Multiple Event Analysis for Large-scale Power Systems through Cluster-based Sparse Coding

Yang Song, Wei Wang, Zhifei Zhang, Hairong Qi and Yilu Liu
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN 37996, USA
Email: {ysong18, wwang34, zzhang61, hqi, liu}@utk.edu

***Abstract*—Accurate event analysis in real time is of paramount importance for high-fidelity situational awareness such that proper actions can take place before any isolated faults escalate to cascading blackouts. For large-scale power systems, due to the large intra-class variance and inter-class similarity, the nonlinear nature of the system, and the large dynamic range of the event scale, multi-event analysis presents an intriguing problem. Existing approaches are limited to detecting only single or double events or a specified event type. Although some previous works can well distinguish multiple events in small-scale power systems, the performance tends to degrade dramatically in large-scale systems. In this paper, we focus on multiple event detection, recognition, and temporal localization in large-scale power systems. We discover that there always exist groups of buses whose reaction to each event shows high degree similarity, and the group membership generally remains the same regardless of the type of event(s). We further verify that this reaction to multiple events can be approximated as a linear combination of reactions to each constituent event. Based on these findings, we propose a novel method, referred to as *cluster-based sparse coding* (CSC), to extract all the underlying single events involved in a multi-event scenario. Experimental results based on simulated large-scale system model (i.e., NPCC) show that the proposed CSC algorithm presents high detection and recognition rate with low false alarms.**

I. Introduction

Conventional US power system planning is typically performed with the consideration of possible N-1 contingency events of intrinsic equipment failure, where N-1 contingency means one component (or one set of closely related components) fails. Occasionally, some regions of the US power system are designed to handle a few critical N-2 or even N-3 contingencies (i.e., simultaneous failures of 2 or 3 components) with the assistance of post-contingency remedial actions. Nevertheless, recent system disturbance reports from the North American Electric Reliability Corporation (NERC) have made it obvious that major disturbances typically involve a number of unlikely, unplanned events, which will lead to N-X operations under emergency, making multiple event analysis of paramount importance.

In power systems, cascading or simultaneous faults/events are common problems that may lead to large area blackout. Recent large scale failures of the grid, such as the August 2003 U.S. northeastern blackouts and the July 2012 India blackouts, indicate that wide-spread blackouts are always initiated from sporadic fault events, causing devastating effect on our everyday life [1]. In order to ensure safe and healthy operations of the power system, the wide-area situational awareness system (WAMS) can essentially provide high-resolution understanding of the power system dynamics, such that disturbance can be better mitigated in time before it escalates to any unexpected cascading blackouts [2] through the butterfly effect.

In recent decades, many works related to disturbance or event detection and recognition have been reported. Various types of signals have been used for analysis purpose, including frequency [3]–[6], power and voltage [7], and phasor angle [8], [9]. There are generally two approaches for event analysis, model-based and classification-based. Model-based methods (e.g., [7]) use many power system parameters to model the power grid, making it inflexible to structure changes and the algorithm might be only applicable to one type of signal since different events cause different effects on voltage, current, frequency and power angle readings. For these reasons, classification-based approaches are more popular [3], [10]–[13]. Nonetheless these conventional classification methods are only suitable for single event classification as the amount of classes may increase exponentially in multi-event problems. In fact, although any single event can be treated as a distinctive pattern, the combination of two or more single events can theoretically generate infinite patterns. Some other methods adopt graph theory, for example [1], [14], but they strongly depend on the structure of the power system.

Existing research for multiple event analysis is limited from several perspectives. For example, [3] only successfully distinguishes cascading events which can be segmented into single events, and [1] provides only the warning of line failures in the transmission system of the power grid. Also [8] can only handle multiple events of the same type. [6] can only perform well in small power systems.

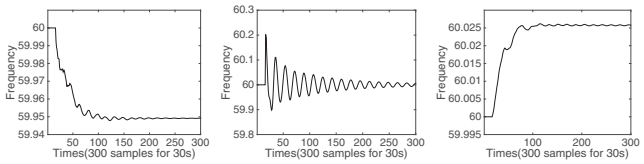
In this paper, we focus on multi-event detection, recognition and temporal localization in large power systems. We choose to analyze the frequency signal because events like generator trips, line trips, or load shedding immediately cause sudden frequency changes within the power system. The rest of the paper is organized as follows: Section II analyzes the challenges of multiple events analysis in large power systems. Section III elaborates on the proposed cluster-based sparse coding algorithm. Section IV demonstrates the effectiveness of this scheme based on simulated data of the NPCC system. Finally, summary and conclusion, as well as future work, are discussed in Section V.

