

基于神经网络的建筑能耗及设备故障预测方法研究

董道国, 朱建龙, 方瑾

(华建数创(上海)科技有限公司)

【摘要】文章在常规建筑能耗和设备故障预测方法的基础上, 根据智慧建筑中BIM模型数据和物联网数据融合形成的数据集, 提出了基于BP神经网络的多任务学习方法, 将建筑能耗和设备故障预测采用同一神经网络模型进行学习和预测, 并阐述了网络结构及损失函数的设计方法, 为建筑能耗和设备故障预测提供了新的思路, 也为智慧建筑运维管理系统中的智能诊断服务提供算法基础, 将人工智能技术真正地应用于智慧建筑运维管理系统之中。

【关键词】智慧建筑; 能耗预测; 故障预测; 关联规则; 神经网络

中图分类号: TU17

DOI: 10.13655/j.cnki.ibci.2022.04.044

Research on Building Energy Consumption and Equipment Failure Prediction Method Based on Neural Network

DONG Dao-guo, ZHU Jian-long, FANG Jin

(Arctron Data & Innovation Technology Co., Ltd.)

【Abstract】Based on the conventional building energy consumption and equipment failure prediction methods, according to the data set formed by the BIM model data and Internet of things data in smart buildings, this paper proposes a multi task learning method based on BP neural network, which uses the same neural network model to study and predict building energy consumption and equipment failure prediction. It also expounds the design method of network structure and loss function, which not only provides a new idea for building energy consumption and equipment failure prediction, but also provides an algorithm basis for intelligent push service in intelligent building operation and maintenance management system, and truly applies artificial intelligence technology to intelligent building operation and maintenance management system.

【Keywords】intelligent building; energy consumption prediction; failure prediction; association rules; neural network

1 引言

目前,智慧建筑在建筑行业内的应用越来越广泛。智慧建筑是为了适应现代信息社会对建筑物功能、环境和高效管理要求,在传统建筑的基础上发展而来^[1]的,因此,基于BIM技术和物联网技术的建筑运维管理系统得到了业内的广泛关注。但是人工智能技术在建筑领域内的应用仍然较少,主要应用方向是以节能为主的能耗系统及故障诊断系统。能耗系统是建筑运维过程中节能减排、节约费用的关键系统,一般基于统计机器学习方法的线性回归(LR)^[2]、支持向量机(SVM)^[3]、时序模型(ARIMA)^[4],以及基于神经网络的BP神经网络^[5]、长短期记忆网络(LSTM)^[6]等进行能耗分析预测。在建筑运维过程中,设备的维护维修同样是一个频繁且重要的工作,对于设备运维期故障的分析预测一般采用关联规则挖掘(apriori)^[7]及BP神经网络算法^[8]。常规的预测方法将能耗预测和故障预测割裂开来,没有考

虑两者的潜在共同影响因素,本文提出一种采用人工神经网络的多任务学习方法,在融合BIM模型数据和物联网数据的基础上,可同时进行建筑能耗和设备故障预测。

2 训练数据

采用基于物联网和BIM模型的智慧建筑,可采集以房间为单位的、大量的设备和传感器数据,因此建筑能耗和设备故障预测的颗粒度可以细化至单个房间,由此产生的大量数据同时解决了机器学习模型基于大数据的问题。通过物联网设备及传感器可以获取每个房间的环境参数和设备信息,环境参数包括某一天的最高温度、最低温度、湿度、天气、人流量等,设备信息包括设备类型、设备数量、设备使用年限,历史上维修次数等。通过BIM模型可获取房间和设备的空间数据,房间的空间数据包括体积、朝向、门窗面积、内外墙面积、楼板面积等,设备的空间信息包括房间内拥有的设备类型、设备数量等。训

练数据的标签分为各种能耗数值及设备故障的有无,每一类设备对应一种设备故障类型。数据集的特征和标签均可根据不同的能耗预测对象及设备故障类型进行增删。

3 Apriori 算法

Apriori 算法是第一个关联规则挖掘算法,也是最经典的算法。它利用逐层搜索的迭代方法找出数据库中项集的关系,以形成规则。项集出现的频率即为项集的支持度,如果某项集满足最小支持度,则称它为频繁项集。根据最小支持度,用 K 项集去探索 $(K+1)$ 项集,首先找出频繁 1 项集的集合记为 L_1 ,基于 L_1 找出频繁 2 项集 L_2 ,依此类推,直到不能找到频繁 K 项集。为了提高频繁项集逐层查找的效率,该算法利用连接(将 L_{k-1} 与自己连接产生 L_k 的候选项集)和剪枝(任何非频繁的 $K-1$ 项集都不可能是频繁 K 项集的子集)减少计算量^[9]。

