On-line Identification of Power System Dynamic Signature Using PMU Measurements and Data Mining

A thesis submitted to The University of Manchester for the Degree of

Doctor of Philosophy

In the Faculty of Engineering and Physical Sciences

2015

Miss Tingyan Guo, BEng

Table of Contents

Word Count: 50467

List of Figures

List of Tables

Acronyms

ANN	Artificial Neural Network
AVR	Automatic Voltage Regulator
APS	Autonomic Power Systems
BCU	Boundary of stability-region-based Controlling Unstable equilibrium point
CART	Classification And Regression Tree
CHAID	CHi-square Automatic Interaction Detector
CPU	Central Processing Unit
DC	Direct Current
DT	Decision Tree
EAC	Equal Area Criterion
EPSRC	Engineering and Physical Sciences Research Council
EDT	Ensemble Decision Tree
EEAC	Extended Equal Area Criterion
GPS	Global Positioning System
HC	Hierarchical Clustering
ICA	Independent Component Analysis
ID3	Iterative Dichotomiser 3
LDC	Load Duration Curve
IEEE	Institute of Electrical and Electronic Engineers
LC	Loading Condition
LF	Loading Factor
MCC	Monitoring and Control Centre
MSVM	Multiclass Support Vector Machine
NETS	New England Test System

NYPS	New York Power System
OMIB	One Machine Infinite Bus
OPF	Optimal Power Flow
PC	Personal Computer
PCA	Principal Component Analysis
PDC	Phasor Data Concentrator
PMU	Phasor Measurement Unit
PNN	Probabilistic Neural Network
PES	Power and Energy Society
PSS	Power System Stabiliser
RBF	Radial Basis Function
RF	Random Forest
SNOC	Self* Network Operation and Control
SIME	SIngle Machine Equivalent
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
TSA	Transient Stability Assessment
TSAT	Transient Security Assessment Tool
TVE	Total Vector Error
UEP	Unstable Equilibrium Point
WAMS	Wide Area Measurement Systems
WECC	Western Electricity Coordinating Council
WGN	White Gaussian Noise

Abstract

On-line Identification of Power System Dynamic Signature Using PMU Measurements and Data Mining

Miss Tingyan Guo, The University of Manchester, May 2015

This thesis develops a robust methodology for on-line identification of power system dynamic signature based on incoming system responses from Phasor Measurement Units (PMUs) in Wide Area Measurement Systems (WAMS). Data mining techniques are used in the methodology to convert real-time monitoring data into transient stability information and the pattern of system dynamic behaviour in the event of instability.

The future power system may operate closer to its stability limit in order to improve its efficiency and economic value. The changing types and patterns of load and generation are resulting in highly variable operating conditions. Corrective control and stabilisation is becoming a potentially viable option to enable safer system operation. In the meantime, the number of WAMS projects and PMUs is rising, which will significantly improve the system situational awareness. The combination of all these factors means that it is of vital importance to exploit a new and efficient Transient Stability Assessment (TSA) tool in order to use real-time PMU data to support decisions for corrective control actions. Data mining has been studied as the innovative solution and considered as promising.

This work contributes to a number of areas of power systems stability research, specifically around the data driven approach for real-time emergency mode TSA. A review of past research on on-line TSA using PMU measurements and data mining is completed, from which the Decision Tree (DT) method is found to be the most suitable. This method is implemented on the test network. A DT model is trained and the sensitivity of its prediction accuracy is assessed according to a list of network uncertainties. Results showed that DT is a useful tool for on-line TSA for corrective control approach. Following the implementation, a generic probabilistic framework for the assessment of the prediction accuracy of data mining models is developed. This framework is independent of the data mining technique. It performs an exhaustive search of possible contingencies in the testing process and weighs the accuracies according to the realistic probability distribution of uncertain system factors, and provides the system operators with the confidence level of the decisions made under emergency conditions. After that, since the TSA for corrective control usually focuses on transient stability status without dealing with the generator grouping in the event of instability, a two-stage methodology is proposed to address this gap and to identify power system dynamic signature. In this methodology, traditional binary classification is used to identify transient stability in the first stage; Hierarchical Clustering is used to pre-define patterns of unstable dynamic behaviour; and different multiclass classification techniques are investigated to identify the patterns in the second stage. Finally, the effects of practical issues related to WAMS on the data mining methodologies are investigated. Five categories of issues are discussed, including measurement error, communication noise, wide area signal delays, missing measurements, and a limited number of PMUs.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or institute of learning.

Copyright Statement

[1] See http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf

[2] See http://www.manchester.ac.uk/library/aboutus/regulations

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof Jovica V. Milanović for his expert guidance and encouragement throughout this research. Prof Milanović has been a tremendous mentor and the best example for me. His commitment to excellence and his noble personal qualities have inspired me to become a better professional and a better human being.

A special acknowledgement must go to the Engineering and Physical Sciences Research Council (EPSRC) and School of Electrical and Electronic Engineering at The University of Manchester who have jointly sponsored this project. I would particularly like to thank all the members in the Self* Network Operation and Control (SNOC) work stream in the Autonomic Power Systems (APS) project for their support.

