Aalborg Universitet

Dynamic Security Assessment of Western Danish Power System Based on Ensemble Decision Trees

Liu, Chengxi; Bak, Claus Leth; Chen, Zhe; Lund, Per

Published in:
Proceedings of the 12th International Conference on Developments in Power System Protection, DPSP 2014

DOI (link to publication from Publisher):
10.1049/cp.2014.0150

Publication date:
2014

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Liu, C., Bak, C. L., Chen, Z., & Lund, P. (2014). Dynamic Security Assessment of Western Danish Power System Based on Ensemble Decision Trees. In *Proceedings of the 12th International Conference on Developments in Power System Protection, DPSP 2014* (pp. 1-6). Institution of Engineering and Technology (IET). https://doi.org/10.1049/cp.2014.0150

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: June 18, 2025

Dynamic Security Assessment of Western Danish Power System Based on Ensemble Decision Trees

Chengxi Liu[†], Claus Leth Bak*, Zhe Chen*, Per Lund[†]*

*Aalborg University, Aalborg, Denmark, [†]Energinet.dk, Frederecia, Denmark

Keywords: Dispersed generation, dynamic security assessment, ensemble decision trees, wide-area measurement.

Abstract

With the increasing penetration of renewable energy resources and other forms of dispersed generation, more and more uncertainties will be brought to the dynamic security assessment (DSA) of power systems. This paper proposes an approach that uses ensemble decision trees (EDT) for online DSA. Fed with online wide-area measurement data, it is capable of not only predicting the security states of current operating conditions (OC) with high accuracy, but also indicating the confidence of the security states 1 minute ahead of the real time by an outlier identification method. The results of EDT together with outlier identification show high accuracy in the presence of variance and uncertainties due to wind power generation and other dispersed generation units. The performance of this approach is demonstrated on the operational model of western Danish power system with the scale of around 200 lines and 400 buses.

Nomenclature

AED	Average Euclidean distance.
CA	Critical attribute.
DT	Decision tree.
EDT	Ensemble decision trees.
LS	Learning set.
N_{OOB}	The number of cases in out-of-bag dataset.
N_{tree}	The number of DTs EDT model.
OC	Operating condition.
OOB	Out-of-bag.
TS	Test set.
$\boldsymbol{X}$	Training datasets from bootstrap sampling.
$\boldsymbol{\Theta}$	Randomly selected sub-vector predictor set.
$h_k(\boldsymbol{X_k}, \boldsymbol{\Theta_k})$	The set of classifiers of *k*th decision tree.
$\hat{C}_{DT}^{k}(x_i)$	Prediction of validation case x_i by the *k*th DT.
$\hat{C}_{EDT}^{N_{tree}}(x_i)$	Prediction of validation case x_i by EDT model containing N_{tree} DTs.
$\hat{M}_{EDT}^{N_{tree}}(x_i, C_i)$	Margin of validation case x_i in EDT model.
$S_{EDT}^{N_{tree}}$	Strength of classifiers of EDT model.
$VI(a_m)$	Relative importance of the *m*th variable
$\Psi(a_m)$	Permutation of the *m*th variable data.
$Prox(x_i, x_j)$	Proximity between case *i* and case *j*.
$OI(x_i)$	Outlier index of case *i*.

1 Introduction

Dynamic security assessment (DSA) provides power system operators with important information about the ability of a certain operating condition (OC) to withstand a defined set of contingencies and to regain a state of operating equilibrium after being subjected to a physical disturbance, with most system variables bounded so that practically the entire system remains intact [1].

Offline DSA performs detailed time-domain (T-D) simulations for all credible contingencies and a variety of OCs, so as to determine the security thresholds of generation and load as well as the limits of power transfer in important interties in power systems. However, even though a large number of simulations are conducted for typical OCs and different contingencies, it is still impossible to cover all OCs, especially for large scale power systems.

Online DSA is able to update current OC to supplement the result of offline DSA for higher reliability. It involves rapid screening process to rank and select the most critical contingencies and intelligent filtering of OCs that should be simulated and updated to the result of offline DSA [2]. Then only limited numbers of traditional T-D simulations are carried out to explore the system security guidelines in terms of different criteria. Finally, the guidelines of offline DSA are updated dynamically with precise prediction of security in case of credible contingencies.

