

面向物联网的时空数据处理算法设计



徐鹤^{1,2} 吴昊¹ 李鹏^{1,2}

1 南京邮电大学计算机学院、软件学院、网络空间安全学院 南京 210023

2 江苏省无线传感网高技术研究重点实验室 南京 210003

(xuhe@njupt.edu.cn)

摘要 随着物联网和5G技术的快速发展,以深度学习为基础的人工智能应用越来越多,使基于时空数据的医疗影像、城市安防、自动驾驶等视觉领域成为物联网方向的研究热点。物联网系统采集到的视频数据、图片数据、温湿度与气体浓度数据同时也急剧增长,最终使得物联网系统的处理速度和反馈速度越来越慢。针对物联网节点采集的时空数据量大且可能存在短暂性异常的问题,文中设计了基于长短记忆网络的EPLSN(Exception Processing Long and Short Memory Network)算法。首先,对输入门的逻辑结构进行设计,并对网络模型结构进行改进,解决了短暂性异常数据与时空数据分类的问题,提高了EPLSN算法对物联网时空数据的分类精度,并能够对异常数据进行数据清洗。其次,依据传感器采集的时空数据特点,将数据存储到不同的数据块中,采用时序数据库对时空数据进行短暂性存储,并提出基于时空数据的物联网搜索架构,加快了物联网系统搜索的速度。

关键词: 物联网;时空数据;异常数据;深度学习;数据清洗;数据分类

中图法分类号 TP301.6

Design of Temporal-spatial Data Processing Algorithm for IoT

XU He^{1,2}, WU Hao¹ and LI Peng^{1,2}

1 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2 Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210003, China

Abstract With the rapid development of Internet of Things (IoT) and 5G technology, there are more and more applications of artificial intelligence based on deep learning, which makes medical imaging, urban security, autonomous driving and other visual fields based on temporal-spatial data become research hotspots in the direction of IoT. At the same time, the video data, picture data, temperature and humidity data and gas concentration data collected by the IoT system also grow rapidly, which eventually makes the processing speed and feedback speed of the IoT system slower and slower. In view of the large amount of temporal-spatial data collected by IoT nodes and the problem of transient anomalies, this paper designs an EPLSN (Exception Processing Long and Short Memory Network) algorithm based on long and short memory network. This paper designs logic structure of the input gate and improves the network model structure, solving the problem of the classification of transient abnormal data and temporal-spatial data, improving the classification accuracy of the IoT temporal-spatial data, and cleaning the abnormal data. According to the characteristics of the temporal-spatial data collected by the IoT sensor, the data is stored in different data blocks. At the same time, a time-series database is used to temporarily store temporal-spatial data, and an IoT search architecture based on temporal-spatial data is proposed. The architecture is suitable for the real-time search system in IoT environment and accelerates the search speed of the IoT system.

Keywords Internet of Things, Temporal-spatial data, Abnormal data, Deep learning, Data cleaning, Data classification

收到日期:2020-04-10 返修日期:2020-07-09 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家重点研发计划项目(2019YFB2103003,2018YFB1003201);国家自然科学基金(61672296,61602261,61872196,61872194,61902196);江苏省科技支撑计划项目(BE2017166,BE2019740);江苏省高等学校自然科学研究重大项目(18KJA520008);江苏省六大人才高峰高层次人才项目(RJFW-111)

This work was supported by the National Key R&D Program of China (2019YFB2103003,2018YFB1003201), National Natural Science Foundation of P. R. China (61672296,61602261,61872196,61872194,61902196), Scientific and Technological Support Project of Jiangsu Province (BE2017166,BE2019740), Major Natural Science Research Projects in Colleges and Universities of Jiangsu Province (18KJA520008) and Six Talent Peaks Project of Jiangsu Province (RJFW-111).