I am extremely appreciative of the excellent academic and social environment provided to me by the Power Quality and Power Systems Dynamics group at the University of Manchester, as well as the Electrical Energy and Power Systems department. The opportunities to develop ideas and to exchange knowledge have been invaluable. Special thanks go to Dr Robin Preece who has selflessly given up countless hours to discuss this research, and to Dr Jairo Quirós-Tortós who introduced me to the volunteer work in the IEEE Student Branch Power and Energy Society (PES) Chapter.

I would also like to congratulate all my friends from North China Electric Power University for recently completing their PhDs in Manchester. Without their good companionship this period would have been far less enjoyable.

Most importantly, I would like to extend my deepest appreciation to Xiaolong Hu, for his patience throughout this research – and particularly during the writing of this thesis. His unconditional love and unwavering belief in me have helped me to get through many difficult times.

To my parents,

for their unlimited love, support and encouragement

during my long lasting education

谨以此论文，献给我的父亲母亲

感谢他们在我漫长的学生生涯中

无止境的爱，支持和鼓励

1 Introduction

As it is widely stipulated, the power systems may become increasingly stressed in the future due to numerous reasons, including increasing load demand, the difficulty to install new transmission lines, changing generation and load types and patterns, and market requirements, etc. Under such conditions the occurrence of disturbances is more likely to push the systems over the stability limits. Therefore, system stability, which has always been one of the most important areas of study in electrical power system design and operation, is currently attracting even more attention than ever before.

This chapter consists of three sections. The first presents a general introduction, the background, and the motivation of this research. The second reviews and summarises the past work in the area. The third section gives the aims and contributions of the research, and describes an overview of the thesis.

1.1 General Introduction

1.1.1 Power System Stability Terms, Definitions and Classification

Stability is a condition of equilibrium between opposing forces. Power system

stability may be broadly defined as the ability of a power system to maintain a stable operation under normal conditions and to regain an acceptable stable operating point after being subjected to a disturbance [1]. The disturbance may be small or large. Small disturbances occur continually on the system because of small variations in loads and generations. The system must be able to adjust itself to operate satisfactorily under these conditions and successfully supply the maximum amount of load. Large disturbances, such as a short circuit on a transmission line, loss of a large generator or load, or loss of a tie between two subsystems, can and do occur in the power systems though less frequently than small disturbances. Under these circumstances, following the operation of protection equipment to remove the faulty component from the network, the remaining system must regain stable operation and restore the supply of power to the affected loads.

The understanding of power system stability is greatly facilitated by the classification of it into various categories.

Rotor Angle Stability indicates the ability of interconnected synchronous machines of the power system to remain in synchronism. In a stable system, the input mechanical torque and output electrical torque of each machine are in equilibrium so that all machines are operated at the same frequency. If the system is perturbed and unbalance between the torques is caused, the rotor of one or more machines would deviate from the synchronous speed. Instability that can result is either a steady increase in rotor angle due to lack of sufficient synchronising torque, or rotor oscillations of increasing amplitude due to a lack of sufficient damping control. The way in which rotor angle stability is maintained is a complex process, depending on both the inherent properties of the rotating machines themselves and the control used

for the regulation of their operation.

According to the type of disturbances, rotor angle stability can be further divided into small-signal (or small-disturbance) stability and transient stability:

- ***Small-signal stability*** is the ability of the power system to maintain synchronism under changes which are considered sufficiently small and so for the purposes of analysis the system equations can be linearized. Since changes such as day to day load fluctuations occur continuously, a practical power system would not be physically operable if it is not small-signal stable. In modern power systems, the main problems surrounding the small-signal stability is to ensure the electromechanical oscillations are well damped [1, 2].
- ***Transient stability***, on the other hand, is the ability of the power system to maintain synchronism when subjected to severe disturbances. System responses would involve significant changes in generator rotor angle, power flows, bus voltages and other system variables, which are influenced by the non-linear characteristics of the system. The post-disturbance operation state is usually different from the pre-disturbance operating state.

Voltage Stability is the ability of a power system to maintain acceptable voltages at all buses. It deals only with the electrical components in the system and is much more localised than rotor angle stability. A system can be considered to be voltage unstable if an increase in reactive power injection at any system bus yields a drop in the voltage at the same bus. Voltages may collapse if a sequence of system events accompanies underlying unknown voltage instability, resulting in unacceptably low voltages for significant parts of the power system. According to the type of disturbances, it can be further divided into small-disturbance voltage stability and

large-disturbance voltage stability.

Mid and Long-Term Stability are concerned with the slow system dynamic responses following severe system disturbances and significant frequency deviations. Study periods typically range from minutes to tens of minutes and incorporate the thermal properties of boilers and generating equipment in addition to electromechanical components.

The classification of power system stability is summarised by Figure 1-1 [1]. As is highlighted in the figure, this research only focuses on transient stability. It is described in detail in the following section.

