距离计算函数，采用余弦距离；*topK*：集合中最小距离索引选择函数

1. 初始化 $\boldsymbol{G}$，$\boldsymbol{G}.size$=[*N*, *N*]
2. FOR *i* IN *range*(*len*($\boldsymbol{S}$)) DO
3. $\boldsymbol{G}_{i,i}=0$ //对角线元素为 0
4. $\boldsymbol{e}=knn(\boldsymbol{S}[i],\boldsymbol{S},k)$ //寻找第 *i* 个样本的 *k* 近邻
5. $\boldsymbol{D}=dis(\boldsymbol{S}[i],\boldsymbol{e}_i)$ //计算第 *i* 个样本与其他近邻的距离
6. $\boldsymbol{D}$=*sort*($\boldsymbol{D}$)//根据距离升序排序
7. ***index*** = *topK*($\boldsymbol{D}$, *N*-1)//记录最近邻的索引
8. For *j* IN *range*(*len*(***index***)) DO
9. $\boldsymbol{G}_{i,j}=dis(\boldsymbol{S}[i],\boldsymbol{index}[j])$ //构建图
10. END FOR
11. END FOR

3.3 模型的构建

3.3.1 双路编码器

为了获得足够有代表性的样本嵌入，双路编码器由特征编码器和图编码器组成，分别对样本的原始特征和样本间的关联性特征进行编码.

（1）特征编码器：为了对样本的原始特征进行编码，特征编码器采用多层感知机（Multi-Layer Perceptron, MLP）进行非线性的特征变换，其中多层感知机由全连接层（Full Connected layers, FC）构成，具体如下：

$$\begin{aligned}\boldsymbol{Z}^{\boldsymbol{X}(l_X)}&=MLP(\boldsymbol{Z}^{\boldsymbol{X}(l_X-1)}\,|\,\boldsymbol{W}^{\boldsymbol{X}(l_X-1)},\boldsymbol{b}^{\boldsymbol{X}(l_X-1)})\\&=\sigma(\boldsymbol{Z}^{\boldsymbol{X}(l_X-1)}\boldsymbol{W}^{\boldsymbol{X}(l_X-1)}+\boldsymbol{b}^{\boldsymbol{X}(l_X-1)})\end{aligned}\tag{2}$$

其中，$\boldsymbol{Z}^{\boldsymbol{X}(l_X-1)}$ $\boldsymbol{Z}^{\boldsymbol{X}(l_X)}$，$\boldsymbol{W}^{\boldsymbol{X}(l_X-1)}$ 和 $\boldsymbol{b}^{\boldsymbol{X}(l_X-1)}$ 分别是输入数据，输出数据，第 l_X 层的权重以及偏差矩阵. $l_X\in\{1,2,...,L_X\}$ 表示网络层数，$\boldsymbol{Z}^{\boldsymbol{X}(0)}=\boldsymbol{X}$ 表示编码器的初始输入. $\sigma(\cdot)$ 表示激活函数，本文使用 $Tanh(x)=\dfrac{2}{1-e^{-2x}}-1$.

（2）图编码器：为了对样本之间的关联性进行编码，图编码器使用一个图注意力层(graph attention layer, GAT)对样本间关联性特征进行聚合，在每个节点上执行一个共享的注意机制：

$$w_{i,j}=attn(\boldsymbol{X}_i,\boldsymbol{X}_j)=\sigma(\boldsymbol{a}^T\cdot[\boldsymbol{W}^c\boldsymbol{X}_i\,\|\,\boldsymbol{W}^c\boldsymbol{X}_j])\tag{3}$$

其中，$w_{i,j}$ 表示节点重要性的权重，*attn*(·)表示被 $\boldsymbol{a}\in\mathbb{R}^D$ 和 $\boldsymbol{W}^c=\mathbb{R}^{\frac{D^c}{2}\times F}$ 参数化的神经网络中的节点的共享权值，D^c 表示 *attn*(·)中的神经元的个数，|| 表示级联操作. 然后通过使用 *softmax* 函数将节点的重要性 $\alpha_{i,j}$ 进行归一化处理：

$$\alpha_{i,j}=\frac{e^{w_{i,j}}}{\sum\limits_{k\in N_i}e^{w_{i,k}}}\tag{4}$$

其中，N_i 表示节点 v_i 的邻居数量，是从邻接矩阵中获得的. 最后，利用学习到的重要性权重 $\alpha_{i,j}$ 与节点的关联矩阵 $\boldsymbol{X}$ 进行加权求和，得到嵌入数据 $\boldsymbol{Z}_i^V$：

$$\boldsymbol{Z}_i^V=\sum_{k\in N_i}\alpha_{i,k}\cdot\boldsymbol{X}_k\tag{5}$$

在学习到嵌入数据 $\boldsymbol{Z}^{\boldsymbol{X}(L_X)}$ 和 $\boldsymbol{Z}^V$ 的基础上，通过一个融合模块，将异构源数据的嵌入融合到一个共享的潜在空间中，通过一个全连接层获得最终的样本嵌入 $\boldsymbol{Z}^f\in\mathbb{R}^{N\times D}$：

$$\boldsymbol{Z}^f=fusion(\boldsymbol{Z}^{\boldsymbol{X}(L_x)},\boldsymbol{Z}^V)=\boldsymbol{Z}^{\boldsymbol{X}(L_x)}\oplus\boldsymbol{Z}^V\tag{6}$$

其中，⊕ 表示矩阵按元素相加.

3.3.2 解码器

解码器的目的是从编码器获得的潜在特征 Z^f 中，重构出原始数据，具体如下：

$$\boldsymbol{Z}^{\hat{\boldsymbol{X}}(l_{\hat{X}})}=MLP(\boldsymbol{Z}^{\hat{\boldsymbol{X}}(l_{\hat{X}}-1)}\,|\,\boldsymbol{W}^{\hat{\boldsymbol{X}}(l_{\hat{X}}-1)},\boldsymbol{b}^{\hat{\boldsymbol{X}}(l_{\hat{X}}-1)})\tag{7}$$

其中 $\boldsymbol{Z}^{\hat{\boldsymbol{X}}(l_{\hat{X}}-1)}\in\mathbb{R}^{N\times D^{l_{\hat{X}}-1}}$，$\boldsymbol{Z}^{\hat{\boldsymbol{X}}^{(l_{\hat{X}})}}\in\mathbb{R}^{N\times D^{l_{\hat{X}}}}$，$\boldsymbol{W}^{\hat{\boldsymbol{X}}(l_{\hat{X}}-1)}\in\mathbb{R}^{D^{l_{\hat{X}}-1}\times D^{l_{\hat{X}}}}$ 和 $\boldsymbol{b}^{\hat{\boldsymbol{X}}(l_{\hat{X}}-1)}\in\mathbb{R}^{D^{l_{\hat{X}}-1}}$ 分别是解码器的第 $l_{\hat{X}}$ 层的输入数据，输出数据，可训练权重和偏置. $\boldsymbol{Z}^{\hat{\boldsymbol{X}}(0)}=\boldsymbol{Z}^f$ 是解码器的初始输入. 最后，通过解码器得到解码后的数据

$$\hat{\boldsymbol{X}}=\boldsymbol{Z}^{\hat{\boldsymbol{X}}(L_{\hat{X}})}\tag{8}$$