通信作者:李鹏(lipeng@njupt.edu.cn)

1 引言

数据分类输出变量通常为离散型^[1],常见的数据分类算法包括神经网络^[2]、随机森林^[3]、贝叶斯^[4]、决策树^[5]、逻辑回归^[6]、KNN (k-Nearest Neighbor)^[7]、SVM (Support Vector Machines)^[8]、ID3 (Iterative Dichotomiser 3)^[9]和 C4.5 决策树^[10]算法等。由于物联网数据的多样性、实时性与多维性,物联网搜索系统的数据采集来源于多源异构的传感器,并且数据采集节点存在不稳定性,容易产生采集的数据集不平衡的问题。在数据分类上,通常使用最近邻算法来解决由于时空数据不平衡所导致的无法准确分类的问题。Fedorov 等^[11]提出了一种基于分层稀疏贝叶斯框架的新型多模稀疏字典学习算法,通过加强字典之间的联合稀疏性,以解决相关的数据分类问题。但上述算法主要针对同一维度、同一种类型的数据进行分类,由于物联网时空数据具有多样性、实时性和多维性的特征,上述算法无法进行处理。Wang 等^[12]提出了一种改进的 BP 神经网络,并应用于物联网数据分类中,相比传统的 BP 神经网络算法,物联网实时采集的数据的分类精度得到了提高。Zhang 等^[13]提出了用于高维海量数据分类的多贝叶斯判别函数,并设计了用于探索贝叶斯估计的类密度分布的递归算法,以及用于选择判别函数同时保持分类器的低复杂度的自动分类方法。但是该方法在处理物联网实时数据时准确度不高,当物联网节点出现异常时,容易出现数据分类错误与数据异常的问题。

针对上述问题,本文将物联网终端节点采集到的时空数据进行实时分类与存储,并将物联网实时采集得到的异常数据进行处理,提出了一种基于长短记忆网络(Long Short-Term Memory, LSTM)的 EPLSN(Exception processing long and short memory network)算法。EPLSN 算法不仅提高了分类的精确度,还可以有效避免物联网节点采集的时空数据出现异常,保证了物联网节点数据的传输拥有较高的稳定性。

2 数据异常与分类算法设计

图 1 所示为 EPLSN 算法对物联网时空数据处理的流程图。

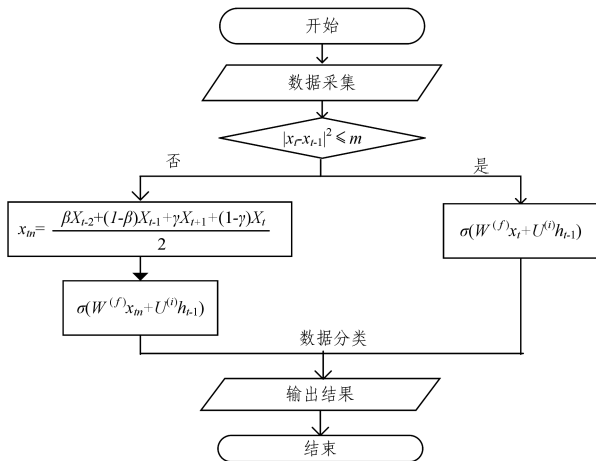


图 1 EPLSN 算法处理数据的流程图

Fig. 1 Flow chart of data processing by EPLSN algorithm

由于网络的不稳定性,以及物联网终端节点采集的数据

经常存在短暂性异常等问题,本文首先采用 EPLSN 算法对物联网终端采集到的异常数据进行优化或删除,然后将清洗过的数据进行分类预测。

当对数据 x_t 进行分类时,算法首先会接收和保留 x_{t-1}, x_t, x_{t+1} 3 个数据,并计算 $|x_t - x_{t-1}|^2$,当 $|x_t - x_{t-1}|^2$ 小于设定的阈值 m 时, x_t 的大小将不会发生变化。当 $|x_t - x_{t-1}|^2$ 大于设定的阈值 m 时,将 x_t 的大小修改为 x_m ,并将 x_m 作为 EPLSN 的输入数据,最后使用 EPLSN 算法对 x_t 进行分类,输出分类结果。

2.1 时空数据异常处理

基于时空数据的物联网搜索系统必须可以对物联网实时采集的短暂性异常时空数据进行纠正或清洗。本文设计的基于 LSTM 的 EPLSN 算法,对物联网实时传输的数据进行处理,相比传统算法,分类精确度得到了提高。

由于基于时空数据的物联网搜索系统所需要的数据存在多样性、数据量大、实时性等特征,同时,物联网实时节点不断增多,并发量不断增大,并且传感器受周围环境的影响较大,传感器数据在传输中容易出现短暂性数据异常或数据误差的情况^[14],本文在 LSTM^[15]的基础上做出改进,提出了 EPLSN 算法,其能有效避免物联网节点的数据异常问题,并保证物联网节点数据的传输能够拥有较高的稳定性。该算法首先针对物联网节点采集数据出现的异常进行清洗,然后对时空数据进行数据分类。