将数据进行选取并规范化处理后,即可采用多元线性回归或神经网络算法,直接对能耗进行回归或者对故障进行分类。但是对于关联规则挖掘,由于 Apriori 算法无法处理连续值的情况,所以必须将连续值进行分箱处理,每一种连续值的分箱方式均需人为指定。同时,Apriori 算法需要设定最小支持度及最小置信度的阈值,将其作为频繁项集的依据。当阈值设定过小时,频繁项集数量会过多,当阈值设定过大时,频繁项集数量会过少甚至没有,因此最小支持度及最小置信度的设置需要一定的经验。此外,在某些情况下频繁项集的挖掘仅仅基于故障和故障之间,不考虑其他的各种因素,因此还需要根据一定的日期间隔及空间间隔对设备故障进行筛选后合并为一条关联数据,间隔的选取同样需要较多的人工经验。对于训练样本较少的设备故障数据而言,机器学习结果本身就存在较大的偶然性,外加采用 Apriori 算法需要较多的人为经验进行调参,因此该算法用于运维系统中的故障预测时使用起来较为麻烦。

4 神经网络

普通的神经网络即 BP 神经网络采用输入层、隐藏层及输出层作为网络结构,如图 1 所示,其中隐藏层的数量可以多于一层。输入的节点数为特征的数量,输出的节点数根据任务需要进行设置,隐藏层的节点数可以任意设置,一般不少于输入节点的数量。

基本 BP 神经网络包括信号的前向传播和误差的反向传播两个过程。以三层神经网络为例,正向传播时,输入值乘上权重矩阵并经过非线性变换后输出至隐藏层,再经过一次权重变换后产生输出信号,若实际输出与期望输出不相符,则转入误差的反向传播过程。误差反向传播是将输出误差通过隐藏层向输入层逐层反传,并将误差分摊给各层所有单元,以此调整权重矩阵。通过调整输入节点与隐层节点的权重值使误差沿梯度方向下降,经过反复学习训练,确定与最小误差相对应的网络参数即权重值,训练结束。神经网络推理时,将与训练数据集格式相同的数据作为输入,网络的输出值可直接使用,神经网络屏蔽了计算细节。

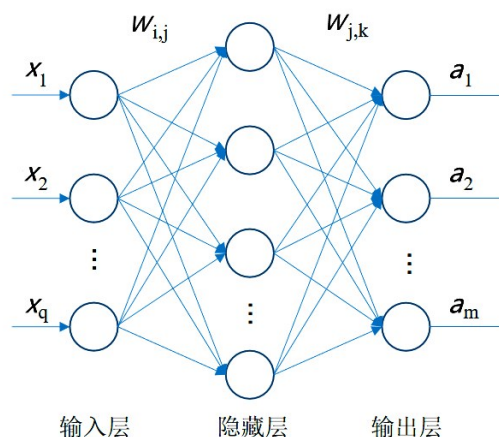


图1 BP神经网络结构图

为了将建筑能耗与设备故障预测在同一个网络中实现输出,需要将图 1 的网络结构进行微调,如图 2 所示。 a_1 至 a_n 为能耗预测的输出项, n 由能耗预测对象的数量决定。 b_{n+1} 至 b_m 为设备故障预测的输出值, m 由设备故障的类型数量决定。

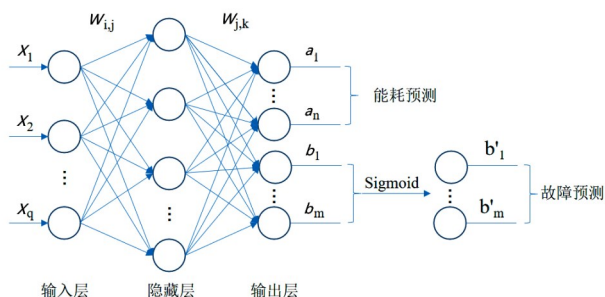


图2 能耗及故障预测BP神经网络结构图

由于设备故障预测类型为二分类,因此还需要经过一层 Sigmoid 函数进行变换,将输出数值转换至 0~1 之间,即故障出现的概率大小。在神经网络的

前向推理过程中,当故障预测部分的某个输出概率大于0.5时,可以认为该输出所对应的故障类型出现的可能性较大,管理系统即向物业管理人员进行提醒。由于每次输入的数据为一个房间的数据,并不能考虑房间与房间之间的关联性,也未能考虑单个房间与整个建筑之间的相互影响,因此对于设备故障的预测存在一定的不准确性,但是作为运维管理系统中的提醒事项已经足够。

