

DOI: 10.19783/j.cnki.pspc.200107

基于长短期记忆网络的PMU不良数据检测方法

杨智伟, 刘灏, 毕天姝, 杨奇逊

(华北电力大学新能源电力系统国家重点实验室, 北京 102206)

摘要: 同步相量测量单元(Phasor Measurement Units, PMUs)因其同步性、快速性和准确性, 已成为复杂电力系统状态感知的最有效工具之一。但是, 现场的复杂环境导致 PMU 数据存在数据丢失、数据损坏、同步异常、噪声影响等质量问题, 严重影响其在系统中的各类应用, 甚至威胁电网安全稳定运行。提出了一种基于长短期记忆(Long Short-Term Memory, LSTM)网络的 PMU 不良数据检测方法。首先分析了 LSTM 在不良数据检测中的优势。然后基于 LSTM 网络对时间序列选择记忆的特性, 构造了一种双层 LSTM 网络架构, 提出了对原始数据的分解重构方法。在此基础上, 定义了两种目标函数, 以获得不同的误差特征。提出了一种基于决策树的不良数据阈值确定方法, 实现了不良数据的有效检测。通过大量仿真与实测数据验证了该方法的可行性和准确性, 可提高 PMU 数据质量, 使其更好地应用于电力系统的各个方面。

关键词: 同步相量测量单元; 数据质量; 不良数据检测; 长短期记忆网络; 决策树

PMU bad data detection method based on long short-term memory network

YANG Zhiwei, LIU Hao, BI Tianshu, YANG Qixun

(State Key Lab of Alternate Electric Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China)

Abstract: Phasor Measurement Units (PMUs) have become one of the most effective tools for state awareness of complex power systems due to their synchronization, speed and accuracy. However, the complex on-site environment causes data loss, data corruption, synchronization anomaly, noise and other quality problems of the PMU data, which seriously affects various applications in the power system and even threatens the safe and stable operation of the grid. This paper proposes a PMU bad data detection method based on Long Short-Term Memory (LSTM) network. First, the advantages of LSTM in bad data detection are analyzed. Based on the characteristics of time sequence selection and memory of the LSTM network, a two-layer LSTM network architecture is constructed, and the decomposition and reconstruction method of original data is proposed. On this basis, two objective functions are defined to obtain different error characteristics. A method for determining the threshold of bad data based on decision tree is proposed, which realizes the effective detection of bad data. The feasibility and accuracy of the proposed method are verified by a large number of simulations and field data. The quality of PMU data is improved, which makes it better applied to all aspects of the power system.

This work is supported by National Key Research and Development Program of China (No. 2017YFB0902900 and No. 2017YFB0902901).

Key words: phasor measurement units (PMUs); data quality; bad data detection; long short-term memory (LSTM) network; decision tree

0 引言

状态感知作为电力系统万物互联的重要环节, 对

保障电力系统安全稳定运行具有重要意义^[1-3]。同步相量测量单元(Phasor Measurement Units, PMUs)因其快速性、同步性和准确性, 可为状态感知提供数据基础^[4-6]。同时, PMU 数据也广泛应用于系统振荡检测、事故分析、闭环控制等方面^[7-10]。目前, 我国已安装投运了 7 500 余台 PMU, 覆盖重要的

基金项目: 国家重点研发计划项目资助(2017YFB0902900, 2017YFB0902901)

220 kV 及以上变电站、主力发电厂和新能源并网汇集站。与此同时,随着大规模分布式电源、储能等不断增长,配电网的同步相量测量技术也在快速发展^[11]。

然而,由于现场环境复杂,PMU 量测数据易受到多种因素影响,存在不同程度的数据质量问题。PMU 不良数据主要包括量测数据丢失、同步信号丢失、量测偏差过大等^[12-13]。据统计,2013 年我国约有 20%~30% 的 PMU 数据存在质量问题^[14],严重影响系统中的各类应用,导致可观性降低,影响状态估计和参数辨识,甚至威胁电网安全稳定运行^[15]。因此,PMU 不良数据检测在提高数据质量和保证准确状态感知方面发挥着重要作用。

针对不同类型的不良数据检测,国内外学者提出了多种应对方法。量测数据和同步信号丢失可采用基于通信协议中数据帧结构^[16]是否异常的方法进行检测。针对量测偏差过大问题,可分为基于系统拓扑模型的状态估计方法和基于数据驱动的检测方法。文献[17]提出了一种基于相量量测的状态估计方法,利用增广状态向量校正相角偏差和量测误差,同时借助冗余量测估计线路参数。文献[18]将时序预测模型与卡尔曼滤波、平滑算法相结合,清洗 PMU 量测中的不良数据。基于状态估计和卡尔曼滤波的不良数据检测方法均需要多台 PMU 量测数据、系统拓扑和线路参数,虽然检测结果精确,但运行速度较慢,当系统拓扑或线路参数存在错误时,易造成误判。同时当系统拓扑未知时,不少学者提出了基于数据驱动的检测方法。文献[19]提出了一种基于逻辑判断的检测方法,一旦量测数据超出阈值,则判定为不良数据。而当系统动态时,量测数据变化复杂,固定阈值无法准确判断不良数据。文献[14]利用多台 PMU 量测数据,提出了一种基于时空相似性的检测方法。然而在某些 PMU 布点较少的地区,多台 PMU 量测数据难以获取,仅依靠单台 PMU 量测数据,该方法难以较好地应用。对此,仅利用单台量测信息,文献[20]提出了一种基于密度聚类的检测方法。文献[21]将三种简易分类器通过集成学习结合,再对不良数据进行检测。文献[22]利用主成分分析法将信号分成低维特征分量和高维噪声分量进行不良数据检测。上述三种方法均利用单台量测信息,在系统准稳态条件下,对不良数据的检测效果较好;但动态条件下,正常数据与不良数据的差异较小,聚类 and 集成学习的方法易造成漏检,主成分分析法则由于不良数据的影响,不能很好地分解信号特征,难以较好地检测不良数据。因此,本文主要解决系统动态条件下基于单台

