

## 基于边缘计算的物联网监测系统中利用 自编码神经网络实现的异常检测

于天琪<sup>1</sup>, 朱咏絮<sup>2</sup>, 王现斌<sup>1</sup>

(1. 西安大略大学电气与计算机工程学院, 安大略 伦敦 N6A 5B9;

2. 拉夫堡大学沃尔夫森机械、电气与制造工程学院, 拉夫伯勒 LE11 3TU)

**摘要:** 物联网以其便于搭建及成本低廉等优点, 被广泛地应用于环境监测中。在大规模物联网监测系统中, 云平台一直被用作远程的数据和控制中心, 然而系统中大量的数据上传和处理给云平台的带宽负载和实时性反馈等方面带来了巨大挑战。因此, 提出了基于边缘计算的物联网监测系统框架。边缘计算作为该系统的中间层, 能够为终端设备提供实时的本地服务, 同时能够通过初步分析分流云平台的计算任务进而降低数据上传量。在此基础上, 进一步提出了一种利用自编码神经网络实现的异常检测方法。采用实地采集的海洋气候监测数据进行仿真分析, 仿真结果表明, 本文提出的基于自编码神经网络的异常检测方法能够充分利用采集数据的空间相关性并准确地检测出异常数据。

**关键词:** 自编码神经网络; 异常检测; 边缘计算; 物联网

**中图分类号:** TP319

**文献标识码:** A

**doi:** 10.11959/j.issn.2096-3750.2018.00076

## Autoencoder neural network-based abnormal data detection in edge computing enabled large-scale IoT systems

YU Tianqi<sup>1</sup>, ZHU Yongxu<sup>2</sup>, WANG Xianbin<sup>1</sup>

1. Department of Electrical and Computer Engineering, Western University, London N6A 5B9, Canada

2. Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough LE11 3TU, U.K.

**Abstract:** Given the advantages of low cost and easy deployment, large-scale Internet of things (IoT) has been deployed for environment monitoring pervasively. Within such systems, cloud platform is typically utilized as a remote data and control center. However, tremendous amount of data uploading and processing induce huge challenges on bandwidth load and real-time data gathering. In order to overcome these challenges, edge computing enabled IoT system architecture was proposed for environmental monitoring. As the intermediate layer, local processing could be supported for end devices with low latency and assist with preliminary analysis to offload computational tasks from cloud and the amount of data uploading could be reduced. Based on this system architecture, an autoencoder neural network-based abnormal data detection scheme was developed newly. Performance evaluation has been conducted based on the practical oceanic atmospheric data. Simulation results indicate that the proposed scheme can accurately detect the abnormal data by fully exploiting the spatial data correlation.

**Key words:** autoencoder neural network, abnormal data detection, edge computing, IoT

### 1 引言

随着物联网技术的发展, 物联网系统以其成本低廉以及易于部署的特性被广泛应用于环境监测

中, 如森林火灾监测、海洋气候监测等。在系统部署过程中, 大量的物联网终端设备被添加到系统中, 随之而来的大规模数据处理分析以及存储, 给基于云平台的物联网系统带来如带宽负载、实时性

反馈以及网络连接稳定性等诸多方面新的困难和挑战<sup>[1]</sup>。在此情况下，边缘计算技术被迅速地纳入大规模物联网监测系统中<sup>[2]</sup>。边缘计算由边缘计算设备提供支持，常用的设备有智能网关、轻量级服务器以及小型基站等。在物联网系统中，边缘计算节点位于系统的中间层，相比于云平台更加贴近物联网终端设备。因此，边缘计算节点能够向终端设备提供本地服务，以提高服务及信息反馈的实时性，进而降低由于终端设备与云平台的远程交互所带来的回程时延。与此同时，边缘计算设备也能够提供数据的初步处理，分流云平台的计算任务，降低上传至云平台的数据量，从而达到减轻主干链路带宽负载的目的。

在本文中，首先提出基于边缘计算的物联网监测系统框架，主要由无线传感器监测节点、边缘计算节点和云平台构成，对每个组成部分的功能及相互之间的通信过程进行详细探讨。在此系统框架的基础上，提出一种利用自编码器神经网络（以下简称“自编码器”）实现的数据异常检测方法。

