FEATURE REDUCTION FOR NEURAL NETWORK BASED SMALL-SIGNAL STABILITY ASSESSMENT

S.P. Teeuwsen
teeuwsen@uni-duisburg.de

I. Erlich
erlich@uni-duisburg.de

M.A. El-Sharkawi
elsharkawi@ee.washington.edu

University of Duisburg, Germany

University of Washington, Seattle, USA

Abstract – This paper introduces several feature extraction and selection techniques. Previous studies showed good results when neural networks are applied to the small-signal stability assessment.
However, the use of reduction techniques can decrease the number of features and thus the number of quantities, which need to be observed, measured, and transmitted, respectively. Moreover, a small number of features allows a quick training of the NN. In this study, three methods are presented: Feature extraction by using a NN as encoder, the principle component analysis, and a selection technique by clustering NN input features. These methods allow a feature reduction up to 75%. Following the reduction, examples are calculated for the prediction of the most important three inter-area modes of the European Interconnected Power System.

Keywords: Small-Signal Stability, Neural Network, Feature Extraction, Principle Components, Clustering

1 INTRODUCTION

The European interconnected power system, also known as UCTE/CENTREL, consists of the western European *Union for the Coordination of Transmission of Electricity* (UCTE) and the *central European power system* (CENTREL), which includes the central European countries Poland, Hungary, the Czech and Slovak Republic. Due to the recent integration of the CENTREL power system, the European network has grown rapidly. Further extensions, e.g. in the Balkan area, are under investigations.
The integration of the two large power systems (UCTE and CENTREL) led to a different stability behavior. Although the European network is strongly mashed, it includes parts with high power concentration, which could swing against each other. Inter-area oscillations are observed particularly when two or more net groups in the power system (i.e. power supply companies) exchange energy. These so-called inter-area oscillations are slow damped oscillations with quite low frequencies.
In the European system, small-signal stability is largely a problem of insufficient damping of these oscillations [1,2].
With the deregulation of the electricity market in Europe, the utilities are allowed to sell their generated power outside their traditional borders and compete directly for customers.
For economical reasons, the operators are often forced to steer the system closer to the stability limits. Thus, the operators need different computational tools for system stability. These tools must be accurate and fast to allow on-line stability assessment.
The small-signal stability method, the modal analysis, is based on the computation of eigenvalues and eigenvectors [3]. The inter-area modes are associated with the swinging of many machines in one part of the system against machines in other parts. In the European case, three global modes (eigenvalues) are of particular importance because when they lack damping the whole system starts to oscillate. For example, load flow situations including large power transits between Spain, Portugal, Poland or some Balkan States lead very often to a weakly damped power system.

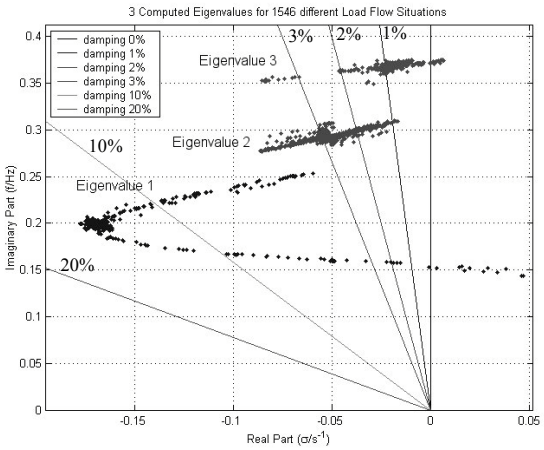

Figure 1: Changes of Dominant Eigenvalues under 1,546 Different Load Flow Situations

Figure 1 shows 3 dominant eigenvalues in the complex plain. These eigenvalues are interesting because they show low frequencies, which identify them as inter-area eigenvalues. The slant lines in the figure characterize constant damping in the range of 0% to 20%. For many different load flow situations, these eigenvalues remain in the stable region, but in some cases they shift to the low damping region and can cause system instability.

2 PROPOSED APPROACH

The computation of the small signal stability is a time consuming process for large networks because it includes the load flow computation, the linearization at the operating point, and the eigenvalue computation. Thus, it is not suitable for on-line applications.