EPLSN 算法主要对物联网节点采集到的短暂性异常数据进行清洗,并对正常数据分类,对于物联网实时传输的数据具有良好的分类性能。但当出现传感器节点持久性损坏、网络断开等情况时,EPLSN 算法将无法对时空数据进行处理。在物联网实时搜索系统中,一般会设立传感器节点和网络异常监控模块,当出现异常时,系统会得到反馈,以便及时地对传感器节点进行维修与管理。

物联网终端节点采集的时空数据是动态变化的,但在实际应用中,某一个物联网终端硬件节点采集的数据种类是固定的,通常相邻的数之间不会有太大的变化。当采集到的物联网时空数据突然发生较大变化时,则此时数据可能为异常值。设置阈值为 m ,当 $|x_t - x_{t-1}|^2 > m$ 时, x_t 和 x_{t-1} 属于同类型的数据。无论数据是否异常,系统都会对采集到的数据进行修正,输入门的逻辑结构如图 2 所示。

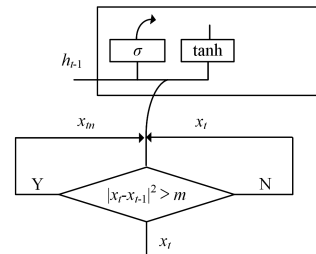


图 2 输入门的逻辑结构设计

Fig. 2 Logic structure design of input gate

此时输入的 x_t 变成 x_m , x_m 的计算如式(1)所示,此时 f_t 的计算方法如式(2)所示:

$$x_m = \frac{\beta X_{t-2} + (1-\beta) X_{t-1} + \gamma X_{t+1} + (1-\gamma) X_t}{2}$$

$$0 < \beta < 0.5, \gamma \infty 1 \quad (1)$$

$$f_t = \begin{cases} \sigma(W(f)x_m + U(f)h_{t-1}), & |x_t - x_{t-1}|^2 \leq m \\ \sigma(W(f)x_t + U(f)h_{t-1}), & |x_t - x_{t-1}|^2 > m \end{cases} \quad (2)$$

此时采集到的 x_t 的大小由 x_{t-2}, x_{t-1}, x_t 与 X_{t+1} 共同决定。 x_{t+1} 为物联网硬件节点或传感器下一次收集的时空数据信息, h_{t-1} 为上一层的隐藏单元, β 为(0, 0.5)区间的常量, γ 表示无穷接近于1的常数,这里将 γ 设置为接近于1的常数。由于物联网终端节点采集到的时空数据是动态变化且连续的,引入 x_{t+1} 是为了解决物联网终端节点采集到的数据突然变大的问题,让 X_m 值在 t 时刻更接近物联网世界的真实值。

2.2 时空数据分类算法设计

循环神经网络的优点在于有记忆性,其通过使用带自反馈的神经元,能够处理任意长度的时序数据。但是这种结构存在间隔较长时间的状态对后续状态没有影响的问题。例如,神经元上一状态对现在状态的影响是0.9,那么神经元上上状态对现在状态的影响就只有0.81,以此类推,间隔了很长时间之后,这些影响都会消失。但在实际运用中,有些有用的信息虽然间隔时间很长,但依然需要被记住,这就需要神经元具有长时间的记忆,而长短记忆网络可以通过构建一种记忆单元来保留这些信息。

针对物联网终端节点采集到的时空数据的特性,算法需要保存过去的物联网时空数据信息,用于决定当前物联网时空数据的分类结果。该分类网络不仅要满足对物联网时空数据的分类效果,对物联网节点采集到的温度数据、湿度数据、气体浓度数据、图像数据进行数据分类,还需要快速处理异常的物联网时空数据,并提高物联网时空数据的分类精度。根据采集到的物联网时空数据类型建立图3所示的基于LSTM的EPLSN算法网络结构图。