在大规模物联网监测系统中，数据异常时常发生。造成数据异常的原因有多种，如监测环境中的突发事件和传感器自身的异常。无论是何种原因，数据异常都直接关联环境以及物联网系统自身的安全。因此，需要及时、准确地对数据异常进行检测。考虑边缘计算设备的能力和在系统中的特殊位置，本文提出了一种基于边缘计算的异常检测方法，一种人工神经网络模型即自编码器，用以捕捉和利用环境监测数据自身的空间相关性。之所以采用自编码器，主要是考虑所监测环境的复杂性，采集的数据虽然具有较强的规律性但可能呈现非简单的线性相关性。自编码器作为一种神经网络，有能力处理复杂的数据模型，既能够对线性相关的数据进行建模分析，也能够应对非线性相关的数据。考虑神经网络模型训练的复杂度，训练过程被部署在云平台进行，所得到的模型及参数将被回传至边缘计算节点用以进行后续的异常检测。在边缘计算节点，利用异常数据在通过自编码器编码与解码过程中所产生的较大波动对其进行检测。针对实地采集的海洋气候监测数据进行分析可知，本文提出的基于自编码器的异常检测方法，能够准确地检测出异常数据。

此外，文章末尾对基于边缘计算的物联网系统中存在的问题和潜在的研究方向进行了简要讨论。

## 2 基于边缘计算的物联网监测系统

基于边缘计算的物联网监测系统如图1所示，具体由3部分组成，包括无线传感器监测节点、边缘计算节点和云计算平台。

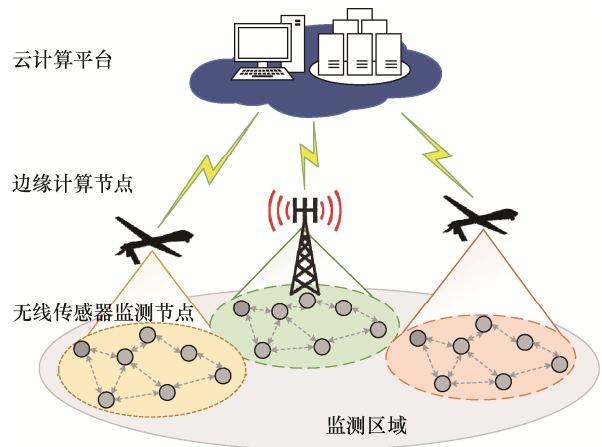


图1 基于边缘计算的物联网监测系统

### 2.1 无线传感器监测节点

在大规模物联网监测系统中，无线传感器节点是系统中最基础、最重要的组成部分。在监测区域，大量的无线传感器节点被随机或有组织地投放出去，用以实时感知并采集环境数据。如在江苏省无锡市部署的森林监测系统 GreenOrbs 中，无线传感器节点被安置在树木上，每一个节点都嵌入了温度、湿度、光照强度以及二氧化碳浓度等传感器，用以监测森林环境并实时检测及预防森林火灾。在基于边缘计算的物联网系统框架中，无线传感器节点主要与边缘计算节点进行通信，与云平台的通信也由边缘计算节点作为中继进行。一方面，可以降低对无线传感器节点通信能力的要求；另一方面，可以利用边缘计算节点的数据处理能力优势。

### 2.2 边缘计算节点

在大规模物联网监测系统中，监测节点周围已部署的小型地面基站可以作为边缘计算的节点。在一些难以进行基站部署的地区，无人机也可以作为移动边缘计算节点提供相应的边缘计算服务。从系统架构角度来看，边缘计算节点作为中间层存在，向无线传感器节点提供本地服务以提高服务质量及信息反馈的实时性，进而降低与云平台的远程交互所带来的时延。与此同时，能够提供数据的初步处理，分流云平台的计算任务，降低上传至云平台的数据量，从而减轻主干链路的带宽负载。

在提出的异常检测方法中，边缘计算节点具有至关重要的作用。在系统初始化阶段，作为中继设备将无线传感器节点采集的环境监测数据上传至云平台。在系统常规运行阶段，对采集的数据进行异常检测。一旦检测出异常情况，边缘计算节点会立即将异常报告给位于云平台的数据和控制中心，同时驱动位于底层的控制器，提供紧急应对方案，边缘计算节点的功能如图 2 所示。