PMU 量测信息的不良数据检测问题。

本文提出了一种仅利用单台 PMU 量测信息,且适用于系统多种状态的 PMU 不良数据检测方法。本文分析了长短期记忆(Long Short-Term Memory, LSTM)网络对 PMU 量测数据学习并记忆序列信息的优势,提出了一种序列分解重构的双层 LSTM 网络架构,并通过设置两种目标函数获取不同误差特征。在此基础上,为避免主观性,提出了一种基于决策树的不良数据阈值确定方法。最后,通过大量仿真与实测数据验证该方法的可行性与准确性,对提高 PMU 数据质量具有重要意义。

1 LSTM 网络在不良数据检测中的应用分析

电力系统发生故障等扰动时会导致电气量发生较大波动,易被识别为不良数据。本文作者提出了一种扰动事件识别方法,在不良数据检测之前应用,可防止将扰动事件数据识别为不良数据。由于该内容不是本文重点,因此未在此提及。

图 1(a)为电力系统准稳态下 PMU 量测数据, t_1 时刻存在不良数据,图 1(b)为电力系统动态下 PMU 量测数据, t_4 时刻存在不良数据。基于聚类和集成学习的不良数据检测方法对图 1(a)中准稳态下不良数据具有良好的检测能力,但对图 1(b)中动态下不良数据易造成误判。这是由于准稳态下不良数据与正常数据偏差较大,二者联系较弱,聚类和集成学习的方法易于区分不良数据;而动态下不良数据处于振荡的波峰与波谷之间,与正常数据差异较小,基于聚类和集成学习的方法不易区分不良数据。基于主成分分析的不良数据检测方法可有效分解信号各成分,通过重构后的数据与原始数据对比,检测出不良数据。当不良数据较少时,该方法可适用于准稳态和动态;但当不良数据较多时,易对分解过程产生影响,导致重构数据不准确,造成误判。

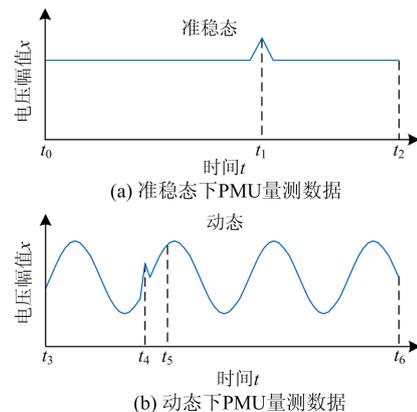


图 1 准稳态和动态下 PMU 量测数据

Fig. 1 Quasi-steady state and dynamic PMU measurement

LSTM 网络作为一种循环神经网络(Recurrent Neural Network, RNN)的特殊变体, 由于其独特的循环结构对时间序列具有强大学习能力, 可将时序中不良信息筛选并选择遗忘, 减小不良数据对数据重构的影响^[23]。同时由于传统 BP 神经网络是输入层到隐含层再到输出层, 层与层之间是全连接的, 每层之间的节点是无连接的。因此传统神经网络无法获得某一序列当前输出与前面信息的关系。而 LSTM 网络通过其内部的遗忘门、输入门、输出门和自身的循环结构可对前面的信息进行选择记忆并应用于当前输出, 即隐藏层之间的节点是有连接的, 并且隐藏层的输入不仅包括输入层的输出还包括上一时刻隐藏层的输出。因此, 考虑长短期记忆网络特有的循环结构, 选用其作为所提方法的主体结构处理 PMU 量测中的不良数据。图 2 为 LSTM 网络时序结构。图中: x_t 为当前时刻单元输入; c_t 为当前时刻单元记忆状态, 负责存储长期记忆特征; h_t 为当前时刻单元输出, 负责短期记忆特征。可以看出, 在 LSTM 网络中, 当前输出 h_t 由上一时刻的状态 c_{t-1} 、上一时刻的输出 h_{t-1} 和当前输入 x_t 共同决定。

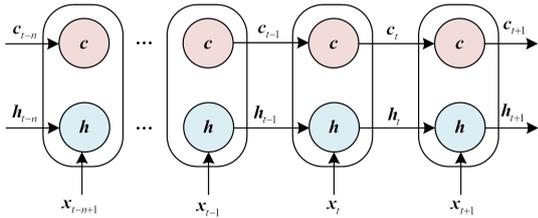


图 2 LSTM 网络时序结构

Fig. 2 LSTM network timing structure

LSTM 网络单元具体结构如图 3 所示。图中 LSTM 网络通过遗忘门 f_t 、输入门 i_t 和输出门 o_t 控制各时刻信息的输入, 因此, LSTM 网络对历史信息具有更好的选择性, 可实现对不良数据的有效筛选。