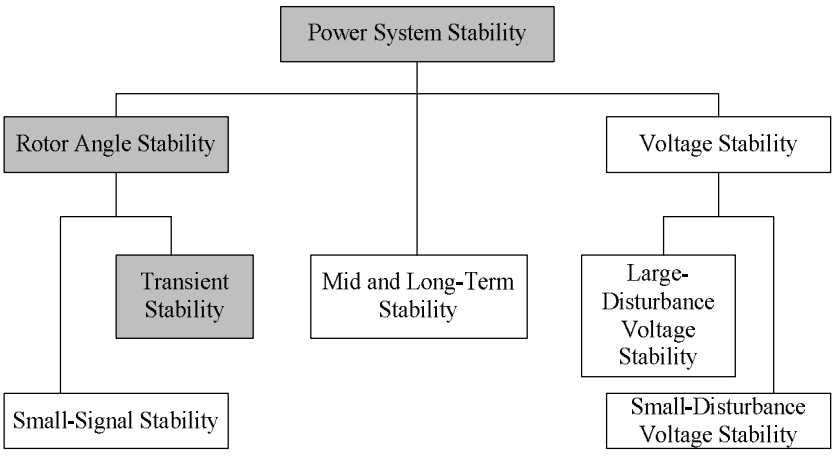


Figure 1-1: Classification of power system stability.

1.1.2 Transient Stability Description

1.1.2.1 Basic Concept

In order to demonstrate the basic concept of transient stability, a simple model consisting of a single generator connected to an infinite bus using a transformer and two parallel transmission lines is considered. Figure 1-2 shows the diagram of the system whilst Figure 1-3 is its reduced equivalent circuit. All resistances are neglected.

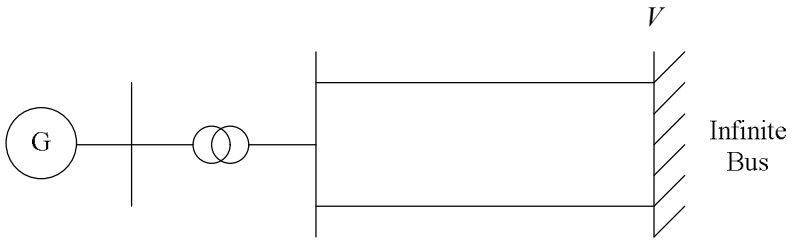


Figure 1-2: Diagram of a simple one machine infinite bus system.

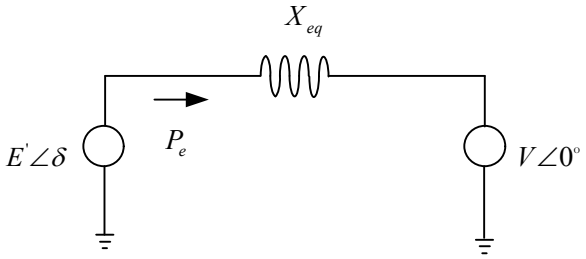


Figure 1-3: Reduced equivalent circuit.

Representing the generator with a voltage E' behind its transient reactance, and assuming the mechanical power P_m is a constant value, the output electrical power of the generator P_e is given by equation (1.1) [1, 3, 4],

$$P_e = \frac{E'V}{X_{eq}} \sin\delta \tag{1.1}$$

where δ is the generator rotor angle, V is the voltage of infinite bus and X_{eq} refers to the equivalent network impedance.

The pre-fault relationship between the electrical power and generator rotor angle can be displayed by the power-angle curve in Figure 1-4. The stable equilibrium point with the rotor angle value of δ_0 indicates the pre-fault steady state operating point A.

Consider a three-phase fault at the bus between the transformer and the transmission lines. As a result of the fault, the terminal voltage of the generator drops to 0. There would be no electrical power produced and hence the power-angle curve becomes 0.

Since there is a continuous input of mechanical power, the generator rotor angle would increase due to the lack of opposing torque resulting from electrical power production. The kinetic energy of the rotor therefore increases during the fault.

After the fault is cleared at δ_{clear} without disconnecting any line, the post-fault power angle curve is the same as the pre-fault one since the system configuration and the value of X_{eq} has not changed. The rotor angle keeps increasing with a progressively declining rate because of the momentum gained during the fault, and reaches its critical value until all kinetic energy is dissipated. Compared to the angle value δ_{limit} of the unstable equilibrium point B, if $\delta_{critical}$ is smaller, as in Figure 1-4 (a), the rotor angle begins to decrease and would oscillate around the original operating point A until damping makes it settle. Otherwise the system would lose transient stability with infinitely increasing rotor angle, as in Figure 1-4 (b).