5 损失函数

对于回归类学习任务,通常采用均方误差(MSE)作为损失函数,

$$MSELoss = \frac{1}{n} \sum_n (y - \hat{y})^2 \quad (1)$$

式(1)中: y 为实际值; $\hat{y}$ 为预测值; n 为能耗预测项目的数量;

对于分类任务,通常采用交叉熵(Cross Entropy)作为损失函数,

$$CELoss = -\frac{1}{m} \sum_m (y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})) \quad (2)$$

式(2)中: y 和 $\hat{y}$ 的含义同上, m 为设备故障类型的数量。

由于训练样本中,正负样本极不平衡,假如以设备出现故障为正样本,那么负样本的数量将远大于正样本的数量,因此需要增加正样本的损失权重,

$$CELoss = -\frac{1}{m} \sum_m (y \cdot \log(\hat{y}) + \alpha \cdot (1 - y) \cdot \log(1 - \hat{y})) \quad (3)$$

式(3)中: α 是小于1的正小数,以此降低负样本的权重。最终的损失应包含能耗预测和故障预测两部分,但由于能耗预测的损失值会大于故障预测,因此需要减少能耗预测的损失,避免优化方向只朝着降低能耗预测部分损失的方向进行。

最终的损失函数为:

$$Loss = \beta \cdot \frac{1}{n} \sum_n (y - \hat{y})^2 - \frac{1}{m} \sum_m (y \cdot \log(\hat{y}) + \alpha \cdot (1 - y) \cdot \log(1 - \hat{y})) \quad (4)$$

式(4)中: β 是小于1的正小数。损失函数拥有之后,即可建立神经网络进行训练。

6 结语

①采用Apriori算法可以挖掘设备故障间的关联

规则,但是需要人为经验进行调参,该算法应用于运维系统中的故障预测时较为麻烦。

②采用神经网络进行的多任务学习可以同时解决建筑能耗和设备故障预测两个问题,即可解决Apriori算法的缺点,又可以挖掘能耗相关参数和设备故障间潜在的关联关系。

③利用BIM模型可以有效利用建筑的空间数据,增加特征数量。缺点是每次输入的是一个房间的数据,未能考虑房间与房间之间的关系、单个房间与整个建筑的影响,因此对于设备故障的预测存在一定的不准确性。

④对于故障预测损失函数需要考虑能耗预测和故障预测两个部分,对于故障预测,需要权衡正样本和负样本数量的不平衡,因此引入两个超参数 α 和 β 对损失进行调节,让能耗预测和故障预测两个部分的优化更加合理。

⑤采用本方法的建筑能耗和设备故障预测可根据目标任务灵活调整网络结构,不足之处在于没有充分利用BIM模型中的空间拓扑关系。

参考文献

- [1] 刘翀.基于城市数字底座的智慧建筑操作系统探析[J].绿色建筑, 12(5): 3.
- [2] 聂子航, 于学军.基于多元线性回归的办公建筑电力能耗评估预测模型的设计[J].电子设计工程, 2016, 24(3): 40-43.
- [3] 吴贤国, 邓婷婷, 陈彬, 陶妍艳, 王雷.基于BIM-DB仿真和LS-SVM的建筑能耗预测[J].土木工程与管理学报, 2020, 37(6): 4-10.
- [4] 周芮锦, 潘毅群, 黄治钟.基于时间序列分析的建筑能耗预测方法[J].暖通空调, 2013, 43(8): 71-77.
- [5] 智伟威, 周新星.神经网络在数据中心能耗模型研究中的应用[J].计算机仿真, 2020(10): 273-277.
- [6] 章挺飞, 罗恒, 刘杭.基于LSTM网络的建筑能耗预测方法[J].苏州科技大学学报(自然科学版), 2020, 37(4): 82-88.
- [7] 刘浩.基于关联规则的高铁列控车载设备故障诊断方法研究[D].北京: 北京交通大学, 2019.
- [8] 吴渊.基于BP神经网络的车载设备故障诊断与预测研究[D].北京: 北京交通大学, 2016.
- [9] 刘华婷, 郭仁祥, 姜浩.关联规则挖掘Apriori算法的研究与改进[J].计算机应用与软件, 2009, 26(1): 146-149.