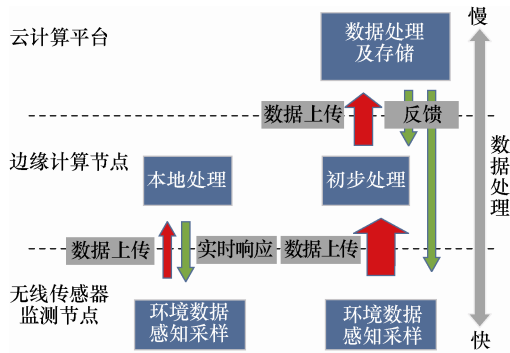


图 2 边缘计算节点的功能

### 2.3 云计算平台

云计算平台以其超强的计算能力和超大的存储空间，在系统中起到远程数据和控制中心的作用，主要用于大规模和高复杂度的数据处理及数据存储。在提出的异常检测方法中，使用自编码器模型以捕捉和利用环境监测数据自身的空间相关性。自编码器作为一种用于特征提取的人工神经网络模型，模型参数的训练具有较高的复杂度。无线传感器节点和边缘计算节点难以提供相应的计算能力以及复杂计算所带来的资源消耗，因此，模型训练过程被部署在云平台进行。生成的模型参数被回传至边缘计算节点，用于实时的异常检测。

## 3 基于自编码器的异常检测方法

本节包含 3 部分内容，首先，对物联网监测系统异常问题进行定义；然后，对自编码器的结构进行介绍；最后，提出基于自编码器的异常检测方法。

### 3.1 异常问题

在物联网监测系统中，异常问题主要是指传感器数据的异常，即在时域或空间域中具有高度规律性的传感器数据中出现不符合规律的数据<sup>[3]</sup>。美国国家海洋和大气管理局在太平洋位于西经 170° 的 7 个不同纬度的监测站所测得的海洋表层温度数据异

常如图 3 所示，温度数据的常规范范围在 28°C ~30°C 且呈现渐变趋势。然而当异常发生时，温度的测量值则突变为-9.99°C。造成数据异常的原因有多种，既可能是监测区域的突发事件如森林火灾，也可能是传感器节点自身的异常如硬件模块损坏、节点电量偏低以及无线通信过程中的干扰等。本文只针对数据异常进行检测，不对造成异常的原因进行分析。

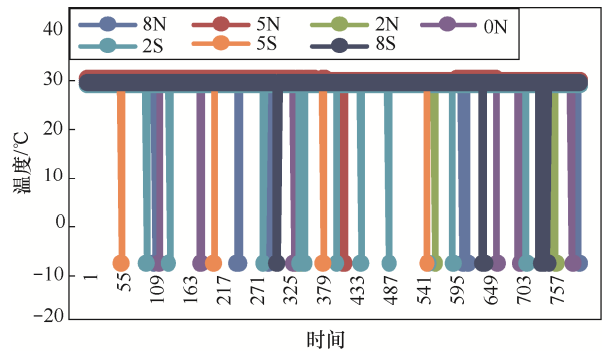


图 3 传感器数据异常

### 3.2 自编码器

自编码器是一种用于特征提取的神经网络模型，由输入层、隐藏层和输出层组成。单隐藏层自编码神经网络结构如图 4 所示。输入层到隐藏层之间的映射被称为编码器（encoder），相应的隐藏层到输出层之间的映射被称为解码器（decoder）。与其他神经网络模型不同的是，自编码器的输出是对输入的还原。编码器的数学表达式如式(1)所示。

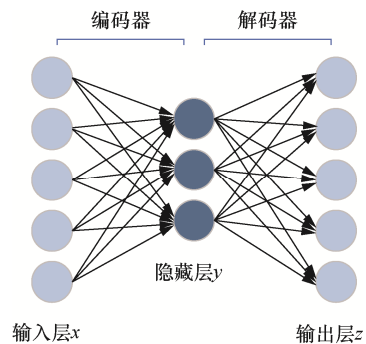


图 4 单隐藏层自编码神经网络结构

$$y = f(W \cdot x + b_f) \tag{1}$$

其中， $x$  为  $n$  维的输入层向量， $y$  为  $k$  维的隐藏层向量， $W_{[k \times n]}$  和  $b_f_{[k \times 1]}$  分别为编码器的加权矩阵和偏差向量。 $f(\cdot)$  是编码器的非线性映射函数，Sigmoid 函数常被用作编码器的映射函数。

解码器的数学表达式如式(2)所示。

$$z = g(V \cdot y + b_g) \quad (2)$$

其中,  $y$  为隐藏层向量,  $z$  为  $n$  维的输出层向量,  $V_{[n \times k]}$  和  $b_{g[n \times 1]}$  分别为解码器的加权矩阵和偏差向量。  $g(\cdot)$  为解码器的映射函数, 常用的解码器映射函数有 Sigmoid 函数和恒等函数。