Nowadays, the increasing integration of renewable energy resources such as wind farms, solar power plants, energy storage units as well as plug-in hybrid electric vehicles further bring more complexity and uncertainties to modern power systems at different voltage levels. These resources have impacts on power system security, but they are not as accessible for direct monitoring and control from control centre as conventional centralized generation. The presence of these uncertain impacts makes the DSA much more challenging for modern power systems.

Many pattern recognition methods, such as support vector machine, artificial neural network and decision tree (DT), are used in online DSA [3,4]. These methods generate a training database offline by T-D simulations and then train a model with decision rules using the training set of label patterns. Fed with online measurements from SCADA or WAMS, the model is then used to determine the security states of current

OCs. However, it is very difficult to achieve such high reliability of three nines (99.97%), which is the required standard for industrial applications [5].

This paper proposes a methodology that uses Ensemble Decision Tree (EDT) for online DSA, which is capable of not only predicting the security states of new OC with high accuracy, but also informing the adaptability of the EDT model by Outlier Index (OI). The OI is calculated by the Average Euclidean Distance (AED) of the new OCs to all the cases existed in the database. The small AED implies the large proximity of the new OC to the model trained by the database, so that the EDT model is sure about the security states of the new OC, and vice versa. Therefore, given a threshold of OI, the misclassified cases can be completely filtered out as unsure cases for further verification. Finally, the new cases together with the existed cases can be used to build an updated EDT model to strengthen the information of the database.

The rest of this paper is organized as follows. Section 2 introduces the algorithms of traditional DT and the novel EDT models; Section 3 describes the details of the proposed approach for online DSA; Section 4 demonstrates the performance of this approach on western Danish power system; finally, conclusions and discussions are provided in Section 5.

2 Algorithms of ensemble decision trees

2.1 Background

A database can be divided randomly to a learning set (LS) and test set (TS). Given a LS and a TS from T-D simulations composed of a number of OCs and their security results in case of contingencies, a DT can be created to find out the critical attributes (CAs) relevant to the security and to predict the security of any OCs in terms of these CAs. DT is a decision support model expressed as single binary tree to predict the possible consequences of the target value by a number of if-then rules. Therefore, the security of a specific case, i.e. either Secure (S) or Insecure (I), represented by a set of measurements (i.e. $a_1, a_2, \dots, a_k$), can be predicted by dropping the measurements of the case downward from the root node to a terminal node of DT, as shown in Figure 1. Details of DT algorithms can be found in [6].

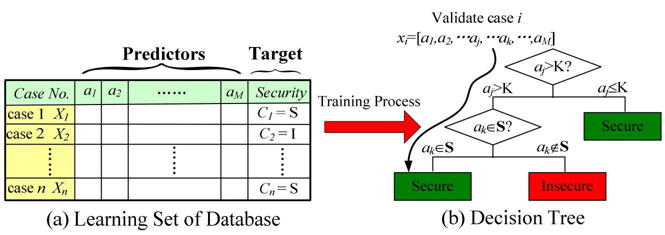

Figure 1: Illustration of database and a simple DT.

The model of EDT, derived from DT, is composed of a multitude of de-correlated DTs such that each node of DTs depends on a sub-vector randomly selected from the full-vector predictors. Bootstrap sampling is also used in EDT by resampling the cases with replacement from the database many times to create a number of datasets with the same size. Each dataset may contain repeated cases because of the replacement. However, it assists in better estimating the distribution of the original database so as to enhance the accuracy. For each DT in EDT model, about one third (i.e. e^{-1} = 36.8%) of the cases are left out from the bootstrap sampling called out-of-bag (OOB) dataset which can be used as TS to give ongoing estimates of the error. The process of creating EDT model is illustrated in Figure 2.