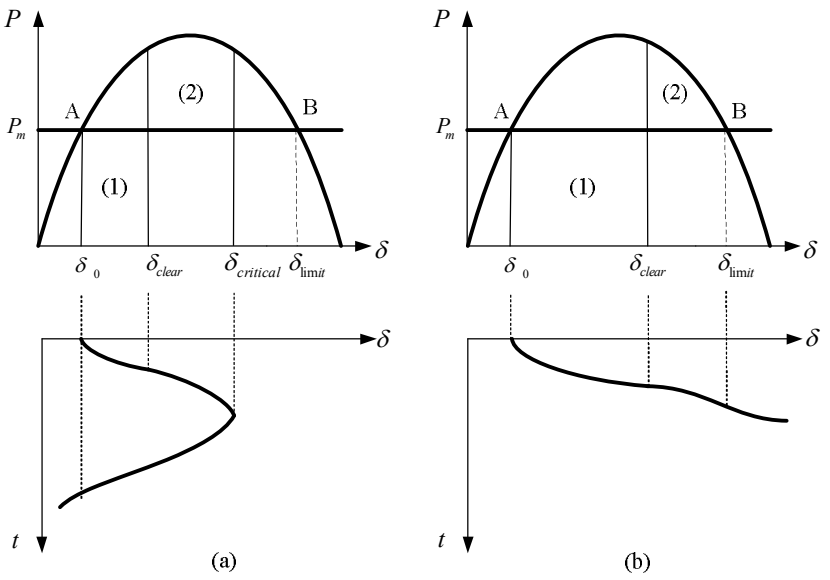


Figure 1-4: Power-angle curve for one machine infinite bus system: (a) short clearing time; (b) long clearing time.

Although the example above is simple, it clearly describes the basic concept of transient stability. The underlying principle is the ability of the power system to

return to the original equilibrium point after large disturbances, or to move from one steady state to another. The swing curve of the generator rotor angle (i.e., evolution according to time) describes the generator's dynamic behaviour.

1) Equal Area Criterion

To determine the transient stability of the system, the simplest method is to utilise the Equal Area Criterion (EAC) [1, 4]. Referring again to Figure 1-4, area (1) represents the amount of kinetic energy the system accrued during the fault whilst area (2) refers to the amount of energy that the system absorbs after the fault. In order for the system to reach the stable equilibrium and maintain synchronism, all the kinetic energy must be absorbed and this can be represented by equating the areas (1) and (2).

2) Influences on Transient Stability

There are many factors that may influence the transient stability of a power system, including [1]:

- Generator(s) loading
- Generator(s) excitation
- Generator(s) parameters
- Fault location
- Type of fault
- Fault clearing time
- Post-fault transmission system impedance

Some of these above factors are related to the disturbance itself whilst the others are related to the operating condition of the system at the time when the disturbance occurs.

1.1.2.2 Types of Instability

In small systems which can be simplified into a single generator infinite bus model as in Section 1.1.2.1, there are two types of unstable cases for the rotor angle response after a larger disturbance. In the first type, the rotor angle continues to grow steadily until synchronism is lost (within 1 to 2 seconds). The transient instability of this type is referred to as *first-swing* instability. In contrast, the system could be first swing stable but then becomes unstable after several seconds as the result of growing oscillations.

In large interconnected power systems, the transient stability problems can be roughly divided into the following situations [4]:

- The generator (or generators) nearest to the disturbance may lose synchronism within the first swing, whilst other generators in the system experience a period of synchronous oscillations until returning to synchronous operation.
- The generator (or generators) nearest to the disturbance loses synchronism at the first place and other generators in the system then follow.
- The generator (or generators) nearest to the disturbance loses synchronism after exhibiting synchronous swings.
- The generator (or generators) nearest to the disturbance exhibits synchronous swings without losing synchronism, but one or more other generators away from the fault lose synchronism with the system.

In transient stability studies, the period of interest is usually 3 to 5 seconds following the disturbance, and can be extended to about 10 seconds due to the configuration of very large systems [1].

1.1.3 Transient Stability Assessment for Different Purposes

Transient Stability Assessment (TSA), as the name suggests, involves the evaluation

of the ability of a power system to maintain synchronism under abnormal conditions. Due to the diversity in power system topology and operation strategies, TSA is carried out for different purposes.

1.1.3.1 Off-line Transient Stability Assessment

Historically, TSA is performed off-line for the purpose of power system planning, *months to years* before the planned system is finally designed, and operation planning where the time horizon is *days or hours*. The primary concern is whether a system in its normal state is able to withstand every possible contingency for a known operating point. A large number of disturbance scenarios must be screened to identify the situations on which the planner should concentrate. Since the time allotted to the process of screening is relatively long, the speed of assessment is not a critical factor.

1.1.3.2 On-line Transient Stability Assessment

As modern power systems are increasingly being operated closer to the boundaries of stability in order to increase the efficiency and economics of their use, TSA needs to be conducted in real-time to provide early warnings and help determine control actions.

1) Preventive Mode

The only on-line TSA scheme used in industry at present time is in preventive mode. Preventive TSA is concerned with forecasting the projected situation [5]. Similar to off-line analysis, a large number of contingencies are screened following each generator re-scheduling in order to identify the potentially harmful ones when the power system is operated in a steady state. If any contingency with high probability of occurrence leads the system to exceed its stability limit, appropriate preventive control action can be deployed. However, in practice only *some (tens of) minutes* are

left for the process of assessing stability of different contingencies in real-time, so that the speed of assessment becomes more crucial.

2) Emergency Mode

Since predicting future disturbances is difficult, preventive TSA essentially aims at balancing the reduction of the probability of losing stability with the economic cost of operation [5]. Because of the combinational nature of the events which might occur, it is very unlikely that control actions can be optimized in the preventive mode even in the context of real-time operation. Ideally, the control decision should be taken for the actual system state, after a large disturbance has actually occurred. This type of emergency (or corrective) control is near optimum since it addresses the real problem. For the need of corrective control, emergency TSA aims at assessing whether the system is in the process of losing stability following an actual disturbance inception. An extremely short time (*fractions of second*) is left to make decisions and take actions and therefore the assessment speed is critical.