II. Challenges

When multiple events occur in cascading fashion, the electromechanical waves generated will interfere with each other,

978-1-4673-8289-2/15/$31.00 ©2015 IEEE

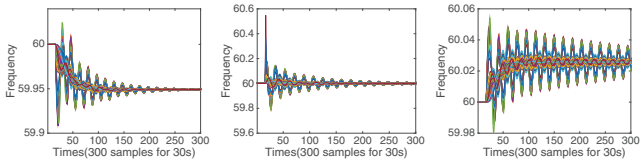
and the frequency measurement taken at a certain bus would more than likely be a mixture of multiple frequency signals. Thus, how to determine the number of events and identify the types of events that occurred with precise estimation of event occurrence time using simply the observed mixture is a very challenging problem. In the following, we describe the challenges from the perspectives of oscillation, intra-class variance, inter-class similarity, the nonlinear characteristics of the power grid, and the large dynamic range of event amplitudes.



(a) GT (b) LT (c) LS

Fig. 1. Frequency signals on one bus corresponding to three typical event types. GT: generator trip, LT: line trip, LS: load shedding.

A typical power system event falls into one of the three categories—generator trip (GT), load shedding (LS), and line trip (LT), as shown in Fig. 1. Generally speaking, these three types of events also come with strong oscillation [15]. For GT and LS, the oscillation can be treated as noise; for LT, however, not only the frequency pattern of oscillation resembles that of LT, the scale of oscillation is also comparable with that of LT. Therefore, oscillation presents great challenge to the detection of LT. From Fig. 2, we also observe that oscillation generally comes with each event type. In addition, oscillation follows various decay rates, making it even harder to distinguish it from an actual LT event.

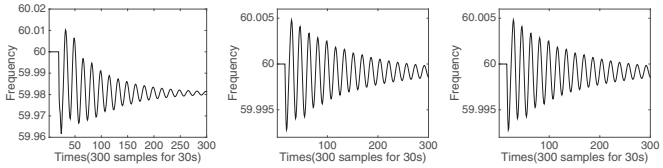


(a) GT (b) LT (c) LS

Fig. 2. Frequency signals on several buses (plotted in different colors) responding to the same single event.

In a small power system, frequency signals of all buses responding to the same event always appear similar. When it comes to a larger and more complicated power system, however, reaction caused by an event can vary largely at different buses. This phenomenon is referred to as the *intra-class variance*, that is, signals belong to the same class show substantial variations, as shown in Fig. 2. This feature of large power systems raises a challenging issue—which signal(s) should be used to represent the behavior of GT? In addition, the same event type caused by different devices at different physical locations may also yield different frequency patterns, adding more difficulty to the event identification problem.

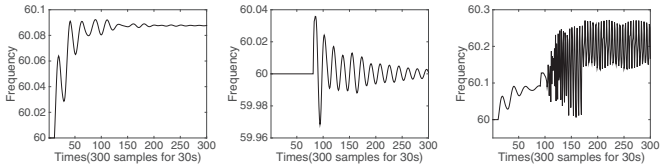
Quite opposite to the intra-class variance, different types of events may result in similar reaction on certain buses, which is illustrated in Fig. 3. This is referred to as *inter-class similarity*, presenting yet another level of difficulty for event identification. Fortunately, not all buses will appear similar for all types of events.



(a) GT (b) LT (c) LS

Fig. 3. Frequency signals display similar patterns responding to different single events.

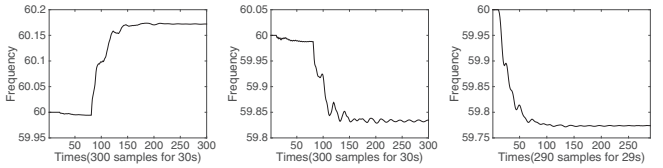
In order to illustrate the nonlinearity of large power systems, Fig. 4 displays two single event cases caused by LS and LT occurred at the 1st and the 8th second, respectively, as well as a concatenated event case resulted from the same LS and LT. Obviously, the signal of the concatenated event is not a strict linear combination of the signals of corresponding single events. In Fig. 4, all signals are averaged frequency signals of all buses. However, if we look into each individual bus, we still discover some linear characteristic, making it reasonable to use linear models to approximate the nonlinear system.