3.3.3 估计网络

SCA-UDADM 基于 GMM 设计一个估计网络，用于对样本的概率密度进行估计，从而完成异常分析与检测. 在文献[22]的启发下，本文采用由多个全连接层 FC 组成的子网络作为估计网络，将重构误差 $\boldsymbol{Z}^r$ 和编码器的输出 $\boldsymbol{Z}^f$ 级联为估计网络输入数据 $\boldsymbol{Z}=\{\boldsymbol{Z}_1,\boldsymbol{Z}_2,...,\boldsymbol{Z}_N\}$，估计样本的混合概率系数矩阵 $\mathbf{P}\in\mathbb{R}^{N\times M}$，*M* 表示 GMM 中混合分布的数量. 然后基于 $\boldsymbol{P}$ 计算输入样本的概率密度并求出样本的能量值：能量值越高，样本异常可能性越大. 具体过程如下：

（1）$\boldsymbol{Z}$ 的计算方法为：

$$\boldsymbol{Z}=[\boldsymbol{Z}^f\,\|\,\boldsymbol{Z}^r],\boldsymbol{Z}^r=[Cos(\boldsymbol{X},\hat{\boldsymbol{X}}),Euc(\boldsymbol{X},\hat{\boldsymbol{X}})]\tag{9}$$

其中，$Cos(\boldsymbol{X},\hat{\boldsymbol{X}})$ 与 $Euc(\boldsymbol{X},\hat{\boldsymbol{X}})$ 分别表示采用余弦距离和 Euclidean 距离来计算重构误差；之后将两种不同的重构误差作为两个不同的特征级联为重构特征 $\boldsymbol{Z}^r$，之后与编码器的输出 $\boldsymbol{Z}^f$ 级联构造为最终的样本嵌入 $\boldsymbol{Z}$. 在得到最终的样本嵌入 *Z* 后，估计网络计算 $\boldsymbol{P}$，具体如下：

$$\boldsymbol{Z}^{M(l_M)} = MLP(\boldsymbol{Z}^{M(l_M)} \mid \boldsymbol{W}^{M(l_M-1)}, \boldsymbol{b}^{M(l_M-1)}) \quad (10)$$

其中，$\boldsymbol{Z}^{M(l_M-1)}$、$\boldsymbol{Z}^{M(l_M)}$、$\boldsymbol{W}^{M(l_M-1)}$和$\boldsymbol{b}^{M(l_M-1)}$分别为估计网络第$l_M$层的输入数据、输出数据、可训练权重和偏置矩阵．$\boldsymbol{Z}^{M(0)} = \boldsymbol{Z}$表示初始数据为样本嵌入$\boldsymbol{Z}$.

（2）$\boldsymbol{P}$的计算公式如下：

$$\boldsymbol{P} = Softmax(\boldsymbol{Z}^{M(L_M)}) \quad (11)$$

（3）利用$\boldsymbol{P}$计算 GMM 的均值向量$\boldsymbol{\mu}$和协方差矩阵$\boldsymbol{\Sigma}$，公式如下：

$$\boldsymbol{\mu}_m = \frac{\sum_{i=1}^{N} p_{i,m}\boldsymbol{Z}_i}{\sum_{i=1}^{N} p_{i,m}} \quad (12)$$

$$\boldsymbol{\Sigma}_m = \frac{\sum_{i=1}^{N} p_{i,m}(\boldsymbol{Z}_i - \boldsymbol{\mu}_m)(\boldsymbol{Z}_i - \boldsymbol{\mu}_m)^T}{\sum_{i=1}^{N} p_{i,m}} \quad (13)$$

其中，$\boldsymbol{\mu}_m$和$\boldsymbol{\Sigma}_m$分别是高斯混合模型预测的第m个分布的样本的均值向量和协方差矩阵.

（4）对样本能量进行评估：样本能量的计算方法如下：

$$E_Z = -\log\left(\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{p_{n,m}}{N}\cdot\frac{\exp\left(-\frac{1}{2}(\mathbf{Z}_n-\boldsymbol{\mu}_m)^T\boldsymbol{\Sigma}_m^{-1}(\mathbf{Z}_n-\boldsymbol{\mu}_m)\right)}{\sqrt{|2\pi\boldsymbol{\Sigma}_m|}}\right) \quad (14)$$

3.3.4　目标函数

SCA-UDADM模型的训练目标函数基于以下损失函数计算：

$$L = \left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|_2^2 + \lambda_1 E_z + \lambda_2\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{1}{(\boldsymbol{\Sigma}_m)_{nn}} + \lambda_3\|\boldsymbol{Z}\|_2^2 \quad (15)$$

其中，第一部分$\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|$表示自编码器的重构误差；第二部分$E_z$表示样本的预测能量，目的是最大化可观测样本的似然函数；第三部分是样本预测的协方差矩阵的惩罚项，为了避免 GMM 中因协方差矩阵对角线元素退化为零而造成训练停滞的问题；最后是正则化项，使正常样本的的潜在空间尽可能的缩小在一个范围内，使异常样本偏离正常样本．λ_1，λ_2，λ_3三个参数分别控制不同项之间的权衡.

SCA-UDADM充分利用双路编码器分别对数据的原始特征和关联性特征进行学习，然后将特征进行融合，产生高质量的样本嵌入，并基于 GMM 作为估计网络进行数据概率分布计算，学习到更准确的样本分布．最后对正常数据分布学习，实现对正常数据和异常数据的分离.

3.4　与现有方法的比较

本节主要讨论所提方法与相关方法之间的差异．本文选取机器学习中的 2 种相关的经典算法：（1）OC-SVM[17]；（2）孤立森林（Isolation Forests, IF）[20]，以及近几年提出的深度学习相关的 4 种代表性方法：（1）Deep Structured Energy Based Models，DSEBM[24]；（2）Deep Autoencoding GMM, DAGMM[7]；（3）Anomaly GAN, AnoGAN[25]；（4）反向学习检测模型（Adversarially Learned Anomaly Detection, ALAD）[40].

OC-SVM：是一种经典的基于核函数的异常检测方法，通过学习正常数据和异常数据的决策边界判断数据是否正常．OC-SVM 对小样本训练集能够得到较为优秀的效果，泛化能力强，但对数据量大、维度高的数据上难以得到较为理想的结果.

