Voltage stability assessment using multi-objective biogeography-based subset selection

Hanieh Mohammadi, Gholamreza Khademi, Maryam Dehghani, Dan Simon

ARTICLE INFO

Keyword:
Biogeography-based optimization
Support vector machine
Feature selection
Multi-objective optimization
Mutual information
Voltage stability assessment

ABSTRACT

We propose a method for voltage stability assessment of power systems using a support vector machine (SVM). We input the information from measurement signals to the SVM. The objectives of this paper are twofold: (1) to select the minimum number of features for training an SVM using multi-objective optimization (MOO); (2) to find the minimum misclassification rate for the SVM. To address these objectives, the training data size is decreased in two stages. First, a mutual information (MI) criterion is used to remove the less significant measurements from the data set. Second, the most relevant features are selected using multi-objective biogeography-based optimization (MOBBO). The minimization objectives are the measurement data size and the misclassification rate of the SVM. The BBO-based MO methods are vector evaluated BBO (VEBBO), nondominated sorting BBO (NSBBO), niched Pareto BBO (NPBBO), and strength Pareto BBO (SPBBO). These four MOO methods are compared using Pareto front normalized hypervolume, relative coverage, and statistical analysis. The methods are applied to a 39-bus benchmark system and a 66-bus real power grid. The results verify that the reduced measurement data using MOO lead to efficient and accurate SVMs. For the real power grid, the trained SVM accurately predicts the voltage stability of the system by using only 8% of the measurement data. The misclassification rates of the SVMs are as low as 2% for the real power grid. Normalized hypervolume, relative coverage, and statistical tests indicate that NSBBO performs better than the other methods.

1. Introduction

Power system voltage stability is the power system’s capability to maintain steady voltages after being subjected to various contingencies [1]. System conditions are associated with power system stability in voltage stability analysis [2]. A power system can face voltage problems and large scale blackouts if the voltage stability is assessed inaccurately and in an untimely manner [3]. Thus, it is essential to use monitoring devices like supervisory control and data acquisition (SCADA) and phasor measurement units (PMUs) in power systems for timely decision making. SCADA has been widely used in real-world power systems but it does not provide the time synchronization of the measured data and suffers from slow sampling time [3]. PMUs overcome the drawbacks of SCADA by providing accurate and fast data acquisition where the sequence of measurements of current and voltage waveforms are synchronized using the global positioning system (GPS). PMUs have been developed over the past few decades for various power system applications, such as transient stability and voltage stability [3–7]. In this paper, we can use either SCADA or PMU measurements since voltage stability assessment can be observed from either type of signal. We use PMU measurements in this research because of its suitability for our intended follow-on research, but the proposed method can be applied with either SCADA or PMU measurements.

Power system voltage stability has been assessed using several methods. Different voltage stability indices have been defined to identify the weak buses and lines such as P-V curve, energy function, maximum power that can be transferred through a line and so on [8]. There are some other methods which use the computational intelligence (CI) methods such as decision trees (DTs) [2,3], neural networks (NNs) [9–12], fuzzy neural networks [13], and support vector machines (SVMs) [14,15] to evaluate the voltage stability in power systems. In order to obtain a suitable knowledge of power systems, different operating conditions (OCs) are considered and in each operating condition possible contingencies are applied to the system. Several parameters such as pre-contingency power flow quantities and topological data or post-contingency data are collected to evaluate the stability status of the system [2,3,7,16]. It has been shown that categorical data provide rich information for classification and improve the

* Corresponding author.

E-mail addresses: h.mohammadi17@csuohio.edu (H. Mohammadi), g.khademi17@csuohio.edu (G. Khademi), mdehghani@shirazu.ac.ir (M. Dehghani), d.j.simon@csuohio.edu (D. Simon).

https://doi.org/10.1016/j.ijepes.2018.06.017
Received 27 July 2017; Received in revised form 28 April 2018; Accepted 5 June 2018
0142-0615/ © 2018 Elsevier Ltd. All rights reserved.
classification accuracy [3,4].

The large amount of collected data from power systems is of concern since training a CI method with all gathered information is not practical and it may lead to a time-consuming procedure with inadequate prediction accuracy. Since there is a highly nonlinear relationship among the attributes of power grids, reducing the number of features must be performed in a systematic way while considering their dependencies [17]. Therefore, different data reduction and feature selection approaches have been utilized [18]. For instance, feature selection methods such as mutual information (MI) and Fisher discrimination techniques and feature extraction methods like principal component analysis (PCA) are implemented in [9–11,17,19] to extract the significant data from the input features.

In our previous work [7], a method was proposed for online power system voltage security assessment based on PCA, correlation analysis, and DTs. We used fault location to improve prediction accuracy, while we reduced the number of training cases to achieve fast prediction. Various types of reduced information can be obtained from the original data depending on the feature selection and dimension reduction techniques and the accuracy that is required from the reduced data. A compromise exists between the simplicity and prediction accuracy of a trained CI method. Thus, multi-objective optimization (MOO) is viewed here as an important step for finding the optimal accuracy of reduced training data while considering both simplicity and prediction accuracy as objective functions. Therefore our research [7] continued to [16], where we minimized the number of principal components and the number of training cases while keeping the error rate as low as possible using several multi-objective methods. We suggested some possible directions to continue our research in [16], which leads us to this paper.

This paper proposes a new SVM-based scheme for power system voltage stability assessment. The original data is reduced to a much lower dimension using MI and MOO. The paper’s main contribution is to develop and implement multi-objective BBO for optimum subset selection of the measurement data. There are two objective functions that need to be minimized: the misclassification rate of the SVM, and the number of features input to the SVM. To address this, vector evaluated BBO (VEBBO), nondominated sorting BBO (NSBBO), niched Pareto BBO (NPBBO), and strength Pareto BBO (SPBBO) are utilized to obtain the best Pareto set.