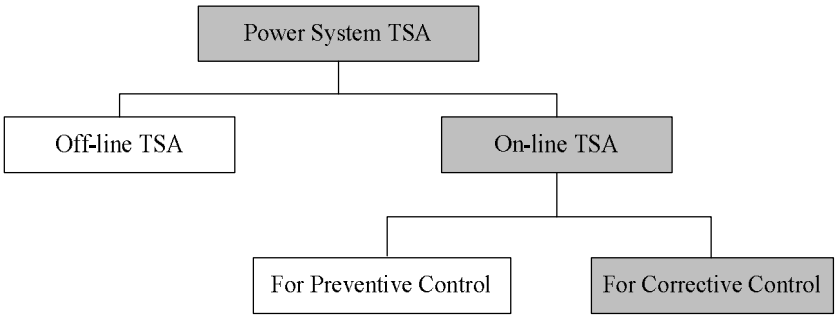


Figure 1-5: Different purposes of TSA.

Different purposes of TSA are illustrated in Figure 1-5. As highlighted, this research only looks into fast on-line TSA in emergency mode, since corrective control and stabilisation is becoming a potentially viable option and needs to be explored in detail to enable safer system operation.

1.1.4 PMUs and Wide Area Measurement Systems

In recent years, Wide Area Measurement Systems (WAMS) have been developed to collect data from various points within large interconnected networks. These measurement systems use Phasor Measurement Units (PMUs) to sample time-stamped phasors which can be used to significantly improve the situational awareness for operational decision making. This opens up the possibilities to predict the system dynamic behaviour after disturbances are cleared in real-time, to help trigger the corrective control schemes.

1.1.4.1 Phasor Measurement Units

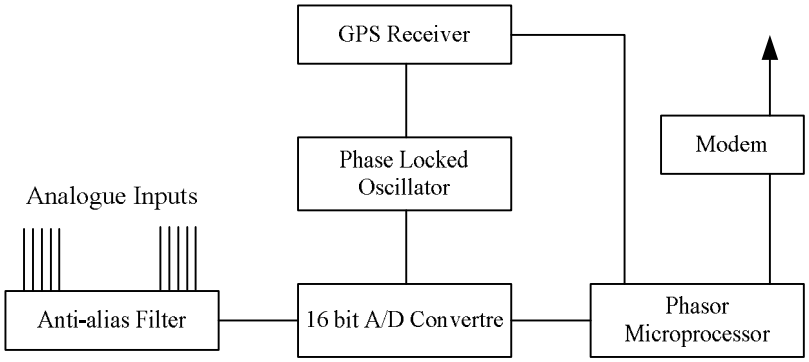


Figure 1-6: Block diagram of PMU.

PMUs were firstly developed at Virginia Polytechnic Institute and State University (Virginia Tech) in 1988 [6, 7]. They are devices which sample voltages and currents waveforms and use signals from the Global Positioning System (GPS) in order to achieve a synchronisation accuracy of 1 μs [7]. Figure 1-6 shows the hardware block diagram of a PMU [8, 9]. The inputs are analogue signals of three-phase voltages and currents measured respectively by potential transformers and current transformers installed in substations. They are sent into an anti-aliasing filter to filter out frequencies above the Nyquist rate, and are converted into digital signals via a 16 bit A/D converter. The phase locked oscillator converts the GPS's one pulse per second

into a sequence of high-speed timing pulses used in the waveform sampling. The phasor microprocessor executes phasor calculations and phasors are finally time stamped and transmitted by means of a modem.

The format of the data files created and transmitted by commercially available PMUs is presently governed by an Institute of Electrical and Electronic Engineers (IEEE) standard (i.e., IEEE Std C37.118.1 - 2011) [10]. PMUs must sample voltage and current measurements with sufficient incidence to ensure the accurate calculation of phasor quantities and are required to report measurements at the system frequency (typically 50 Hz or 60 Hz), though faster reporting rates are encouraged.

The difference between the information from a PMU that describes a phasor and the true phasor itself is measured by the Total Vector Error (TVE), which combines the magnitude and angle error bands into a single error quantity [10]. The IEEE C37.118.1 - 2011 standard establishes a criterion for the TVE to be less than 1%.

1.1.4.2 Generator Rotor Angle Measurement

Although synchronised measurements of voltage and current phasors are a set of signals required to track the dynamic performance of power systems, there is another important signal, i.e., the rotor angle of the synchronous generator. This parameter is mechanical which cannot be directly measured through PMUs. However, it can be developed using electrical phasor measurements together with other measurement techniques.

As indicated in Annex F of [10], the techniques can be categorised into two types. The first one, *Electrical Calculation Method*, derives generator rotor angles from the knowledge of the direct-axis reactance X_d, the quadrature-axis reactance X_q and PMU measurements representing the terminal voltage and current. This method may lead to

errors because the values of X_d and X_q might vary with the generator operating conditions [4]. It is, therefore, usually not adopted in the industry application. The other one, *Rotor Position Measurement Method*, calibrates generator rotor angles against the rotor position (monitored by optical or magnetic means) and terminal voltage. This method is generally accurate and is suitable for real time rotor angle measurement when the power system is subject to a disturbance. The measurement error of the generator rotor angle in this case comes from the Total Vector Error (TVE) of the voltage phasor, and the error introduced in the process of developing the electrical phasor measurements into the mechanical one.