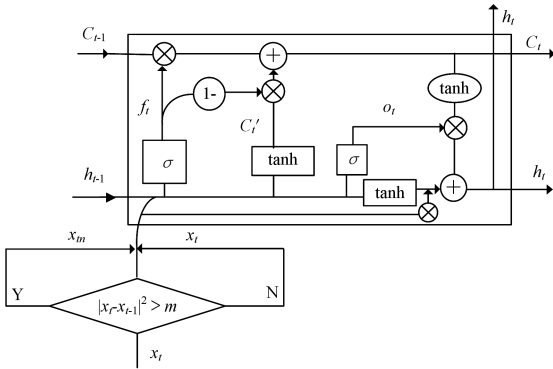


图3 EPLSN算法网络结构图

Fig. 3 Network structure diagram of EPLSN algorithm

不同于传统的LSTM网络,遗忘门和需要添加的新的时空数据信息共同决定 C_t 的大小。如式(3)所示, C_t 的输出结果是由 $(1-f_t)$ 控制的临时记忆单元 c'_t 和遗忘门 f_t 控制的过去的 $t-1$ 时刻的记忆单元 C_{t-1} 组成, C_t 的结果即为对物联网时空数据最终分类的结果。

$$C_t = (1-f_t) \cdot C'_t + f_t \cdot C_{t-1} \quad (3)$$

把输入门与遗忘门进行组合,生成一个单一的更新门。同时,EPLSN算法中混合了细胞状态和隐藏状态,根据物联网实时系统的数据传输特点,某一个物理节点收集的数据前

后差距不大。为了增强物联网实时分类的准确率,对输出的 h_t 做出相应的改变,增大了 h_{t-1} 的作用,改进的EPLSN网络结构如图3所示。 h_t 的计算公式如式(4)、式(5)所示。

$$y = x_t \tanh(h_{t-1}) \quad (4)$$

$$h_t = o_t \tanh(c_t) + y \quad (5)$$

其中, x_t 为输入的新的值, h_{t-1} 为上一层的隐藏单元。

EPLSN算法的处理流程如算法1所示。

算法1 EPLSN算法

输入: prev_ht // 先前的隐藏状态

input // 输入数据

输出: 输出隐藏状态

1. def EPLSN(prev_ct, prev_ht, input):
2. if input-preinput <= m: // 阈值比较
3. combine = prev_ht + input
4. else:
5. input = (a * input_{j2} + (1-a)input_{j1} + b * input_{p1} + (1-b) * input)
6. combine = prev_ht + input
7. ft = forget_layer(combine)
8. candidate = candidate_layer(combine)
9. it = input_layer(combine)
10. c_t = pre_ct * ft + (1-ft) * i_t
11. o_t = output_layer(combine)
12. y = x_t * tanh(h(h-1))
13. h_t = o_t * tanh(c_t) + y
14. return h_t, c_t

与LSTM不同的是,EPLSN首先需要将当前输入的值与之前输入的值相减。若差值大于设定的阈值,则对输入的数据进行修正,作为当前的输入值,并同时修正过的值与之前的隐藏状态进行组合连接;若差值小于阈值,则可以直接将输入值与先前的隐藏状态进行组合连接。

3 数据存储设计

3.1 逻辑结构设计

传统的物联网实时搜索系统把物联网实时采集的数据统一传输到中央服务器的数据库中,容易造成数据冗余,并且系统健壮性差,导致搜索效率低。为此,本文构建了图4所示的基于InfluxDB^[16]的物联网时空数据逻辑结构图。

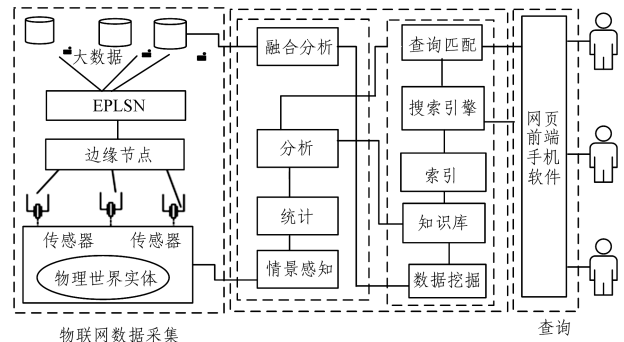


图4 搜索系统逻辑结构图

Fig. 4 Search system logic structure diagram

EPLSN算法将物联网实时数据先存储到边缘节点,接着

将实时采集的数据在边缘节点上进行计算和数据分类,同时对实时采集到的异常物联网数据进行清洗,然后将不同类型的物联网数据存储到不同的 InfluxDB 数据块中。系统采用分布式的存储架构以加快索引的速率。