Although other MOO methods could be implemented on our problem [20,21] we select the aforementioned methods as representative EA/ MOO combinations [22,23]. Note that the combination of BBO and the MOO methods (VE, NS, NP, and SP) have not been used for feature selection in voltage stability assessment before now although the individual components are well-known and widely used in different areas [22,24,25].

The rest of this paper is organized as follows. Section 2 discusses the basic theory of SVM, MI, and MOO. Section 3 explains the proposed voltage stability assessment method using MI and MOBBO for subset selection of SVM input data. SVM is used to classify the data into stable cases and unstable cases. MI is used as a feature selection method for removing irrelevant features from the measurement data. After elimination of irrelevant features, MOO is used as an optimum feature selection method to select the best subset from the relevant features. Section 4 discusses simulation results for both the 39-bus benchmark system and a 66-bus real power grid. The results verify that the MOBBO methods select the most significant features from the measurement data and lead to a remarkable reduction of data size. The best selected set of features with minimum classification error are systematically compared using relative coverage, normalized hypervolume, and statistical tests. These comparisons show that NSBBO performs better than other MOO methods for both the test case and the real-world case in this paper. Section 5 presents the conclusion and suggests future work.

2. Background

In this section, SVM equations are reviewed. This classifier will be used for voltage stability assessment in Section 3. Then, the MI method is presented. MI is used for pre-selection of the measurement data. Finally, MOBBO is briefly discussed. This MOO method will be used for subset selection of the data while keeping the classification error rate as low as possible.

2.1. Support vector machines

Support vector machines (SVMs) are effective and well-known learning algorithms used for binary and multi-class classification [26–29]. In our research, SVM is used for binary classification. In binary SVM, the two class labels are \( Y_j \in \{ -1, +1 \} \). The training data is defined as follows.

\[
X_{\text{m}\times\text{n}} = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1n} \\
  x_{21} & x_{22} & \cdots & x_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{m1} & x_{m2} & \cdots & x_{mn} 
\end{bmatrix}, \quad Y_{\text{m}\times1} = \begin{bmatrix}
  Y_1 \\
  Y_2 \\
  \vdots \\
  Y_m 
\end{bmatrix}
\]

(1)

where \( m \) is the number of training cases and \( n \) is the number of features.

As illustrated in Fig. 1(a), the objective of SVM is to find an optimal linear hyperplane with maximum margin and bounded error that divides the training data into two separate classes. The linear hyperplane is defined as

\[
\omega^T \Phi(X_i) + b = 0 \quad \omega \text{ and } X_i \in \mathbb{R}^{n\times1}, \quad b \in \mathbb{R}
\]

(2)

where \( \Phi() \) maps the input pattern into the feature space. \( \omega^T \) is the weight vector and \( b \) is the bias. The optimal hyperplane with maximum margin is expressed as the following convex quadratic optimization problem.

\[
\min_{\omega,b} \frac{1}{2} ||\omega||^2 + C \sum_{j=1}^{m} \xi_j \\
Y_j(\omega^T \Phi(X_i) + b) \geq 1 - \xi_j \quad \xi_j \geq 0, \quad j = 1, \ldots, m
\]

(3)

Fig. 1. (a) Simple linear SVM classifier, (b) SVM classifier with kernel.
where $\xi_i$ is the smallest value satisfying the above inequality and $C > 0$ is a weight constant that sets a trade-off between the number of misclassified input patterns and maximal separation of the remaining patterns.

Eq. (3) is the primal optimization problem. We have two simultaneous objectives in the cost function. The first term provides a hyperplane with the maximum margin, whereas the second term is a regularizing term which makes the SVM more robust to outliers and the second term ensures that the majority of the training patterns have sufficient margin.

To efficiently solve the optimization problem, especially in high-dimensional space, it is recommended to apply the Lagrange function. Therefore, Eq. (3) simplifies to the following optimization problem, which is referred to as the dual optimization problem.

$$
\max_a \sum_{j=1}^{m} \alpha_j - \sum_{j,k=1}^{m} \alpha_j \alpha_k Y_j Y_k \varphi(X_j, X_k)
$$

(4)

where $\alpha_j, \alpha_k \geq 0$, $0 \leq C$, $j = 1, \ldots, m$, and $\varphi(X_j, X_k)$ is the kernel function used to map input variables into the feature space as shown in Fig. 1(b). Various optimization techniques exist to solve the dual problem in Eq. (4) [30]. In this paper, we use sequential minimal optimization (SMO) to solve the quadratic programming problem of Eqs. 4 and 5 [30]. $\alpha^*$ is the optimum solution for the dual optimization problem in Eq. 4. Parameters of the optimal linear hyperplane in the feature space $\omega^*$ and $b^*$ are obtained by the following equation:

$$
\omega^* = \sum_{j=1}^{m} Y_j \alpha_j^* \varphi(X_j), \quad b^* = \frac{1}{2} \min_{|Y_j = 1|} \omega^T \varphi(X_j) + \max_{|Y_j = -1|} \omega^T \varphi(X_j)
$$

(6)

Note that $\alpha_j^* \neq 0$ only for the training patterns shown on the dotted margins in Fig. 1(a), which are called the support vectors. A given test input $x^* = (x_1, \ldots, x_n)$ is classified based on the kernel function:

$$
S = \text{sign}(\omega^T x + b^*) = \text{sign} \left( \sum_{j=1}^{m} Y_j \alpha_j^* \varphi(x_j, x) + b \right)
$$

(7)

where $x$ is assigned to the positive class if $S > 0$, and is otherwise assigned to the negative class. Various kernels such as linear, polynomial, RBF, and sigmoid can be used in SVMs.

### 2.2. Multi-objective biogeography-based optimization

BBO is a recently developed population-based evolutionary optimization algorithm [31]. BBO incorporates a mathematical model describing how species emigrate, immigrate, distribute themselves among islands [32]. The BBO algorithm is initialized with a randomly generated population. The probability of sharing features (independent variables) between islands (candidate solutions) has a direct relationship to the emigration rate and immigration rate of each island, which are based on their cost function values. After migration among all individuals, mutation is used to diversify the population. The individuals with the lowest cost function values are kept as elites for the next generation [31].