最终的目标函数定义为输入向量和输出向量的平方误差, 如式(3)所示。

$$\begin{aligned} J_{\theta, \theta'} &= \frac{1}{m} \sum_{i=1}^m \|z^{(i)} - x^{(i)}\|_2^2 \\ &= \frac{1}{m} \sum_{i=1}^m \|g_{\theta'}(f_{\theta}(x^{(i)})) - x^{(i)}\|_2^2 \end{aligned} \quad (3)$$

其中,  $m$  为用于模型训练的数据量。通过目标函数最小化可以得到优化的模型参数  $\theta = \{W, b_f\}$  和  $\theta' = \{V, b_g\}$ 。小批量梯度下降算法(mini-batch based gradient descent)被用于模型训练从而得到相应的优化参数。

### 3.3 基于自编码器的数据异常检测方法

在本节中, 将详细地介绍基于自编码器的数据异常检测方法。图 5 简要展示了所提出的异常检测方法流程, 如图 5 所示, 该方法分为 3 个主要阶段, 分别为系统初始化阶段、模型训练阶段以及异常检测阶段。

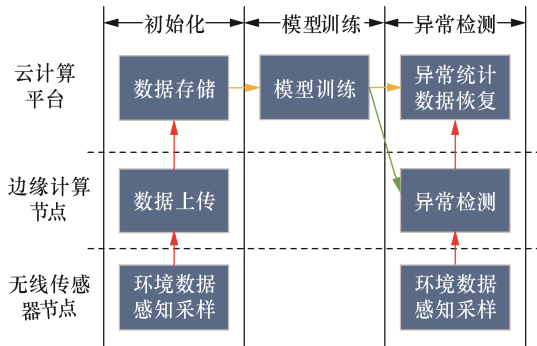


图 5 基于自编码器的数据异常检测方法流程

#### 3.3.1 系统初始化阶段

在系统初始化阶段, 无线传感器节点通过广播 beacon 信号的方式表明自身的存在, 并自组织形成无线传感器网络。组网后, 无线传感器节点周期性地感知并采集环境数据, 节点  $i$  在  $t$  时刻所生成的数据向量为

$$x_i^{(t)} = [x_{i,1}^{(t)}, x_{i,2}^{(t)}, \dots, x_{i,p}^{(t)}]^T \quad (4)$$

其中,  $p$  为节点  $i$  所监测物理变量的数量。

该数据通过距离节点最近且能够进行无线通信的边缘计算节点规律性地上传至云平台, 并在云平台进行存储。在此阶段, 边缘计算节点只具有数据中继的作用, 负责收集底层传感器数据, 并直接上传至云端, 不进行本地处理或其他预处理操作。

#### 3.3.2 模型训练阶段

模型训练阶段的操作均在云平台进行, 主要考虑神经网络训练的复杂度、物联网中各个组成部分的运算能力以及资源储备。该阶段操作基于第一阶段存储的历史数据进行, 在时间  $T$  内存储的无线传感器节点  $i$  的数据矩阵为

$$X_i = [x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(T)}] \quad (5)$$

首先, 对数据进行归一化处理, 以消除不同标度对数据所带来的影响。

$$\bar{x}_{i,j}^{(t)} = \frac{x_{i,j}^{(t)} - \min(x_{i,j})}{\max(x_{i,j}) - \min(x_{i,j})} \quad (6)$$

其中,  $x_{i,j} = [x_{i,j}^{(1)}, x_{i,j}^{(2)}, \dots, x_{i,j}^{(T)}]$  为节点  $i$  监测的物理变量  $j$  在时间  $T$  内存储的历史数据向量,  $\min(x_{i,j})$  和  $\max(x_{i,j})$  为  $x_{i,j}$  的最小值和最大值。

然后, 基于传感器数据的空间相关性, 利用聚类算法如自适应 K-均值算法<sup>[4]</sup>, 将无线传感器节点分簇, 由此生成  $K$  个簇。在分簇结果的基础上, 针对每一个簇建立自编码器模型, 并进行模型参数训练。对于任一簇  $k, k = 1, 2, \dots, K$ , 会得到一对参数  $\theta_k = \{W, b_f\}$  和  $\theta'_k = \{V, b_g\}$ 。分簇的结果以及自编码器模型训练所得到的参数将回传至边缘计算节点, 用于后续的异常检测。