3.2 时空数据存储设计

物理世界中很多物体的表面都装有传感器,如汽车、雷达、卫星、衣服、电话、微波炉,甚至人体等,而这些物体都要求传感器采集的数据可以实现实时流式传输。因此,物联网系统不仅需要考虑时间和测量值(带有时间戳的数据),还需要考虑位置信息。由于基于时空数据的搜索系统需要不断采集时间和空间数据,所以在数据处理上,采用带有时间戳的 InfluxDB 数据库^[17]。InfluxDB 具有如下 3 个方面的特点:

(1)Time Series(时间序列)。InfluxDB 可以直接利用数据库自带的与时间相关的函数(如最小、均值、求和等)。

(2)Metrics(度量)。InfluxDB 可以实时对大量数据进行计算。

(3)Events(事件)。InfluxDB 支持任意的事件数据。

InfluxDB^[18]是开源时间序列数据库,旨在处理较高的写入和查询负载,并提供一种称为 InfluxQL 的类似于 SQL 的查询语言,用于与数据进行交互。InfluxDB 支持物联网系统的原因在于每秒可以进行数百万次的读入,可以满足最大的监控和 IoT 部署的需求。InfluxDB 具有如下功能:

(1)监控与追踪。对物联网节点的数据实施监控与追踪,例如监控硫化氢气体浓度、监控场所的温湿度以及进行实时目标追踪等。

(2)行动与控制。对物联网节点可以进行快速控制,可以实现在无人看守干涉的情况下,对采集到的数据实时地进行反馈与操作。例如,当系统检测到石油泄漏时,则自动关闭泵;当系统监测到风速改变时,风力涡轮机的方向会发生变化。当物联网系统检测到监控环境中硫化氢浓度过高时,会自动报警,及时通知安全管理负责人以及有关安全生产管理部门。

(3)分析。提供实时数据分析功能,可以自动地对物联网节点数据进行实时分析与预测操作。如实时天气数据预测、实时物联网数据分类等。

与 MySQL,Oracle 等数据库相比,InfluxDB 有很大的不同,基于时空数据库设计的逻辑架构如图 5 所示。

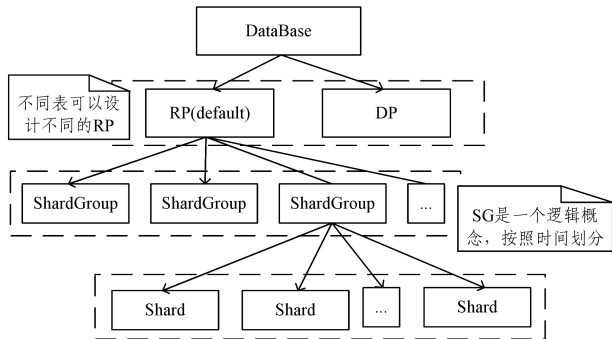


图 5 时序数据库逻辑结构图

Fig. 5 Logical structure diagram of InfluxDB

其主要功能如下:设定数据的副本的具体个数、保存时间、截止日期和设置时间范围。多个 Shard 组成一个 ShardGroup 集群,其中各个 ShardGroup 是相互独立的,分别存储系统开发人员预先规定的不同时间段的数据。例如,将 2019 年 9 月 1 日采集到的物联网时空数据(如温湿度、气体、时间、地理位置等)信息存在 ShardGroup0 模块中,而 2019 年 9 月 9 日采集到的数据存在 ShardGroup1 模块中。但 ShardGroup 在实际的存储中只是多个 Shard 模块集群的统称,真正用于存储的数据只是各个 Shard。

Shard 又称作 TSM(Time Sort Merge Tree)Engine,是一种数据库存储引擎,由预写日志模块、缓存模块和数据文件模块组成。其中,预写日志模块 WAL(Write-Ahead Logging)主要应用于数据的持久化,恢复和还原由于系统错误、断网等未及时保存到 TSM 中的数据;缓存 Cache 主要使用 Shard 中用于缓存数据的文件;数据文件模块 Data File 用于存储数据的文件。图 6 为 Shard 存储结构图。其中数据存储的 Shard 策略采用两层分片,上层采用范围分片,下层使用 Hash 分片。由于 InfluxDB 是基于时序的数据库,采用范围分片十分合理,但是如果只采用范围分片的方法,会出现某个点是热点数据时,将热点数据持续存储到新的节点中,而旧的节点将不再使用,从而导致资源浪费。为了解决这一问题,对 key 值进行 Hash 求值,使数据能够平均散列到各个数据库,有效解决了数据分布不平衡与热点的问题。