In a multi-objective optimization problem, the solution is a set of non-dominated points called a Pareto set [32, Chap. 20]. The combination of BBO with different MOO approaches results in several different MOBBO algorithms, such as VEBBO, NSBBO, NPBBO, and SPBBO. In VEBBO, the next generation is created by selecting one objective function at a time for recombination of the individuals. In NSBBO, the individuals are grouped into subpopulations depending on their non-domination rank. The best rank is assigned to the first non-dominated set, which are more probable to share information for recombination. In NPBBO, two individuals are randomly selected and compared to a randomly selected subset of the population. Each candidate that dominates the subset is used for recombination. In SPBBO, elitism is used explicitly by assigning strength values to each individual based on the number of individuals they dominate. All non-dominated individuals are kept in an archive and are more probable to be used for recombination. These four MOO methods are described more completely in [32].

### 2.3. Mutual information for feature selection

Mutual information (MI) provides a measure of mutual dependence between two random variables. In this research, we use MI to pre-select the most informative subset of features and give this reduced data to MOBBO for final feature selection. We suppose that $X$ is the input feature and $Y$ is the output class. MI estimates the amount of information provided by the input feature about the output class. MI between two random variables is calculated as [33, 34]

$$
I(X;Y) = I(Y;X) = H(Y) - H(Y|X)
$$

(8)

where $H(Y)$ is the entropy measuring the uncertainty associated with output class $Y$, and is defined as

$$
H(Y) = - \sum_{y \in Y} p(y) \log p(y)
$$

(9)

and $H(Y|X)$ is the uncertainty of output class $Y$ knowing the input feature $X$. MI in Eq. (8) can be reformulated as

$$
I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right)
$$

(10)

where $p(x,y)$ is the joint pdf, and $p(x)$ and $p(y)$ are the marginal pdfs. Eq. (10) can be estimated by histogram approximation of the joint and marginal pdfs. In this approach, the probability density function is approximated by the number of samples in a particular interval divided by the total number of data samples. For more details about MI estimation, refer to [33, 34]. MI is zero if and only if the feature and the output class variables are independent ($p(x,y) = p(x)p(y)$) and the feature does not provide any information about the output class. MI has a maximum value when the output class is a deterministic function of the input feature.

### 3. Voltage stability assessment using MI, MOBBO, and SVM

This paper proposes an efficient method for power system voltage stability assessment using SVM, MI, and MOBBO. MI is used to omit insignificant features from the PMU data. MOBBO selects the most informative features from the remaining features in such a way that the misclassification rate of the SVM, and the number of features input to the SVM are minimized. The algorithm includes the steps that are explained in Sections 3.1, 3.2, 3.3.

Fig. 2 shows the outline of the proposed method. First, we generate the data base as discussed in Section 3.1 by collecting data from PMUs and generating $X_{\text{max}}$ and $Y_{\text{max}}$. Second, we reduce the number of measurements in each fault case using MI as discussed in Section 3.2. That is, $X_{\text{max}}$ is reduced to $X_{\text{max}m}$, where $n_1 \leq n$. Third, the reduced data set $X_{\text{max}m}$ is input to MOBBO for optimum selection of a subset of features as discussed in Section 3.3. MOBBO outputs the reduced set of data $X_{\text{max}m, m}$, where $n_2 \leq n_1$. SVM is then trained with this reduced set of features and the misclassification rate is computed. The total number of features $n_1$ and the prediction error are the two minimization objectives. These objective values are iteratively input to MOBBO to find improved independent variable values for minimizing the cost functions. As Fig. 2 shows, first we collect the data from all buses and branches of the system based on the observability of the system. Our method needs all this data for the mutual information block and the MOBBO block, which are used to reduce the amount of data as much as...
possible. Finally, the SVM is trained with the selected variables. Thus, we need to pass all available information to the feature selection blocks to select the most important features for the SVM.

3.1. Data base

Several operating conditions (OCs) are considered for the power network based on load pattern changes. Certain contingencies such as faults on all branches or a generator outage are accounted of each OC. For system characterization in each OC, certain pre-contingency variables such as MVAR flows from bus $i$ to bus $j$ and vice versa ($Q_{ij}$, $Q_{ji}$), loading and current magnitudes of all branches ($I_{ij}$, $I_{ji}$), and squared voltage magnitudes ($V^2$) and voltage phase angles ($\delta$) of all buses, are calculated from PMU measurements. These fault-independent predictors can also be collected from SCADA measurements in voltage stability analysis, if the SCADA devices are placed in such a way that the power system is fully observable. Thus the proposed method for voltage stability analysis can be applied with either SCADA or PMU measurements. The reason that we use PMU data in this paper is to enable future directions in this research (mentioned in the Conclusion), in which we are going to assess other types of power system stability, such as transient stability in which the transient post-contingency data is collected and the sampling rate is important. This follow-on research will require PMU measurements.