#### 3.3.3 异常检测阶段

异常检测过程主要依托于边缘计算节点进行, 主要考虑边缘计算所带来的计算任务卸载以及数据下沉, 有利于提高异常检测的实时性, 并减轻与云平台相连的主干链路的带宽压力。

自编码器在对通过编码器后的数据进行解码还原时, 会产生一定的误差。如果原始数据中存在异常, 生成的误差相比于常规误差有较大波动, 本文所提出的异常检测方法主要利用该误差波动。在  $t$  时刻, 原始采样数据和还原值的平方误差定义如式(7)所示

$$\varepsilon = \|g_{\theta'}(f_{\theta}(\bar{x}^{(t)})) - \bar{x}^{(t)}\|_2^2 \quad (7)$$

其中,  $\bar{x}^{(t)}$  为  $t$  时刻进行归一化后的原始采样数据,  $g_{\theta'}(f_{\theta}(\bar{x}^{(t)}))$  为通过编码器和解码器后的数据。

异常检测算法的具体步骤如下。首先, 对数据进行归一化处理, 用以消除不同标度对数据所带来的影响。然后, 利用存储的自编码器模型参数对归一化后的数据进行平方误差的计算。如果平方误差值超过了预先设定的阈值, 则判定该时刻的采样数据中存在异常, 并将原始数据和异常警告同时上传至云平台做后续处理。如果平方误差值低于阈值, 则将通过编码器后的数据上传至云平台, 并在云平台进行数据解码还原, 以此减少数据的上传量。基于自编码器的数据异常检测方法的伪代码总结如算法 1 所示。

**算法 1** 基于自编码器的异常检测方法

- 1: 初始化阶段
- 2: 上传传感器数据至云平台并存储
- 3: 模型训练阶段
- 4: 数据归一化  $X \Rightarrow \bar{X}$
- 5: 无线传感器节点分簇
- 6: 模型训练  $\theta = \{W, b_f\}$ ,  $\theta' = \{V, b_g\}$
- 7: 异常检测阶段
- 8: 上传  $X^{(t)}$  至边缘计算节点, 归一化  $\bar{X}^{(t)}$
- 9: 计算  $\varepsilon = \|g_{\theta'}(f_{\theta}(\bar{X}^{(t)})) - \bar{X}^{(t)}\|_2^2$
- 10: if  $\varepsilon > \xi$  then
- 11: 检测为异常
- 12: 上传  $X^{(t)}$  以及异常警告至云平台
- 13: else
- 14: 上传  $f_{\theta}(\bar{X}^{(t)})$  至云平台
- 15: end if

**3.4 仿真结果**

美国国家海洋和大气管理局监测热带海洋气候 (TAO, tropical atmosphere ocean) 的数据被用于分析基于自编码器的异常检测方法的检测准确性, 具体的传感监测站分布如图 6 所示。

选取位于西经 170° 的 7 个监测站在 2018 年 8 月

21 日至 8 月 30 日期间每隔 10 min 所采集的海洋温度进行相关的异常检测分析。每个监测站选取海下深度分别为 25 m、50 m、75 m、100 m、125 m、150 m、175 m、200 m、300 m 和 500 m 处的温度测量值。根据地理位置, 将所选取的 7 个监测站划分为一个簇, 其中, 每一个监测站收集位于不同海下深度的 10 个温度值, 所以在每一个采样时刻, 可以收集来自这一簇的包含 70 个变量的数据向量。在选取的采样时间内, 每隔 10 min 采集一次数据, 在 10 天内共收到 1 440 条采样数据, 前 1 000 条数据用于模型训练。根据数据的特征和数量, 选取单隐藏层的自编码器模型, 编码器采用的映射函数为 Sigmoid 函数, 解码器选取的映射函数为 Relu 函数。小批量梯度下降算法被用于具体的模型参数训练, 选取第 1 001~1 100 条数据用来测试异常检测的准确性。在该 100 个采样时刻内数据还原生成的平方误差如图 7 所示。据统计, 在选取的 100 条采样数据中, 共有 44 条数据的 70 个变量中存在至少一处数据异常。由图 7 可以看到, 异常的出现会在数据还原时带来较大的平方误差, 选取适当的阈值 (图 7 中的红色线) 则可以进行相关的异常检测。