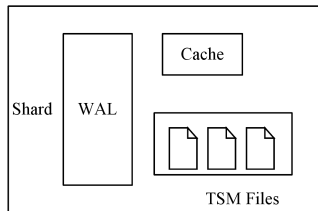


图 6 Shard 存储结构

Fig. 6 Shard storage structure

4 实验结果与分析

本文针对物联网数据的特点,提出了 EPLSN 算法对物联网数据进行分类,将分类后的结果集存储到不同的数据块中,方便用户进行搜索,从而加快检索效率。实验中采用 EPLSN,GRU(Gated Recurrent Unit)^[19],RNN^[20]3 种算法对摄像头采集拍摄的 mnist 数据集^[21]进行数据分类,将数据类型分为 0-9 这 10 类情况,并对训练过程中的准确率进行可视化,准确率对比结果如图 7 所示。

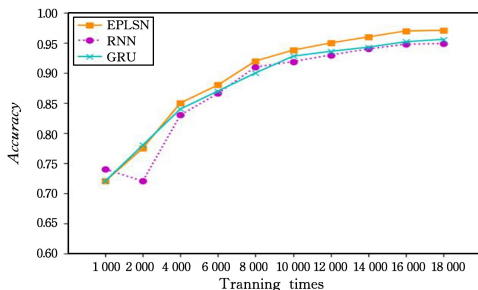


图 7 分类的准确率

Fig. 7 Classification accuracy

RP(Retention Policy)是一种针对数据存储的保留策略,

图 7 中,横坐标代表 3 种网络模型的训练时间,单位为秒;纵坐标代表 3 种算法对物联网实时数据的分类准确率。随着迭代次数的不断增多与训练时间的增长,3 种算法的分类准确率逐渐提高。EPLSN 的最终记忆单元是由遗忘门和需要添加的新的信息共同决定的,且能够对短暂时异常数据进行修正,因此针对物联网数据,EPLSN 算法的分类准确率可以达到 0.971,相比 RNN 的 0.949 与 GRU 的 0.956,分类准确率得到了有效提高。

