

A Novel Feature Selection Method for the Dynamic Security Assessment of Power Systems Based on Multi-Layer Perceptrons

Peyman BEYRANVAND¹, Cavit Fatih KUCUKTEZCAN²
Zehra CATALTEPE^{1,4}, Veysel Murat Istemihan GENÇ^{*,3}

Accepted : 22/01/2018 Published: 30/03/2018

Abstract: In this study, the effect of feature selection methods on the performance of multi-layer perceptrons used for the dynamic security assessment of electric power systems is investigated. The existence of many measurable parameters (features) characterizing the power system security status complicates the use of multi-layer perceptron both in terms of prediction accuracy and training time. In this paper, the dynamic security of a power system subject to a number of critical contingencies is assessed as the critical clearing time of any credible fault is predicted by a multi-layer perceptron. In addition to the study of two different feature selection methods, which are Minimum Redundancy Maximum Relevance (mRMR), and Regression Relief (RRelief), a novel multi-layer perceptron based feature selection method is proposed to be applied in the prediction of security indices. The performance of the feature selection methods on the dynamic security assessment is investigated on a 16-generator, 68-bus test system.

Keywords: Dynamic security assessment, feature selection, neural networks, transient stability

1. Introduction

Power systems are being operated under more and more stressed operating conditions with reduced security margins and unplanned power flow patterns within the network as the consumption of power, enforcement of economics and energy markets and the uncertainties in generation due to the utilization of renewable sources increase. In such cases, transient instabilities caused by credible contingencies could become an important factor in determining the security level of the system and in taking the necessary control actions when they are needed. Therefore, fast and accurate dynamic security assessment (DSA) methods involving the study of transient stability have always been important for a safe and reliable operation of power systems [1]. The proposed methods for DSA can be divided into a number of classes [2]. Using numerical integration methods [3] is an approach mostly followed for off-line applications, while it is considered as computationally intensive for on-line applications. Direct methods [4] utilize Lyapunov functions to assess the stability instead of using a numerical integration, but their efficiency depends on the simplicity of the models that are used. Probabilistic methods that determine the probability distributions for the stability of the system are suitable for system planning due computational requirements [5]. On the other hand, pattern recognition methods such as artificial neural networks (ANNs) assess the security of an operating point (OP) as they develop a mapping between the features representing the states of the system and its security through the use of a knowledge base (KB) generated off-line [6-14].

As a specific implementation of ANNs, multi-layer perceptrons (MLPs) can be used for the prediction of security indices during

the DSA of a large power system. However, in these applications, a large number of measurements (features) that characterize the power system and its security brings a complexity to the training process of MLPs. Using a large number of features has two potential adverse effects. The time complexity of training a neural network increases dramatically with the increasing number of input features [15]. In addition, the existence of irrelevant features increases the number of inputs without providing new information. Thus, exponentially increasing number of training samples are required to train the ANN effectively [17]. To mitigate these adverse effects on the training of ANN and to increase its prediction performance, which directly affects the success in DSA, an effective feature selection must be applied.

Various feature selection methods adopting the measures of feature quality such as sensitivity index [15], divergence [16] and Fisher discrimination [17] have been used for the DSA of power systems based on MLPs. In this study, the effect of using filter type feature selection methods, such as Minimum Redundancy Maximum Relevance (mRMR) [18] and Regression Relief (RRelief) [19], on the final performance of MLP for predicting the power system security indices is investigated. The security indices selected in this work are the critical clearing time (CCT) of the faults and the minimum damping ratio for the electromechanical oscillations occurring in the system.

In addition to the use of mRMR and RRelief, this paper proposes a new methodology for feature selection based on the MLP weights. The proposed MLP Weight Based Method (MLPW), uses training weights of MLP to draw conclusions on the correlations of input features and identify an optimal subset of features. Neural network based feature selection has been reported in [20], however, only the output of ANN is used for sensitivity analysis and feature selection, in contrast to MLPW that uses training information for feature selection. Performance of DSA using MLPW method is compared with the mRMR and RRelief methods on the dynamic security assessment for a 16-generator, 68-bus test system.