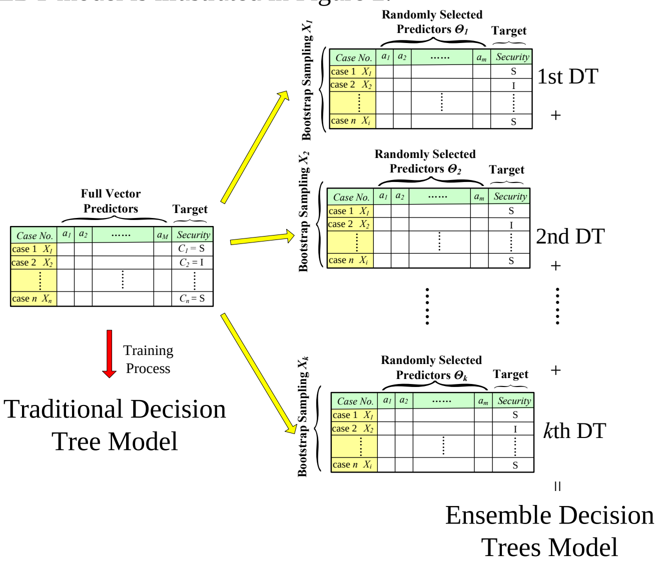

Figure 2: The process of an EDT model.

The classification output (i.e. Secure or Insecure) of EDT model is the majority voting result (largest fraction) from a large number of DTs. Although each DT is unpruned and over-fitted, the overall EDT model can benefit from aggregated-base variance reduction. The structure of EDT is shown in Figure 3. Details of EDT algorithm can be found in [7].

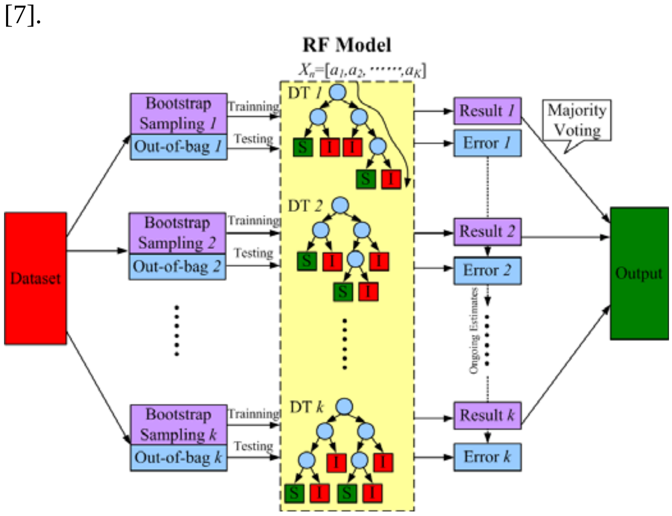

Figure 3: The illustrative structure of an EDT model.

Assuming we have a database containing N cases, each case is represented by M predictors $X_n = \{a_1, a_2, \dots, a_M\}$

representing the measurements of case n and one target C_n = {S or I} standing for the secure or insecure state for case n which is obtained by T-D simulations. Defining that there is an EDT model composed of a number of N_{tree} trees, the process of EDT training is that for each node in the kth tree ($k \leq N_{tree}$), the random sub-vector Θ_j is selected, which is independent from the past random vector $\Theta_1, \Theta_2, \ldots, \Theta_{j-1}$ but with the same distribution. The kth tree is grown using the training set $\boldsymbol{X_k}$ from the kth bootstrap sampling and $\boldsymbol{\Theta_k}=\{\Theta_1, \Theta_2, \ldots, \Theta_j, \ldots\}$ resulting in a number of classifiers $h_k(\boldsymbol{X_k}, \boldsymbol{\Theta_k})$. Therefore, an EDT model is composed of a collection of N_{tree} tree-structured classifiers $DT_k=\{h_k(\boldsymbol{X_k}, \boldsymbol{\Theta_k}),\ k = 1, \ldots, N_{tree}\}$ and each DT casts a unit vote for the most popular class at input x_i, which is defined in (1)

$$\hat{C}_{EDT}^{N_{tree}}(x_i) = \text{majority voting}\left\{\hat{C}_{DT}^{k}(x_i), k = 1, 2, \cdots\cdots, N_{tree},\ x_i \in \mathbf{X}\right\} \quad (1)$$

where $\hat{C}_{DT}^{k}(x_i)$ is the prediction of case x_i by the kth DT in the EDT model.