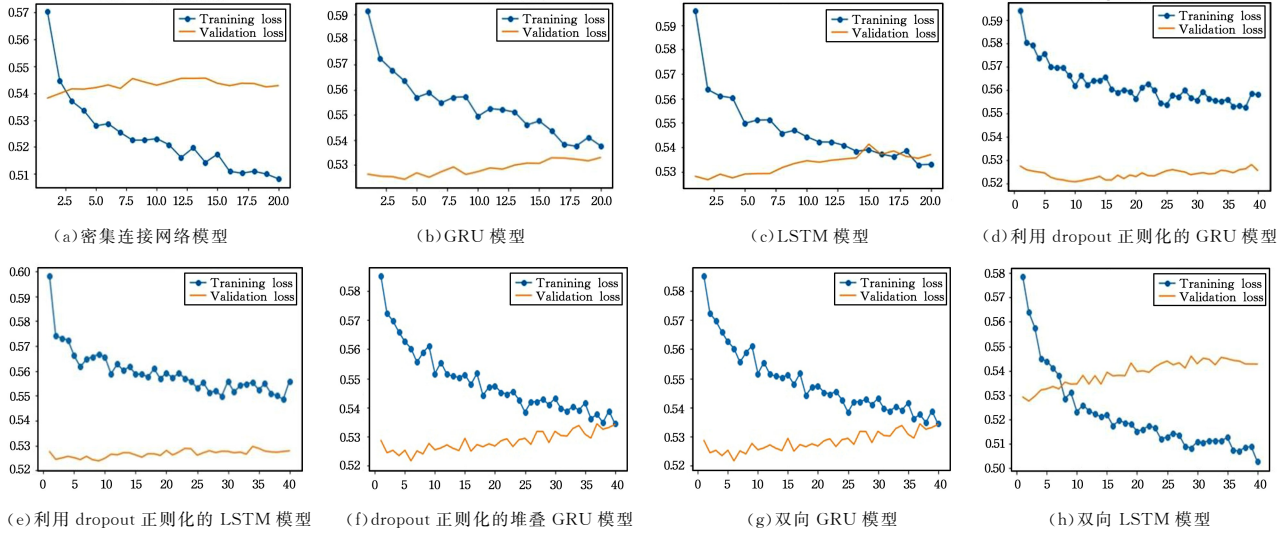


图 8 LSTM 与 GRU 模型的训练损失和验证损失图

Fig. 8 Training loss and verification loss of LSTM and GRU models

从图 8 可以看出,在精度方面,GRU 领先于 LSTM 网络,证明了循环神经网络与序列展平的密集网络相比在分类上具有优势,但从训练和验证曲线可以明显看出,两种模型都会发生过拟合现象;经过 GRU 与 LSTM 模型训练几轮之后,验证损失虽然趋于稳定,但训练损失和验证损失开始显著分离,且模型的预测准确率较差。

EPLSN 算法的训练集损失和验证集损失如图 9 所示,每迭代 5 次计算一次训练损失与验证损失的结果,可以看出,随着迭代次数的不断增加,当迭代次数接近 3 时,验证损失趋于稳定,训练损失开始下降。将图 9 中的 EPLSN 算法与图 8 中的 LSTM 与 GRU 算法进行比较可以得出,EPLSN 在训练集和验证集的预测精确度上比 LSTM 和 GRU 都有较大的提高。

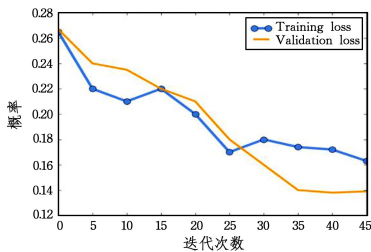


图 9 EPLSN 算法的训练损失和验证损失

Fig. 9 Training loss and verification loss of EPLSN algorithm

另外,从物联网的 InfluxDB 数据库中选取一天内测量的温度信息,用于验证 EPLSN 模型对异常数据的纠错能力。实验中共有 17280 条数据,数据会根据传入数据库的时间顺序保存下来。在这段数据中,采用 random 函数随机生成 50

实验使用 jena_climate_2009_2016.csv 数据集,通过在数据集中加入物联网传感器节点来获得周围实时的天气情况,将数据集分别在密集连接网络模型、LSTM 模型与 GRU 模型上进行模型训练与分类。天气的预测分类结果有晴天、阴天、多云、雨与雪这 5 类天气情况。图 8 给出了随着网络模型迭代次数的增加,训练损失与验证损失的精度变化情况。其中,横坐标代表各算法模型的迭代次数,纵坐标代表损失概率。

个异常数据,替换原数据库中的数据。再将这些数据经过 EPLSN 算法分类、过滤、存储到数据库中。最后将两组数据进行对比,发现有 17277 条数据与原数据库中的数据相同,只有 3 条数据发生了改变,其改变范围在 2°C 之内,其他 47 条数据得到了有效纠正,纠正率达 94%。由此可见,EPLSN 算法能够及时处理物联网实时传输的异常数据,保证了物联网采集到的数据能够稳定正确地传输。

结束语 本文提出了能够处理实时异常数据的 EPLSN 算法。EPLSN 算法包括数据异常处理和数据分类两个模块,既可以对实时获取到的物联网时空短暂时异常数据进行修正,也可以对时空物联网数据进行预测与分类。此外,本文还提出了针对基于时空数据的物联网搜索系统的数据存储结构,利用 InfluxDB 数据库的特性,对物联网实时数据进行存储,同时保存时间戳字段,以便后续的算法可以利用存储的时空数据进行分类与预测。实验结果表明,本文提出的算法不仅能有效减少异常的物联网数据,还具有较高的数据分类性能,可以有效地提高物联网搜索系统的效率,去除无效数据,保证实时数据的有效性。本文目前还没有对物联网搜索系统的搜索、索引机制、架构等进行研究,研究面向物联网时空数据的处理算法最终是为了构建物联网搜索系统,但建立快速反馈的物联网搜索系统仍然还有很多工作要做。为了设计一个可以快速反馈的物联网搜索系统,下一步将对算法进一步优化,并根据采集到的物联网时空数据的特性,设计合理的物联网搜索和索引机制,以加快物联网搜索系统的数据搜索速度。