IF：该算法一般用于结构化数据的异常检测，它将异常点定义为“容易被孤立的离群点”，即分布稀疏、且距离高密度群体较远的点．在数据空间里，若一个区域内只有分布稀疏的点，表示数据点落在此区域的概率很低，因此可以认为这些区域的点是异常的．因此，IF 的两个前提条件（局限性）为异常数据占总样本量的比例很小以及异常点与正常点的特征值具有较大差异．即 IF 对于非具有该类特征的数据集效果不甚理想.

DSEBM：是一种基于深度学习的异常检测方法，在不同的网络层之间积累能量，并依此判断数据是否为异常．它使用了 2 种不同的标准 DSEBM-r 和 DSEBM-e，分别以能量和重构误差作为异常评分．DSEBM 充分利用了训练过程中的信息来检测异常，但需要正常数据和异常数据具有较大差异.

DAGMM：是一种基于自编码器的异常检测方法，由自编码器构成的数据压缩网络和基于 GMM 的密度估计网络组成．前者通过自编码器来学习数据的嵌入表示，后者对学到低维数据进行密度估计，最终模型通过比较估计的样本能量和预先设定的阈值来进行异常检测．DAGMM 能够通过计算重构误差作为部分特征弥补数据压缩过程中丢失的信息，但模型需要质量较高的训练集进行训练.

AnoGAN：是 GAN 应用在异常检测的开山之作：在训练阶段使用正常样本训练一个 DCGAN（Deep Convolutional GAN）用于学习样本潜在空间的概率分布；测试阶段输入异常样本，通过梯度下降法求样本在潜在空间的表示，利用潜在表示在生成器中生成的数据和原始样本进行异常检测．AnoGAN 的核心是模型的对抗过程，对于维度较高

的数据，模型难以很好训练.

ALAD：也是一个基于 GAN 的异常检测模型. 模型训练一个从数据的原始特征空间到数据潜在空间的编码器，并对输入数据 $\boldsymbol{X}$ 进行编码得到 $\boldsymbol{Z}$. 同时从某个分布中采样 $\boldsymbol{Z}'$ 并解码得到假数据 $\boldsymbol{X}'$. 将 $(\boldsymbol{X},\boldsymbol{Z})$ 和 $(\boldsymbol{X}',\boldsymbol{Z}')$ 输入到判别器进行判别，以此优化编码器，稳定 GAN 的训练. 最后模型使用重构误差来确定数据样本是否异常. 该方法能够处理复杂的高维数据，但模型处理过程比较复杂，对于小数据集训练不充分，效率不理想.

综合以上分析可以看出，传统的机器学习方法在数据量大、维度高的数据上表现效果不好，且对训练数据集的质量要求很高；而相关的深度学习方法在其中的某些方面也存在不足之处. 本文的 SCA-UDADM 深入分析产生这些问题的内在原因，重点研究样本间关联特征，并与原始特征进行融合，从而有效解决这些问题. 以下实验可佐证以上分析结论，并证明本文方法的优势.

4 实验结果与分析

4.1 实验环境和数据集

实验基于 Ubuntu 环境，使用 Python 语言，在 tensorflow 环境下进行模型搭建. 平台主要硬件参数为：CPU 使用 Inter Core i7-9700，64G 内存，GPU 使用双路 NVIDIA 2070，8G 显存.

实验选取本领域 3 个代表性的公开数据集，分别为（1）数据量大、数据维度较高的 KDDCUP99 数据集；（2）数据量小、数据维度高的 Arrhythmia 数据集；（3）数据量适中，异常数据比例高的 Satellite 数据集. 相关统计如表 1 所示.

KDDCUP99[41]：是异常检测通用的一个经典数据集. 本文使用其一个 10%的数据子集进行实验，包括 494021 个样本，每条样本包括 41 种特征，其中 34 种连续特征，7 种分类特征（离散型数据）. 实验采用独热编码（One-hot）的方式对 7 种分类特征进行处理，将离散型的取值扩展到欧式空间[7]，最终得到 120 维的数据. 另外，由于数据集中大部分样本是异常点，所以本实验将原始数据中的异常点作为正常点，正常点作为异常点.

Arrhythmia[42]：是离群值检测数据集(Outlier Detection DataSets, ODDS)中的一个数据集，包含 279 种特征的多类分类数据集. 本实验丢弃了原始数据中的 5 种无用特征（处理后的数据共有 274 种特征），将条数最少的 8 类（第 3、4、5、7、8、9、14、15 类）视为异常数据，其他类被视为正常数据.

Satellite[42]：与 Arrhythmia 数据集一样，都属于 ODDS，也是一个多分类数据集，包括 36 种特征. 本实验将条数最少的 3 类（第 2、4、5 类）视为异常数据，其他类被视为正常数据.

表 1 数据集的统计信息

数据集	维度	数据条数	异常数据比例
KDDCUP99	120	494021	0.2
Arrhythmia	274	452	0.15
Satellite	36	6435	0.32

4.2 评价指标

与文献[4, 18]等本领域相关成果一样，本文主要采用精确率(Precision)、召回率(Recall)和 F1-score 等评价指标，其统计来源为异常检测的分类混淆矩阵，如表 2 所示. 通常，我们期望这些评估指标的值尽可能大.

精确率：体现了检测结果为正常样本中正常类的比例，计算方法如下：

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

召回率：体现了正常类被正确识别的比例，计算方法如下：

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

F1-score：基于精确率和召回率两项指标计算，其作用在于当精确率与召回率都无法比较模型的综合性能时（例如：精确率高，但召回率低），F1-score 作为精确率与召回率的一种折中方式来比较模型的综合性能. 其计算方法如下：

$$\text{F1-}score = \frac{1}{\alpha/Precision + (1+\alpha)/Recall} \tag{18}$$

其中，α的改变可以实现对精确率和召回率的折中，一般情况下，α 取值为 0.5[43].

表 2 异常检测分类混淆矩阵

真实类别	检测正常	检测异常
正常类	*TP*	*FN*
异常类	*FP*	*TN*

4.3 实验参数

经过基于不同参数的大量实验，SCA-UDADM 在不同数据集下的基本运行参数设定如表 3 所示：

表 3　不同数据集下基本运行参数设置

数据集	Iteration	Batchsize	K	M	λ_1	λ_2	λ_3
KDDCUP99	300	1024	15	4	0.1	0.005	10
Arrhythmia	20000	128	5	2	0.1	0.005	0.001
Satellite	3000	512	15	4	0.1	0.005	0.005

SCA-UDADM在不同数据集下的具体实现架构如表 4 所示. 其中，$FC(D_{in},D_{out},\sigma)$表示具有 D_{in} 个输入神经元、D_{out} 个输出神经元以及激活函数为 σ 的全连接多层神经网络. 同样地，$GAT(D_{in},D_{out},\sigma)$表示具有 D_{in} 个输入神经元、D_{out} 个输出神经元以及嵌入层激活函数为 σ 的图注意力层. 对于作为对比实验的其他方法，参数设定与文献[7]相同. 每次实验独立运行 10 次，取平均值作为最终结果.