In addition, some fault-dependent features, such as the location of the faults, are collected. Some PMU measurements might vary over time after a contingency due to the severity and location of fault (e.g., rapid decrease in the PMU voltage signals after an unstable case, which shows that the cases classified as unstable are effectively unstable). However, we are not interested in collecting the post-contingency data for training SVMs since we want to predict a probable instability before it occurs and we want to collect as much informative pre-contingency data as possible to have an accurate assessment of the power system voltage stability. Pre-contingency variables are chosen to convey the critical information about the power system and identify voltage stability problems accurately and intuitively. $\delta$, and $I_{ij}$, $I_{ji}$ are good measures to indicate the stress level in each OC; $Q_{ij}$, $Q_{ji}$ provide a better representation of the voltage profiles than active power flows; and $V^2$ reflects the voltage stability more accurately than voltage magnitude since it has a direct relation with reactive power [3]. Moreover, certain fault-dependent variables like the location of the fault (loc) are recorded, which are categorical predictors. We need to check an index to label each data point as stable or unstable. We choose the convergence of the load flow as the voltage stability index since it has been widely used for two-group classification and it has been verified for the assessment of voltage stability [1,3,7,35,36]. We could choose other indices, such as loading index or profile index, for voltage stability assessment, but the indices that we choose provide sufficient information when we want to classify the data into more than two groups of stability cases [7]. The convergence of the load flow implies that the power system is stable and the index 1 is assigned to this case; otherwise the power system is unstable and the index 0 is assigned. We confirmed that all the operating conditions (OCs) result in convergence of the load flow before a contingency is applied, and we collected pre-contingency data as numerical features for the SVM, so the contingencies are the only reason for divergence of the load flow.

The input matrix $X_{m,n}$ and the output vector $Y_{m,1}$ are generated by collecting all the aforementioned information from the system. $X_{m,n}$ includes all of the measurements corresponding to different fault locations, and $Y_{m,1}$ includes the stability index (0 or 1) of the system after each contingency. The components of $X$ and $Y$ are described as follows [16].

\[
X_{m,n} = \begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_m \\
\end{bmatrix}, 
Y_{m,1} = \begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_m \\
\end{bmatrix}
\]

(11)

\[
x_t = \{ \text{loc}_t, I_{ijk}(t), \ldots, I_{ij}, I_{ji}(t), \ldots, I_{ij}, I_{ji}(t), Q_{ij}(t), Q_{ji}(t), Q_{ij}, Q_{ji}(t), Q_{ij}, Q_{ji}(t), V_i^2(t), V_j^2(t), \delta_i(t), \delta_j(t), \}
\]

(12)

\[
y_i = \text{Stability index in case } t, \quad y_i \in [0, 1]
\]

(13)

where $I_{ijk}(t)$ is the current injected from bus $i$ to bus $j$ through branch $k$ in case $t$, $Q_{ijk}(t)$ is MVAR flow from bus $i$ to bus $j$ through branch $k$ in case $t$, $V_i^2(t)$ is squared voltage of bus $i$ in case $t$, $\delta_i(t)$ is the angle of bus $i$ in case $t$, the number of branches and the number of buses are defined as $n_b$ and $n_b$, respectively.

The total number of cases $m$ and the number of data points $n$ are obtained as

\[
m = N_c n_b
\]

(14)

\[
n = n_{f} n_b + n_{r} n_b + n_{u} n_b + 1
\]

(15)

where the number of OCs is defined as $N_c$; the number of measurements, including the branches information, is defined as $n_{f}$, and the number of measurements, including bus information, is defined as $n_{u}$. The addition of 1 in Eq. (15) accounts for the categorical variable.

For example, in a 39-bus test case power grid, there are 39 buses and 46 branches; see Fig. 3, where the candidate buses for placing the PMUs are circled in red. Optimal PMU placement is based on integer linear programming to achieve full system observability [37,38]. Based on [38], the full observability of the 39-bus system is achieved by placing nine PMUs on the buses shown in Fig. 3. By considering this topology for placing the PMUs, there is no restriction in obtaining the information needed to create the initial data set from 39 buses and 46 branches of the system. We considered 11 different OCs for this power system, based on the generated load patterns for five zones of the system from our previous work [7]. In each OC, all 46 branches are out-of-service to create $m = 11 \times 46 = 506$ cases. $I_{ijk}$ is the current injected from bus $i$ to bus $j$ through branch $k$. Since there are 46 branches in this system, the total number of available measurements for $I_{ijk}$ in each case is equal to $(k = 1, 2, \ldots, 46)$. The same holds for $I_{ijk}$, $Q_{ijk}$, and $Q_{ijk}$. The squared voltages and angles of all 39 buses in the system are collected from PMUs. Therefore, there are $n = 4 \times 46 + 2 \times 39 + 1 = 263$ available measurements for each case. It is clear that a huge number of features exist in large power systems. A time-consuming procedure with insufficient results may be the consequence of training SVMs with this...
amount of information. Therefore, reducing the data size while still maintaining fault prediction accuracy is an important step in constructing classifiers.

### 3.2. Mutual information for pre-selection of features

MI is widely used as a powerful performance metric for feature selection [39, 40]. Thus, in this paper, we use this metric for preliminary feature selection to remove irrelevant features. MI evaluates the quality of a subset of features with respect to the output class, and it is computationally efficient. The problem statement of the MI method for the pre-selection of features is to remove all the features that do not significantly contribute to the stability assessment of the power system. This is done by calculating the MI for each feature using Eq. (10). A threshold \( \theta \) is selected to remove all of the features having MI below the specified threshold. In other words, MI is used to eliminate the irrelevant features from the data and reduce the number of measurements in each fault case. This step is important since it allows more informative data to be given to the MOO method so that it can find better Pareto fronts, which are the sets of data with minimum number of features and the minimum classification error. There is a relationship between the number of removed features and the threshold \( \theta \). If we increase \( \theta \), the number of removed features will increase and we will be more likely to remove features that can contribute to the voltage stability assessment of the power system. If we set \( \theta \) to a smaller value, then the number of irrelevant features that are removed increases and it will make the process of selecting an effective subset from the features more difficult. In order to tune \( \theta \), one method is to gradually increase \( \theta \) from 0, and for each \( \theta \), separate the entire data set into two categories: the pre-selected and the pre-removed features. Then, we train the SVM once with the entire data set and next with the pre-selected features. We increase \( \theta \) until the results of the trained SVM with the entire data set and the results of the trained SVM with the selected features have a considerable difference. That is how we chose 0.2 in this study; to make sure that we keep the important features for the next step while still removing the irrelevant features from the initial set of data. After pre-selection of the data with MI, the input matrix \( X_{mn} \) decreases to \( X_{mn1} \), where \( n1 \leq n \). Algorithm 1 shows the outline of this step.