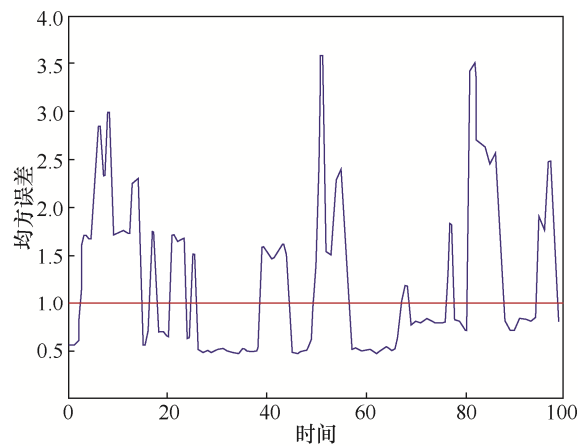


图 7 在选取的采样时刻内数据还原生成的平方误差

在设定阈值的情况下, 异常检测结果由真阳性率 (TPR, true positive rate) 和伪阳性率 (FPR, false

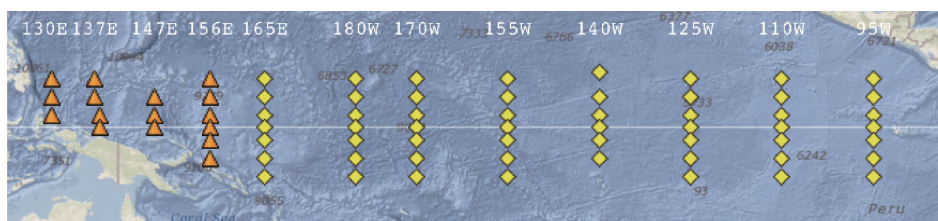


图 6 传感监测站分布

positive rate) 进行评估, 其中 TPR 为所有实际为异常的样本中, 被正确地判断为异常的比率; FPR 为所有实际为正常值的样本中, 被错误地判断为异常的比率。选取不同阈值可得到多组 TPR 和 FPR, 由此可以绘制接收者操作特征曲线 (ROC 曲线, receiver operating characteristic curve), 其中横轴为 FPR, 纵轴为 TPR。ROC 曲线下的面积 AUC (area under the curve) 常被用于衡量异常检测方法的优劣, AUC 的范围在 0~1, AUC 越接近 1, 则越接近理想检测器。

图 8 绘制了在单隐藏层选取不同维度时, 所得到的多条异常检测的 ROC 曲线。由图 8 可以看出, 3 条 ROC 曲线重合且 AUC 均为 1, 表明无论维度取值如何, 总可以找到一个合适的阈值, 使得该异常检测方法的 TPR 为 1, 同时保证 FPR 为 0。在选取不同隐藏层的维度时, 所得到的适宜阈值也不同, 当维度分别为 70、50 和 35 时, 阈值的适宜取值分别为 0.9、1.4 和 2。这是由于随着隐藏层的维度降低 (从 70 降至 35), 通过编码器的数据压缩程度会增加, 数据解码还原的准确性将随之下降, 即数据还原的误差上升。误差的整体升高不会影响异常检测的准确性, 但最终会导致适宜阈值的取值随之升高。尽管数据压缩程度的增加会使数据还原的准确性下降, 但是可以进一步降低上传至云平台的数据量, 两者之间可以通过具体应用的要求达成一定的平衡。