表 4　不同数据集下实现框架设置

模型	KDDCUP99	Arrhythmia	Satellite
双路编码器	FC(120,64,Tanh) FC(64,32,Tanh) GAT(120,32,Tanh) FC(32,8,None)	FC(272,32,Tanh) GAT(272,32,Tanh) FC(32,2,None)	FC(36.16,Tanh) GAT(36,16,Tanh) FC(16,2,None)
解码器	FC(8,32,Tanh) FC(32,64,Tanh) FC(64,120,Tanh)	FC(2,10,Tanh) FC(10,274,Tanh)	FC(2,16,Tanh) FC(16,36,Tanh)
估计网络	FC(10,20,Tanh) FC(20,8,Tanh) FC(8,4,Softmax)	FC(4,10,Tanh) FC(10,2, Softmax)	FC(4,10,Tanh) FC(10,4, Softmax)

4.4　实验结果与分析

4.4.1　异常检测实验

实验根据异常在数据集中的比例确定阈值，(如假设异常样本比例为 20%，则能量最高的 20%样本被视为异常)，将能量高于阈值的样本认定为异常. 数据集的分割与文献[7]中的设置相同：50%的随机正常样本用于训练，其余用于测试. 结果如下：

KDDCUP99：实验结果如表 5 所示. 从中可以看出，SCA-UDADM 在 KDDCUP99 这种大型数据集上的精确率为 96.01%、召回率为 97.53%，F1-score 为 97.46%.（1）对比机器学习方法（OC-SVM 和 IF），在这种大型数据集上，机器学习方法在可以得到很好训练的情况下，由于受到数据维度和数据集规模的影响，IF 的各项指标比 SCA-UDADM 均低于3.8%以上，OC-SVM的各项指标比SCA-UDADM均低于 20%以上.（2）对比深度学习方法（DSEBM、DAGMM、AnoGAN 和 ALAD），SCA-UDADM 的各项指标比其中效果最好的 ALAD 均高出 1.76%以上. 总体而言，在大型数据集下，深度学习方法的效果比机器学习方法优秀，而且在考虑了数据关联性后的 SCA-UDADM 效果最优. 这就说明，由于 DSEBM、DAGMM、AnoGAN 和 ALAD 等模型未考虑数据的关联性从而导致数据训练不够充分，结果不如 SCA-UDADM. 相反，SCA-UDADM 通过样本之间的关联性分析从而可以利用数据集中更多的隐含信息进行训练，不仅提高了训练效率，而且达到了较好的预期效果.

表 5　KDDCUP 数据集异常检测结果

模型	精确率	召回率	F1-score
OC-SVM	74.57%	85.23%	79.54%
IF	92.16%	93.73%	92.94%
DSEBM-r	85.21%	64.72%	73.28%
DSEBM-e	86.19%	64.66%	73.99%
DAGMM	92.97%	94.42%	93.69%
AnoGAN	87.86%	82.97%	88.56%
ALAD	94.27%	95.77%	95.01%
SCA-UDADM	**96.01%**	**97.53%**	**97.46%**

Arrhythmia：实验结果如表 6 所示. 从中可以看出，SCA-UDADM 在数据量较小但数据维度较大的 Arrhythmia 数据集上的精确率为 56.41%、召回率为 57.89%，F1-score 为 57.14%.（1）对比机器学习方法（OC-SVM 和 IF），由于受到数据量小造成的学习能力不足和数据高维度影响，导致效果较好的 IF 的各项指标均低于 SCA-UDADM.（2）对比深度学习方法（DSEBM、DAGMM、AnoGAN 和 ALAD），其中效果最好的 ALAD 的各项指标比 SCA-UDADM 均低出 4.76%以上，甚至不如传统的机器学习方法. 这就说明，在数据规模小、样本维度高的数据集下，DSEBM、DAGMM、AnoGAN 和 ALAD 等深度学习方法由于无法得到充分训练，导致结果最差. 而 SCA-UDADM 通过考虑样本之间的关联性，能够充分利用数据集中更多的隐含信息进行训练，可以很好地弥补这一缺点，并在高维度数据集下也优于机器学习方法.

表 6　Arrhythmia 数据集异常检测结果

模型	精确率	召回率	F1-score
OC-SVM	53.97%	40.82%	45.81%
IF	51.47%	54.69%	50.03%
DSEBM-r	15.15%	15.13%	15.10%
DSEBM-e	46.67%	45.65%	46.01%
DAGMM	49.09%	50.78%	49.83%
AnoGAN	41.18%	43.75%	42.42%
ALAD	50.00%	53.13%	51.52%
SCA-UDADM	**56.41%**	**57.89%**	**57.14%**

Satellite：实验结果如表 7 所示. 从中可以看出，SCA-UDADM 在 Satellite 这种异常数据较多的数据集上的精确率为 81.99%、召回率为 82.75%，*F*1-score 为 82.37%.（1）对比机器学习方法（OC-SVM 和 IF），SCA-UDADM 的精确率比效果较好的 IF 高 21.18%，*F*1-score 高 7.33%. 虽然 IF 的召回率较高表明其有较好的识别正常样本的能力，但对于异常检测而言，无法保证较高的检测精确率是无法接受的.（2）对比深度学习方法（DSEBM、DAGMM、AnoGAN 和 ALAD），SCA-UDADM 的各项指标比其中效果最好的 DAGMM 均高出 1.15%以上. 总体而言，在异常样本较多的情况下，一般的深度学习方法要优于最好的机器学习方法，而且在考虑了数据关联性后的 SCA-UDADM 效果最优. 这就说明，在异常数据较多的数据集下，SCA-UDADM 在充分发挥深度学习优势的基础上，深入挖掘了数据的隐含关联性，能够正确识别绝大部分的异常.

表 7 Satellite 数据集异常检测结果

模型	精确率	召回率	*F*1-score
OC-SVM	52.42%	59.99%	61.07%
IF	60.81%	94.89%	75.04%
DSEBM-r	67.84%	68.56%	68.22%
DSEBM-e	67.79%	68.61%	68.18%
DAGMM	80.77%	81.60%	81.19%
AnoGAN	71.19%	72.03%	71.59%
ALAD	79.41%	80.32%	79.85%
SCA-UDADM	**81.99%**	**82.75%**	**82.37%**

综合以上实验结果可以看出，SCA-UDADM 在不同特点的数据集下，能有效利用数据间的隐含关系来更高效地训练模型，解决机器学习和其他深度学习模型由于数据的相似性、数据量小和异常比例大而造成的训练不充分问题，从而验证了 SCA-UDADM 模型的有效性和优势.