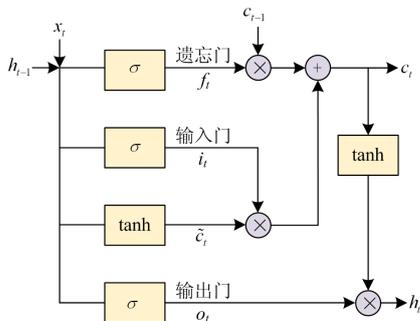


图 3 LSTM 网络单元结构

Fig. 3 LSTM network unit structure

遗忘门 f_t 会根据每一时刻的输入 x_t 和上一时刻的输出 h_{t-1} 决定上一时刻的单元记忆状态 c_{t-1} 中

某些长期记忆特征是否需要被遗忘。遗忘门 f_t 计算如式(1)所示。

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

式中: h_{t-1} 包含上一时刻量测数据的特征; W_f 为遗忘门权重矩阵; b_f 为遗忘门偏置项; 当 f_t 中表示 t 时刻数据特征的维度取值接近 0 时, 该特征从长期记忆状态 c_{t-1} 中被剔除, 若取值接近 1 则会被保留; σ 为激活函数。

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

以图 1(b)为例进行说明。当 t_5 时刻数据输入时, 遗忘门 f_{t_5} 中表示 t_4 时刻不良数据特征的维度取值会接近 0, 该特征从长期记忆状态 c_{t_4} 中被剔除。

LSTM 网络 t_5 时刻在长期记忆中遗忘不良数据的特征后, 还需在长期记忆中加入新的特征。输入门 i_{t_5} 根据 x_5 和 h_{t_4} 决定哪些信息加入长期记忆状态 c_{t_5} 中生成新状态 c_{t_5} 。输入门和新状态的计算为

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (4)$$

$$c_t = f \odot c_{t-1} + i_t \odot \tilde{c}_t$$

式中: \tilde{c}_t 为 t 时刻输入的单元记忆状态; W_i 、 W_c 分别为输入门、状态权重矩阵; b_i 、 b_c 分别为输入门、状态偏置项; \tanh 为激活函数, 如式(5)所示。

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

输入门在 t_5 时刻, 消除了 t_4 时刻对数据序列的影响, 将整体振荡状态的特征保留在 c_{t_5} 中。LSTM 网络得到新的记忆状态 c_{t_5} 后, 还需产生 t_5 时刻输出, 即此时序列的短期特征 h_{t_5} 。输出门 o_{t_5} 根据新状态 c_{t_5} 、上一时刻输出 h_{t_4} 和当前的输入 x_5 来决定该时刻的输出 h_{t_5} 。其计算公式为

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

式中: W_o 为输出门权重矩阵; b_o 为输出门偏置项。

2 PMU 不良数据检测

2.1 PMU 数据分解重构方法

对 PMU 数据分析与重构是不良数据检测的基础。电力系统准稳态和动态下 PMU 数据特性差别较大, 本文拟分别对准稳态和动态下 PMU 数据进行分解重构。因此, 首先需对 PMU 数据进行准稳态和动态的辨识, 该方法在文献[24]中有详细论述,

并非本文重点, 在此不进行介绍。

假设某时间 T 内的 PMU 量测幅值数据 $D=X_1, X_2, \dots, X_{nk}$, 数据量 $N=nk$, 其中未包含不良数据, 如图 4 所示。

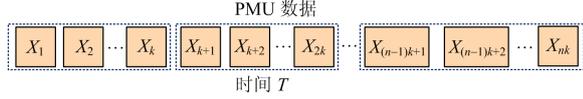


图 4 PMU 正常量测数据示意图

Fig. 4 PMU normal measurement data diagram

图 4 中, 橘色方框表示正常数据 x_i , 每 k 个数据分为一组, 作为独立训练样本, 共有 n 组。PMU 数据分解重构网络框架如图 5 所示。首先利用单层网络可多输入单输出的特点, 构造网络 LSTM-d 对正常量测数据分解, 得到原始数据的特征向量 h_k^d 。再利用单层网络可单输入多输出的特点, 构造网络 LSTM-r 将上层网络的输出向量 h_k^d 作为本层的初始化输入, 对正常量测数据重构。从图 5 可知, 序列分解重构过程中, 需对 LSTM-d、LSTM-r 网络进行训练, 对两层网络中遗忘门 f_i 、输入门 i_i 、输出门 o_i 的各参数 W_f 、 b_f 、 W_i 、 b_i 、 W_o 、 b_o 不断优化, 使重构数据与原始数据尽量一致。以第一个样本中的 PMU 幅值数据 x_1, x_2, \dots, x_k 为例, 将其作为第一层 LSTM-d 网络的输入。每一时刻的 x_i 和上一时刻的输出 h_{i-1}^d 一同决定该时刻的输出 h_i^d 。第二层网络 LSTM-r 将上层网络的向量输出 h_k^d 作为其初始输入, 逆序重构。这使得原始数据中大部分信息均可用于重构, 保证了重构的准确性, 重构序列为

$$x_i^r = W_i^r h_i^r + B_i^r \quad (8)$$

式中, W_i^r 、 B_i^r 分别为 LSTM-r 网络中待训练的权重矩阵和偏置项。

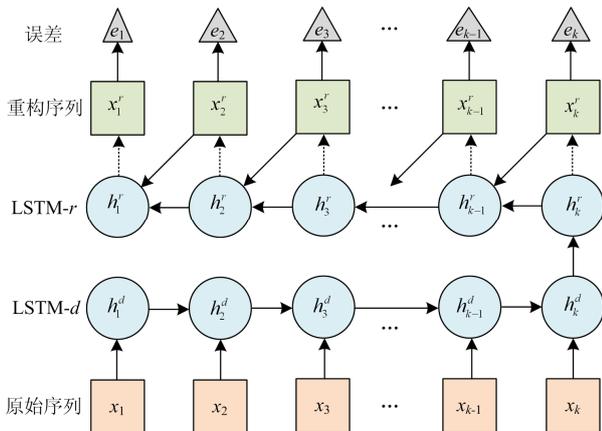


图 5 PMU 数据分解重构框架

Fig. 5 PMU data decomposition and reconstruction framework

另外, 为保证所得模型更加准确, 训练时, LSTM-r 网络中每一时刻的 h_i^r 由上一时刻的输出 h_{i+1}^r 和原始序列的 x_i 共同决定。测试时, LSTM-r 网络中每一时刻的 h_i^r 由上一时刻的输出 h_{i+1}^r 和重构序列的 x_i^r 共同决定。通过大量数据训练, 可得重构序列 x_i^r , 则误差为