#### Algorithm 1: MI for pre-selection of the features

For each feature vector \( X_i, i = 1, \ldots, n \), where \( X_i \) is the \( i \)-th column of \( X_{mn} \) in Eq. 1

1. Calculate MI between the output and each feature using Eq. 10
2. If MI \( \leq \theta \) then
   - Remove the feature \( X_i \) from the data set \( X_{mn} \)
3. Next feature

### 3.3. Multi-objective optimization for optimal feature selection

Evolutionary and swarm intelligence methods have been used for feature selection, including particle swarm optimization (PSO) [41], binary firefly algorithm [42], bare bones PSO [43], and BBO [16]. However, multi-objective methods have not yet been used for power system stability analysis feature selection. In this section, the goal is to use MOBBO for finding the best feature sets from the reduced data \( (X_{mn1}) \) obtained in the previous section. MOBBO minimizes the

![Fig. 3. A 39-bus test case power grid. Brown dashes divide the system into five zones for generating different OCs. Red circled buses are the candidate buses for PMUs that make the system fully observable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
following two objectives,

1. The misclassification rate of training SVMs with the reduced data:
   \[ E = \frac{N_{mc}}{N_{ts}} \]  
   where \( N_{mc} \) is the number of misclassified test cases and \( N_{ts} \) is the total number of test cases.

2. The total number of features \( n_2 \).

There is a compromise between these two objectives. A lower misclassification rate usually requires a higher dimensional feature set, and a lower dimensional data results in an SVM with a lower prediction accuracy.

The steps of MOBBO for feature selection are as follows. First, we initialize the population with 50 individuals, that is, different sets of features \( (X_{mn,i}) \) that are selected randomly. \( n_i \) is the number of selected features in each individual. Next, we evaluate the two cost functions mentioned above and use VEBBO, NSBBO, NPBBBO, or SPBBO to recombine the individuals. Each method uses a different approach to rank the individuals, as discussed in Section 2.2. The individuals with better ranks are more likely to share their information with other individuals, based on the emigration rate and immigration rate assigned to each individual, and mutation is applied to diversify the population and create a new set of individuals. The best individuals in each generation are kept as elites and the process continues for a specific number of generations. The final solution is reported as the best set of features and is denoted as the Pareto set.

After solving the minimization problem, NSBBO provides a set of non-dominated solutions (feature subsets) known as the optimal Pareto points. Note that all Pareto points are equally preferable apart from subjective prioritization. MOBBO finds the optimal set of features, the data set \( X_{mn,i} \) is reduced to \( X_{mn,i} \) where \( n_2 \leq n_1 \).

4. Simulation results

In this section, the simulation results for 39-bus and 66-bus power grids are given. The simulation results for the 39-bus benchmark is presented in Section 4.1. We use four MOBBO methods to find the optimum data size for this power grid. The Pareto fronts found by each method are compared using several methods, such as hypervolume, relative coverage, and statistical tests. Then we show the results for training the SVM with the most significant features obtained from all four methods. Similar simulations are implemented on the 66-bus real test case in Iran in Section 4.2.

For the following simulations, linear SVMs are used since nonlinear kernels did not improve the prediction error noticeably. Ten-fold cross-validation is used for training and testing the SVMs. Thus, the error rate is reported as the mean error of ten-fold cross-validation ± the standard deviation of the estimation error.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>MOBBO parameters for 39-bus benchmark power system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>50</td>
</tr>
<tr>
<td>Number of generations</td>
<td>400</td>
</tr>
<tr>
<td>Number of elites</td>
<td>2</td>
</tr>
<tr>
<td>Mutation rate</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Fig. 4. Pareto fronts obtained for 39-bus power grid.
4.1. Voltage stability assessment for the 39-bus power system

The proposed method is implemented on a 39-bus test system. This system has 39 buses and 46 branches (see Fig. 3 and the accompanying discussion for more details). 506 different cases are generated for this system. In each case, a fault is considered on a branch and the stability status of the system is checked to see whether the case is secure or not. The initial data is \(x_{\text{initial}}\), 56% of the data set are secure cases while the other 44% are insecure cases. Due to uniform random distribution of cases into ten fold for cross validation, this proportion of secure and insecure cases is preserved in the training and testing data sets. Fig. 2 shows the flowchart of the proposed methodology.

After generating the data base, MI is used to remove insignificant features from the data. We select the threshold \(\theta = 0.2\) so that the features with MI less than 0.2 are removed. In this step, the 262 numerical features are reduced to 105 features.

Next, we implement four MOBBO methods to find a set of features that minimizes two important objectives. Table 1 shows the parameters used in the MOBBO algorithm. We obtained these parameters by running MOBBO several times to find the best results. The number of generations is large enough for MOBBO to converge; a larger number of generations only increased the computation time but did not noticeably improve the results.

Fig. 4 shows the Pareto fronts that were obtained using VEBBO, NSBBO, NPBBO, and SPBBO. The results show that the classification error rate is in the range 5–11%. The SVM trained with all 105 features has a mean 6.73% prediction error. Some of the Pareto fronts found for the 39-bus system have a lower mean error rate (5.96%) with a much lower number of features (as low as 11). This shows that an optimally selected subset of features is even more informative than all of the features for voltage stability assessment. That is because the SVM trained with all 105 features leads to overfitting, while the optimally selected features provides a bias-variance trade-off and improves the prediction accuracy [44].

To determine which MOO algorithm performs better than the other methods, we analyzed normalized hypervolume, relative coverage metrics, and statistical tests. In general, a smaller normalized hypervolume with more Pareto front points indicates better performance.

Also, a smaller value for relative coverage indicates better performance.

For the statistical test, if the p-value for each pair of MOO methods is less than a specified significance level, those MOO methods are significantly different from each other. Each metric is briefly introduced and is followed by the comparison results below.