¹ Computer Engg., Istanbul Technical University, Istanbul - 34469, TURKEY

² Electrical & Electronics Engg., Bahcesehir University, Istanbul - 34469, TURKEY

³ Electrical Engg., Istanbul Technical University, Istanbul - 34469, TURKEY

⁴ tazi.io Machine Learning Solutions, Istanbul - 34469, TURKEY

* Corresponding Author Email: gencis@itu.edu.tr

2. Feature Selection for Neural Networks

Finding the optimal subset of features serves a central role in predicting the security index, fast and accurately, for a real power system whose dynamic performance can be characterized in various ways involving a large number of variables. Consider a set of N instances of operating points x_k and the corresponding CCT value y_k constituting the KB $\{x_k, y_k\}, k = 1, \dots, N$. Each operating point is an instance itself, consisting of D features $F_i (i = 1, \dots, D)$. The task of feature selection involves identifying an optimal subset of features $S = \{F_j: j \in \{1, \dots, m\}\}$ for the DSA. mRMR and RRelief feature selection methods, which have been applied in many different types of problems, have also been used in this study. In addition to these methods, an MLP based method has been introduced and used.

In all feature selection methods investigated in this study, m is a design parameter which represents the size of the optimal features subset S . In addition, none of these methods can compute the optimal m . For computing the optimal m , a parameter selection step is used and the details of this step are provided in Section 3.

2.1. mRMR

mRMR [18] is a filter type feature selection method. The candidate subset S provided by mRMR algorithm is independent of the prediction step and MLP is just used for predicting the security index of the system. The main criterion implemented in mRMR algorithm maximizes the relevance of features to the target output while minimizing the redundancy of selected features. By maximizing the relevance of features and the output, features that contain useful information about the target output are kept in S . In addition, minimizing the redundancy prevents the selection of features that depend on each other, hence does not add any new information about the target. This two-step algorithm provides a candidate optimal subset of features S [18].

2.2. RReliefF

RReliefF [19] is another filter type feature selection method. Similar to the previous approach, the MLP is used as a predictor in the mRMR method. However, here, the main criterion to select candidate feature subsets is different than the one in mRMR. The main idea is to rank features by their qualities in discriminating the values y_k corresponding to instances x_k that are close to each other, hence emphasizing on local information of instances. What makes RReliefF a viable option for feature selection is its ability to estimate the dependency between features, while most other methods, including mRMR, assume independence between features. In a domain that there is a strong dependency between features, RReliefF usually performs better. RReliefF generates a ranking of features based on their importance in predicting the target parameter [19].

2.3. MLP Weight Based Method

As filter type methods that do not require classifier or predictor training, mRMR and ReliefF are considered as fast feature selection methods. However, their performances depend on the predictor which, later, is used on the candidate optimal features set. Although there are other feature selection methods that use the predictor in feature selection steps, such as backward and forward feature selection, they are computationally intensive and unfit for online DSA. In MLP weight (MLPW) based method, the behavior of MLP predictor during the sample training is used.

MLP is a general function estimator which has the form $f(W_i, W_h, x) \in R$, where x denotes the vector of input variables,

W_i and W_h are the matrices of weights of the input layer and the second layer of the MLP, respectively. The process of training comprises of repeated estimation and fine-tuning of network weights W_i and W_h to minimize the discrepancy between the actual target output y and the prediction of MLP, $f(W_i, W_h, x)$.

In order to determine the correlated features, the behavior of input layer weights W_i are observed during training. It is expected that the correlation between input features are embodied in the correlation of weights of the network during the training phase. If the weights connecting two features behave similarly during training, then it is assumed that these features should have higher correlation.