An alternative method is to use a neural network (NN) trained with off-line data for different load flow conditions. By using NN, a fast computation of the eigenvalues is possible, providing that the network is properly designed. For on-line applications, the NN predicts the dominant eigenvalues based on the current operating conditions.

The off-line data can be generated by simulating various load flow situations using a model of the UCTE/CENTREL power system. Hereby, the generation of net groups in the power system is changed to create diverse load flows between the different net groups.

Each new load flow situation in the network provides a new pattern for NN training and the basic challenge is to simulate load flow cases that are highly correlated with the system stability.

Another advantage of the NN is that it can be properly trained with few input features. This is also important considering that due to increasing competition utilities may not share essential information. Only very few features are commonly available such as the transmitted power or the generation of complete net groups. Information about single generators or transmission lines is usually not available.

Once it is trained, the NN can predict the eigenvalues within milliseconds. However, the key issue is to find the best input features that describe the system under study. These input features have to be measurable and need to contain as much information as possible about the small-signal stability.

3 PRE-SELECTION OF NN INPUT FEATURES

The principal applicability of NN for stability prediction has been proven in a previous work [4]. The study is for the large-scale dynamic model of the UCTE/CENTREL power system and Figure 2 shows the accuracy of the three inter-area eigenvalues for different power flow scenarios by using the NN.

The entire data for the UCTE/CENTREL system include features for power equipment such as the transmission lines, transformers, generators, and loads. Hence, there is a large number of features in such an extensive power system. The size of this feature set creates the bottleneck problem for NN training. Therefore, feature extraction or selection techniques are indispensable for NN based small-signal stability assessment.

First, a pre-selection is performed by engineering judgment, whereby only the available and measurable features are used. After pre-selection, the size of these feature sets can be reduced using a reduction technique.

In this study, the selected features are:

- Total generated real power in each net group/utility
- Total generated reactive power in each net group/utility
- Real power transmitted between neighboring net group/utility
- Reactive power transmitted between neighboring net group/utility

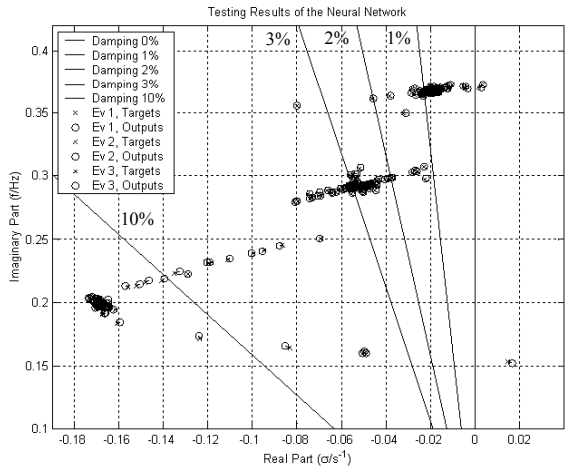

Figure 2: Testing Results of the NN trained with Real Transmission Power Features

The power flow between two neighbored net groups is the sum of power over all transmission lines between them. Voltages are not used in this study because the load flow calculation is based on PV generator nodes that provide a constant voltage level network independently from the current load flow situation.
The total number of all pre-selected features is 211, which is extensive for NN training. Figure 2 shows previous results for NN when trained with real transmission power as selected feature [4]. The eigenvalues marked with crosses are the ones used as targets. The circles are the NN outputs. The targets and the NN outputs are connected by lines. The results are very accurate but further reduction can provide improvement in terms of NN training, which is the focus of this paper.

4 FEATURE SELECTION AND EXTRACTION

The process of finding features that meet given constraints out of a large group of features is called feature reduction. In literature one can find many different concepts for feature reduction. These concepts can be divided into feature selection and feature extraction techniques. Figure 3 illustrates the basic idea.