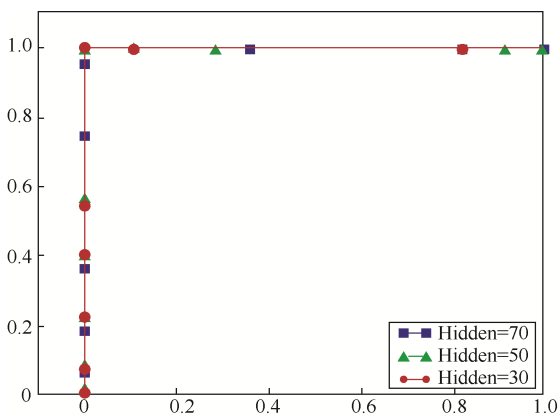


图 8 异常检测的 ROC 曲线

## 4 未来研究方向

### 4.1 基于人工智能的数据处理

随着人工智能研究的推进, 人工神经网络以及深度学习逐渐被应用于各个领域, 特别是计算机视

觉和大数据分析领域。随着物联网系统的大规模部署, 物联网系统中传感及控制数据量快速增长, 人工智能算法逐渐被引入物联网系统中用于数据的处理和分析<sup>[5]</sup>。人工智能算法的引入能够带来以下三方面的优势。首先, 传统的统计分析方法在面对海量物联网数据时效率较低, 人工智能算法能够提高数据处理的效率。其次, 人工智能算法能够通过模型训练过程学习数据的规律, 提高数据 (特别是非线性数据) 分析的准确性, 如数据分类、数据聚合、数据特征提取及时间序列预测等。此外, 相较于传统的分析方法, 人工智能算法能够更有效地对用户行为和周围环境进行分析, 并对用户行为和环境需求进行预判, 更好地提供定制化服务<sup>[6]</sup>。表 1 按功能划分并总结了一些常用人工智能算法。

表 1 常用人工智能算法

功能	算法
分类	决策树、随机森林、支持向量机、朴素贝叶斯、感知器、K-近邻、神经网络等
回归	线性回归、逻辑回归、K-近邻、回归树等
聚类	K 均值聚类、图论聚类、层次聚类、DBSCAN 等
降维	主成分分析、独立成分分析、典型相关分析、神经网络等
决策	强化学习 (Q 学习、SARSA 算法、时间差分学习等)

在 4.2 节任务卸载中, 将进一步讨论在面对如人工智能算法训练等高复杂度计算时, 如何合理地利用系统的资源分配计算任务。

### 4.2 任务卸载

任务卸载是指针对特定的计算任务, 通过决策算法选择系统内完成该任务的主要设备或平台。任务卸载决策算法的主要优化目标是 minimized 完成计算任务所消耗的时间或网络资源, 时间和资源包括不同设备平台间数据的通信以及完成计算任务本身所带来的全部消耗。

任务卸载是基于边缘计算的物联网系统的潜在研究方向之一<sup>[7]</sup>。基于边缘计算的物联网系统中的任务卸载如图 9 所示, 计算任务可以在终端设备进行, 也可以卸载到边缘计算设备或者云平台进行, 需要综合考虑三者的计算能力和三者之间的通信消耗。如以任务完成时间为优化目标时, 需要比较边缘计算设备完成计算任务的时间, 任务卸载到云平台时数据上传、云平台计算以及结果反馈的总时间, 进而决定是否将边缘计算的计算任务卸载至

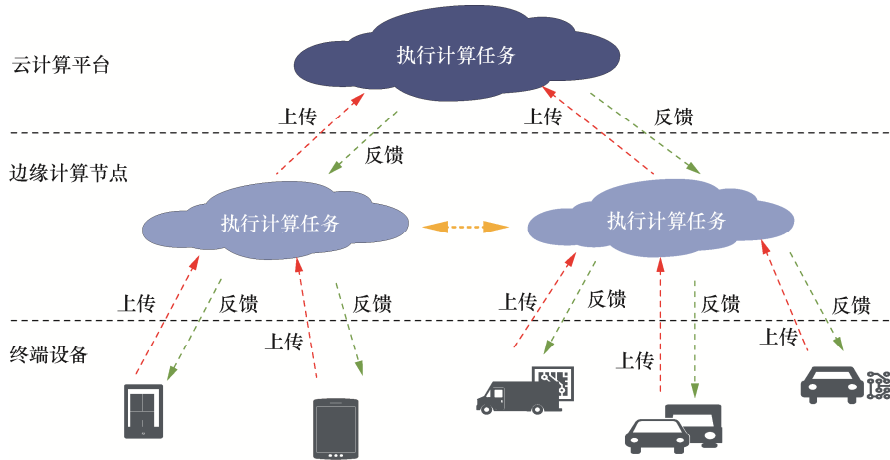


图 9 基于边缘计算的物联网系统中的任务卸载

云平台。

### 4.3 接入控制

接入控制是指通过验证允许或拒绝用户及设备访问系统资源的方法。在传统基于云平台的物联网系统中，云平台通常由可信任的第三方公司或组织提供，如谷歌云、亚马逊 AWS、微软 Azure 以及阿里云等。物联网的终端设备也由系统的控制中心进行中心化的设备验证进而接入系统。