Suppose that a Pareto front \(\tilde{F}_1 = f(x_1)\) has \(M\) points, where \(f(x_1)\) is a \(k\)-dimensional function. The following equation shows the normalized hypervolume metric.

\[
S(\tilde{F}_1) = \frac{\sum_{j=1}^{N} \prod_{i=1}^{M} f(x_j)}{M} \tag{17}
\]

where an MOO algorithm with better performance has a smaller value of \(S(\tilde{F}_1)\). Table 2 shows the normalized hypervolume comparisons using Eq. 17. The best Pareto front in this comparison is the one with the smallest normalized hypervolume. The results show that NSBBO is the best method for this MOO problem.

Another metric to compare the performance of two Pareto fronts is relative coverage. In this metric, the number of weakly dominated individuals in the first Pareto set (relative to at least one individual in the second Pareto set) is computed [32]. Denote the two Pareto fronts as \(\tilde{F}_1(1)\) and \(\tilde{F}_1(2)\). The following equation shows the relative coverage metric.

\[
C(\tilde{F}_1(1), \tilde{F}_1(2)) = \frac{|\{a_j \in \tilde{F}_1(2) : \exists a_i \in \tilde{F}_1(1); a_i \succeq a_j\}|}{|\tilde{F}_1(2)|} \tag{18}
\]

where the notation \(a_i \succeq a_j\) indicates that \(a_i\) weakly dominates \(a_j\). The relative coverage has a value between 0 and 1. If none of the individuals in \(\tilde{F}_1(2)\) are weakly dominated by individuals in \(\til\tilde{F}_1(1)\), then \(C(\tilde{F}_1(1), \tilde{F}_1(2)) = 0\), and if every individual in \(\tilde{F}_1(2)\) is weakly dominated by at least one individual in \(\tilde{F}_1(1)\), then \(C(\tilde{F}_1(1), \tilde{F}_1(2)) = 1\).

Table 3 shows the comparison between the MOO methods using Eq. (18). Each entry \(x/y\) in the table shows the relative coverage of method \(x\) versus method \(y\). For instance, the relative coverage of VEBBO to NSBBO is 4/6 in the table. This indicates that four points out of six total VEBBO Pareto front points are weakly dominated by at least one point in the NSBBO Pareto front. The total percentage of dominated points for SPBBO is 100% which indicates that SPBBO is weakly dominated by all other methods, and is therefore not effective for this problem. The total percentage of dominated points for NSBBO is only 8% which shows that other methods rarely dominate NSBBO points and is therefore the most suitable method for this problem.

Another technique to assess MOO methods is with a statistical analysis of the optimization results [45,46]. We use normalized hypervolume as the comparison metric to perform statistical tests. Each MOO algorithm is run for 10 independent trials to provide a sufficient number of samples for the statistical tests. We use the nonparametric Wilcoxon rank sum test to statistically compare the MOOs [47]. The test returns the p-value for a two-sided Wilcoxon rank sum test. The null hypothesis of the test is that the differences between two MOO algorithms come from a distribution with a zero mean at the specified level of significance. If the p-value is less than the level of significance, we reject the null hypothesis in favor of the hypothesis that the two algorithms are statistically significantly different. If the algorithms are not statistically different, we do not reject the null hypothesis because the p-value is greater than the level of significance. The Wilcoxon test is conducted here at the significance level of 0.05.

Table 4 shows the p-values obtained by Wilcoxon rank sum tests when comparing the four MOO methods for the 39-bus power system. According to Table 4, all of the four MOO methods are statistically different. The NSBBO normalized hypervolume mean over 10 runs is

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Comparison between each pair of MOBBO methods for the 39-bus power system using relative coverage.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VEBBO</td>
</tr>
<tr>
<td>VEBBO</td>
<td>–</td>
</tr>
<tr>
<td>NSBBO</td>
<td>4/6</td>
</tr>
<tr>
<td>NPBBO</td>
<td>0/6</td>
</tr>
<tr>
<td>SPBBO</td>
<td>0/6</td>
</tr>
<tr>
<td>Total Percentage of Dominated points</td>
<td>22%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>p-values obtained by Wilcoxon rank sum tests when comparing four MOO methods for the 39-bus power system. Asterisks indicate that the lower triangular half of the table is equal to its upper triangular half.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VEBBO</td>
</tr>
<tr>
<td>VEBBO</td>
<td>–</td>
</tr>
<tr>
<td>NSBBO</td>
<td>*</td>
</tr>
<tr>
<td>NPBBO</td>
<td>*</td>
</tr>
<tr>
<td>SPBBO</td>
<td>*</td>
</tr>
</tbody>
</table>
smaller than the other methods, and so we conclude that NSBBO is statistically significantly better than the other algorithms. SPBBO has a normalized hypervolume mean over 10 runs larger than the other methods so we conclude that SPBBO is statistically significantly worse than the other algorithms.

We conclude from the results of the normalized hypervolume, relative coverage, and statistical test that the best MOO method for this system is NSBBO and the worst MOO method is SPBBO.

4.1.1. Best compromise solution in the 39-bus power system

After Pareto points are obtained, we need to subjectively choose a single solution for voltage stability assessment of the 39-bus power grid. To take advantage of all four MOBBO methods, we create a set that consists of Pareto points from VEBBO, NSBBO, NPBBO, and SPBBO. We then find the non-dominated points in this set to obtain a combined Pareto front. Fig. 5 illustrates the combined Pareto front for 39-bus power system. Fig. 5 shows that NSBBO and VEBBO, with four and one points respectively, provide the maximum contribution to the combined Pareto front, while NPBBO and SPBBO provide no contribution.