2.2 Margins

Margin of a case x_i measures the extent to which the average proportion of votes by $\hat{C}_{DT}^{N_{tree}}(x_i)$ for the true class C_i exceeds the average proportion of votes for any other false classes. As defined in (2), the margin of a case is the proportion of votes for true class minus the maximum proportion of votes for the remaining false class. Margin is a measure used to evaluate the confidence of the classification model. Hence, the larger the margin for case i implies the higher of confidence of the classification by EDT model.

$$\hat{M}_{EDT}^{N_{tree}}(x_i, C_i) = \underset{k}{avg}\left(\hat{C}_{DT}^{k}(x_i) = C_i\right) - \max\left(\underset{k}{avg}\left(\hat{C}_{DT}^{k}(x_i) \neq C_i\right)\right) \quad (2)$$

Therefore the strength of classifiers in EDT is defined as the average margin of all cases in the dataset, as defined in (3).

$$S_{EDT}^{N_{tree}} = \frac{1}{N}\sum_{n=1}^{N}\hat{M}_{EDT}^{N_{tree}}(x_n, C_n) \quad (3)$$

2.3 Variable Importance

Compared with single DT, an EDT model has the advantage that it gives each variable the chance to appear in the different context with different covariates, so as to better reflect its potential effect on the response. The importance of variables in EDT model is computed to assess the contribution of the variable to grow the EDT model and the relevance of each variable over all DTs in the EDT model.

There are basically two approaches to measure the relative importance of the mth variable $VI(a_m)$ in the prediction vector $X_n=\{a_1, a_2, \ldots, a_m, \ldots, a_M\}$. The traditional approach is used by default for a single DT, which is calculated by aggregating the improvement indices of splitting variable and the surrogate variables for every node and sort the result descendingly [6]. Then, the relative value of variable importance is calculated by scaling these improvement indices such that the largest value is 100%.

Although the traditional approach is intrinsic, it is subjected to the over-fitting issues in EDT. Hence, the novel approach is used to assess the importance of the mth variable $VI(a_m)$, as defined in (4):

a) Assume that we already know the OOB performance P_0 of a given EDT model;
b) Randomly permute all values of variable m;
c) Score the data and compute the new OOB performance P_m;
d) Compute the decrease of performance P_0-P_m;
e) Sort and scale the performance decrease.

$$\begin{aligned} VI(a_m) &= P_0 - P_m \\ &= \frac{1}{N_{OOB}}\sum_{n=1}^{N_{OOB}}\left\{P_{S_{OOB}}\left(x_n, T_k\left(\mathbf{X}, \mathbf{\Theta}_k\right)\right) - P_{S_{OOB}}\left(x_n, T_k\left(\mathbf{X}_{\psi_m}, \mathbf{\Theta}_k\right)\right)\right\} \end{aligned} \quad (4)$$

where N_{OOB} is the case number in OOB dataset and $P_{S_{OOB}}(x_n, DT_k)$ is the one-zero function for testing the nth case of OOB in EDT model, defined in (5)

$$P_{S_{OOB}}\left(x_n, DT_k\left(\mathbf{X}, \mathbf{\Theta}_k\right)\right) = \begin{cases} 1 & \hat{C}_{EDT}^{N_{tree}}(x_n) = C_n \\ 0 & \hat{C}_{EDT}^{N_{tree}}(x_n) \neq C_n \end{cases}, \forall x_n \in S_{OOB,k}. \quad (5)$$

$\boldsymbol{X}$ and $\boldsymbol{X}_{\Psi m}$ are predictor vector of database before and after permutation of the mth variable a_m, as defined in (6) and (7), respectively.

$$\mathbf{X} = \left\{a_1, \cdots, a_{m-1}, a_m, a_{m+1}, \cdots, a_M\right\} \quad (6)$$

$$\mathbf{X}_{\psi_m} = \left\{a_1, \cdots, a_{m-1}, \psi(a_m), a_{m+1}, \cdots, a_M\right\}. \quad (7)$$

where $\Psi(a_m)$ is the result of permutation of variable m.