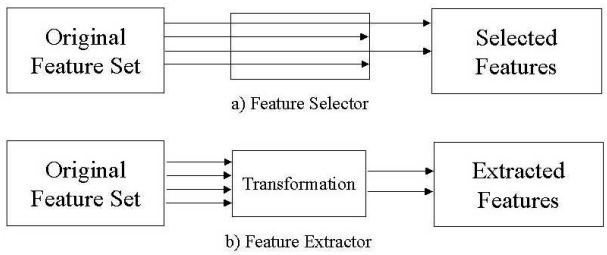

Figure 3: Basic Idea of Feature Selection (a) and Feature Extraction (b)

By feature selection methods, only independent features that provide quality information about the system will be selected. The physical meaning of the features are not changed in any way.
The feature extraction methods work in a different way. Hereby, the features are projected onto a set of reduced order feature space by a transformation function. This transformation changes the physical meaning of the features. The transformation function is an analytical function and the challenge is to find the best function for the given feature set.
Another methods for feature extraction include the encoder technique and the principle component analysis (PCA). These methods are explored in this paper for the small-signal stability of power systems. In addition, the paper describes the use of a cluster algorithm for feature selection.
The neural networks used in this study are designed as multilayer feedforward networks with one hidden layer. The number of inputs depends on the number of used features. The number of outputs is 6 for the real and imaginary parts of 3 dominant eigenvalues. The networks are trained by the backpropagation algorithm.
During the NN training, the error function decreases but could be trapped in a local minimum. To improve the training, the error is observed and the weights are perturbed randomly when the error decreases below a defined rate.
Another tendency of neural networks behavior is memorization. This is an over fitting of training data [5]. Usually, the training error is very small in this case, but the error for the testing patterns is much larger. To prevent a NN from memorizing the training data, the training patterns are shuffled after several iterations. The accuracy of the NN during testing is evaluated, and the training process is stopped when the testing error starts to substantially increase.

4.1 Encoder Network

This section presents the results of feature extraction using an encoder network applied to a pre-selected feature set. The encoder NN is designed to replicate its input as shown in Figure 4.

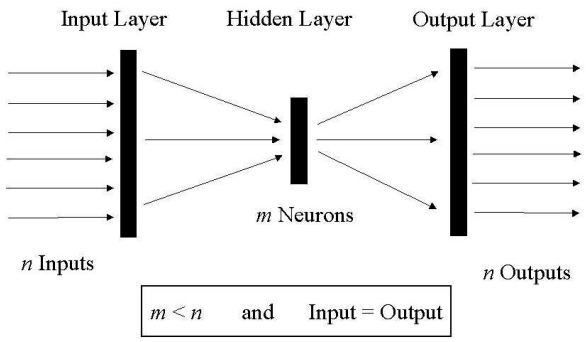

Figure 4: Neural Network as Encoder

The original features are presented to the NN input, passed through a hidden layer and replicated at the output layer. Hence, the number of neurons in the input layer is identical to the number of neurons in the output layer.
For this architecture, the output of the hidden layer neurons can be expanded to the high order input space. Since the hidden layer has a smaller number of neurons than the input layer, the hidden layer represents the feature reduction space.

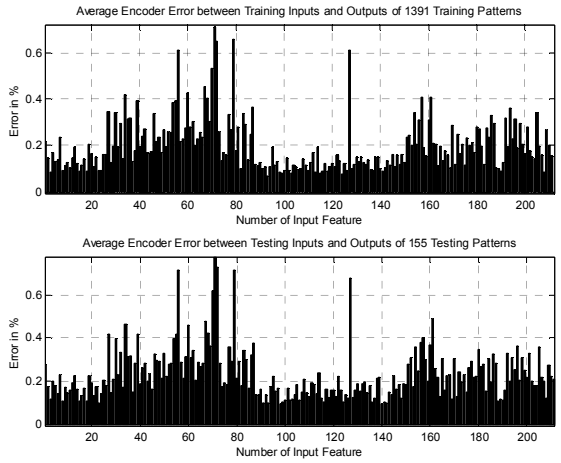

Figure 5: Average Encoder Error for Training (Top) and Testing (Bottom)