(a) LS (1s) (b) LT (8s) (c) LS (1s) + LT (8s)

Fig. 4. Illustration of the nonlinear characteristic: (a) single event case: an LS occurred at the 1st sec; (b) single event case: an LT occurred at the 8th sec; (c) concatenated event with the same LS occurred at the 1st sec and the LT at the 8th sec

In multi-event cases, a small disturbance accompanied with large ones is referred to as an unbalanced event, which usually results in missed detections since those small disturbances are easy to be considered as noise. Here, "small" and "large" denote the amplitude of frequency disturbance caused by an event. From the three cases shown in Fig. 5, small scale disturbance will be submerged by the large scale disturbance.



(a) GT (2s) + LS (8s) (b) GT (1s) + GT (8s) (c) GT (1s) + LT (7s)

Fig. 5. Three cases with unbalanced events: (a) concatenated events with a small GT occurred at the 2nd sec and a large LS at the 8th sec; (b) concatenated events with small GT at the 1st sec and large GT at the 8th sec; (c) concatenated events with a large GT at the 1st sec and a small LT at the 7th sec.

III. METHODOLOGY

Before we describe the proposed cluster-based sparse coding (CSC) algorithm for multi-event analysis, we first briefly introduce the sparse coding theory to set up the context.

A. Sparse coding

A signal caused by multi-events can be considered as a linear combination of signals from each individual events [6], [16]. Sparse coding, which models data vectors as sparse linear combinations of basis elements, is widely used to solve decomposition problems in many areas, such as machine learning, signal processing, and statistics [17]. For multiple events analysis, sparse coding aims to solve optimization problem in Eq. 1, where s is the multi-event signal. The matrix $D = [d_1, d_2, \cdots, d_k]$ is a an overcomplete dictionary, in which each column vector $d_i (1 < i < k)$ denotes a representative single event signal that is referred to as "root pattern". A multi-event signal s is supposed to be represented by a linear combination of a small set of root patterns, thus the coefficient $\mathbf{a}$ should be sparse. Usually, L_1 norm is employed as the penalty term to enforce a sparse $\mathbf{a}$ [18]. In addition, based on the physical meaning of this application, the coefficient vector $\mathbf{a}$ should be non-negative.

$$\begin{aligned} &\arg\min_{\mathbf{a}} \left\{ \|\mathbf{s} - D\mathbf{a}\|^2 + \lambda \|\mathbf{a}\|_1 \right\} \\ &\text{s.t. } \mathbf{a} \succeq 0 \end{aligned} \tag{1}$$

B. Problem formulation

Wang et al. [6] interpreted multiple event formation as a linear combination and used Eq. 1 to identify the constituent single events as well as their occurrence time. However, their approach suffers severe performance degradation when applied to large-scale power systems. A new approach, CSC, is thus proposed here to solve the challenging issues present in large-scale power systems that usually consist of hundreds or even thousands of buses. Assume there are n buses in a large power system, and m single event tests are implemented to obtain prior knowledge of disturbance. In each individual test, frequency signals from n buses are recorded in the format as Eq. 2, where $\mathbf{S}_i$ $(i = 1, 2, \cdots, m)$ is a matrix that denotes the signals of all buses in the ith test, and $\mathbf{s}_{ij}$ $(j = 1, 2, \cdots, n)$ is a column vector that denotes the signal on the jth bus in the ith test.

$$\mathbf{S}_i = \begin{bmatrix} \mathbf{s}_{i1} & \mathbf{s}_{i2} & \cdots & \mathbf{s}_{in} \end{bmatrix} \tag{2}$$

After m tests, all signals are collected and organized as Eq. 3, where $\mathbf{b}_j = \{\mathbf{s}_{ij}\}_{i=1}^{m} = \{s_{1j}, \cdots, s_{ij}, \cdots s_{mj}\}^T, (i = 1, \cdots, m)$. Obviously, $\mathbf{b}_j$ represents all the signals of m tests on the jth bus. $\mathbf{b}_j$ is still a column vector formed by concatenating all $\{\mathbf{s}_{ij}\}_{i=1}^{m}$. $\mathbf{B}$ represents the m tests on n buses, which is helpful to compare characteristic of reaction between buses.

$$\begin{aligned} \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_m \end{bmatrix} &= \begin{bmatrix} \mathbf{s}_{11} & \mathbf{s}_{12} & \cdots & \mathbf{s}_{1n} \\ \mathbf{s}_{21} & \mathbf{s}_{22} & \cdots & \mathbf{s}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_{m1} & \mathbf{s}_{m2} & \cdots & \mathbf{s}_{mn} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} = \mathbf{B} \end{aligned} \tag{3}$$

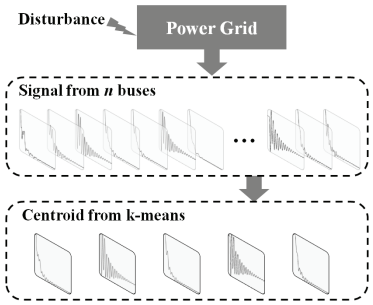


Fig. 6. Cluster-based signal extraction method for large power systems.