2.4 Proximities and Outliers

Proximity matrix ***Prox*** is introduced by EDT to evaluate the Euclidean distance between every two cases in the spaces of observation. The proximity between case i and case j in the database is the number of times that they occur in the same terminal node of trees, then these counts are normalized by dividing the total number of DTs, i.e. N_{tree}, as defined in (8)

$$\text{Prox}(x_i, x_j) = \frac{1}{N_{tree}}\sum_{n=1}^{N_{tree}}\left\{C_{i,j}\left(DT_k\left(\mathbf{X}, \mathbf{\Theta}_k\right)\right)\right\} \in [0, 1] \quad (8)$$

where $C_{i,j}$ (DT_k) is the one-zero function for counting the times that x_i and x_j going to the same terminal node, as defined in (9)

$$C_{i,j}\left(DT_k\left(\mathbf{X}, \mathbf{\Theta}_k\right)\right) = \begin{cases} 1 & \text{if } x_i, x_j \text{ in the same terminal node of } DT_k \\ 0 & else \end{cases}. \quad (9)$$

Therefore, the diagonal elements in the proximity matrix are 1.0 which are self-proximities. Cases that are alike will have proximity close to 1.0; while dissimilar will have proximities close to 0.0.

Then, outliers are able to be detected as cases having small proximities to all other cases belonging to the same target class. The outlier index (OI) of case i is defined in (10), in which N_k is the number of cases that the classification is the same with case i.

$$OI(x_i) = 1 \Big/ \sum_{k=1}^{N_k} \left(\text{Prox}(x_i, x_k) \right)^2, \forall C_k = C_i \tag{10}$$

3 Proposed approach

The flowchart of the proposed EDT-based online DSA is shown in Figure 4. The proposed approach is executed in the following stages.

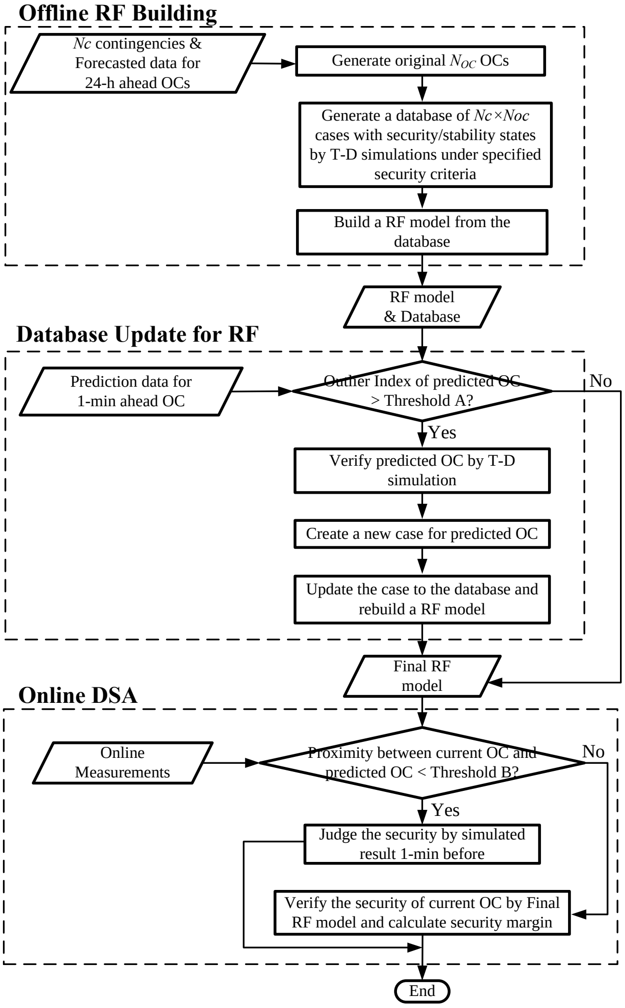

Figure 4: The flowchart of EDT based online DSA scheme.

3.1 Stage I: Offline EDT building

The original database is prepared offline by detailed T-D simulations of $Nc \times Noc$ cases for Nc critical contingencies and Noc OCs based on 24-hour horizon forecasted data of the system, such as load forecast, wind forecast, unit-commitment-based generation plan, network topology as well as the unavailability of system elements due to scheduled maintenance, etc. Then, the original contingency-oriented EDT model is trained based on the prepared database.