4.4.2 噪声数据的影响

本节实验使用 KDDCUP99 数据集分析噪声数据对 SCA-UDADM 训练的影响. 该数据集数据量较大、数据维度较高、异常数据也较多，是最典型的异常检测数据集. 具体数据处理方案如下：随机选取 50%的数据样本用于测试，在剩余数据样本中分别注入 1%-5%的异常数据（噪声）构造等量的训练集. 同时我们选择经典的机器学习方法 OC-SVM 和结果较好的深度学习方法 DSEBM-e 和 DAGMM 进行对比. 结果如下：

OC-SVM：实验结果如表 8 所示. 从中可以看出，随着噪声数据的增加，OC-SVM 的性能显著下降，尤其在加入 5%的噪声之后，精确率下降到 11.55%. 这是因为机器学习方法本身对高维数据的特征学习能力较差，对噪声数据更加敏感. 整体而言，经典的机器学习方法对数据集的要求较高，需要高质量的训练数据才能达到优秀的效果.

表 8 OC-SVM 噪声实验

噪声比例	精确率	召回率	*F*1-score
1%	71.29%	67.85%	69.53%
2%	66.68%	52.07%	58.47%
3%	63.93%	44.70%	52.61%
4%	59.91%	37.19%	45.89%
5%	11.55%	33.69%	17.20%

DSEBM-e：实验结果如表 9 所示. 从中可以看出，随着噪声数据的增加，3 个指标也呈下降趋势. 虽然下降趋势比机器学习方法 OC-SVM 小，但在加入 5%的噪声之后，各项指标仍然下降了 15.5%以上. 这表明噪声对该模型的影响较大，包含噪声的低质量数据集对 DSEBM-e 模型训练会产生较大的负面影响，从而影响模型的最终效果.

表 9 DSEBM-e 噪声实验

噪声比例	精确率	召回率	*F*1-score
1%	68.95%	71.35%	70.65%
2%	67.80%	68.76%	68.27%
3%	62.13%	63.67%	62.89%
4%	57.04%	58.13%	57.58%
5%	53.45%	53.75%	53.60%

DAGMM：实验结果如表 10 所示. 从中可以看出，随着噪声数据的增加，3 个指标也呈下降趋势. 虽然 DAGMM 下降趋势比机器学习方法 OC-SVM 和深度学习模型 DSEBM-e 小，但在加入 5%的噪声之后，各项指标仍然下降了 6.97%以上. 这表明噪声对 DAGMM 的训练影响较大，模型的鲁棒性差. 其主要原因是噪声的加入导致 DAGMM 模型对样本概率密度的学习不准确.

表 10 DAGMM 噪声实验

噪声比例	精确率	召回率	*F*1-score
1%	92.01%	93.37%	92.68%
2%	91.86%	93.40%	92.62%
3%	91.32%	92.72%	92.01%
4%	88.37%	89.89%	89.12%
5%	85.04%	86.43%	85.73%

SCA-UDADM：实验结果如表 11 所示. 从中可以看出，随着噪声不断注入，模型的整体性能表现稳定. 加入 5%噪声之后，模型精确率、召回率和 *F*-score 分别为 94.35%、96.04%、95.30%，模型的各项指标之差稳定在 1.18%以内，效果很好. 这表明，相比于 DAGMM，SCA-UDADM 在注入噪声后，仍能利用样本间的关联性来消除噪声数据对样本概率密度学习的影响，从而保持理想的效果.

表 11　SCA-UDADM 噪声实验

噪声比例	精确率	召回率	*F*1-score
1%	95.53%	97.04%	96.28%
2%	95.32%	96.82%	96.06%
3%	94.83%	96.33%	95.58%
4%	94.62%	96.12%	96.36%
5%	94.35%	96.04%	95.30%

综合以上实验结果可以看出，OC-SVM、DSEBM-e 和 DAGMM 这 3 种模型对噪声数据的敏感性强，也就说明这 3 种模型对训练数据的健康程度要求较高. 相反，SCA-UDADM 能够充分利用样本的关联性来消除噪声数据的影响，这就说明 SCA-UDADM 可以在低质量的训练数据上依然表现出优秀的实验结果，具有优秀的稳定性.

4.4.3　参数敏感性

本节实验对构图过程中的参数 *K* 的不同取值进行实验以验证参数 *K* 对模型的影响. 经过测试后，实验将 *K* 的取值区间设定为[5,19]，分别在 3 个数据集上进行实验. 每组实验运行 10 次，取结果的平均值. 结果如下：

KDDCUP99：实验结果如图 5 所示. 从中可以看出，在不同的 *K* 值下，精确率稳定在[95.00%, 96.50%]内，浮动范围保持在 1.50%以内；召回率稳定在[96.50%, 97.60%]内，浮动范围保持 1.10%以内；*F*1-score 稳定在[95.90%, 97.10%]，浮动范围保持在 1.20%以内. 这就说明，在 KDDCUP99 这种大型数据集下，SCA-UDADM 对参数 *K* 的设定不敏感.

Arrhythmia：实验结果如图 6 所示. 从中可以看出，在不同的 *K* 的设定下，精确率稳定在[54.00%, 56.41%]内，浮动范围保持在 1.41%以内；召回率稳定在[56.01%, 57.89%]内，浮动范围保持 1.88%以内；*F*1-score 稳定在[55.90%, 57.14%]，浮动范围保持在 1.24%以内. 这就说明，在 Arrhythmia 这种数据量较小、维度高的数据集下，参数 *K* 对 SCA-UDADM 的影响较小.

Satellite：实验结果如图 7 所示. 从中可以看出，在不同的 *K* 的设定下，精确率稳定在[81.69%, 82.06%]内，浮动范围保持在 0.37%以内；召回率稳定在[82.29%, 82.80%]内，浮动范围保持 0.51%以内；*F*1-score 稳定在[82.03%, 82.37%]，浮动范围保持在 0.34%以内. 这就说明，在 Satellite 这种数据量适中、维度较高且异常数据较多的数据集下，SCA-UDADM 更加稳定.