C. Cluster-based signal extraction

Here, we discuss which of the n buses we should use to form matrix $\mathbf{B}$. The simplest way is to keep signals from all buses. However, this method is not computationally efficient. In addition, due to the large intra-class variance and inter-class similarity, it seems that certain smoothing should be conducted to reduce intra-class variance as well as inter-class similarity. We hypothesize that buses within close geographical or electrical vicinity should react similarly to the same event. We have conducted comprehensive experiments that validate our hypothesis–a set of geographical or electrical clusters does exist and within each cluster, the buses show large inter-class variance and large intra-class similarity.

Therefore, we can confidently cluster the columns of matrix $\mathbf{B}$, which consists of single event signals of n buses from m tests. Assume k ($k \leq n$) clusters are obtained, and the indicator matrix $\mathbf{X}$ (Eq. 4) is generated according to the k-means result. All $\mathbf{x}_j$ $(j = 1, 2, \cdots, k)$ are orthonormal with each other. In other words, one and only one element is 1 in each row of $\mathbf{X}$, that indicates the cluster label of the corresponding bus.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \end{bmatrix}_{n \times k} \tag{4}$$

With the indicator matrix $\mathbf{X}$, each raw signal $\mathbf{S}_i$ can be simplified by averaging the signals within that cluster and use the averaged output to represent reaction of buses in that cluster. The cluster-based signal extracting method can be expressed in Eq. 5, where $\mathbf{e}_{1\times n}$ denotes a $1 \times n$ all-one vector. The function "diag()" constructs a square diagonal matrix with elements of the input vector on the main diagonal.

$$\mathbf{S}_i' = \mathbf{S}_i \mathbf{X} \times \text{diag}^{-1}\left(\mathbf{e}_{1\times n} \mathbf{X}\right) \tag{5}$$

Eq. 5 reduces or projects the number of signals from n to k, which is much smaller than the original number of buses. This largely reduces computational complexity and at the same time preserves intra-class characteristics yet enhances inter-class differences, leading to potentially higher recognition accuracy. The extraction procedure is visualized in Fig. 6, where $k = 5$.

D. Dictionary construction

Now, we can use $\mathbf{S}'_i$ obtained from Eq. 5 to build the dictionary. $\mathbf{S}'_i$ is an $l \times k$ matrix, where l is the length of a signal sequence. Note that l is the same as the length of the column vectors ($\mathbf{s}_{ij}$) in Eq. 2. First of all, each $\mathbf{S}'_i$ is normalized using Eq. 6.

$$s'_{ij} = \frac{s'_{ij} - s'_{1j}}{\sigma_j} \tag{6}$$

where s'_{ij} denotes the ith ($i = 1, 2, \cdots, l$) row and jth ($j = 1, 2, \cdots, k$) column of $\mathbf{S}'_i$, and σ_j is the standard deviation of the jth column. Normalization makes all k signals in $\mathbf{S}'_i$ start from zero, which will increase the accuracy of events detection and identification. Then, the k-means algorithm is employed again to cluster the $m \times k$ extracted signals. Specifically, the extracted signals are separated into three groups according to their event types, then each group is clustered. Finally, the cluster centers of all clusters are collected to build the dictionary. Note that an extracted signal $\mathbf{S}'_i$ is a matrix with k columns. Thus, a root pattern in the dictionary is a two-dimensional matrix, and the dictionary is a stack of several two-dimensional matrices. After clustering, main features of each type of event is preserved, and redundant signals are ignored. A compact dictionary could significantly speed up the event detection. Furthermore, cluster-based signal extraction and dictionary building can efficiently keep most characteristics of both the power system and events.

E. Cluster-based Sparse Coding

The cluster-based sparse coding is introduced aiming at overcoming the challenges mentioned in Sec. II. Each root pattern is a signal cluster here, whose scheme is illustrated in Fig. 7.

