Generalized Graph Laplacian Based Anomaly Detection for Spatiotemporal \(\mu\)PMU Data

Mingjian Cui, Senior Member, IEEE, Jianhui Wang, Senior Member, IEEE, Anthony R. Florita, Member, IEEE, and Yingchen Zhang, Senior Member, IEEE

Abstract—This letter develops a novel anomaly detection method using the generalized graph Laplacian (GGL) matrix to visualize the spatiotemporal relationship of distribution-level phasor measurement unit (\(\mu\)PMU) data. The \(\mu\)PMU data in a specific time horizon is segregated into multiple segments. An optimization problem formulated as a Lagrangian function is utilized to estimate the GGL matrix. During the iterative process, an optimal update is constituted as a quadratic program (QP) problem. To perform the \(\mu\)PMU-based spatiotemporal analysis, normalized diagonal elements of GGL matrix are proposed as a quantitative metric. The effectiveness of the developed method is demonstrated through real-world \(\mu\)PMU measurements gathered from test feeders in Riverside, CA.

Index Terms—Anomaly detection, distribution PMU (\(\mu\)PMU), graph Laplacian matrix, spatiotemporal analysis.

I. INTRODUCTION

With the increasing development of distribution-level phasor measurement units (\(\mu\)PMUs), the real-time monitoring of distribution systems has been significantly improved with advanced sensor technologies. Three-phase voltage and current phasors with corresponding magnitude and angle information can be provided by GPS-synchronized measurements [1]. Synchronized voltage and current measurements at higher resolution and precision are capable of facilitating a high level of visibility for distribution systems [2]. Synchronized \(\mu\)PMU data-driven anomaly detection can contribute to numerous applications in distribution system operations. Farajollahi et al. [1] developed a method to locate the source of events in power distribution systems by using distribution-level \(\mu\)PMUs. Jamei et al. [3] detected the abnormal behavior in the control perimeter by focusing on \(\mu\)PMUs.

Unlike conventional \(\mu\)PMU event detection techniques, such as the CUSUM algorithm [3] which is highly dependent on a pre-defined threshold obtained by training a semi-supervised behavior-based detector, this letter attempts to capture affinity similarities and patterns among \(\mu\)PMU data by using a generalized graph Laplacian (GGL) matrix and the three-sigma rule. The \(\mu\)PMU data is represented by weighted graphs that can offer general and flexible representations for modeling affinity similarity among anomalies in \(\mu\)PMU data. Essentially, the similarity of \(\mu\)PMU data under normal conditions would be broken when anomaly events occur. Based on this concept, this letter seeks to address two critical questions for \(\mu\)PMU-based anomaly detection. (i) Is it possible to quantitatively represent the similarity between anomalies and the normal \(\mu\)PMU data with numerical metrics? (ii) Can users deploy the spatiotemporal similarity to identify system-wide anomalies? Toward this end, this letter develops a GGL based anomaly detection method for the spatiotemporal analysis and visualization. The main technical contributions of this letter are that: (i) using graph learning techniques to detect \(\mu\)PMU-based anomalies for the first time and (ii) analyzing spatiotemporal characteristics of anomalies using high-resolution \(\mu\)PMU data.

The organization of this letter is as follows. In Section II, the developed GGL method is briefly introduced with the pseudocode of generating GGL matrix. Case studies and result analysis performed on real-time \(\mu\)PMU data are discussed in Section III. Concluding remarks are summarized in Section IV.

II. GGL METHODOLOGY

The increasing graph learning techniques pave a distinctive way for the anomaly detection using the streaming \(\mu\)PMU data with high resolution. Compared with the combinatorial graph Laplacians (CGLs), GGLs are shown to be more useful and practical [4]. GGL can maintain all the edges with positive weights and introduce additional connectivity due to negative weights [5]. To estimate the GGL matrix, the optimization problem using a Lagrangian function is constituted as:

\[
\min L(\Theta) = \|\Theta \odot H\|_1 + \|\Theta \odot M\|_1 - \logdet(\Theta)
\]

s.t.