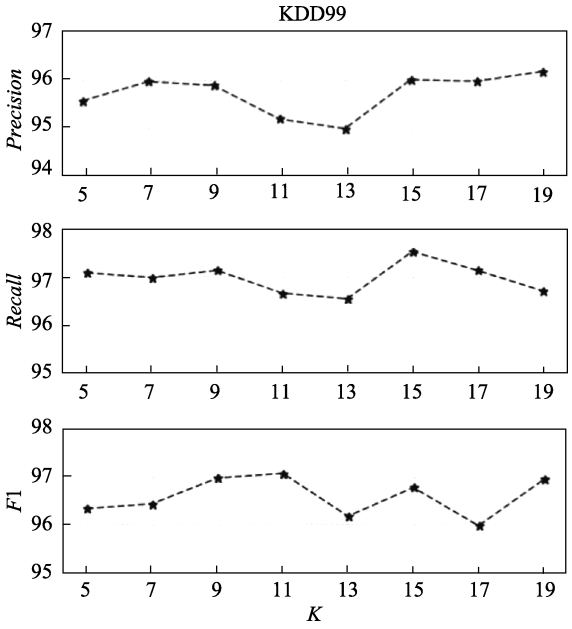

图 5　KDDCUP99 数据集参数敏感性实验

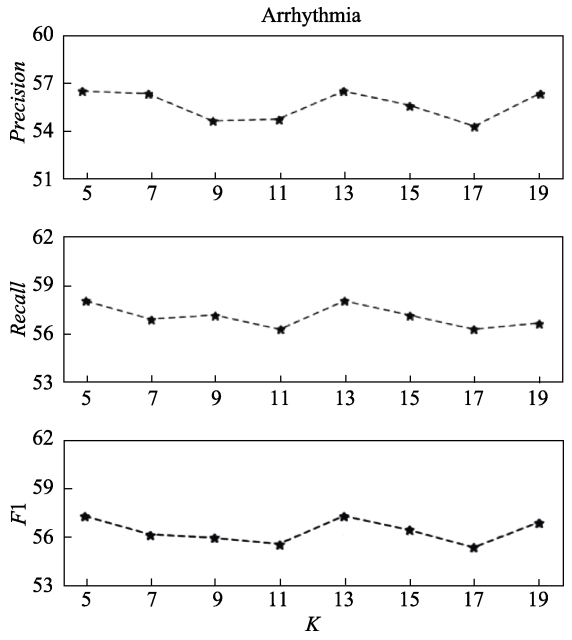

图 6　Arrhythmia 数据集参数敏感性实验

综合以上实验结果可以看出，参数 *K* 的变化对 SCA-UDADM 模型的性能影响很小，实验结果稳定. 这就体现了数据关联性建模参数没有对模型的

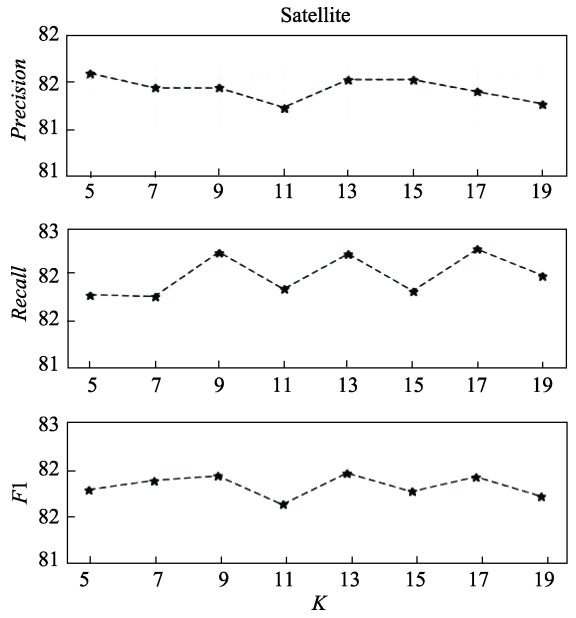

图 7 Satellite 数据集参数敏感性实验

结果产生较大影响，即模型对参数 K 不敏感.

4.4.4 模型框架消融实验

为了验证模型中各部分的有效性，本节使用不同的模型框架在 KDDCUP99 数据集上进行消融实验，实验参数设置同 4.3 节. 不同的模型框架如下：

（1）SCA-UDADM(only dualAE)，未使用估计网络以及正则化项. 使用自编码器的重构误差作为损失函数，并且将重构误差作为样本的能量；

（2）SCA-UDADM(dualAE & regular)，未使用估计网络，在损失函数中添加了正则化项，其余同（1）模型；

（3）SCA-UDADM(without GAT)，采用仅包含特征编码器的单路自编码器，其余同本文模型；

（4）SCA-UDADM(only GAT)，采用仅包含图编码器的单路自编码器，其余同本文模型；

（5）SCA-UDADM 模型，本文最终模型.

实验结果如表 12 所示. 从表 11 中可以看出，模型(2)的各项指标要优于模型(1), 其中 $F1$-score 比模型（1）高 2.49%；但各项指标均不如最终模型（5），其中 $F1$-score 分别低 4.58%和 3.09%. 这表明了本文最终模型中，正则化项和估计网络模块能够有效提升模型检测性能. 对于两个单路自编码器模型（3）和（4），模型（3）的各项指标比模型（4）的各项指标均低 1%左右，二者的各项指标都比本文最终模型（5）低，其中 $F1$-score 分别低 1.7%和 1.0%. 这表明了最终模型中，双路自编码器模块能够提升模型检测的综合性能.

表 12 SCA-UDADM 消融实验

模型	精确率	召回率	$F1$-score
SCA-UDADM(only dualAE)	91.81%	91.89%	91.88%
SCA-UDADM(dualAE & regular)	94.35%	94.37%	94.37%
SCA-UDADM(without GAT)	94.76%	94.77%	94.76%
SCA-UDADM(only GAT)	95.45%	95.46%	95.46%
SCA-UDADM	**96.01%**	**97.53%**	**97.46%**

综合以上实验结果表明，SCA-UDADM 中的各模块对模型的性能提升都有正面作用. 其中双路自编码器模块以及数据关联模块的使用，能够帮助模型挖掘出样本中更有效的隐含信息，从而使模型能够获得更好的检测效果.

4.4.5 特征融合实验

模型设计实现中，公式（6）采用基于相加的特征融合方式(ADD)，而特征融合还包括级联方式(CONCAT). 因此，对两种融合方式进行对比，实验参数设置同 4.3 节，实验结果如表 13 所示. 从中可以看出，CONCAT 比 ADD 在精确度上高 0.74%，在召回率上低 0.78%，在 $F1$-score 上低 1.31%. 因此，ADD 比 CONCAT 的综合性能更优. 这主要是由于模型在学习相同维度的嵌入表示时，CONCAT 需要双路自编码器学习两个更低维的子嵌入，可能会导致少部分的信息丢失；而 ADD 却不会有此问题. 因此，本文最终模型选取了 ADD 融合方式.

表 13 SCA-UDADM 特征融合实验

模型	精确率	召回率	$F1$-score
SCA-UDADM(CONCAT)	96.74%	96.75%	96.75%
SCA-UDADM(ADD)	96.01%	97.53%	97.46%

4.4.6 数据可视化

为了探索学习过程中数据嵌入的质量，本节实验对不同方法的样本嵌入进行可视化比较. 具体而言，实验将 DSEBM、DAGMM 和 SCA-UDADM 学习到的样本的样本嵌入作为 t-SNE 工具的输入，从 KDDCUP99 的测试集中随机选择 40000 个数据样本，然后生成二维空间的样本嵌入可视化，结果如图 8、图 9 和图 10 所示，其中深色对应于正常类别，浅色对应于异常类别.