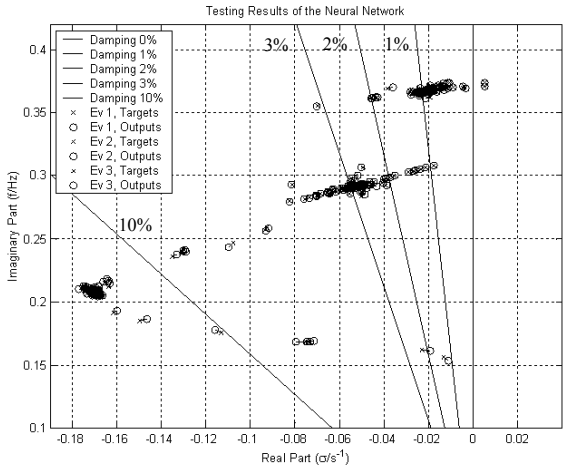

Figure 6: Testing Results of the NN after Feature Extraction by a NN Encoder

The NN encoder designed in this study has 211 features (inputs). The hidden layer includes 53 neurons, which reduces the feature vector by 75%. The number of hidden neurons is obtained from experience with NN training. If the number is smaller than 53, the encoder error will increase extremely and the reduced feature set cannot be used for eigenvalue prediction.

Figure 5 shows the average errors of the NN training and testing, respectively. The average error is computed by equation (1), whereby i represents the feature number and p the number of patterns.

$$E_i = \frac{1}{p} \cdot \sum_{j=1}^{p} \left| output_i - input_i \right| \qquad (1)$$

After the feature extraction process, a different NN was trained by the new reduced feature vector to predict the stability of the power system. The testing results of this NN are shown in Figure 6.

4.2 Principle Component Analysis

The principle component analysis, also known as Karhunen-Loeve expansion, is a linear feature extraction technique [6].

Let **F** be a matrix of dimension $p \times n$, whereby n is the number of the original feature vectors and p is the number of patterns.

$$\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n] \quad (2) \qquad \mathbf{f}_i^{\mathrm{T}} = [f_{i1}, f_{i2}, ..., f_{ip}] \quad (3)$$

The feature matrix is standardized. In other words, the mean is subtracted from the feature vectors and then the result is divided by the sample standard deviation.

Thus, the correlation matrix **C** of the feature matrix **F** can be written as

$$\mathbf{C} = \mathbf{F}^{\mathrm{T}} \cdot \mathbf{F} \qquad (4)$$

The eigenvalues of **C** are determined by solving the following equation

$$\det(\mathbf{C} - \lambda \cdot \mathbf{I}) = 0 \qquad (5)$$

and the eigenvectors, which form an orthogonal principal coordinate system, are given by

$$\mathbf{C} \cdot \mathbf{v}_i = \lambda_i \cdot \mathbf{v}_i \qquad i = 1, ..., n \qquad (6)$$

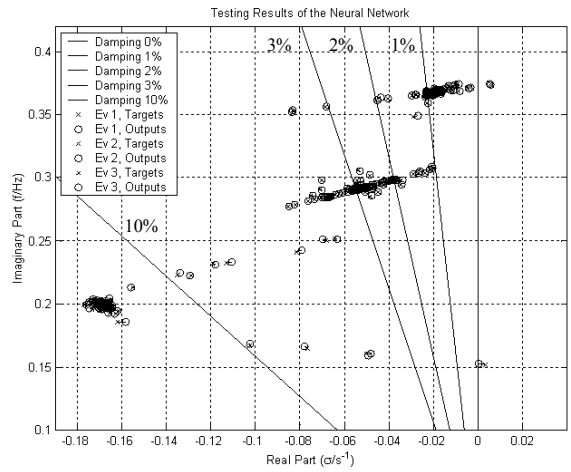

Figure 7: Testing Results of the NN after Feature Extraction by PCA

However, not all principal components must be taken into account, but only those corresponding to the largest m eigenvalues. This is evident considering the fact that the eigenvalues are the variances of the principal components. Therefore, small eigenvalues can be neglected and the remaining percentage of total variation in the data set is computed according to [7] as follows

$$t_m = 100 \cdot \sum_{i=1}^{m} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i \qquad (7)$$

The user can determine the reduction level according to equation (7). However, for comparison purposes, we reduced the feature set to

53, which is results in a percentage of total variation larger 99.9999%.
Finally, the projection of the original data set onto the new space leads to *m* extracted features given by

$$\boldsymbol{\alpha}_i = \mathbf{F} \cdot \mathbf{v}_i \qquad i = 1,...,m \qquad (8)$$

The stability NN was trained by using the 53 features and the results are given in Figure 7.