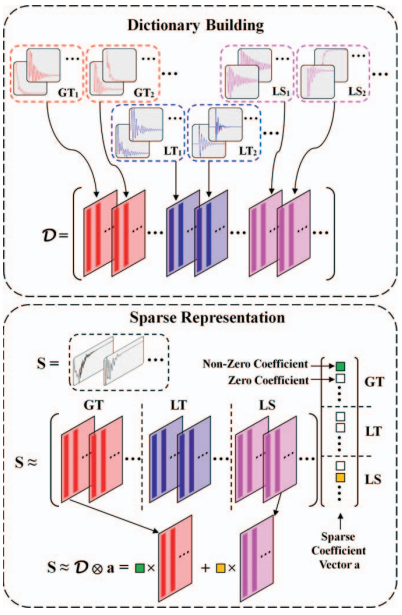


Fig. 7. Overview of the cluster-based sparse coding method. The top one: the diagram of dictionary building. The bottom one: the diagram of sparse representation.

Given $\mathbf{S}_{\text{in}}$ that denotes the raw signals of n buses from a new test. The exactly same method in Eq. 5 can be applied to extract signals. Eq. 7 rewrites the extraction method as

$$\mathbf{S}_{\text{out}} = \mathbf{S}_{\text{in}}\mathbf{X} \times \text{diag}^{-1}\left(\mathbf{e}_{1\times n}\mathbf{X}\right) \tag{7}$$

where $\mathbf{S}_{\text{out}}$ represents the extracted signals. Then, $\mathbf{S}_{\text{out}}$ is normalized using the same way as in Eq. 6, and the normalized signals are denoted as $\mathbf{S}$. Then, the sparse cording algorithm can be employed to fit $\mathbf{S}$ using root patterns in the dictionary. Since there are k individual signals in $\mathbf{S}$, they will be fitted simultaneously in the sparse coding procedure. Fitting a cluster at the same time can enhance the robustness and accuracy of event detection and recognition.

It is necessary to describe the organization of the dictionary. Assume the number of root patterns is r, and each root pattern is denoted as $\mathbf{D}_i \in \mathbb{R}^{l\times k}$ ($i = 1, 2, \cdots, r$). Since the dictionary is stacked by two-dimensional matrices $\mathbf{D}_i$, it can be represented as $\mathcal{D} = \{\mathbf{D}_i\}_{i=1}^{r} \in \mathbb{R}^{l\times k\times r}$. The signal needed to be detected is $\mathbf{S} \in \mathbb{R}^{l\times k}$. The final goal is to find the sparse coefficients $\mathbf{a} \in \mathbb{R}_+^r$ that can well reconstruct $\mathbf{S}$ as shown in Eq. 8.

$$\mathbf{S} \approx \hat{\mathbf{S}} = a_1\mathbf{D}_1 + a_2\mathbf{D}_2 + \cdots + a_r\mathbf{D}_r \tag{8}$$

where $\hat{\mathbf{S}}$ is the reconstructed signal matrix, and a_i ($i = 1, 2, \cdots, r$) is the ith element of the column vector $\mathbf{a}$. We need to minimize $\|\mathbf{S} - \hat{\mathbf{S}}\|_F$, where $\|.\|_F$ is the Frobenius norm. Denote $\sum_{i=1}^{r} a_i\mathbf{D}_i$ as $\mathcal{D}\otimes\mathbf{a}$, we can write the objective function of cluster-based sparse coding as in Eq. 9,

$$\begin{aligned} &\arg\min_{\mathbf{a}} \left\{\|\mathbf{S} - \mathcal{D}\otimes\mathbf{a}\|_F^2 + \lambda\|\mathbf{a}\|_1\right\} \\ &\text{s.t.} \quad \mathbf{a} \succeq 0\,, \quad \lambda > 0 \end{aligned} \tag{9}$$

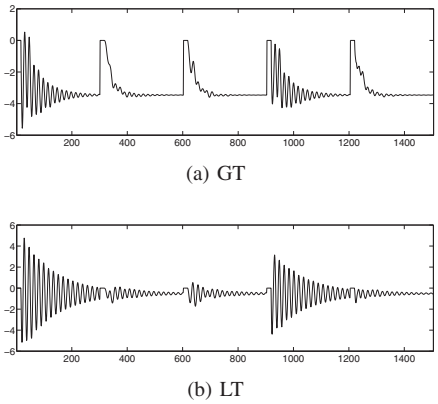
where $\lambda > 0$ is the sparse parameter to adjust the weight between the sparsity of $\mathbf{a}$ and the reconstruction error. Eq. 9 can be solved by many methods, such as BP [19], OMP [20], Lasso [18], etc. Based on the feature-sign search algorithm [21], we use a nonnegative feature-sign search method to ensure a nonnegative sparse coefficient vector $\mathbf{a}$. Vectorization method is used here to significantly simplify this tensor sparse coding procedure. The vectorization of $\mathcal{D}$ is to vectorize each root pattern matrix into a vector by concatenating the columns together. In the same token, matrix $\mathbf{S}$ is also vectorized into a vector.