$$e_i = |x_i - x_i^r| \quad (9)$$

训练目标分别设为如下函数的最小值,

$$M = \min \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (10)$$

$$G = \min \frac{1}{N} \sum_{i=1}^N e_i \quad (11)$$

将原序列的 n 个样本依次输入图 5 所示的网络结构中进行训练, 由两个目标函数可得到两个参数不同的网络模型。重构序列 x_i^r 均对应重构误差 e_{iM} 和 e_{iG} 。若重构误差较大, 则判定为不良数据; 较小则为正常数据。因此, 误差阈值的设置是实现不良数据检测的关键。人为设置误差阈值, 主观性较强, 可能存在部分不良数据漏判, 基于决策树的不良数据阈值设置方法可避免这一现象。

2.2 PMU 不良数据阈值确定

将相同 PMU 另一时间 T' 内的量测数据作为测试数据 D' , 其中包含不良数据, 如图 6 所示。

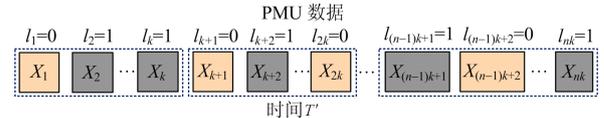


图 6 测试数据示意图

Fig. 6 Test data diagram

图 6 中, 灰色方框表示不良数据。正常数据和不良数据的标签分别设置为 0 和 1。将测试数据分别输入上述训练得到的两个网络, 可得各数据的重构误差分别为 e_{iM} 、 e_{iG} 。利用 C4.5 决策树对误差阈值进行判定^[25-26], 从而实现不良数据的检测。将 e_M 、 e_G 作为特征属性, 数据类型作为类别属性, 具体步骤如下。

1) 将 80% 的测试数据用于训练决策树, 剩余数据用于验证检测精度。

2) 计算样本的信息熵, 如式(12)所示。

$$E(D') = - \sum_{i=1}^2 p_i \log_2 p_i \quad (12)$$

式中: p_1 表示正常数据在样本中的比例; p_2 表示不良数据的比例。

3) 计算两类误差特征的信息增益和增益率。假

设特征 e_M 作为划分特征, 其存在 v 个从小到大排列的取值 $\{e_M^1, e_M^2, \dots, e_M^v\}$, 将各区间 $[e_M^i, e_M^{i+1})$ 的中位点 $\frac{e_M^i + e_M^{i+1}}{2}$ 作为候选划分点, 即有 $v-1$ 个候选划分点。

$$Q = \{q_i = \frac{e_M^i + e_M^{i+1}}{2} | 1 \leq i \leq v-1\} \quad (13)$$

各候选划分点的信息增益为

$$Z(D', e_M, q_i) = E(D') - \frac{|D'_{q_i^-}|}{|D'|} E(D'_{q_i^-}) - \frac{|D'_{q_i^+}|}{|D'|} E(D'_{q_i^+}) \quad (14)$$

式中: $|D'|$ 表示训练的样本个数; $|D'_{q_i^-}|/|D'|$ 表示 $e_M^i \leq q_i$ 个数所占比例; $|D'_{q_i^+}|/|D'|$ 表示 $e_M^i > q_i$ 个数所占比例。

4) 选择所有候选划分点中信息增益最大的作为误差特征 e_M 的信息增益, 由于信息增益偏重可取值数目较多的特征, 需计算增益率作为最终划分标准, 计算公式如为

$$R(D', e_M, q_i) = \frac{Z(D', e_M, q_i)}{I(e_M)} \quad (15)$$

$$I(e_M) = - \sum_{\beta \in \{-, +\}} \frac{|D'_{q_i^\beta}|}{|D'|} \log_2 \frac{|D'_{q_i^\beta}|}{|D'|} \quad (16)$$

式中, $I(e_M)$ 为特征 e_M 的固有属性, 其可能取值个数 v 越多, 则 $I(e_M)$ 通常越大。

5) 重复上述步骤 3)、4) 即可构造如图 7 所示决策树。

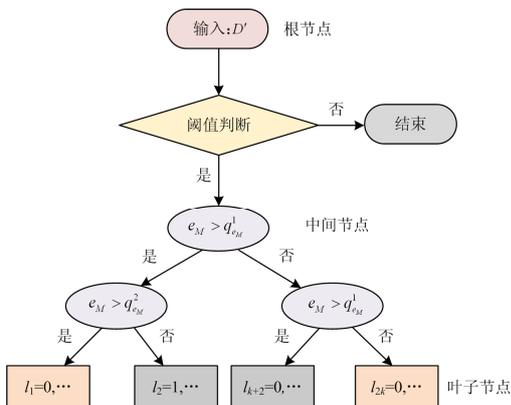


图 7 决策树示意图

Fig. 7 Decision tree diagram

图 7 中: 叶子节点表示划分结果为正常或不良数据, 其中橘色框中为正常数据, 灰色框中为不良数据; 中间节点表示训练样本中对应的误差特征划分, 其中紫色圈为不同阶段对特征的划分。首先, 选择两类误差特征 e_M 、 e_G 中较大的信息增益率与给