The general framework of the proposed methodology consists of training a sample MLP with all features once, recording the training weight, performing the correlation analysis and feature ranking. These training weights are recorded in column vectors. For each feature $x_i, i = 1, \dots, D$, and for each epoch $t_j, j = 1, \dots, T$ of batch training, MLP has h , the number of hidden units, weights corresponding to the weights of the connections between feature i and each of the h hidden units. Then, these weights of different training epochs are concatenated, so that it ends up with D columns of length $T \times h$. In other words, the result is a D -by- $(T \times h)$ matrix of training weights, $NetW \in R^{D \times (T \times h)}$. The next step is to calculate the correlations between different columns of the $NetW$ matrix. The absolute value of the Pearson correlation is used to form a matrix of pairwise correlations between network weights of different features. The general steps of the proposed feature selection method are represented in the algorithm below:

Algorithm (MLPW Feature Selection):

Input= NetW // matrix of network weights

Output=featureRank (Ranking of features by their importance)

1. counter=1
2. corr=correlation(NetW)
3. while eliminatingFeatures
4. imVec=rowSum_of_corr
5. referenceFeature=max(imVec)
6. IdxCF=max(corr(referenceFeature,i)/ imVec (i))
7. featureRank(1, counter)= IdxCF
8. counter=counter+1
9. Remove IdxCF entry from R
- 10.end

The proposed algorithm gets the matrix $NetW$ of training weights, then produces a list of features, $featureRank$, which records the features in order, so that irrelevant features can be eliminated to reach an optimal subset of the feature set. In this sense, the proposed algorithm gives a feature ranking similar to the ReliefF algorithm. Line 5 of the MLPW algorithm is an important step where the feature that minimizes the following is obtained:

$$\max_i \frac{R(\text{referenceFeature}, i)}{\text{Sum}(R(i, :))} \quad (1)$$

Thus, the feature that is highly correlated with the $referenceFeature$, from the point of view of MLP, and at the same time, that has the lowest correlation with all the features, including the $referenceFeature$, is found.

3. Methodology

3.1. Neural Networks for DSA

DSA can be formulated as a classification or a regression problem depending on the security information required for the application. In the classification task, classifiers can classify the security status of the system at a particular operating point (OP) as secure or insecure, whereas in the regression task, the regression tools can predict a security index that measures how far the system is away

from becoming unstable if the contingencies do occur. In this study, the MLPs developed for DSA are trained through a supervised learning algorithm which requires a KB composed of various examples of OPs and the security status of the system operating at them.

The method starts with the task of a contingency scan over a wide range of OPs as the critical contingencies are distinguished. The critical contingencies are selected from the ones leading angular instabilities after the occurrence of three-phase faults.

A crucial step in the method is to generate the training data of representative instances that are properly chosen. For a number of different topologies and loading levels, a large set of OPs is created using the power flow solutions. Then, at each OP, the security status of the system is determined by means of the post-contingency dynamics through time-domain simulations. When the problem is formulated as a classification problem, the security of the system at a particular OP can be represented as secure or insecure by a straightforward calculation of the angle stability index,

$$\eta = \frac{360 - \delta_{max}}{360 + \delta_{max}} \times 100 \quad , \quad -100 < \eta < 100 \quad (2)$$

where δ_{max} is the maximum angle separation in degrees of any two generators at the same time in the post-contingency response and OP is labeled as secure if the index is positive. In the regression problem, the security is quantified by a security index, the CCT, which is the maximum time allowed for clearing the fault without causing any angular instability. In addition to the assessment for transient stability using the aforementioned indices, the dynamic security assessment of the system is extended with the assessment for small-signal angle stability. For all operating conditions, the system must also be secure in the sense that both inter-area and intra-area oscillations are well damped in a power system when it is subjected to disturbances. This can be assessed by computing the eigenvalues associated with the electromechanical modes of oscillations and distinguishing the ones with the minimum damping ratio.

In this study, a multilayer perceptron (MLP) is trained and developed to determine the security status of the system operating at a given OP for each critical contingency. The KB contains a large number of OPs from a wide range of operating conditions, each of which is defined by a particular schedule of generation and distribution of the load demand, a particular network topology and a loading level. Having such a large KB, a feature selection process becomes crucial to attain better performances of the MLPs predicting the security as well as to reduce the size and the complexity of the MLPs without losing too much information. In the next section, the framework for feature selection and the selection of the best MLP structure are given in more detail.