然而随着边缘计算设备的引入以及物联网终端设备数量的急速增长，如何对这些设备在接入物联网系统时进行验证和管理也是亟待解决的问题之一<sup>[8]</sup>。与可信任的云平台不同，边缘计算设备更加轻量级也更加普及，可以由不同公司或组织提供，数量较多。就终端设备而言，种类和数量更庞大，预计在 2022 年将达 180 亿。如此高量级的设备数量会给传统的中心化验证和管理方式带来巨大压力。因此，为防止恶意设备模仿或伪造身份入侵系统，去中心化的高可靠性接入控制方法如新兴的区块链技术将成为潜在的解决方案之一，但相关技术并不成熟，仍需要更多研究精力的投入。

### 4.4 隐私与安全

物联网系统如智能家居系统、智能健康系统和车联网系统等，不仅注册了用户个人信息，还收集了用户的日常生活数据，这些数据直接关系用户的个人隐私以及人身和财产安全。然而，随着物联网系统的大规模部署，大量的边缘计算设备以及物联网终端设备应用其中。这些设备存在较大的性能差异，其中一些低成本、低顽

健性的设备容易遭受来自多方的、不同种类的安全攻击，将对物联网用户造成极大的隐私与安全威胁<sup>[9]</sup>。

除了已经提到的接入控制问题以及传统网络中常见的 DoS 攻击、路由攻击和欺骗攻击等，数据泄露、数据篡改等数据相关的攻击在物联网系统中将更频繁地出现。如何确保数据的机密性、完整性和可用性愈发重要，一方面，可以在设备生产过程中对硬件的质量及顽健性进行加强；另一方面，需要尽快研发适用于物联网系统的轻量级数据加密算法和信任管理机制。

## 5 结束语

本文首先提出了基于边缘计算的物联网监测系统架构，由云平台、边缘计算节点以及无线传感器监测节点组成。对各组成部分的功能以及相互之间的信息交互方式均进行了详细介绍，并且在此基础上提出了利用自编码器实现的数据异常检测方法。应用美国国家海洋和大气管理局提供的 TAO 监测数据对该异常检测方法进行了分析，所得到的 ROC 曲线表明，该异常检测方法接近理想检测器。最后，对基于边缘计算的物联网系统中仍存在的问题和潜在的研究方向进行讨论。

### 参考文献:

[1] SHI W, CAO J, ZHANG Q, et al. Edge computing: vision and challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646.

[2] 杨昶, 李凯, 徐海东, 等. 雾计算赋能机器人同步定位与地图构建[J]. 物联网学报, 2018, 2(2):33-40.

YANG Y, LI K, XU H D, et al. Fog computing-enabled robot simula-

neous localization and mapping[J]. Chinese Journal on Internet of Things, 2018, 2(2):33-40.

- [3] ZHANG Y, MERATNIA N, HAVINGA P J M. Outlier detection techniques for wireless sensor networks: a survey[J]. IEEE Communications Surveys and Tutorials, 2010, 12(2):159-170.
- [4] BHATIA S K. Adaptive K-means clustering[C]//AAAI Flairs Conference Menlo Park, CA, USA. 2004:695-699.
- [5] MAHDAVINEJAD M S, REZVAN M, BAREKATAIN M, et al. Machine learning for Internet of things data analysis: a survey[J]. Digital Communications Networks, 2018, 4(3):161-175.
- [6] SEZER O B, DOGDU E, OZBAYOGLU A M. Context-aware computing, learning and big data in Internet of things: a survey[J]. IEEE Internet of Things Journal, 2018, 5(1):1-27.
- [7] ZHU Q, SHI B, YANG F, et al. Task offloading decision in fog computing system[J]. China Communications, 2017, 14(11):59-68.
- [8] NOVO O. Blockchain meets IoT: an architecture for scalable access management in IoT[J]. IEEE Internet of Things Journal, 2018, 5(2):1184-1195.
- [9] FRUSTACI M, PACE P, ALOI G, et al. Evaluating critical security issues of the IoT world: present and future challenges[J]. IEEE Internet of Things Journal, 2018, 5(4):2483-2495.

#### [作者简介]



于天琪(1991-),女,西安大略大学电气与计算机工程学院博士生,主要研究方向为物联网、边缘计算、机器学习、数据分析等。



朱咏絮(1988-),女,博士,拉夫堡大学研究员,主要研究方向为无线通信性能分析、无线边缘缓存、基于毫米波的无线传输、无人机通信等。



王现斌(1971-),男,博士,西安大略大学电气与计算机工程学院教授、加拿大工程院院士、IEEE 院士,主要研究方向为下一代宽带无线移动通信传输理论、5G、通信安全、定位技术、协同通信、异构网融合、边缘计算等关键技术的理论及标准化研究。