定阈值比较, 如黄色框所示。若小于, 则停止计算, 所有数据均为正常或不良数据, 该树为单节点树; 若大于, 则进行划分。假设 $q_{e_M}^1$ 作为划分点时信息增益率最大, 如第一阶段紫色圈所示, 特征 $e_M > q_{e_M}^1$ 为一类, 剩余为一类。依次递归重复, 直至划分出的数据均属于同一类别, 如图 7 中叶子节点所示。

最后, 将剩余测试数据的误差 e_{iM} 、 e_{iG} 依次输入决策树, 即可实现不良数据检测, 并验证模型准确性。结果在算例分析中给出。

2.3 算法流程

综上, 本文提出的 PMU 不良数据检测方法如图 8 所示。

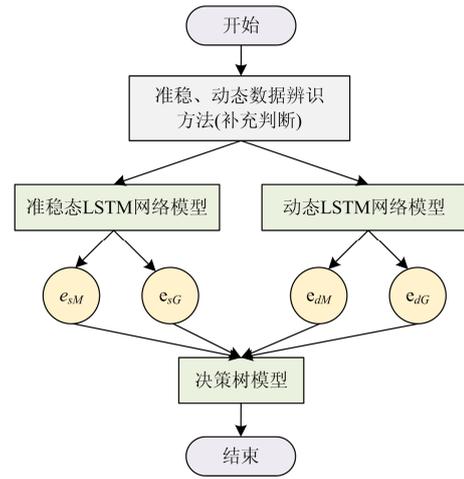


图 8 算法流程示意图

Fig. 8 Flow chart of algorithm

首先对待测数据进行准稳态和动态辨识; 将准稳态和动态数据分别输入训练好的稳态和动态 LSTM 网络模型中, 可得准稳态数据的重构误差 e_{sM} 、 e_{sG} 和动态数据的重构误差 e_{dM} 、 e_{dG} ; 最后, 利用决策树模型对重构误差进行判断, 即可有效检测 PMU 不良数据。

需特别指出, 当发生瞬时性故障时, PMU 测量数据会出现少量突变的数据, 与不良数据行为极为类似。为减少误判, 本文利用瞬时故障时电压与电流的关系, 在利用本文方法识别出坏数据后额外加一个判据。即若电压幅值减小的同时电流幅值增大, 则有发生了瞬时故障的可能, 需对识别结果进行标注, 再通过其他测量装置的数据进行验证, 确定是否为不良数据。若不满足上述条件, 则直接判断为不良数据。

3 仿真与实际数据验证

本节分别对电力系统稳态和动态信号进行仿真测试, 并利用现场 PMU 实测数据验证, 与文献[20]

中的聚类方法、文献[21]中的集成学习法和文献[22]中的主成分分析法对比检测结果。其中, LSTM-d 的隐藏层维度为 64, LSTM-r 的隐藏层维度为 128, 样本长度为 100。

3.1 仿真测试

本文计算环境为: 计算机 CPU 为 Core i7-8700K, 主频为 3.7 GHz, 内存为 16 GB, GPU 为 NVIDIA GTX 1080Ti, 在 Matlab 2018b 平台上编制计算程序。

1) 系统稳态信号仿真

理想条件下, 电力系统无外界干扰, 通常处于稳态。其信号的一般表达式为

$$x(t) = \sqrt{2}X_m \cos(2\pi f_0 t + \varphi_0) + n(t) \quad (17)$$

式中: X_m 为相量幅值; f_0 为工频; φ_0 为初相角; $n(t)$ 为噪声部分。 $X_m=57.73$ V; $f_0=50$ Hz; $\varphi_0=0^\circ$; $n(t)$ 的信噪比为 60 dB。

在上述信号中设置幅值偏差范围 0.3%~5% 的不良数据, 对比三种方法的检测结果, 如图 9 所示。

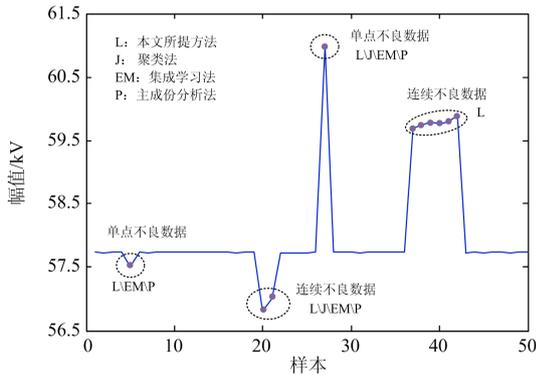


图 9 稳态下不良数据检测结果

Fig. 9 Bad data detection result in steady state

图 9 中, 紫色标记为不良数据, 包括单点不良数据和连续不良数据。L 表示本文所提方法, J 表示文献[20]的聚类法, EM 表示文献[21]的集成学习法, P 表示文献[22]的主成分分析法。从图 9 可以看出, 本文所提方法可以检测出不同类型的不良数据, 聚类法可检测偏差较大的单点不良数据, 集成方法对单点不良数据检测准确, 两者对连续数据的检测能力均较差。这是因为偏差较小的单点不良数据与正常数据联系紧密, 聚类方法易造成误判。同时, 当连续不良数据个数较多, 且幅值接近时, 聚类法和集成学习法难以与正常数据区分。由于多点连续不良数据对成分分解过程影响较大, 主成分分析法对其检测能力较弱。