3.2. Feature Selection

The feature selection step consists of determining two parameters: the size of the optimal feature subset, m , and the size of the hidden layer of the MLP, h . First, the size of the optimal subset of features m is calculated. For this purpose, a set of initial guesses for m is required, then, for each of them, the best subset of features are calculated by using mRMR and RReliefF. This optimal subset of features is fed to the MLP and the accuracy of prediction is evaluated for different number of hidden units, h . Finally, the best combination of m and h is chosen based on the prediction error. The steps of the proposed feature selection and training approach are shown in Fig. 1.

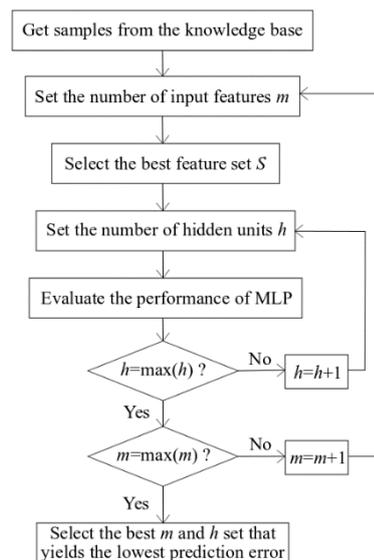


Fig. 1. Feature selection steps.

Increasing the number of the units in the hidden layer of MLP always increases the training accuracy. However, the training performance represents only the ability of the MLP in predicting instances that exist in the training set while the actual performance of the MLP is tested on instances that have not been seen yet. Therefore, increasing the value of h will have a negative effect on the prediction performance on the unknown instances as it decreases the generalization performance. To overcome this effect, a portion of data is set aside for evaluation only and named as validation set. The value of h is determined based on the performance of the MLP on the validation set. By increasing h , a decrease in the prediction error is expected at first, however, then the prediction error will eventually increase. Additionally, to eliminate any adverse effect of choosing a fixed validation set on the final performance of prediction, 10 fold cross-validation is applied. In this method, the KB is divided into 10 validation and training sets, each validation set consisting of 10% of the knowledge base, and the remaining 90% is designated as the corresponding training set.

By combining the feature selection step and model selection step, it is possible to identify the best subset of features that in combination with the MLP yields the best prediction performance. This is important since the choice of MLP as the predictor will affect the size of the optimal feature subset, m .

4. Results

4.1. Test System

The proposed methodology is applied on a 16-generator, 68-bus test system [21] shown in Fig. 2. Each generating unit is modeled with the 6th order two-axis synchronous generator model, a static exciter model of order 1, a power system stabilizer model and a speed governor turbine unit model of order 3. Through a contingency scan over a wide range of operating condition, 12 critical contingencies (Fig. 2), which lead to instabilities, are found. Each critical contingency is a three-phase fault at one end of a transmission line cleared by the removal of the faulted line well after its CCT.

For each OP, the power flow computations and DSA via time-domain simulations and eigenvalue analysis are performed. The KB including instances (OPs) at various loading levels (75% - 125%) with 45 different topologies is generated by DSATools™

[22]. Instances are represented by 136 features, which are the magnitudes and phase angles of the bus voltages, and the stability index values (the CCTs related to the critical contingencies and the minimum damping ratios at the pre-fault conditions).

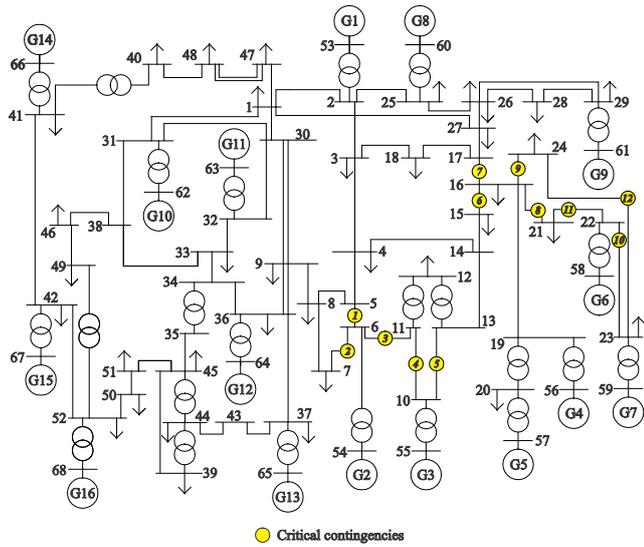


Fig. 2. 16-generator-68-bus power system.