One can pick any solution from the combined Pareto front, depending on the engineer’s preference of the problem objectives. However, our preference here is to select the best compromise solution (BCS) that creates a trade-off between the two conflicting objectives. To make selection of the BCS convenient, we use a pseudo-weight approach given as follows [48]:

$$\omega = \frac{f_i^{\max} - f_i}{f_i^{\max} - f_i^{\min}} \quad i = 1, 2$$

where $W$ is a pseudo-weight vector assigned to each solution, and $f_i^{\max}$ and $f_i^{\min}$ are the maximum and minimum of the $i$-th objective function $f_i$ among all the Pareto points in Fig. 5. The sum of the elements of weight vector $\omega_1$ and $\omega_2$ is 1, and they indicate the relative importance...
of the two objective functions in each solution. For instance, a solution with pseudo-weight \( W = (0, 1) \) indicates maximum importance of the misclassification error \( f_1 \) and minimum importance of the number of features \( f_2 \). Fig. 5 shows the pseudo-weight for all Pareto points. A solution is indicated as the BCS if the relative importance of objective functions are almost equal (i.e., \( |\omega_1 - \omega_2| \) is minimum) [49]. Therefore, the solution with pseudo-weight \( W = (0.5, 0.5) \) is selected as the BCS. The trained SVM using the BCS has a prediction error rate 6.74%.

4.1.2. Common features among Pareto fronts in 39-bus power system

In this section, we investigate which features are most frequently selected in the Pareto fronts found by the four MOBBO methods. There are 105 features input to MOBBO and the frequency of each feature being selected is shown in Fig. 6 for VEBBO, NSBBO, NPBBO, and SPBBO. For instance, in VEBBO (see Fig. 4(g)) six Pareto front points are found. Each Pareto front offers a set of features to be passed to SVM to predict the voltage stability status of the system. The frequency of selection is calculated for all 105 features to detect which features are more likely to be selected by VEBBO. The bar plot in Fig. 6(a) shows how many times each feature is selected. As the figure shows, the feature with index 17 is selected in 5/6 of the Pareto points, which indicates the significance of this feature in training the SVM.

After calculating the frequency of selection for all features in each MOO method, each feature that is selected in more than 50% of the Pareto points is labeled as one of the most significant features. These features are then input to the SVM to find the error rate of prediction. Table 5 shows the results of the SVM trained with this set of features. The error rate is at the same level as using all features. For instance, in NSBBO four features are commonly used in 50% of the Pareto points (three numerical features and one categorical feature). The three numerical features are the data measured from the PMUs placed at buses (three numerical features and one categorical feature). The number of features is the same as the error of one of the Pareto points with 11 features (see Fig. 4(b)). Compared to the BCS obtained for the 39-bus power system in previous section, this solution with fewer number of features provides smaller error rate. Thus, we can reduce our initial data \( X_{66 \times 263} \) to \( X_{66 \times 28} \) and still have a reasonable prediction error rate.

### 4.2. Voltage stability assessment for the 66-bus real power grid in Iran

This section considers a real-world power grid in Iran consisting of 66 buses and 73 branches. Load patterns are taken over 15 days, and 26 OCs are defined [7]. To obtain a fully observable system, 24 PMUs were placed in this power grid using the method proposed in [38]. The same voltage stability assessment scheme as described above for the 39-bus test system is applied in this section. Possible contingencies are considered for all branches of the system and \( m = 26 \times 73 = 1898 \) cases are generated. For the real 66-bus power system, 80% of the data are secure cases and the other 20% are insecure cases. Due to the uniform random distribution of cases into ten folds for cross-validation, this proportion of secure and insecure cases is preserved in the training and testing data sets. The 66-bus system is significantly more secure than the 39-bus system. This could be due to the fact that the defined OCs for 66-bus system are based on the real loading patterns of the system over a year once every 15 days plus peak-load and light-load information. In the 39-bus system we generate the loading patterns as high or low as we wanted and we were able to create more cases that have the potential to be insecure in the presence of a contingency such as branch outage. Thus, the proportion of stable and unstable cases are different in both systems. The input matrix \( X_{66 \times 424} \) is produced, with the number of measurements in each case equal to \( n = 4 \times 73 + 2 \times 66 + 1 = 425 \) (424 numerical features and 1 categorical feature).

After generating the data base, MI is used to remove insignificant features from the data. We select the threshold \( \beta = 0.2 \) so that features with MI less than 0.2 are removed. In this step, the 424 numerical features are reduced to 314 features.

Next, we implement four MOBBO methods to find a set of features that minimizes two important objectives. Table 6 shows the parameters used in the MOBBO algorithm. We obtained these parameters by running MOBBO several times to find the best results. The number of generations is large enough for MOBBO to converge; a larger number of generations only increased the computation time but did not noticeably improve the results.

Fig. 7 shows the Pareto fronts that were obtained using VEBBO, NSBBO, NPBBO, and SPBBO. The results show that the classification error rate is in the range 1–2%. The SVM trained with all 314 features has a mean prediction error of 2.53%. Some of the Pareto fronts found for the 66-bus system have a lower mean error rate (1.26%) with lower number of features (as low as 96). This shows that an optimally selected subset of features is even more informative than all of the features for voltage stability assessment. That is because the SVM trained with all 314 features will lead to overfitting, while the optimally selected features provide a bias-variance trade-off that improves prediction accuracy [44].

In order to determine which MOO algorithm performs better than the other methods, normalized hypervolume, relative coverage metrics, and statistical tests are examined. In general, a smaller normalized hypervolume and relative coverage indicates better performance. In statistical tests, if the p-value for each pair of MOO methods is less than an specified significance level, those MOO methods are significantly different from each other.