2) 系统动态信号仿真

采用含有 3 种间谐波的信号表征系统次同步振

荡现象, 信号表达式为

$$x(t) = \sqrt{2}X_m \cos(2\pi f_0 t + \varphi_0) + \sqrt{2} \sum_{i=1}^3 X_{mi} \cos(2\pi f_i t + \varphi_i) + n(t) \quad (18)$$

式中: X_{mi} 为某一间谐波相量幅值; f_i 为其谐波频率; φ_i 为其初相角; $n(t)$ 的信噪比为 60 dB。

$$X_{m1}=2 \text{ V}; f_1=20 \text{ Hz}; \varphi_1=0^\circ;$$

$$X_{m2}=3 \text{ V}; f_2=40 \text{ Hz}; \varphi_2=0^\circ;$$

$$X_{m3}=4 \text{ V}; f_3=65 \text{ Hz}; \varphi_3=0^\circ;$$

在上述调制信号中设置幅值偏差范围 2%~15% 的不良数据, 对比四种方法的检测结果, 如表 1 所示。

表 1 四种方法对不同偏差不良数据的检测结果

Table 1 Detection results of four methods of different deviations of bad data

| 方法 | 2% | 5% | 10% | 15% |
|--------|----|----|-----|-----|
| 本文所提方法 | ✓ | ✓ | ✓ | ✓ |
| 聚类法 | × | × | ✓ | ✓ |
| 集成学习法 | × | × | ✓ | ✓ |
| 主成分分析法 | ✓ | ✓ | ✓ | ✓ |

设置不同的不良数据比例, 对比四种方法的检测结果, 如表 2 所示。

表 2 四种方法对不同比例不良数据的检测结果

Table 2 Detection results of four methods of different proportions of bad data

| 方法 | 1% | 5% | 10% | 20% |
|--------|----|----|-----|-----|
| 本文所提方法 | ✓ | ✓ | ✓ | ✓ |
| 聚类法 | ✓ | ✓ | × | × |
| 集成学习法 | ✓ | ✓ | × | × |
| 主成分分析法 | ✓ | ✓ | × | × |

由表 1、表 2 可知, 本文所提方法的检测范围大于另外三种的方法。当不良数据偏差较小时, 聚类法和集成学习法不能很好地检测不良数据, 主成分分析法受偏差影响较小; 当不良数据比例较多时, 聚类法、集成学习法和主成分分析法的检测能力较差。

3.2 仿真测试

中国西部某新能源汇集地区已多次发生因间谐波导致的次同步振荡事件。本文在某次次同步振荡时 PMU 监测所得数据的基础上进行人为设置不良数据验证本文方法。其中振荡数据约为 3 000 个, 结合临近线路的振荡数据与历史数据, 在各数据集中选取等量样本数, 各样本均为 100 个连续振荡数据, 训练样本个数为 1 000 个。人为设置连续不良数据和单点不良数据, 检测结果如图 10 所示。

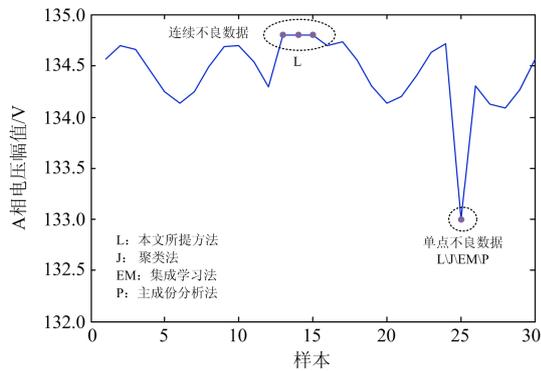


图 10 新能源地区 PMU 实测数据检测结果

Fig. 10 Detection results of PMU measurement data in a renewable energy area

从图 10 可知, 动态下, 四种方法均可有效检测单点不良数据。而当连续不良数据存在时, 仅本文方法可有效检测。由于连续不良数据的偏差较小, 聚类和集成学习的方法不能较好地检测, 而成分分解又受其影响导致主成分分析法对其检测能力较弱。

在不同时间窗下对比四种检测算法的运行时间, 其中, L 表示本文所提方法, J 表示聚类法, EM 表示集成学习法, P 表示主成分分析法。结果如表 3 所示。可知, 随着时间窗的增加, 四种方法的运行时间均会增加。由于集成学习法较为复杂, 因此它的运行时间较其他方法而言更长。聚类法运行速度较快, 本文所提方法与主成分分析法运行时间相近。

表 3 四种方法运行时间对比

Table 3 Comparison of running time of four methods

| 时间窗(数据量) | 时间 L/s | 时间 J/s | 时间 EM/s | 时间 P/s |
|-----------|---------|---------|---------|---------|
| 1 s (50) | 0.006 4 | 0.004 9 | 0.011 | 0.005 8 |
| 2 s (100) | 0.015 | 0.013 | 0.026 | 0.012 |
| 4 s (200) | 0.026 | 0.028 | 0.057 | 0.027 |

当不良数据比例一定时, 改变不良数据与正常数据的偏差范围, 对比四种方法的检测精度, 结果如表 4 所示。

由表 4 可知, 当不良数据比例不同时, 本文所提方法对不同偏差范围下的不良数据均具有较好的检测效果。随着不良数据比例的增加, 四种方法的检测精度均不同程度地降低。同时, 检测精度随着不良数据偏差范围的增加而提高, 聚类方法受不良数据比例和偏差范围的影响最大。

3.3 实际系统应用效果验证

为验证所提方法在实际应用中的效果, 将该方法在课题组开发的源网荷全景同步测量系统中进行了应用, 实时检测不良数据。测量系统运行环境:

国产高性能服务器 8 台, CPU 为 Xeon E5-2609v4, 主频 1.70 GHz, 内存为 16 G。该系统在 2019 年 6 月 17 日 11 时 14 分 55 秒至 11 时 14 分 56 秒间, 有装置因时间同步丢失, 造成幅值测量结果跳变, 产生不良数据, 利用本文算法检测不良数据, 结果如图 11 所示。

表 4 四种方法检测精度对比

Table 4 Comparison of detection accuracy of four methods

| 不良数据比例/ 偏差范围 | 方法 | 准确性 |
|-----------------|----|--------|
| 1.0%/10% | L | 99.93% |
| | J | 94.21% |
| | EM | 99.13% |
| | P | 97.54% |
| 2.5%/10% | L | 99.52% |
| | J | 91.10% |
| | EM | 97.69% |
| 5.0%/10% | L | 98.81% |
| | J | 88.05% |
| | EM | 95.78% |
| | P | 93.07% |

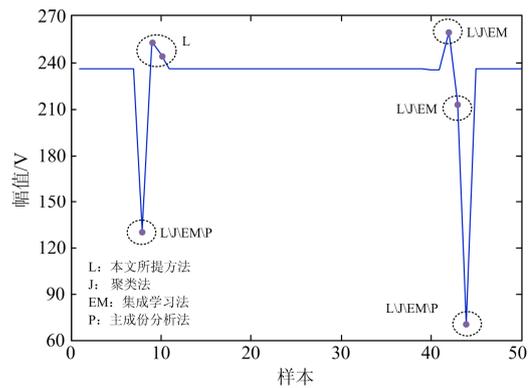


图 11 全景同步量测系统测量数据检测结果

Fig. 11 Detection results of distribution network PMU measurement data

由图 11 可得, 本文所提方法可有效检测全部不良数据。聚类法、集成学习法和主成分分析法对连续不良数据的检测效果不好, 由于连续不良数据中存在与正常数据幅值较为接近的数据, 这些数据易影响聚类法和集成学习法。同时, 连续不良数据对成分分解的影响较大, 导致其检测结果出现漏判。因此, 本文所提方法可有效检测系统静动下的多种类型不良数据, 将不良数据检测与相应修正算法结合, 可显著提高 PMU 数据质量。

4 结论

本文提出了一种基于长短期记忆网络的 PMU 不良数据检测方法,该方法可准确检测出多种因素导致的不良数据,提高了 PMU 数据质量,为状态感知提供了数据基础,得到如下结论:

1) 利用 LSTM 网络对序列选择记忆的特点,提出了一种 PMU 数据分解重构的双层 LSTM 网络架构。

2) 构造两种目标函数,获取不同误差特征检测不良数据,保证了检测的准确性。

3) 与基于监督学习的决策树相结合,避免了人为设定不良数据阈值的主观性。

4) 所提方法在系统处于不同状态下,面对多种类型的不良数据均可有效检测,弥补了现在方法在动态条件下检测精度不高的缺陷。

本文未考虑量测装置算法问题,下一步将结合地理位置临近的多台 PMU 量测信息和系统拓扑,对 PMU 不良数据的检测和修正进行深入研究。

参考文献

- [1] “透明电网”是智能电网发展的极致体现[EB/OL]. [2019-05-27]. http://www.ccdi.gov.cn/yaowen/201905/t20190527_194627.html.
“Transparent Power Grid” is the ultimate embodiment of the development of smart grid[EB/OL].[2019-05-27]. http://www.ccdi.gov.cn/yaowen/201905/t20190527_194627.html.
- [2] 张宁, 杨经纬, 王毅, 等. 面向泛在电力物联网的 5G 通信:技术原理与典型应用[J]. 中国电机工程学报, 2019, 39(14): 4015-4025.
ZHANG Ning, YANG Jingwei, WANG Yi, et al. 5G communication for the ubiquitous internet of things in electricity: technical principles and typical applications[J]. Proceedings of the CSEE, 2019, 39(14): 4015-4025.
- [3] 董飞飞, 俞登科. 我国电网技术的回顾与展望[J]. 智慧电力, 2018, 46(4): 6-10.
DONG Feifei, YU Dengke. Retrospect and prospect of china power grid technology[J]. Smart Power, 2018, 46(4): 6-10.
- [4] 袁明军, 江浩, 黎强, 等. 四统一四规范同步相量测量装置关键技术研究[J]. 电力系统保护与控制, 2019, 47(4): 188-193.
YUAN Mingjun, JIANG Hao, LI Qiang, et al. Research on key technology of four unified four standard phasor measurement unit[J]. Power System Protection and Control, 2019, 47(4): 188-193.
- [5] HUANG Can, LI Fangxing, ZHOU Dao, et al. Data quality issues for synchrophasor applications part I: a review[J]. Journal of Modern Power Systems and Clean Energy, 2016, 4(3): 342-352.
- [6] 刘洁波, 黄纯, 江亚群, 等. 基于强跟踪泰勒-卡尔曼滤波器的动态相量估计算法[J]. 电工技术学报, 2018, 33(2): 433-441.
LIU Jiebo, HUANG Chun, JIANG Yaqun, et al. Dynamic phasor estimator based on strong tracking Taylor-Kalman filter[J]. Transactions of China Electrotechnical Society, 2018, 33(2): 433-441.
- [7] 高海翔, 伍双喜, 苗璐, 等. 发电机组引发电网功率振荡原因及其抑制措施研究综述[J]. 智慧电力, 2018, 46(7): 49-55.
GAO Haixiang, WU Shuangxi, MIAO Lu, et al. Overview of reasons for generator-induced power oscillations & its suppression measures[J]. Smart Power, 2018, 46(7): 49-55.
- [8] 李珊, 何怡刚, 项胜, 等. 基于故障电压比较与多信息融合的广域后备保护算法[J]. 电力系统保护与控制, 2015, 43(22): 92-98.
LI Shan, HE Yigang, XIANG Sheng, et al. A wide backup protection algorithm based on fault voltage comparison and multi-information fusion[J]. Power System Protection and Control, 2015, 43(22): 92-98.
- [9] 李施, 丁仁杰. 一种基于信号相关性的低频振荡辨识方法[J]. 电力系统保护与控制, 2018, 46(11): 46-54.
LI Shi, DING Renjie. An identification method for low-frequency oscillation based on signal correlation[J]. Power System Protection and Control, 2018, 46(11): 46-54.
- [10] ZHANG Fang, SUN Yuanzhang, CHENG Lin, et al. Measurement and modeling of delays in wide-area closed-loop control systems[J]. IEEE Transactions on Power Systems, 2015, 30(5): 2426-2433.
- [11] 王宾, 孙华东, 张道农. 配电网信息共享与同步相量测量应用技术评述[J]. 中国电机工程学报, 2015, 35(增刊 1): 1-7.
WANG Bin, SUN Huadong, ZHANG Daonong. Review on data sharing and synchronized phasor measurement technique with application in distribution systems[J]. Proceedings of the CSEE, 2015, 35(S1): 1-7.
- [12] HUANG Can, LI Fangxing, ZHAN Lingwei, et al. Data quality issues for synchrophasor applications part II: problem formulation and potential solutions[J]. Journal of Modern Power Systems and Clean Energy, 2016, 4(3): 353-361.