IV. Experiment

In this section, a series of experiments based on simulated data will demonstrate the effectiveness of the proposed CSC algorithm for large-scale power systems. These simulations are done based on the "NPCC" testbed which is a reduced model of the real system, using Power System Simulator for Engineering (PSS/E) [22]. The reduced model still keeps the characteristic of the real system. The "NPCC" testbed is based on a 48-machine (140 buses) system of 28 GW of load. This model represents the NPCC region covering the whole or parts of ISO-NE, NYISO, PJM, MISO and IESO [23].

A. On intra-class variance and inter-class similarity

In the proposed algorithm, clustered signals are vectorized into a longer sequence. According to the existing training set, the number of clusters is set as 5 in k-means through



(a) GT

(b) LT

Fig. 8. Vectorized extracted signals: (a) single event of GT; (b) single event of LT

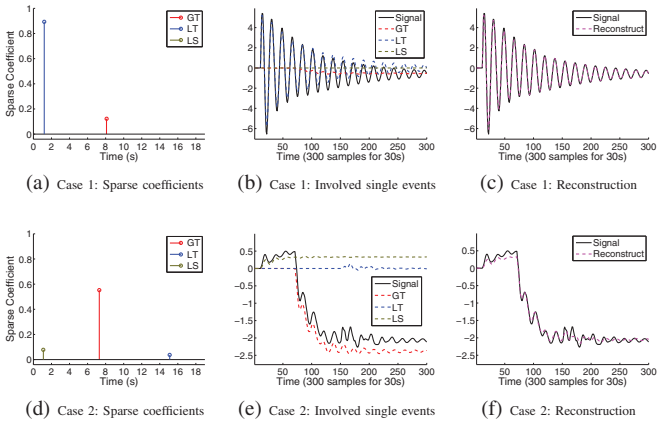
empirical study. Fig. 8 shows some extracted signal sequences. Each sequence consists of five sequences obtained from cluster centers, and they are concatenated together. From Fig. 8, it is easy to observe the intra-class variance since each of the five segments in a sequence significantly differ from each other. If we compare (a) with (b), their first segments appear similar, that illustrates inter-class similarity.

B. Detection of unbalanced events

When unbalanced events occur, the events with relatively small scale tend to be ignored because it is easy to be considered as noise. So far, there is no previous work can well detect unbalanced events when more than two events occur within a short period of time (e.g., 20 seconds). In this experiment, two unbalanced test cases are listed to demonstrate that the CSC can accurately detect small-scale events. For example, the GT in Fig. 9(c) and LT in Fig. 9(d) are small disturbances. The detection result of case 1 is a large LT happened at the 1.1959 sec and a small GT happened at the 8.0973 sec as shown in Fig. 9 (a) and (c). In case 2, a small LS occurred at the 1.1000 sec, then a larger GT struck at the 7.2936 sec, and finally a small LT came at the 15.1078 sec. Note that our event signal should contain 5 clusters, like the signals shown in Fig. 8. Here, for simplicity, only one cluster signal is shown in Fig. 9.

C. Comparison with state-of-the-art

This experiment compares the proposed CSC method with the NSEU method in [6] on a large-scale power system. Two multi-event cases are detected using these two methods as shown in Fig. 10 and Fig. 11. Case 3 is a triple-event signal—a GT happened at the 1st sec, LT at the 8th sec and a GT at the 14th sec. Fig. 10(a)-(c) shows the detection results from the NSEU method, which results in a false GT at the 18th sec. By contrast, the CSC method precisely identifies all involved single events shown as Fig. 10(d)-(f). Case 4 is another triple-event signal—a GT happened at the 1st sec, LT at the 8th sec and another LT at the 15th sec. Fig. 11(a)-(c) illustrates that the NSEU method misses the LT around the 8th sec. The CSC can accurately detect all events as shown in Fig. 11(d)-(f). Note that the CSC method uses all five clustered signals, and only the fifth cluster signal is plotted for simplicity. In addition, for a better comparison, only one cluster signal is plotted in these

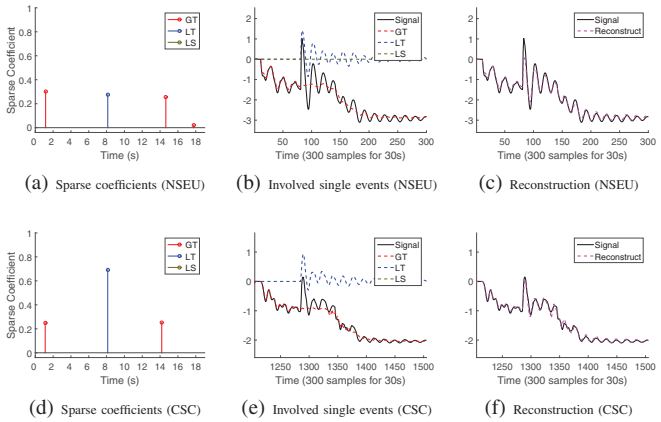


(a) Case 1: Sparse coefficients (b) Case 1: Involved single events (c) Case 1: Reconstruction