Table 7 shows the normalized hypervolume comparisons using Eq. (17). The MOO methods are ranked based on their normalized hypervolumes. The best Pareto front in this comparison is the one with smallest normalized hypervolume. The results show that NSBBO is the best method for this MOO problem. Table 8 compares the MOO methods using Eq. (18). For instance, the entry 3/4 in the second row and first column shows the relative coverage of VEBBO to NSBBO. This means that all VEBBO Pareto front points are weakly dominated by at least one point in the NSBBO Pareto front. It is clear from the table that SPBBO is weakly dominated by all of the other methods, and is therefore not suitable for this problem. The table also shows that NSBBO points are never dominated by other methods and is therefore the most suitable method for this problem. Another technique to assess MOO methods is with a statistical analysis of the optimization results. We use normalized hypervolume as the comparison metric to perform statistical tests. Each MOO algorithm is run for 10 independent trials to provide a sufficient number of samples for the statistical tests. If the p-value is less than the level of significance, we reject the null hypothesis.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>SVM results trained with 50% common features for the 39-bus power system.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Features</td>
<td>Prediction Error (%)</td>
</tr>
<tr>
<td>VERBO</td>
<td>5</td>
</tr>
<tr>
<td>NSBBO</td>
<td>4</td>
</tr>
<tr>
<td>NPBBO</td>
<td>17</td>
</tr>
<tr>
<td>SPBBO</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6</th>
<th>MOBBO parameters for 66-bus real power grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>100</td>
</tr>
<tr>
<td>Number of generations</td>
<td>2000</td>
</tr>
<tr>
<td>Number of elites</td>
<td>3</td>
</tr>
<tr>
<td>Mutation rate</td>
<td>0.04</td>
</tr>
</tbody>
</table>
in favor of the hypothesis that the two algorithms are statistically significantly different. The Wilcoxon test is conducted here at the significance level of 0.05.

Table 9 shows the p-value obtained by Wilcoxon rank sum tests when comparing the four MOO methods for the 39-bus power system. According to Table 9, all of the four MOO methods are statistically different. The NSBBO normalized hypervolume mean over 10 runs is smaller than the other methods, and so we conclude that NSBBO is statistically significantly better than the other algorithms. SPBBO has a normalized hypervolume mean over 10 runs larger than the other methods so we conclude that SPBBO is statistically significantly worse than the other algorithms.

We conclude from the results of the normalized hypervolume, relative coverage, and statistical test that the best MOO method for this system is NSBBO and the worst MOO method is SPBBO.

4.2.1. Best compromise solution in the 66-bus power system

In this section, we use the pseudo-weight approach as discussed in Section 4.1.1 to choose the BCS for voltage stability assessment of the 66-bus power grid. Similar to Section 4.1.1, we find the combined Pareto front obtained from the four MOBBO methods, as illustrated in Fig. 8.

The pseudo-weight is calculated for each solution in the combined Pareto front using Eq. (19), and is shown in Fig. 8. The solution with
pseudo-weight $\omega = (0.53, 0.47)$ provides the minimum value for $|\omega_1 - \omega_2|$. Therefore, we pick this solution as the BCS. The trained SVM using the BCS has a prediction error rate of 1.317%.

4.2.2. Common features among Pareto fronts in 66-bus power system

In this section, we investigate which features are most frequently selected in the Pareto fronts found by the four MOBBO methods. There are 314 features input to MOBBO and the selection frequency of each feature is calculated. The features that are selected in at least 75% of the Pareto fronts are considered the most significant features for that method. Moreover, the features that were selected in all Pareto front points (100%) are analyzed. We thus selected two levels for the common features (75% and 100%) to investigate how much we can reduce the number of features while still obtaining accurate predictions. These features are input to the SVM to find the prediction error rate. Table 10 shows the results of the SVMs trained with the features that were common for each method. The error rates are in the same level as when using all measurements. For instance, in NSBBO 33 features are used in 75% of the Pareto points. The trained SVM using this set of features has a prediction error rate 1.317 $\pm$ 0.560%. This error is the same as the BCS with 63 features (see Section 4.2.1). Thus, we can reduce our initial data $X_{39\times425}$ to $X_{39\times33}$ and still have low prediction error rate.

5. Conclusion

A large amount of PMU data is available to represent the state of a power system, and it is not practical to use all of the data for fault classification. An online method is proposed for power system voltage stability assessment exploiting MI, MOBBO, and SVM. The key contribution of this paper is optimum subset selection from the PMU data set in such a way that the reduced data represents the system state after any credible event. This goal is accomplished by utilizing various MOBBO methods to minimize two important objective functions: (1) the misclassification rate of the SVM; (2) the total number of features input to the SVM. The proposed method is implemented on two power grids: a 39-bus test system with 11 different OCs, and a 66-bus real-world power grid with 26 different OCs. SVMs are trained for stability assessment using the reduced-size database. The Pareto fronts obtained from MOBBO are systematically compared using hypervolume, relative coverage, and statistical tests. The results show that NSBBO performs best for both the 39-bus and the 66-bus power system. The data size for the 39-bus system is reduced from $X_{39\times262}$ to $X_{39\times64}$ with the reasonable mean prediction error of 5.957%. The data size for the 66-bus system is reduced from $X_{39\times425}$ to $X_{39\times33}$ with the low mean prediction error of 1.317%. In future work, other voltage stability indices can be investigated and its performance will be compared with the method we used in this paper. Transient stability can be assessed with a reduced number of features. Moreover, other feature selection and dimension reduction techniques can be studied and compared. It will also be important to study other classification methods for this problem, and transient stability problem such as DTs, random forests, neural networks, and so on. Other MOO methods could be compared to investigate which ones are more suitable for different power system stability analysis under various conditions.

The source code that was used to generate the results in this paper is available for download at [http://embeddedlab.csuohio.edu/power-optimization](http://embeddedlab.csuohio.edu/power-optimization). Due to the policy of regional electric company in Iran the online source code is available only for the standard IEEE 39-bus system.

Acknowledgment

This research was supported by US National Science Foundation grants 1344954 and 1536035. Many thanks to Fars Regional Electric Company, Shiraz, Iran for providing the data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [http://dx.doi.org/10.1016/j.ijepes.2018.06.017](http://dx.doi.org/10.1016/j.ijepes.2018.06.017).

References