\[
\mathcal{L}(\Theta) = \left\{ \Theta \in \mathcal{L} \left| \begin{array}{c} (\Theta)_{ij} \leq 0 \quad \text{if} \quad (A)_{ij} = 1 \\ (\Theta)_{ij} = 0 \quad \text{if} \quad (A)_{ij} = 0 \end{array} \right. \right\}
\]

where \(H\) is the regularization matrix and \(H = \alpha(I - \Pi)\). \(I\) is an identity matrix. \(\Pi\) is an all-ones matrix. \(\alpha\) is the regularization parameter. \(\Theta\) is the estimated GGL matrix. \(\mathcal{L}\) is the target set of graph Laplacians. \(A\) is the similarity matrix. \(\odot\) means the element-wise multiplication of two matrices. \(\|\cdot\|_1\) means the sum of absolute values of all elements (\(\ell_1\)-norm). \(\logdet(\cdot)\) means the natural logarithm of a determinant. \(M\) is the Lagrange multiplier matrix. The similarity matrix \(A\) is calculated by the difference of \(d\text{diag}(\Theta) - \Theta\):

\[
A = d\text{diag}(\Theta) - \Theta
\]

The derivative of (1) with respect to \(\Theta\) is set to an all-zeros matrix \(O\), given by:

\[
dL(\Theta)/d\Theta = -\Theta^{-1} + K + M = O
\]

where \(K = S + H\) and \(S\) is the covariance matrix empirically obtained from \(\mu\)PMU data samples.

Letting the inverse matrix \(C = \Theta^{-1}\) and using the formula in (4), the optimality condition for the \(u\)th row/column of \(\Theta\), given by:

\[
-c_u + k_u + m_u = o
\]

\[
-c_u + k_u + m_u = 0
\]
where $c_u$, $k_u$, and $m_u$ are vectors obtained from the $u$th row (or column) of matrices $C$, $K$, and $M$, respectively, $c_u$, $k_u$, and $m_u$ are scalars obtained from the $u$th diagonal of matrices $C$, $K$, and $M$, i.e., $c_u = (C)_{uu}$, $k_u = (K)_{uu}$, and $m_u = (M)_{uu}$, respectively.

Based on the above optimality conditions, an optimal update is designed for the $u$th row/column of $\Theta$ as a quadratic program (QP) problem, given by:

$$
\min_{\beta} \beta^T Q \beta / 2 - \beta^T p
$$

(7)

where $\beta = - (\theta_u)_S$, $p = (k_u/k_u)_S$, and $Q = (\Theta^{-1})_{SS}$. Vectors/matrix $\beta$, $p$, and $Q$ are selected from the original variable vectors/matrix $\theta_u$, $k_u/k_u$, and $\Theta^{-1}$ based on the index set $S$.

The pseudocode in Algorithm 1 is shown to solve the desired GGL matrix. It iteratively updates each row/column of the estimated GGL matrix $\hat{\Theta}$ and its estimated inverse matrix $\hat{C}$ by optimizing the QP problem formulated in (7). The convergence criterion is defined as the sum of squared values of GGL matrix elements in the $u$th iteration:

$$
\frac{||\hat{\Theta}_u - \hat{\Theta}_{u-1}||_F}{||\hat{\Theta}_{u-1}||_F} = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} (\hat{\theta}_{j,v} - \hat{\theta}_{i,v})^2} - \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} (\hat{\theta}_{j,v-1} - \hat{\theta}_{i,v-1})^2}
$$

(8)

III. CASE STUDIES AND RESULTS

A. μPMU Data Description and Preprocessing

The performance of our developed method is validated using μPMU data from four real-world distribution feeder and building locations in Riverside, CA [1]. Distribution μPMU devices are used to collect the synchrophasor data of the voltage and current data with a 60 Hz sampling rate. In this letter, it is assumed that each data segment spans approximately 20 seconds and the total number of segments is 37, which corresponds to an approximately 97% true positive rate (TPR) of detecting anomalies. The μPMU data is preprocessed to filter the ambient noise using $\ell_1$ trend filtering, which is a simple and tractable technique for generating piecewise linear fits to the original data by solving a convex program [7].

B. Detected Anomalies in Temporal Domain

Fig. 1 shows the visualization of current data of Phase $a$ at Location 1 with two anomaly events. As shown in the bottom subfigure, the preprocessed current data without ambient noise (orange line) is used to detect the anomalies. The first anomaly (27th segment) is shown as a smooth-curve event [8] and challenging to detect for operators. It is mainly because this smooth-curve anomaly presents a smooth and continuous coupling together with adjacent points of current data. The second anomaly (30th segment) is shown as a small pulse event [8]. The green line indicates the diagonal elements of estimated GGL matrix $\hat{\Theta}$. As shown in the top subfigure, a deeper color means a higher similarity between two segments. Both segments 27 and 30 almost do not have similarity with any other segments (see lines connected with Seg-27 and Seg-30 denoted by the lack of colors) due to the occurrence of anomalies. By comparing subfigures in Fig. 1, the accuracy of detected anomalies using GGL can be validated.