4.2. Prediction of CCT

In this study, a KB of 3692 instances of operating points and their corresponding CCT values is generated. 20% of the KB has been randomly selected and allocated to evaluate the final performance of the MLP-based CCT predictor and is not used in any step of the feature selection or the MLP model selection. The remaining 80% is used for validation and training sets. Thus, it is guaranteed that the reported prediction error is as close to a test experiment as possible. For the training of MLP, scaled conjugate gradient back propagation method is used. In addition, to prevent the MLP from over-fitting, early stopping is applied. The proposed procedure for all feature selection methods is independent of the initial selection of critical contingencies. Therefore, it is applicable to other contingencies.

The measure for evaluating the performance of MLP is chosen to be mean squared error (MSE). MSE is a widely used performance criterion for prediction problem in MLP. For a set of instances and target values in form of $\{x_k, y_k\}$ with $(k = 1, \dots, K)$ and a set of corresponding K prediction \hat{y}_k , the MSE is defined as

$$MSE = \frac{\sum_{k=1}^K (\hat{y}_k - y_k)^2}{K} \quad (3)$$

where y_k and \hat{y}_k are the actual and the predicted values of CCT, respectively.

The evaluation step has been repeated 10 times to ensure an accurate representation of the actual prediction error and the average MSE is reported. At each repetition of the evaluation, the MLP is initialized with a different set of weight so that the small change in the final performance of the MLP is considered.

In order to choose the best number of hidden units, for a fixed number of input features m , after feature selection step, MLP is trained and evaluated by using cross-validation, for different number of hidden units. For each size of the hidden unit, cross-validation provides us with two performance measures, one for training and one for validation. The validation error is used to choose h . Thus, it is ensured that the MLP does not over-fit the training set. After choosing h , the MLP with h hidden units and for different values of m is trained and its performance on the test data

is evaluated. The number of hidden units is changed from 5 to 60, while the number of input features is changed from 1 to 136.

Fig. 3 shows the performance analysis graph of the proposed framework on the contingency 2. For this contingency, it is observed that the proposed feature selection method, MLPW, results in a small MSE using less features than the other methods. The performance of the CCT prediction using MLPW has a lower average MSE. In addition to this, the standard deviation (SD) of the error should also be considered. A sizable SD is an indication of low quality feature subset with unstable regression performance. In other words, it signifies that the regression performance could vary significantly using the same set of features. It should be noted that, for different contingencies, different feature selection methods may yield the best performance results. For instance, in Fig. 4, MLPW outperforms mRMR and RRelief and the best result is achieved using 40 features selected by MLPW. As seen in Fig. 3 and Fig. 4, using a proper feature selection method will improve the performance prediction.

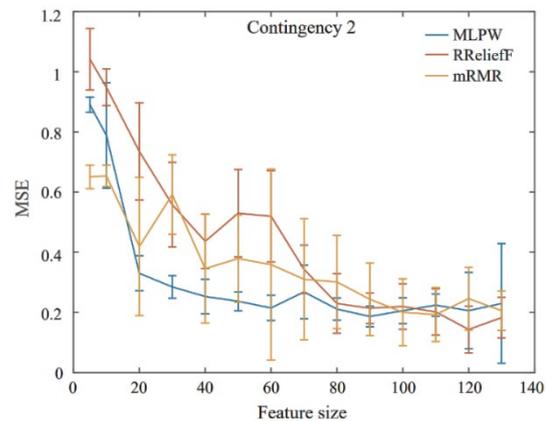


Fig. 3. Test MSE for CCT prediction, contingency 2.