- [13] BHARGAV A, DUSMANTA K M. A review on synchrophasor communication system: communication technologies, standards and applications[J]. Protection and Control of Modern Power Systems, 2018, 3(3): 383-399. DOI: 10.1186/s41601-018-0110-4.
- [14] WU Meng, XIE Le. Online Detection of low-quality synchrophasor measurements: a data driven approach[J]. IEEE Transactions on Power Systems, 2017, 32(4): 2817-2827.
- [15] 罗深增, 李银红, 石东源. 广域测量系统可观性概率评估及其在 PMU 优化配置中的应用[J]. 电工技术学报, 2018, 33(8): 1844-1853.
LUO Shenzeng, LI Yinong, SHI Dongyuan. Wide area monitoring system observability probabilistic evaluation and its application in optimal PMU placement[J]. Transactions of China Electrotechnical Society, 2018, 33(8): 1844-1853.
- [16] JIM F, BRETT A. A data quality filter for PMU measurements: description, experience, and examples[C] // 2016 IEEE Power and Energy Society General Meeting (PESGM), July 17-21, 2016, Boston, MA, USA, 2016: 1-5.
- [17] SCOTT G, JOE Chow, GEORGE S, et al. Phasor-measurement-based state estimation for synchrophasor data quality improvement and power transfer interface monitoring[J]. IEEE Transactions on Power Systems, 2014, 29(2): 881-888.
- [18] KEVIN J, ANAMITRA P, JAMES T. Methodology for performing synchrophasor data conditioning and validation[J]. IEEE Transactions on Power Systems, 2015, 30(3): 1121-1130.
- [19] QIANG Z, VAITHIANATHAN V. Synchrophasor time skew: formulation, detection and correction[C] // 2014 North American Power Symposium (NAPS), 2014, Pullman WA: 1-6.
- [20] RANGANATH V, ANUPAM M, PRANKASH R. A density-based clustering scheme for situational awareness in a smart-grid[C] // 2015 IEEE International Conference on Electro/Information Technology (EIT), May 21-23, 2015, Dekalb, IL, USA: 346-350.
- [21] ZHOU Mengze, WANG Yuhui, SRIVASTAVA A, et al. Ensemble-based algorithm for synchrophasor data anomaly detection[J]. IEEE Transactions on Smart Grid, 2019, 10(3): 2979-2988.
- [22] MAHAPATRA K, CHAUDHURI N, KAVASSERI R. Bad data detection in PMU measurements using principal component analysis[C] // 2016 North American Power Symposium (NAPS), September 18-20, 2016, Denver, CO, USA: 1-6.
- [23] HUANG P, KIM M, HASEGAWA-JOHNSON M, et al. Joint optimization of masks and deep recurrent neural networks for monaural source separation[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2015, 23(12): 2136-2147.
- [24] YANG Zhiwei, LIU Hao, BI Tianshu, et al. An adaptive PMU missing data recovery method[J]. International Journal of Electrical Power & Energy Systems, 2020, 116, 105577.
- [25] ZHAO Huimin, RAM S. Constrained cascade generalization of decision trees[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(6): 727-739.
- [26] QUINLAN R. C4.5: programs for machine learning[M]. San Francisco: Morgan Kaufmann Publishers Inc, 1992.

收稿日期: 2020-01-23

作者简介:

杨智伟(1994—), 男, 博士研究生, 主要研究方向为 PMU 数据质量评估; E-mail: hsyangzhiwei@163.com

刘 灏(1985—), 男, 通信作者, 副教授, 主要研究方向为同步相量测量技术; E-mail: hliu@ncepu.edu.cn

毕天姝(1973—), 女, 教授, 博士生导师, 主要研究方向为电力系统保护与控制、广域同步相量测量技术及应用。E-mail: tsbi@ncepu.edu.cn

(编辑 周金梅)