DSEBM 和 DAGMM 的可视化结果中，正常数据分布虽然有部分集中于一个范围内，但是仍然有很多数据分布散乱. 而 SCA-UDADM 的可视化结果中，正常数据和异常数据分布在不同的空间范围内，异常数据进行了有效聚集. 对比可以看出，SCA-UDADM 由于考虑了数据样本间的关联特征，将数

据进行了更有效地分离，从而可以将异常数据进行有效识别，为后续的异常检测提供了一个很好的先决条件. 这也是 SCA-UDADM 比 DSEBM 和 DAGMM 有更好的检测效果的直接原因.

图 8　DSEBM 数据可视化

图 9　DAGMM 数据可视化

图 10　SCA-UDADM 数据可视化

综合以上所有实验可以表明，SCA-UDADM 在异常检测上具有更好的性能表现：异常检测实验体现了模型的有效性；噪声数据实验表明了模型的训练对数据质量要求较低，模型训练简单；参数敏感性实验体现了模型对数据关联建模参数不敏感；数据可视化实验验证了在加入数据关联特征后，模型能够学习到更加优秀的数据特征嵌入，更好地完成正常与异常的分离与区分.

5　总结与展望

深度学习模型应用于异常检测是当前相关应用研究的热点之一. 在异常检测中，数据间通常没有显性的关系，大量隐形的非线性关系隐藏在数据样本中. 然而，当前的诸多研究成果中很少有涉及数据间的关联性分析，使得检测效果在不同的应用背景下、在不同的性能指标上表现差强人意. 基于此，本文提出了 SCA-UDADM 模型，重点研究了数据样本之间的关联性并利用深度学习模型进行训练，将数据样本间的信息和数据原始特征空间的信息充分利用，从而完成异常检测任务：采用图结构对样本进行数据关联性建模，采用双路自编码器对原始特征空间信息和图结构特征空间信息进行特征提取与融合，充分利用特征空间的各种有效信息，最终基于高斯混合模型设计估计网络学习样本的概率分布来完成异常检测.

本文在本领域典型的 3 种公开数据集上进行了大量实验，结果证实了深入挖掘数据样本间的关联性并加以利用可显著提升模型的分类效果，验证了该方法的可行性. 今后，我们将继续针对不同应用背景，如图像、视频等，深入挖掘数据间的各种隐性关系，并关注于如何建立一个动态模型以适应更复杂多变的应用背景.

参　考　文　献

[1] Murali S, Jamalipour A. A lightweight intrusion detection for Sybil attack under mobile RPL in the Internet of things. IEEE Internet of Things Journal, 2020, 7(1): 379-388

[2] Zhao C H, Yao X F. Progressive line processing of global and local real-time anomaly detection in hyperspectral images. Journal of Real-Time Image Processing, 2019, 16(6): 2289-2303

[3] Nawaratne R, Alahakoon D, De Silva D, et al. Spatiotemporal anomaly detection using deep learning for real-time video surveillance. IEEE Transactions on Industrial Informatics, 2020, 16(1): 393-402

[4] Fernandes G, Rodrigues J J P C, Carvalho L F, et al. A comprehensive survey on network anomaly detection. Telecommunication Systems, 2019, 70(3): 447-489

[5] Yuan F N, Zhang L, Shi J T, et al. Theories and applications of auto-encoder neural networks: a literature survey. Chinese Journal of Computers, 2019, 42(01): 203-230 (in Chinese)

(袁非牛, 章琳, 史劲亭等. 自编码神经网络理论及应用综述. 计算机学报, 2019, 42(01): 203-230)

[6] Fu X, Shen Y T, Fu L H, et al. An optimized sparse auto-encoder network based on feature clustering. Acta Electronica Sinica, 2018, 46(05): 1041-1046 (in Chinese)
(付晓, 沈远彤, 付丽华等. 基于特征聚类的稀疏自编码快速算法. 电子学报, 2018, 46(05): 1041-1046)

[7] Zong B, Song Q, Min M R, et al. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection //Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018: 1-19

[8] Amarasinghe K, Kenney K, Manic M. Toward explainable deep neural network based anomaly detection//Proceedings of the 11th International Conference on Human System Interaction. Gdansk, Poland, 2018: 311–317

[9] Tran C P, Tran D K. Anomaly detection in POSTFIX mail log using principal component analysis//Proceedings of the 10th International Conference on Knowledge and Systems Engineering. Ho Chi Minh City, Vietnam, 2018: 107-112

[10] Wang M X, Zhou HC, Chen J. A moving window principal components analysis based anomaly detection and mitigation approach in SDN network. KSII Transactions on Internet and Information Systems, 2018, 12(8): 3946-3965

[11] Tang K, Liu T Z, Xi, X Y, et al. Power transformer anomaly detection based on adaptive kernel fuzzy C-means clustering and kernel principal component analysis//Proceedings of 2018 Australian and New Zealand Control. Melbourne, Australia, 2018: 318-323

[12] O'Reilly C, Gluhak A, Imran M A. Distributed anomaly detection using minimum volume elliptical principal component analysis. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(9): 2320-2333

[13] Meng Q, Catchpoole D, Skillicom D, et al. Relational autoencoder for feature extraction//Proceeding of 2017 International Joint Conference on Neural Networks. Anchorage, USA, 2017: 364-371

[14] Feng Q S, Dou Z, Li C M, et al. Anomaly detection of spectrum in wireless communication via deep autoencoder//Proceedings of the 8th International Conference on Computer Science and its Applications/11th International Conference on Ubiquitous Information. Bangkok, Thailand, 2017: 259-265

[15] Chen R, Zhang F B, Xi L. Anomaly detection algorithm based on FCM with improved krill herd//Proceedings of 2018 International Symposium on Power Electronics and Control Engineering. Xian, China, 2018, Article Number: 042028: 1-11

[16] Ding N, Ma H X, Gao H B, et al. Real-time anomaly detection based on long short-Term memory and Gaussian Mixture Model. Computers and Electrical Engineering, 2019,79: Article Number: UNSP 106458: 1-11

[17] Li K L, Huang H K, Tian S F, et al. Improving one-class SVM for anomaly detection//Proceedings of the 2nd International Conference on Machine Learning and Cybernetics. Xi'an, China, 2003,5: 3077–3081

[18] Miao X D, Liu Y, Zhao H Q, et al. Distributed online one-class support vector machine for anomaly detection over networks. IEEE Transactions on Cybernetics, 2019, 49(4): 1475-1488

[19] Liu K, Liu W, Ma H D, et al. Generalized zero-shot learning for action recognition with web-scale video data. World Wide Web-Internet and Web Information Systems, 2019, 22(2): 807–824

[20] Zhang K, Kang X, Li S. Isolation forest for anomaly detection in hyperspectral images//Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama, Japan, 2019: 437–440