In addition to CCTs, the proposed methodology is applied to predict the minimum damping ratio of electromechanical oscillations. Fig. 5 shows the results of the analysis. In this case, Rrelieff surpasses all the other methods in terms of the feature selection performance. However, the results of the different parts of the graph may be different. For example, if for any reason only the top 20 features are used, mRMR provides a better performing subset of features, with both smaller error and SD of error. However, for feature subsets of size larger than 20, MLPW and Rrelieff consistently outperform mRMR. To validate the observations, the dependency of features and outputs to each other is also investigated. For this purpose, mutual information measure is used.

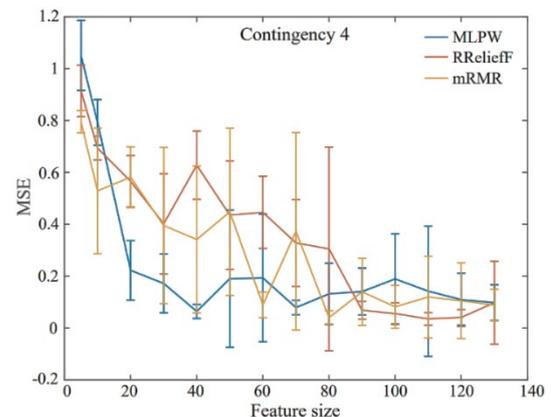


Fig. 4. Test MSE for CCT prediction, contingency 4.

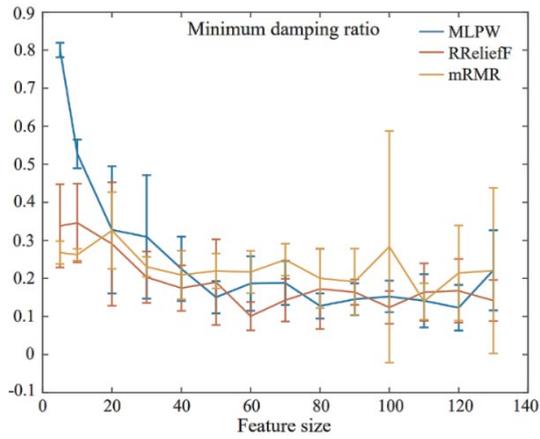


Fig. 5. Test MSE for minimum damping ratio prediction.

By definition, mutual information meters the amount of information obtained from one variable, using another variable. Fig. 6 shows the plot of mutual information score for different features, measured against contingency 2. The observation is that most of the features have the same amount of information about the CCT value of contingency 2. Therefore, it is validated that all features can be equally informative about the output variable.

5. Conclusion

In this study, the effect of feature selection algorithms in the DSA based on MLPs is investigated. The results show that a comparable prediction performance is achievable using only a subset of power system parameters which have an important effect on simplifying the process of security assessment.

The other side of the argument is the dependency of features to each other; how much information a feature can provide about a different feature. In this case, high mutual information score indicates that one of the two features can be excluded without much loss of overall information provided to the predictor. Fig. 7 provides the pairwise mutual information of the features. The results show that some of the features are relatively informative of each other. This observation is in line with the error rate graphs of Fig. 3. For contingency 2, at around 60 features, the rate of decrease of the prediction error drastically reduces. In other words, the added useful information for prediction selecting more than 60 features is miniscule.

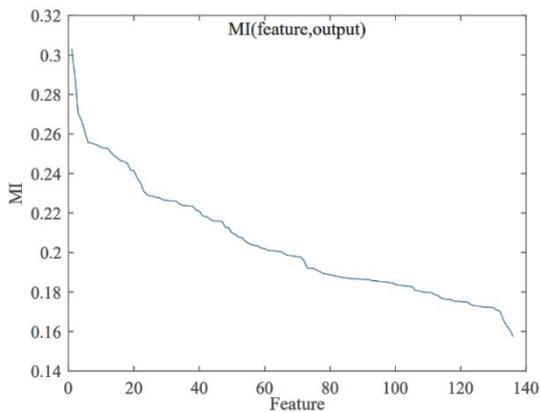


Fig. 6. Mutual information of features and output for contingency 2.