1.1.4.3 Wide Area Measurement Systems

The synchronised signals produced by the PMUs are collected at Phasor Data Concentrators (PDCs), and then transformed into Monitoring and Control Centre (MCC), both using high speed data communication networks. Different system parameters such as active and reactive power and generator rotor speed can then be computed, providing crucial information to monitoring and control applications. Since the devices are geographically widespread, the WAMS usually have a hierarchical structure as shown in Figure 1-7.

Transmitting signals from the remote locations to MCC incurs some time delays. These depend on the physical distances involved as well as the communication media, and have been reported as ranging from 7 – 185 ms for fibre optic cables [11-13] and between 100 – 500 ms for satellite communications [12, 14]. When the satellite links are used, the delays can be randomly increased and the signals can even be completely lost. Furthermore, communication noises in the signals are unavoidable during the data transfer process.

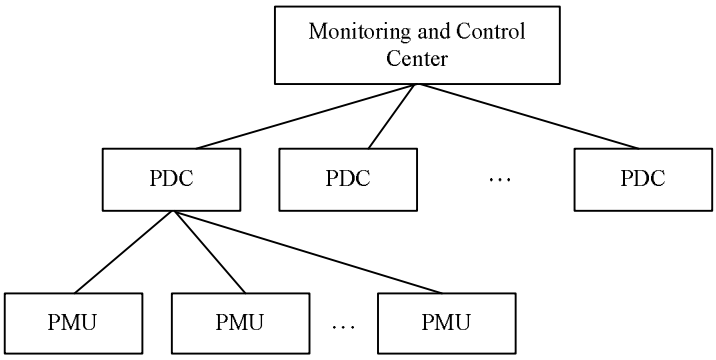


Figure 1-7: General WAMS structure.

1.1.4.4 Recent Implementations

A large number of WAMS projects are being or have been implemented all over the world, with an increasing number of PMUs installed. In China, by the end of 2006, ten WAMS projects were completed or partly completed, i.e., 70 PMUs were put into service, and another 35 PMUs were under construction. They were deployed in five regional systems (including Northern, North-eastern, Central, Southern, and Eastern China power grids) and five provincial systems (including Jiangsu, Henan, Guangdong, Yunnan, and Guizhou provincial power grid). By the end of 2005, the Western Electricity Coordinating Council (WECC) system in North America had reached the size of 11 PDCs (operated by 9 data owners), 53 integrated PMUs and 7 stand-alone PMUs. Hydro-Québec had 8 PMUs installed in 2004 [15]. In 2011 the Finnish WAMS consisted of 12 PMUs and one PDC [16]. More examples and reports can be found for Brazil [17], Mexico [18], Great Britain [19], Denmark [20], Norway [21], Switzerland [22], Japan [23], etc.

With all of these projects, variations in voltage, current, power flows and generator rotor angle and their changing rates at various points within the networks can be monitored. A wide range of novel applications have been made with the incoming measurements, among which a very important stream is to assist the assessment of

different types of system stability and suggestion of control.

1.1.5 Scope of the Thesis

Under this context, this thesis aims to look at the methodologies which can assess power system transient stability in real-time after a disturbance has been cleared, using the incoming monitoring data measured by WAMS and PMUs. The decision making process aims to be as fast as possible, so that time can be saved for deploying corrective control.

1.2 Review of the Past Work

1.2.1 State-of-the-Art of Methodologies for Transient Stability Assessment

In power system engineering, TSA has always been one of the most important and most active theoretical research areas. Many practical implementations of TSA algorithms have also been recently developed or are currently under development. The literature on this topic is extremely extensive. In this section, an overview of the state-of-the-art of TSA methodologies is presented. This overview does not attempt to make a critical survey nor an exhaustive description of all existing approaches. Rather, it attempts to briefly introduce the main streams of methodologies with the purpose for which they are usually applied (i.e., off-line, on-line preventive and on-line corrective). Apart from several fundamental books and papers, the majority of the references have been published within the last 20 years.

1.2.1.1 Time Domain Method

The most straightforward and conventional approach to determine the transient stability of a power system is the time domain simulation [1, 4, 5]. As shown in Figure 1-8, the synchronous generators, including their controls (Power System

Stabiliser, Automatic Voltage Regulator, excitation, turbine and governor), induction motors and other dynamic devices are modelled with differential equations as (1.2) whilst the transmission network and static load are modelled with algebraic equations as (1.3),

$$\dot{x} = f(x, y, p) \tag{1.2}$$

$$0 = g(x, y, p) \tag{1.3}$$

where x represents the vector of states of all dynamic devices, y is the vector of non-integrable inputs and p represents the vector of other parameters that influence the system dynamic.