(d) Case 2: Sparse coefficients (e) Case 2: Involved single events (f) Case 2: Reconstruction

Fig. 9. Detection of unbalanced events: Case 1: ground truth is an LT occurred at the 1st sec and GT at the 8th sec; Case 2: ground truth is an LS happened at the 1st sec, a GT at the 7th sec, and an LT at the 15th sec.

two cases since the NSEU method is based on the average signal of all buses or signal from one bus.

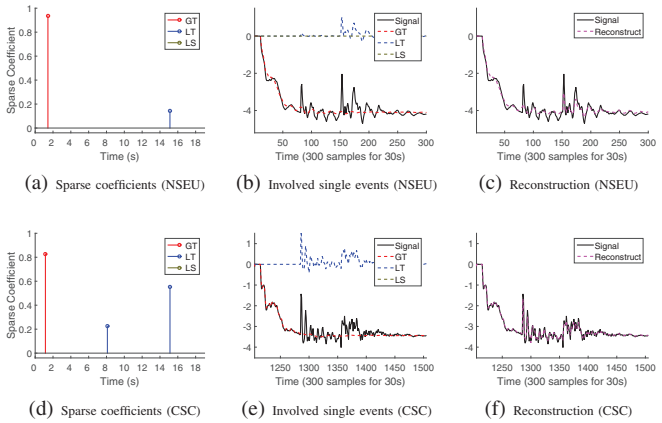


(a) Sparse coefficients (NSEU) (b) Involved single events (NSEU) (c) Reconstruction (NSEU)

(d) Sparse coefficients (CSC) (e) Involved single events (CSC) (f) Reconstruction (CSC)

Fig. 10. Case 3: triple events case. The NSEU method detection result: a GT at the 1.16 sec, an LT at the 8.131 sec, a GT at the 14.64 sec, and GT at the 17.76 sec. The CSC detection result: a GT happened at the 1.113 sec, an LT at the 8.127 sec, and a GT at the 14.18 sec. The ground truth is a GT at the 1st sec, an LT at the 8th sec, and a GT at the 14th sec.

We also compare these two methods on NPCC test dataset which contains single event cases (S1C), double event cases (M2C) and triple event cases (M3C). Roughly, over 100 testing samples are created for each type of case. Four metrics are defined according to [6] in order to evaluate and compare the performance of different methods. The four metrics are detection accuracy (DA), false alarm rate (FA), root-pattern recognition rate (PRR) and occurrence time delay (OTD). The experimental results are shown in Table I. In each test, the regularization parameter λ of CSC is adaptively selected from $[10, 100]$. Comparing with the NSEU method, the CSC method improves the detection and recognition accuracy and decreases the false alarm for all these three types of multi-event cases at the same time. In detail, for single event detection, the CSC method gets perfect 100% DA and 0% FA. For multiple events detection, DA is above 90% with relatively low FA. However, the NSEU method can only achieve 76% DA with 25% FA for M2C cases, and 58% DA with 17% FA for M3C.



(a) Sparse coefficients (NSEU) (b) Involved single events (NSEU) (c) Reconstruction (NSEU)

(d) Sparse coefficients (CSC) (e) Involved single events (CSC) (f) Reconstruction (CSC)

Fig. 11. Case 4: triple events case. The NSEU method detection result: a GT at the 1.43 sec and an LT at the 15.11 sec. The CSC detection result: a GT happened at the 1.147 sec, an LT at the 8.1 sec, and an LT at the 15.122 sec. The ground truth is a GT at the 1st sec, an LT at the 8th sec, and a GT at the 15th sec.