Due to space limitations, only the top 5 most important features obtained in each algorithm for the contingencies, ctg. 2, ctg. 4, ctg.8, and ctg. 10, are provided in Table 1, where the voltage magnitude and its phase angle at bus i of the test system are

denoted by V_i and θ_i , respectively. As it is expected, the closer features to the location of each contingency are found as important. Incorporating smaller set of network parameters reduces the time complexity and model complexity of MLP based DSA and prove to be critical in large power systems in which fast security assessment and control are critical to deliver high quality service.

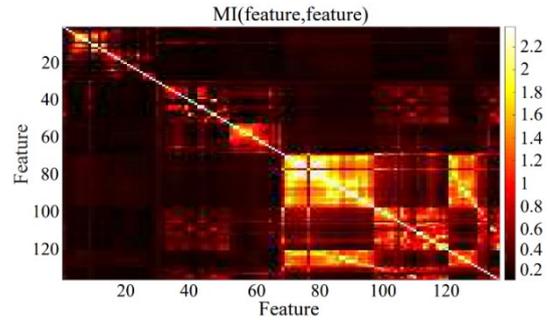


Fig. 7. Pairwise mutual information of features.

Table 1. Top 5 most important features determined by the feature selection methods for each contingency

Method	Top 5 most important features for each contingency				
	Ctg. 2	Ctg. 4	Ctg. 6	Ctg. 8	Ctg. 10
MLPW	V_24	V_12	V_35	θ_8	V_29
	θ_{11}	V_28	V_65	θ_{19}	θ_{10}
	V_50	V_60	V_56	θ_{20}	θ_{39}
	θ_{40}	θ_{54}	θ_{33}	V_45	θ_{66}
	θ_{41}	V_25	θ_{16}	θ_{68}	θ_{57}
RReliefF	V_42	V_42	V_42	V_42	V_42
	V_8	V_4	V_23	V_23	V_23
	V_41	V_10	V_21	V_21	V_21
	V_7	V_12	V_24	V_24	V_24
	V_12	θ_{65}	V_41	V_12	V_22
mRMR	V_54	θ_{55}	θ_{58}	θ_{58}	V_58
	θ_{55}	V_10	V_30	V_30	V_42
	V_55	V_29	V_53	V_53	θ_{59}
	V_6	V_55	θ_{59}	θ_{59}	V_27
	V_53	V_53	V_42	V_58	θ_{58}

In addition to complexity reduction, another effect of feature selection is to measure less number of variables in order to predict the security status of the system. This can be a critical factor in practical situations when measuring or estimating some of the features is costly or time consuming. Therefore, by means of a proper feature selection method and identifying the important features, the effort for prediction and online DSA of the system can be optimized.

Acknowledgements

This research was supported by The Scientific and Technical Research Council of Turkey (TUBITAK) project no. 114E157.

References

- [1] Y. V. Makarov, P. Du, S. Lu, T. B. Nguyen, X. Guo, J. W. Burns, J. F. Gronquist and M. A. Pai, "PMU-Based wide-area security assessment: concept, method, and implementation," *IEEE T. Smart Grid*, vol. 3, no. 3, pp. 1325-1332, 2012.
- [2] P. Sauer, K. L. Tomsovic and V. Vittal, "Dynamic Security Assessment," in *The Electric Power Engineering Handbook*, 2nd ed., Vol. 5, Chapter 15, London: Taylor and Francis, 2007.
- [3] H. W. Dommel and N. Sato, "Fast Transient Stability Solutions," *IEEE T. Power Ap. Syst.*, vol. 91, no. 4, pp. 1643-1650, 1972.