3.2 Stage II: Database update for EDT model

System OC of 1 minute ahead can be predicted with much higher accuracy by short-term power flow prediction algorithms, such as short-term load forecast and short-term wind forecast, etc. Then Outlier Index of 1-min ahead OC is calculated according to (10) to evaluate the AED of the predicted OC to all the cases already existed in the database. Given a threshold of OI, the predicted OC with OI larger than the specified threshold will be filtered out as unsure case for immediate verification by T-D simulation. Then the original database will be updated by the new case of predicted OC together with the existed cases. Finally, an upgraded EDT model will be built straightaway with strengthened information of the updated database.

3.3 Stage III: Online dynamic security assessment

Fed by online measurement from WAMS and SCADA system, the proximity between current OC and predicted OC is calculated by (8). If the proximity is smaller than the specified proximity threshold, which implies that the current OC matches the predicted OC, then the security of the current OC is decided by detailed T-D simulation conducted 1-min before. If the proximity is larger than the specified proximity threshold, the security of the current OC is judged by final EDT model built in Stage II and security margin for current OC is calculated by the equation in (2).

4 Case study

The approach proposed in this paper is demonstrated on western Danish power system. The detailed model of western Danish power system with about 400 buses, 200 lines is developed in DIgSILENT/PowerFactory and provided by Energinet.dk. 8 central power plant units and 2 offshore wind farms are connected to the power grid at the transmission system level. Besides, 150 combined and heat power (CHP) plants and more than 200 onshore wind power plants are integrated to the distribution system. The geographical map of western Danish power system is shown in Figure 5.

As mentioned in Section 2, the training process of EDT model should be based on a database which is prepared by offline T-D simulation considering the critical "*N-1*" contingencies and "*N-k*" contingencies from Danish transmission system operator's (Energinet.dk) historical record and experience. The "*N-1*" contingencies are the three-phase faults at the terminals of 400/150kV transmission lines with clearing time of 0.12sec (6cycles). The database contains not only the measurement data as predictor values, but also the target values which are the results of T-D simulation secure (S) or insecure (I) based on criteria as given below:

a) Transient stability: The system is considered as transient instable for a given contingency, if the systems transient stability index (TSI) defined by (11) is lower than 10%, in which $\Delta\delta max$ is the maximum angle separation of any two rotor angles in degree.

$$TSI = \frac{360 - \Delta\delta_{\max}}{360 + \Delta\delta_{\max}} \times 100\% \tag{11}$$

b) Short-term voltage security: The system is considered to be insecure if the duration of any bus voltage going out of range from 0.8pu to 1.1pu is longer than 0.5sec.

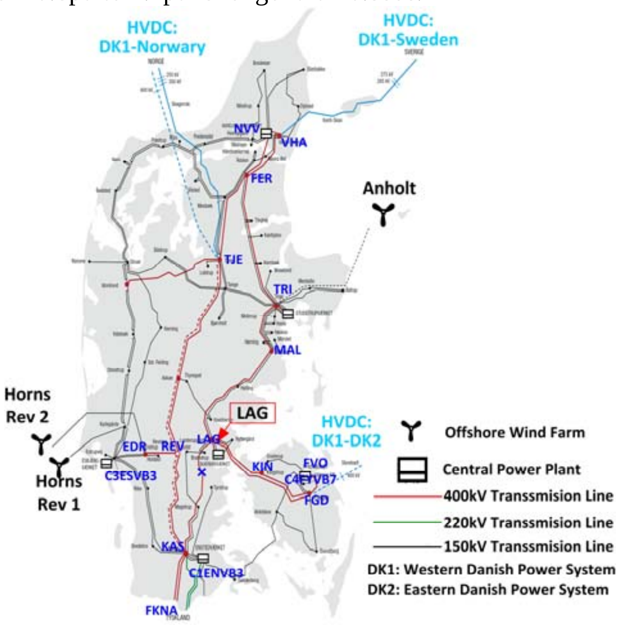

Figure 5: Map of western Danish power system [8].

In order to demonstrate the proposed approach with a typical scenario, the OC with peak load with one 400kV backbone KAS_400_LAG out of service is selected as the tested scenario. The contingency is a 3-phase short circuit in 400kV overhead line FGD_400_LAG close to the substation of LAG followed by the line trip. The short circuit place is marked as a red squre in Figure 5. Then a database with 660 cases is created considering the variance of dispersed generation, such as wind farms and CHP plants.