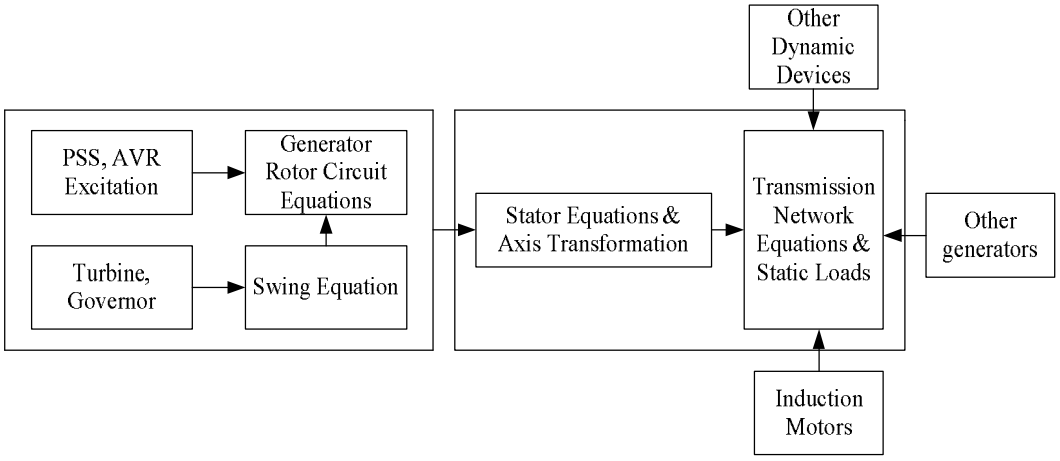


Figure 1-8: Power system model for TSA.

For a given disturbance, the set of non-linear equations are solved using numerical integration methods to compute the generator swing curves (i.e., rotor angle evolution with time) as well as other important system parameters in both during-fault and post-fault configurations.

This approach is very common in practice and yields very accurate and reliable results provided that the modelling of a power system is appropriate and system

parameters accurately known. However, the approach involves extremely intensive computation and it is very time consuming, and therefore it is traditionally used in off-line studies. In the 1990s, much effort was put into the increase of computational speed through methods like parallel processing [24-29]. With the increased Central Processing Unit (CPU) speeds of high performance workstations in recent years, the time required for a single simulation with high order models of a power system has also dramatically shrunk from half an hour to a few seconds [5], and time domain simulations become even "faster than real-time" in many contemporary applications. Furthermore, time domain simulations are essential for the design of many other modern TSA methods which will be presented later in this section.

1.2.1.2 Direct Method

Another famous method for TSA is the direct method (or generally referred to as energy function method) based on Lyapunov's stability theory and the energy-type function. Proposed in the late 1940's to 1960's [30-32], this method has always been one of the most academically appealing topics in power system research. An impressive number of contributions have been made to the field, which is still being fed by new publications.

Physically, using this approach, the assessment of system stability is done by comparing the kinetic energy gained during disturbances and potential energy dissipated in the post-fault duration, instead of solving any differential-algebraic equations. From a mathematical point of view, it consists of the construction of an energy-type Lyapunov's function $V(x)$ to describe the transient energy in the dynamic system. The stability of the system is based on computing $\Delta V = V_{clear} - V_{critical}$ in which $V_{critical}$ is the value of $V(x)$ on the boundary of stability

(i.e., the critical energy required for the system to lose synchronism), and V_{clear} is the system energy at the instant of fault clearing. The system will be considered as unstable if ΔV is bigger than zero. For a simple two-machine system, the direct method is equivalent to the EAC as described in Section 1.1.2.1.

The main advantages that this method can provide, compared to the time domain method, are: i) it is much faster than full scale simulations and thus more suitable for on-line (preventive) applications; ii) it allows the computation of a stability margin which shows how far away the state of the system is at the instant of fault clearing from the stability boundary and facilitates quantitative as well as qualitative assessment of stability. It is extremely difficult, however, to construct a good energy-type function for a multi-machine power system unless over-simplified (actually unacceptably simplified) modelling is used. Different types of Lyapunov's functions have been investigated [33-35]. The analysis has been mostly limited to power system modelling with synchronous generators represented by classical model and loads represented as constant impedance, although efforts have been put into the incorporation of more detailed models [36-39]. Furthermore, it is also difficult to determine a good value of $V_{critical}$, i.e., the critical energy on the stability boundary. Different methods are used such as the closest Unstable Equilibrium Point (UEP) approach, the controlling UEP approach, the Boundary of stability-region-based Controlling UEP (BCU) approach and sustained-fault approach. Extensive references related to these methods can be found in [1, 40]. As a result of these difficulties, the direct method is usually considered to be too mathematically complicated and less accurate than time domain simulation.

1) Extended Equal Area Criterion

The Extended Equal Area Criterion (EEAC) is related to the direct method, and offers the same advantages and disadvantages as previously described. It can be considered as the EAC designed for the multi-machine power system. The essence of this approach is to decompose the system generators into two groups, based on the approximations about the machine coherency and their (over)simplified models at the instance of disturbance inception, transfer them into two equivalent machines and further into a One Machine Infinite Bus (OMIB) system, so that the EAC can be applied. The method has also been extensively investigated in the literature [5, 40-43].