TABLE I. COMPARISON RESULTS WITH THE NSEU METHOD AND THE CSC METHOD

	Number of tests	DA (%)	FA (%)	RPR (%)	OTD (s)
Method	The NSEU method [6]				
S1C	144	93.3	6.67	100	0.251
M2C	115	75.34	25.13	91.34	1.245
M3C	138	58.33	17.34	86.7	0.942
Method	The CSC method				
S1C	144	100	0	100	0.123
M2C	115	95.65	2.17	98.64	0.193
M3C	138	91.55	0.97	98.15	0.202

V. CONCLUSIONS AND FUTURE WORKS

A novel cluster-based sparse coding algorithm was proposed to analyze multi-event problem in large power systems. The challenges caused by the complexity of large-scale systems are thoroughly analyzed. Then, the cluster-based sparse coding method was discussed in detail aiming at tackling those challenges. Finally, experiment results based on simulated data of the NPCC system revealed that the proposed scheme can generate highly accurate detection and recognition results with very low false alarm rates for large-scale power systems.

In the future, data from real power systems and other readings, such as voltage will be used to test the effectiveness of this work. In addition, we hope to be able to generalized the findings of region-based homogeneous reaction to solve other problems or other application domains.

REFERENCES

[1] S. Soltan, D. Mazauric, and G. Zussman, "Cascading failures in power grids: analysis and algorithms," in *Proceedings of the 5th International Conference on Future Energy Systems*. ACM, 2014, pp. 195–206.

[2] H. Qi, Y. Liu, F. Li, J. Luo, L. He, K. Tomsovic, L. Tolbert, and Q. Cao, "Increasing the resolution of wide-area situational awareness of the power grid through event unmixing," in *44th Hawaii International Conference on System Sciences (HICSS)*, 2011, pp. 1–8.

[3] A. Bykhovsky and J. H. Chow, "Power system disturbance identification from recorded dynamic data at the northfield substation," *International Journal of Electrical Power & Energy Systems*, vol. 25, no. 10, pp. 787–795, 2003.

[4] Y. Zhang, P. Markham, T. Xia, L. Chen, Y. Ye, Z. Wu, Z. Yuan, L. Wang, J. Bank, J. Burgett *et al.*, "Wide-area frequency monitoring network (FNET) architecture and applications," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 159–167, 2010.

[5] L. Liu, J. Chai, H. Qi, and Y. Liu, "Power grid disturbance analysis using frequency information at the distribution level," in *IEEE International Conference on Smart Grid Communications*, 2014, pp. 523–528.

[6] W. Wang, L. He, P. Markham, H. Qi, Y. Liu, Q. C. Cao, and L. M. Tolbert, "Multiple event detection and recognition through sparse unmixing for high-resolution situational awareness in power grid," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1654–1664, 2014.

[7] M. J. Smith and K. Wedeward, "Event detection and location in electric power systems using constrained optimization," in *IEEE Power and Energy Society General Meeting*, 2009, pp. 1–6.

[8] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215–2224, 2012.

[9] J. E. Tate and T. J. Overbye, "Double line outage detection using phasor angle measurements," in *IEEE Power and Energy Society General Meeting*, 2009, pp. 1–5.

[10] T. Bi, X. Song, J. Wu, and Q. Yang, "Novel method for disturbance identification in power systems," in *IEEE Power Engineering Society General Meeting*, 2006, pp. 5–pp.

[11] R. Agrawal and D. Thukaram, "Identification of fault location in power distribution system with distributed generation using support vector machines," in *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*. IEEE, 2013, pp. 1–6.

[12] T. Ji, Q. Wu, L. Jiang, and W. Tang, "Disturbance detection, location and classification in phase space," *IET Generation, Transmission & Distribution*, vol. 5, no. 2, pp. 257–265, 2011.

[13] N. Zhang and M. Kezunovic, "Improving real-time fault analysis and validating relay operations to prevent or mitigate cascading blackouts," in *IEEE Transmission and Distribution Conference and Exhibition*, 2006, pp. 847–852.

[14] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 342–351, 2011.

[15] R. M. Gardner, W. Li, J. West, J. Dong, Y. Liu, and G. Zhang, "Power system frequency oscillation characteristics," in *IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1–7.

[16] J. Dong, J. Zuo, L. Wang, K. S. Kook, I.-Y. Chung, Y. Liu, S. Affare, B. Rogers, and M. Ingram, "Analysis of power system disturbances based on wide-area frequency measurements," in *IEEE Power Engineering Society General Meeting*, 2007, pp. 1–8.

[17] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[19] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[20] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44.

[21] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," *Advances in Neural Information Processing Systems*, pp. 801–808, 2006.

[22] Introduction of PSS@E. [Online]. Available: http://w3.siemens.com/smartgrid/global/en/products-systems-solutions/software-solutions/planning-data-management-software/planning-simulation/pages/pss-e.aspx

[23] Introduction of large scale testbed at CURENT. [Online]. Available: http://curent.utk.edu/files/3713/6811/8508/Fact_sheet_tomsovic.pdf