Table I sorts the metric of the GGL matrix’s diagonal elements, $\text{diag}(\hat{\Theta})$, with the first four segments. The threshold of this metric is determined by the empirical three-sigma rule [9]. Based on experimental experience in this case, we use the $1.2$-sigma rule to obtain thresholds. If the metric is less than the threshold, anomalies are detected. As can be seen, segments 27 and 30 with anomalies show the smallest values (see the

![Fig. 1. Visualization of current data anomalies of Phase $a$ at Location 1.](image-url)

Table I: Metrics of First Four Segments of Phase $a$ at Location 1

<table>
<thead>
<tr>
<th>Sort of Metrics</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>diag($\hat{\Theta}$)</td>
<td>0.0176 0.0680 0.1549 0.1698</td>
</tr>
<tr>
<td>Segments</td>
<td>Seg-27 Seg-30 Seg-22 Seg-5</td>
</tr>
</tbody>
</table>
red area). This observation verifies that the developed method based on GGL matrix is accurate for anomaly detection in μPMU data.

To compare our method with the CUSUM method in [3], the TPR and false positive rate (FPR) are used for illustration. Table II shows the quantitative results of TPR and FPR. Larger TPR and smaller FPR mean a better performance of the detection method. As shown in this table, the diag(Θ) metric validates the effectiveness of the proposed method.

Table II shows the quantitative results of TPR and FPR using CUSUM and GGL. The proposed method has a higher TPR and lower FPR compared to CUSUM, validating the effectiveness of the proposed method.

![Diag(Θ) results with different ambient noises on 21 time series of μPMU data.](image1)

![Spatiotemporal visualization of anomalies using the normalized diagonal elements of GGL matrix Θ.](image2)

Fig. 2. diag(Θ) results with different ambient noises on 21 time series of μPMU data.

Fig. 3. Spatiotemporal visualization of anomalies using the normalized diagonal elements of GGL matrix Θ.

To show the robustness of the developed approach, 21 time series of μPMU data are collected and added by a Gaussian random noise with a zero mean and a standard deviation of a proportion λ of the corresponding μPMU data. λ is increased from 1% to 3% with a step of 0.1% (i.e., λ ∈ [1%, 3%]). Fig. 2 shows the results of diag(Θ) with different ambient noises. As seen from top to bottom in the vertical axis, the similarity (GGL diagonal) among μPMU segments is gradually reduced with ambient noises denoted by the increasing proportion λ. However, anomalies on segments 27 and 30 are entirely detected for all 21 time series. This observation can validate the robustness of the developed approach with respect to ambient noises.

C. Spatiotemporal Analysis of Detected Anomalies

The three-phase current and voltage data at four locations are used for the spatiotemporal analysis of detected anomalies. Other information of μPMU data is introduced in Section III-A. The normalized diagonal element of GGL matrix, diag(Θ), is proposed as a metric for quantifying spatiotemporal relationship, given by:

\[
diag(Θ) = \frac{\text{diag}(Θ) - \min(\text{diag}(Θ))}{\max(\text{diag}(Θ)) - \min(\text{diag}(Θ))}
\]

Fig. 3 visualizes the spatiotemporal relationship of μPMU data by using the metric diag(Θ). As seen in this figure, each phase has the same metric value (color) among three phases under most circumstances. However, for Phase b current at Location 2 in Seg-34, it is possible that one single-phase current event occurs. We can also see several three-phase current events that occur at Location 3 and 4. For Location 1 and 2, there are two current/voltage events in Seg-21 and Seg-27, respectively. In Seg-7, there is a system-wide voltage-stability event that occurs at all four locations with a deeper color. Theoretically, voltage-stability anomalies routinely occur due to reactive power events caused by distribution generators, capacitors, or reactors. Current events are primarily caused by distribution generator trips, line faults, etc.

IV. CONCLUSION

In this letter, we develop a novel μPMU-based anomaly detection method by using a generalized graph Laplacian (GGL) matrix. The developed detection method utilizes a Laplacian optimization function to estimate the GGL matrix. A quadratic program technique is used to design the optimal update during the computation process. In addition, the normalized diagonal elements of GGL matrix are proposed as a metric for the spatiotemporal analysis of μPMU data. Some observations are shown as follows:

(i) The developed method is able to accurately and robustly detect the anomalies in the temporal domain.

(ii) The spatiotemporal affinity of system-wide anomalies can be clearly characterized and visualized for operators.

ACKNOWLEDGMENT

This work was authored in part by Alliance for Sustainable Energy, LLC, the Manager and Operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Grid Modernization Lab Consortium. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

REFERENCES