4.3 Selection by Clustering

Another way of reducing the dimension of a feature set is to eliminate redundant features. The redundancy can be identified by a cluster algorithm. Hereby, the entire set of features is divided into groups called clusters. A cluster consists of features that are similar statistically. Once the set is divided into clusters, one feature from every cluster can be used as a NN input. Because of the similarity between the features within a cluster, one of them can be selected and the others can be treated as redundant information.

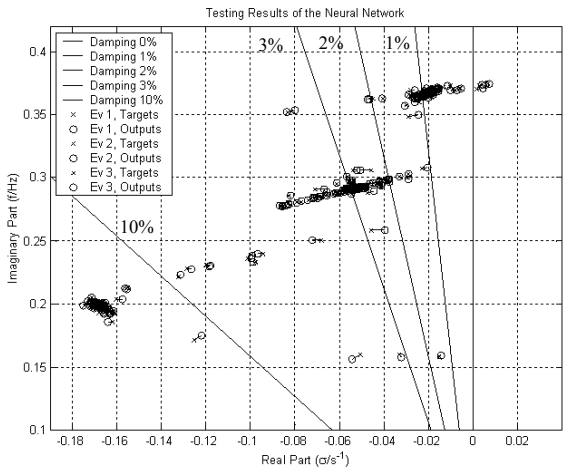

Figure 8: Testing Results of the NN after Feature Selection by Clustering

However, the first step of the clustering technique is to compute the distances between the feature vectors, which can be the Euclidean distance. The Euclidean distance *dist* between two given feature vectors $\mathbf{f}_i$ and $\mathbf{f}_j$ is defined as

$$dist(\mathbf{f}_i, \mathbf{f}_j)^2 = (\mathbf{f}_i - \mathbf{f}_j) \cdot (\mathbf{f}_i - \mathbf{f}_j)^{\mathrm{T}} \qquad (9)$$

The features can then be clustered basing on the largest distance between any two clusters. If feature vector $\mathbf{f}_i$ is the *i*-th object in cluster *r* and $\mathbf{f}_j$ is the *j*-th object in cluster *s*, the distance *d* between cluster *r* and *s* needs to be maximized:

$$d(r,s) = \max(\, dist\,(\mathbf{f}_i, \mathbf{f}_j)\,) \qquad (10)$$

This algorithm is also known as furthest neighbor algorithm and complete linkage, respectively.
Once the process of clustering is finished, the second step is evaluated. Hereby, one feature from each group is selected to form the new feature set. The selection was performed by using the maximal correlation between features and targets.
Figure 8 shows the testing results after the NN is trained using the selected feature set. In this case, the data set was clustered into 53 clusters to reach a reduction by 75%.

4.4 Comparison of the Results

For comparison purposes, an error function needs to be defined. In this study, the error E_λ for a given eigenvalue $\lambda = \sigma + j\omega$ determines the normalized distance between targets and outputs for one pattern. It is defined by the following equation:

$$E_\lambda = \frac{\sqrt{(\sigma_{\text{output}} - \sigma_{\text{target}})^2 + (\omega_{\text{output}} - \omega_{\text{target}})^2}}{\sqrt{\sigma_{\text{target}}^{\;2} + \omega_{\text{target}}^{\;2}}} \qquad (11)$$

This error can be computed for all testing patterns. Table 1 shows the mean and the standard deviation for 3 eigenvalues and all patterns.

	Mean Error	Standard Deviation
Encoder	0.0633 %	0.1646 %
PCA	0.0432 %	0.1011 %
Cluster	0.0920 %	0.1727 %

Table 1: Comparison of the Applied Techniques using Mean Error and Standard Deviation of the Error Function defined by Equation (11)

Table 1 allows to compare the different techniques, but the main criterion for the stability assessment is the prediction of the damping coefficient. Therefore, a second error function E_ξ for the damping coefficient can be defined as follows by equation (12) and (13):

$$E_\xi = \frac{\left|\xi_{\text{output}} - \xi_{\text{target}}\right|}{\xi_{\text{target}}} \qquad (12) \qquad\qquad \xi = \frac{-\sigma}{\sqrt{\sigma^2 + \omega^2}} \qquad (13)$$

Table 2 shows the mean and the standard deviation for 3 eigenvalues and all patterns using an error function regards the damping coefficient defined by equation (12).