A data mining software Salford Predictive Miner is adopted to train an EDT model base on the database. The EDT model contains 500 deep-grown DTs, and 10 out of 202 predictors are randomly selected as the predictors within each node of every DT. Figure 6 shows the training process, which shows that the performance of EDT model (Error Rate) does not improve obviously after the 150th DTs. Figure 7 shows the relative variable importance, calculated by (4).

Figure 8 shows the proximity matrix map of the 660 cases in the training database, in which the blue colour shows the secure cases and red colour shows the insecure cases. The brightness of the points indicates the value of proximity (i.e. 0~1) between each 2 cases of the same class in the dataset, which is calculated by (8).

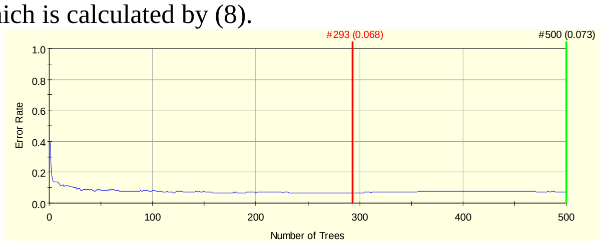

Figure 6: Training process of EDT model.

Variable	Score
FER_400_TJE	100.00
C1ENVB3	82.33
NVV_400_VHA	76.61
FER_400_NVV	75.69
FER_400_VHA	58.00
FKNA400_KAS1	45.55
KAS_400_LAG	37.60
C4FYVB7	32.26
EDR_400_REV	28.05
FGD_400_FVO	27.29
MAL_400_TRI	19.51
KAS_400_REV	12.22
FGD_400_KIN	9.45

Figure 7: Relative variable importance trained by EDT model.

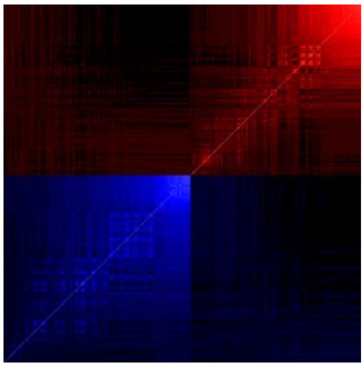

Figure 8: Proximity matrix map of 660 cases in the training database.

After the training process of EDT model, another database with 440 cases with different variance parameters, independent from the previous database is generated to validate the created EDT model.

Figure 9(a) shows the 440 cases in the test set with respect to their OI, which is then sorted in a descendent order in Figure 9(b). The misclassified cases are all within the 30 cases with the highest OI. The blue cases are the correct ones, while the green and red cases are the misclassified ones.

Given a threshold of OI, the predicted OC with OI larger than the specified threshold will be filtered out as unsure case for immediate verification by T-D simulation. As shown in Figure 10(a), the original accuracy of EDT model is tested as 92.27%. If the OI threshold is given as 6.15, and all the cases with OI larger then 6.15 are verified by further T-D simulation, the accuracy can be increased to 95.68%. Whilst, if stricter (more conservative) threshold is given (i.e. 1.54), the accuracy can reach to even 100%.

Figure 10(b) shows the percentage of cases that need for further verification. Only a percentage of 3.82% out of 440 cases are needed to be verified if the threshold is given as 6.15. Nevertheless, a percentage of 11.36% out of 440 cases are needed to be verified if stricter threshold is given as 1.54, and within these cases, a percentage of 7.50% are the mismatched cases.

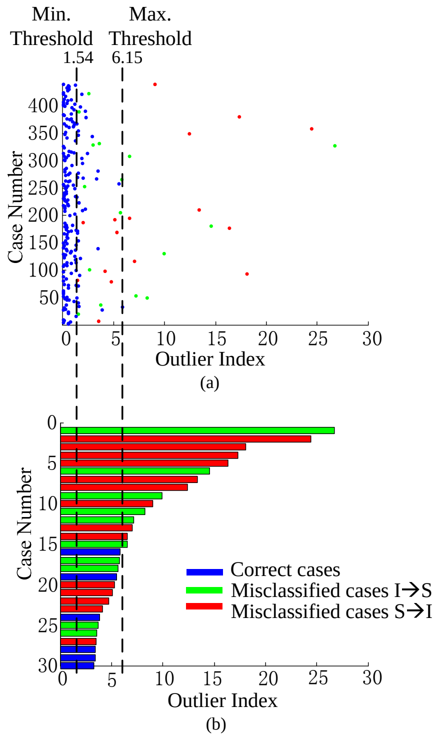

Figure 9: Cases in the validation set w.r.t. their OI.