[21] Singh K, K James Mathai K J. Performance comparison of intrusion detection system between deep belief network (DBN) algorithm and state preserving extreme learning machine (SPELM) algorithm// Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies. Coimbatore, India, 2019: 1–7

[22] Brown A, Tuor A, Hutchinson B, et al. Recurrent neural network attention mechanisms for interpretable system log anomaly detection//Proceedings of the 1st Workshop on Machine Learning for Computing Systems. Tempe, USA, 2018: 1-8

[23] Greff K, Srivastava R K, Koutn ík J, et al. LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 2016, 28(10): 2222-2232

[24] Zhai S F, Cheng Y, Lu W N, et al. Deep structured energy based models for anomaly detection//Proceedings of the 33rd International Conference on International Conference on Machine Learning. Florida, USA, 2016: 1100-1109

[25] Schlegl T, Seeböck P, Waldstein S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery//Proceedings of the 2017 International conference on information processing in medical imaging. Boone, USA, 2017: 146-157

[26] Zenati H, Romain M, Foo C S, et al. Adversarially learned anomaly detection//Proceedings of the 2018 IEEE International Conference on Data Mining. Singapore, 2018: 727-736

[27] Huyan K, Fan X, Yu L T, et al. Graph based neural network regression strategy for facial image super-resolution. Journal of Software, 2018, 29(04): 914-925 (in Chinese)
(呼延康, 樊鑫, 余乐天等. 图神经网络回归的人脸超分辨率重建. 软件学报, 2018, 29(04): 914-925)

[28] Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains//Proceedings of the 2005 IEEE International Joint Conference on Neural Networks. Montreal, Cananda, 2005: 729–734

[29] Luo C, Wang J, Wang P F, et al. Coarse-grained pooled features learning in convolutional autoencoders. Acta Electronica Sinica, 2017, 45(10): 2390-2401 (in Chinese)
(罗畅, 王洁, 王鹏飞等. 卷积自编码器中粗粒度池化特征提取研究. 电子学报, 2017, 45(10): 2390-2401)

[30] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs//Proceedings of the 31st Annual Conference on Neural Information Processing Systems. Long Beach, USA, 2017: 1024-1034

[31] Ahmad T, Mao H, Lin L, et al. Action recognition using attention-joints graph convolutional neural networks. IEEE Access, 2020, 8: 305-313

[32] Ma F, Gao F, Sun J P, et al. Attention graph convolution network for image segmentation in big SAR imagery data. Remote Sensing, 2019, 11(21), Article Number: 2586: 1-21

[33] Zhang C H, James J Q, Liu Y. Spatial-temporal graph attention networks: a deep learning approach for traffic forecasting. IEEE

Access, 2019, 7: 166246-166256
[34] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks//Proceedings of the 15th European Semantic Web Conference. Crete, Greece, 2018: 593-607
[35] Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention//Proceedings of the 32 nd International Conference on Machine Learning. Lille, France, 2015, 2048-2057
[36] Zhong J X, Li N N, Kong W J,et al. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2019: 1237--1246
[37] Ding K, Li J, Bhanushali R, et al. Deep anomaly detection on attributed networks//Proceedings of the 2019 SIAM International Conference on Data Mining. Calgary, Canada, 2019: 594-602
[38] Fan, H Y, Zhang, F B, Li, Z Y. AnomalyDAE: Dual autoencoder for anomaly detection on attributed networks //Proceedings of the 45th International Conference on Acoustics, Speech, and Signal Processing. Barcelona, Spain, 2020: 5685-5689
[39] Pektaş A, Acarman T. Deep learning for effective Android malware detection using API call graph embeddings. Soft Computing, 2020, 24(2): 1027-1043
[40] Zenati H, Romain M, Foo C S, et al. Adversarially learned anomaly detection// Proceedings of the 2018 IEEE International Conference on Data Mining. Singapore, 2018: 727-736
[41] Siddique K, Akhtar Z, Khan F A, et al. KDD Cup 99 data sets: a perspective on the role of data sets in network intrusion detection research. Computer, 2019, 52(2): 41-51
[42] Ting K M, Chuan J T S, Liu F T. Mass: A new ranking measure for anomaly detection. IEEE Transactions on Knowledge and Data Engineering, 2009: 1-13
[43] Sepulveda J, Velastin S A. F1 score assessment of Gaussian mixture background subtraction algorithms using the MuHAVi Dataset//Proceedings of the 6th International Conference on Imaging for Crime Prevention and Detection. London, UK, 2015: 1-6

XI Liang, Ph.D., associate professor, master supervisor. His research interests include artificial intelligence and its application, network and information security, IoT security, software defined network, data mining, deep learning, etc.

WANG Rui-Dong, M.S., candidate. His research interests include network and information security, deep learning, etc.

FAN Hao-Yi, Ph.D. candidate. His research interests include artificial intelligence, data mining, deep learning, etc.

ZHANG Feng-Bin, Ph.D., professor, Ph.D. supervisor. His research interests include artificial intelligence, network and information security, etc.

Background

Anomaly detection models have been applied in many fields, such as network security, medical image diagnosis, video monitoring, and so on. The keys are the completeness of the detection rule set and the efficiency of the detection method. At present, relevant scholars have introduced many methods of artificial intelligence and machine learning, so that anomaly detection systems can effectively absorb the latest abnormal features in real time and update rule sets to ensure effective detection performances. Deep learning is the further development of machine learning and artificial intelligence technology. The recent anomaly detection models based on deep learning model shows a more satisfying performance.

Therefore, this paper focuses on the unsupervised deep learning model and its application to anomaly detection, basing on the data-correlation analysis between the samples which has been neglected before, combining with the feature of traditional analysis methods, proposes a sample-correlation-aware unsupervised deep anomaly detection model. Firstly, the samples and their edges are in forms of undigraph structure, to analyze the sample-correlation. Secondly, a dual-encoder with traditional feature encoder and graph encoder is employed to encode both the sample's traditional features and correlation information among samples. Then, a decoder is designed for data reconstruction. Finally, a estimation network with Gaussian Mixture Model is utilized to estimate the density of samples and compute the their energy to detect the anomalies. A large number of experiments were conducted on three representative data sets, and the result comparisons of the latest representative models in different experimental backgrounds shows that the model proposed in this paper has better effects on Precision, Recall, f1-score and data visualization, which prove its feasibility and validity.

This work has been supported in part by the Natural Science Foundation of Heilongjiang Province (No. F2018019). It is a further in-depth study after introducing deep learning models on the basis of the existing achievements of these projects. Our research group has done lots of research on artificial intelligence, deep learning and their applications to anomaly detection and related application areas. Recent related works have been published in international journals and conferences, such as, Neural Computing and Applications, Journal of Biophotonics, Journal of Computer Research and Development, Journal of Software, Control and Decision, The 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2020), and so on.