	Mean Error	Standard Deviation
Encoder	0.65 %	1.82 %
PCA	0.67 %	4.50 %
Cluster	1.94 %	8. 27 %

Table 2: Comparison of the Applied Techniques using Mean Error and Standard Deviation of the Error Function defined by Equation (12)

Another type of comparison is the time performance. While the PCA method is very fast, the use of an encoder requires much time depending on the total number of features.
Table 3 provides time performance information for the different techniques.

	Reduction Time
Encoder	5 h
PCA	2 sec
Cluster	< 5 min

Table 3: Time Performance Table for the Different Applied Reduction Techniques

5 CONCLUSIONS

The paper discussed different feature extraction and selection techniques for small-signal stability assessment. After application of these techniques, a NN was trained with the reduced data sets. The results are very promising. The eigenvalues are approximated with good accuracy and the stability of the system can be accurately predicted. Although the presented techniques show good results, each has its own merits and drawbacks. The disadvantage of the encoder technique is the time necessary for training. Due to the large number of inputs and targets, the NN structure can be extensive and the training time could be excessive.
In contrast to the encoder, the extraction by PCA is not only highly accurate but also fast. In both methods, the physical meaning of the new patterns is lost. The clustering technique removes redundant features. The introduced cluster solution is accurate and applicable. One drawback of this method results from the fact, that the features are evaluated as single objects. Therefore, it is possible that the n best features selected in this way are not necessarily the best n features for characterizing the whole system.
However, the PCA method and the introduced cluster technique are not only highly accurate but also fast and applicable. Therefore, future research should focus mostly on these two methods.
The impacts of a changing network topology on the NN results were not investigated in this paper. Thus, further studies will take this problem into account.

REFERENCES

[1] U. Bachmann, I. Erlich and E. Grebe, “Analysis of interarea oscillations in the European electric power system in synchronous parallel operation with the Central-European networks”, IEEE PowerTech, Budapest 1999
[2] H. Breulmann, E. Grebe, M. Lösing, W. Winter, R. Witzmann, P. Dupuis, M.P. Houry, T. Margotin, J. Zerényi, J. Dudzik, J. Machowski, L. Martín, J.M. Rodríguez, E. Urretavizcaya, “Analysis and Damping of Inter-Area Oscillations in the UCTE/CENTREL Power System”, CIGRE 38-113, Session 2000
[3] P. Kundur, Power System Stability and Control, McGraw-Hill, New York, 1994
[4] S.P. Teeuwsen, A. Fischer, I. Erlich, M.A. El-Sharkawi, “Assessment of the Small Signal Stability of the European Interconnected Electric Power System Using Neural Networks”, LESCOPE 2001, Halifax, Canada, June 2001
[5] M.A. El-Sharkawi, “Neural network and its ancillary techniques as applied to power systems”, IEE Colloquium on Artificial Intelligence Applications in Power Systems, 1995, Page(s): 3/1 - 3/6
[6] M.B. Zayan, M.A. El-Sharkawi, N.R. Prasad, “Comparative Study of Feature Extraction Techniques for Neural Network Classifier”, Proceedings of the International Conference on Intelligent Systems Applications to Power Systems, ISAP’96, 1996, pp. 400-404
[7] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986
[8] Y. Mansour, E. Vaahedi, M.A. El-Sharkawi, “Dynamic security contingency screening and ranking using neural networks”, IEEE Transactions on Neural Networks, Volume: 8 Issue: 4, July 1997, pp. 942-950
[9] L.S. Moulin, M.A. El-Sharkawi, R.J. Marks, A.P. Alves da Silva, “Automatic Feature Extraction for Neural Network Based Power Systems Dynamic Security Evaluation”, ISAP 2001, No. 21, Budapest, Hungary, June 2001