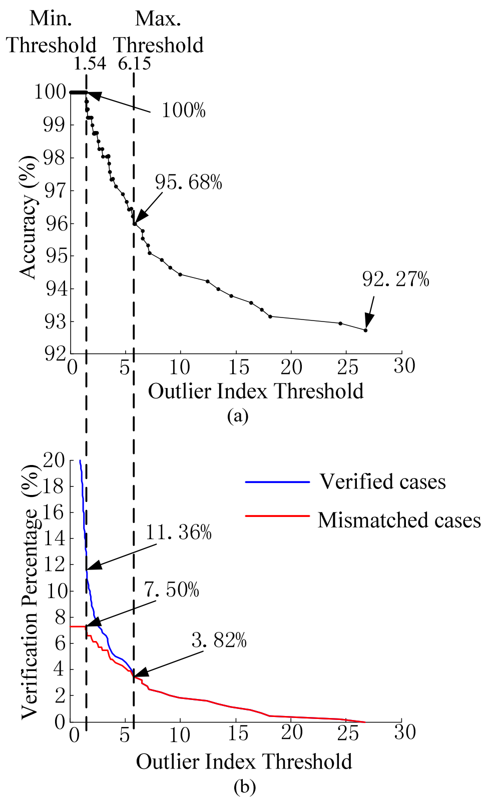

Figure 10: Accuracy and verification percentage w.r.t. different OI thresholds.

Afterward, the database can be periodically updated by the new cases together with the existed cases. Finally, a stronger EDT model can be created straightaway with strengthened information of the updated database. Therefore, the percentage of filtered out cases for verification should be significantly decreased if more cases are added into the database for upgrading the EDT model.

5 Conclusions and discussions

This paper proposes an online DSA approach based on EDT. This approach is demonstrated on western Danish power system considering the uncertainties of wind power and other forms of dispersed generation. Given a threshold of OI, misclassified cases can be filtered out as outliers for further verification 1min ahead of the real time. Practically, this outlier identification method is able to improve the accuracy of online DSA to even 100%.

This paper is a first step towards a goal that to achieve online DSA with 100% accuracy theoretically in the presence of uncertainties of renewable energy. The exact value of OI thresholds to be given should have strong relationship with the value of variance or uncertainties of wind or other DG units in mathematics, which is being pursued in future study.

Acknowledgements

This work was supported in part by Danish Strategic Research Centre under DSF 09-067255, “Development of a Secure, Economic and Environmentally-friendly Modern Power System” (SEEMPS). Authors would like to thank Danish TSO—Energinet.dk for providing the data of western Danish power system.

References

[1] Working group on Dynamic Security Assessment, PES Committee, “Dynamic security assessment practices in North America,” *IEEE Trans. Power Syst.*, vol. 3, no. 3, pp. 1310-1321, (1998).

[2] P. W. Sauer *et al.*, “Dynamic security assessment” in *Power System Stability and Control*, 2nd ed, CRC Press, Boca Raton, FL, USA, 1985, ch. 15.

[3] K. Sun, S. Likhate, V. Vittal *et al.*, “An online dynamic security assessment scheme using phasor measurement and decision trees,” *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1935-1943, (2007).

[4] C. Liu, K. Sun, Z. H. Rather *et al.*, “A system approach for dynamic security assessment and the corresponding preventive control scheme based on decision trees,” *IEEE Trans. Power Syst.*, accepted for future issue.

[5] Electric Advisory Committee: ‘Smart grid: enabler of the new energy economy’. US Department of Energy, (2008).

[6] L. Breiman *et al.*, *Classification and Regression Trees.* Belmont, CA, Wadsworth, 1984.

[7] L. Breiman, “Random forest,” *Mach. Learn.* vol. 45, pp. 5-32, (2001).

[8] *System plan of Danish power system in 2010.* Energinet.dk, 2010.