参 考 文 献

- [1] ZHANG P N. Research on entity state matching estimation method for Internet of things search service [D]. Beijing :Beijing University of Posts and Telecommunications,2017.
- [2] ALMA Y A. Electricity Prices Forecasting using Artificial Neural Networks [J]. IEEE Latin America Transactions, 2018, 16(1):105-111.
- [3] XIA X, TOGNERI R, SOHEL F, et al. Random Forest Classification based Acoustic Event Detection Utilizing Contextual-Information and Bottleneck Features [J]. Pattern Recognition, 2018, 81:1-13.
- [4] WEBB G I, BOUGHTON J R, ZHENG F, et al. Learning by extrapolation from marginal to full-multivariate probability distributions; decreasingly naive Bayesian classification [J]. Machine Learning, 2012, 86(2): 233-272.
- [5] WANG S, FAN C, HSU C, et al. A Vertical Handoff Method via Self-Selection Decision Tree for Internet of Vehicles [J]. IEEE Systems Journal, 2016, 10(3): 1183-1192.
- [6] KOLETSI D, PANDIS N. Ordinal logistic regression [J]. American Journal of Orthodontics and Dentofacial Orthopedics, 2018, 153(1): 157-158.
- [7] ZHANG S, LI X, ZONG M, et al. Efficient kNN Classification With Different Numbers of Nearest Neighbors [J]. IEEE Transactions on Neural Networks & Learning Systems, 2018, 5(29): 1-12.
- [8] GU B, SHENG V S, TAY K Y, et al. Cross Validation Through Two-Dimensional Solution Surface for Cost-Sensitive SVM [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1103-1121.
- [9] HAN J, SEO H, CHOI Y, et al. Expression and regulation of inhibitor of DNA binding proteins ID1, ID2, ID3, and ID4 at the Maternal-Conceptus Interface in Pigs [J]. Theriogenology, 2018, 108: 46-55.
- [10] CHERFI A, NOUIRA K, FERCHICHI A. Very Fast C4. 5 Decision Tree Algorithm [J]. Applied Artificial Intelligence, 2018, 3(3): 1-19.
- [11] FEDOROV I, RAO B D, NGUYEN T Q. Multimodal sparse Bayesian dictionary learning applied to multimodal data classification [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017: 2237-2241.
- [12] WANG F, NIU L. An Improved BP Neural Network in Internet of Things Data Classification Application Research [C] // 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference. 2016: 805-808.
- [13] ZHANG J, WANG S, CHEN L, et al. Multiple Bayesian Discriminant Functions for High - Dimensional Massive Data Classification [J]. Data Mining and Knowledge Discovery, 2017, 31(2): 465-501.
- [14] GAO Y D. Research on Internet of Things Entity Information Description and Search Technology based on Web [D]. Beijing: North China Electric Power University, 2016.
- [15] GREFF K, SRIVASTAVA R K, JAN KOUTNÍK, et al. LSTM: A Search Space Odyssey [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(10): 2222-2232.
- [16] MORABITO R, COZZOLINO V, DING A Y, et al. Consolidate IoT Edge Computing with Lightweight Virtualization [J]. IEEE Network, 2018, 32(1): 102-111.
- [17] LEIGHTON B, COX S, CAR N, et al. A Best of Both Worlds Approach to Complex, Efficient, Time Series Data Delivery [J]. IFIP Advances in Information & Communication Technology, 2015(448): 371-379.
- [18] BALIS B, BUBAK M, HAREZLAK D, et al. Towards an Operational Database for Real-time Environmental Monitoring and Early Warning systems [J]. Procedia Computer Science, 2017, 108: 2250-2259.
- [19] ZHANG J, DU J, DAI L. A GRU-Based Encoder-Decoder Approach with Attention for Online Handwritten Mathematical Expression Recognition [C] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). 2017: 902-907.
- [20] ILIA K, MAXIM H, ANATOLII Z, et al. Putting Hands to Rest: Efficient Deep CNN-RNN Architecture for Chemical Named Entity Recognition with No Hand-crafted Rules [J]. Journal of Cheminformatics, 2018, 10(1): 1-10.
- [21] BU L Z, WANG H D, ZHU M Q, et al. Multi-source Digital Recognition based on Improved Convolutional Neural Network [J]. Journal of Computer Applications, 2018, 38(12): 3403-3408.



XU He, born in 1985, associate professor, master supervisor, is a member of China Computer Federation. His main research interests include Internet of Things (IoT) technology and applications.



LI Peng, born in 1979, Ph.D, professor, master supervisor, is a member of China Computer Federation. His main research interests include computer communication networks, cloud computing, and information security.