- [4] A. A. Fouad and V. Vittal, *Power System Transient Stability Analysis Using the Transient Energy Function Method*. NJ: Prentice Hall, 1992.
- [5] M. P. Anderson and A. Bose, "A Probabilistic Approach to Power System Stability Analysis Power Engineering Review," *IEEE T. Power Syst.*, vol. 3, no. 8, pp. 25-26, 1983.
- [6] C. F. Kucuktezcan and V. M. I. Genc, "Dynamic security assessment of a power system based on Probabilistic Neural Networks," in *Proc. IEEE PES Innovative Smart Grid Technologies Conference Europe*, Gotenburg, Sweden, 2010, pp. 1-6.
- [7] E. M. Voumvoulakis, A. E. Gavoyiannis and N. D. Hatziaargyriou, "Application of Machine Learning on Power System Dynamic Security Assessment," in *Proc. IEEE International Conference on Intelligent Systems Applications to Power Systems*, Kaohsiung, Taiwan, 2007, pp. 1-6.
- [8] A. Hoballah and I. Erlich, "Real-time operation of deregulated electricity market: An integrated approach to dynamic stability assurance," in *Proc. IEEE International Energy Conference and Exhibition*, Manama, Bahrain, 2010, pp. 262-267.
- [9] J. Shang, J. Zhang, W. Zhao and J. Liu, "ANN based dynamic voltage security assessment for a practical power system," in *Proc. IEEE International Power Engineering Conference*, Singapore, 2007, pp. 794-798.
- [10] A. Al-Masri, M. Z. A. A. Kadir, H. Hizam, J. Jasni and N. Mariun, "Transient security assessment for power system stability: A review on artificial intelligence approach," in *Proc. IEEE Student Conference on Research and Development*, Malaysia, 2009, pp. 371-375.
- [11] K. S. Swarup, and K. V. Prasad Reddy, "Neural network based pattern recognition for power system security assessment," in *Proc. IEEE International Conference on Intelligent Sensing and Information Processing*, Piscataway, New Jersey, USA, 2005, pp. 234-239.
- [12] N. Srilatha, G. Yesuratnam and M. S. Deepthi, "RBF Neural Network approach for security assessment and enhancement," in *Proc. IEEE PES T&D Conference and Exposition*, Chicago, Illinois, USA, 2014, pp. 1-5.
- [13] C. F. Kucuktezcan and V. M. I. Genc, "A comparison between ANN based methods of critical clearing time estimation," in *Proc. IEEE 8th International Electrical and Electronics Engineering*, Bursa, Turkey, 2013, pp. 132-136.
- [14] S. P. Teeuwsen, I. Erlich and M. A. El-Sharkawi, "Robust oscillatory stability assessment for large interconnected power systems," in *Proc. IEEE PES General Meeting*, Denver, Colorado, USA, 2004, pp. 1871-1876.
- [15] H. Sawhney and B. Jeyasurya, "A feed-forward artificial neural network with enhanced feature selection for power system transient stability assessment," *Ele. Pow. Sys. Res.*, vol. 76, no. 12, pp. 1047-1054, 2006.
- [16] K. R. Niazi, C. M. Arora and S. L. Surana, "Power system security evaluation using ANN: feature selection using divergence," *Ele. Pow. Sys. Res.*, vol. 69, no. 2, pp. 161-167, 2004.
- [17] C. Jensen, M. El-Sharkawi and R. J. Marks, "Power system security assessment using neural networks: feature selection using Fisher discrimination," *IEEE T. Power Syst.*, vol. 16, no. 4, pp. 757-763, 2001.
- [18] H. Peng, F. Long and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE T. Pattern Anal.*, vol. 27, no. 8, pp. 1226-1238, 2005.
- [19] M. Robnik-Šikonja, and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," in *Proc. IEEE International Conf. on Machine Learning*, Nashville, Tennessee, USA, 1997, pp. 296-304.
- [20] J. N. Fidalgo and J. Lopes, "ANN sensitivity analysis for identification of relevant features in security assessment," in *Proc. IEEE Power Tech Proceedings*, Porto, Portugal, 2001, pp. 1-6.
- [21] G. Rogers, *Power System Oscillations*. Norwell, MA: Kluwer, 2000.
- [22] DSATools™ Dynamic Security Assessment Software, by Powertech Labs Inc. (2014) [Online]. Available: <http://www.dsatools.com>