1.2.1.3 Hybrid Method

To combine the advantages of time domain simulation (like accuracy and reliability) with the benefits that can be obtained from the direct method (such as the computation of stability margins and other stability indices), hybrid methods have been developed. A large number of publications have been devoted to the hybrid methods, and they generally can be grouped into two families.

The first family is of the multi-machine type. The idea is to construct a Lyapunov's function for a multi-machine system from the post-fault trajectories of related system parameters gained through the step by step computation of a time domain programme with the desired detailed network model [44, 45]. The resulting energy function actually becomes path-depending and is not a true Lyapunov's function, and is usually used as an early stopping criterion for the time domain simulation.

The second family of hybrid method is of the single-machine equivalent type [5, 46]. For a given disturbance, based on the post-fault trajectories of system parameters resulting from the time domain simulation, the generators in a multi-machine system are divided into a group of "critical machine" and another group of the remaining

generators. A one-machine equivalent system is then constructed from these two groups and the study of its stability is based on the EAC. The most typical example of this family of hybrid method is the Preventive SIME (for SIngle Machine Equivalent) [5, 47-50].

All the methodologies that have been reviewed so far can only be used in preventive mode. Most of the currently existing commercial programmes for on-line TSA use one of the above mentioned methods or combinations of them [40].

1.2.1.4 Emergency SIME

As previously mentioned, with an increasing level of situational awareness provided by WAMS and PMU based measurements, the emergency TSA that is able to quickly recognise the potentially dangerous condition of a post-fault power system and allow sufficient time to take corrective control actions is crucial to be developed, and is the focus of this research.

One of the most studied methods is the Emergency SIME [5, 51]. When a large disturbance actually occurs in a power system, instead of time domain simulation results, the post-fault swing curves coming from the network are sent as input to the SIME assessment procedure. At each sample time, all generators are sorted in decreasing order based on the measured rotor angle values and the group of critical machines, which are above the largest angular distance between two successive machines, are identified. The One Machine Infinite Bus (OMIB) analysis is then performed to decide whether the system remains stable and the corresponding stability margin is calculated.

It has been stated that the Emergency SIME method can identify the impending instability of the system within a very short time (less than 0.5 s) after the clearance

of a large disturbance [5] so that the corrective control actions can be taken early enough. Furthermore, the stability margin can also be provided as the benefit of using the energy related method. However, as with all methods that utilise the OMIB approach, the massive simplification made in the process of transforming a multi-machine system into its equivalent make this method much less reliable. More importantly, only the transient stability status of the system as a whole can be predicted without any information of the dynamic behaviour of individual generators.

1.2.1.5 Data Mining Method

Another category of innovative methods that has also been extensively studied and considered promising as the solution to on-line TSA using PMU measurements is data mining [52-81].

Generally, data mining is an analytic process designed to automatically explore large amounts of data to extract previously unknown interesting patterns or systematic relationships between variables. These patterns or relationships may then be used for future prediction. The process of data mining usually consists of the initial preparation of a database, model building or pattern identification with validation, and maybe the application of the model to new data in order to make predictions. In recent years, it has been used widely in the area of business, science and engineering, etc.

In the particular context of power system on-line TSA, an overall description of the data mining approach can be illustrated by Figure 1-9. A database of contingencies needs to be generated in the off-line stage using time domain simulation, which covers a wide range of prospective operating conditions and disturbances. These contingencies are recorded as a series of parameters (i.e., *predictors*) with classes of

the transient stability status of their resulting system (i.e., *targets*). A classification algorithm is then applied to train a classifier that learns the underlying relationships between the predictors and targets. In the on-line stage, the monitoring data coming from the PMUs in the network will be used as input to the classifier. Intelligent decisions regarding the stability of the system are expected to be made with a high level of speed and accuracy.

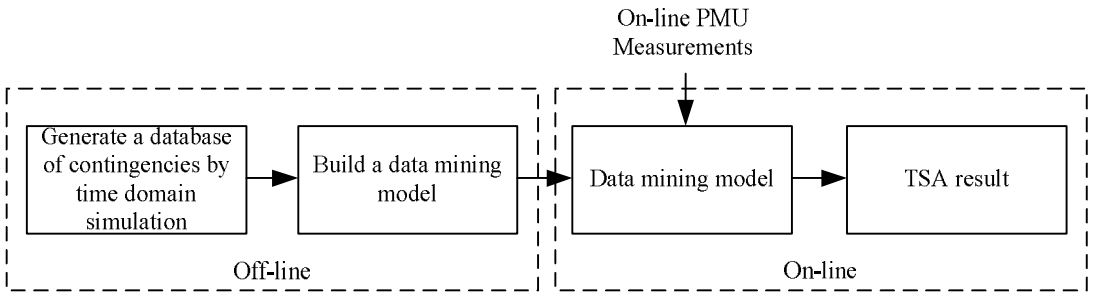


Figure 1-9: Overall description of data mining approach for on-line TSA.

In the literature, the data mining approach has been applied in both preventive and corrective modes. The main difference for building the model in the off-line stage is the selection of parameters to describe the contingencies. For preventive control [52, 53, 56, 57, 61, 63, 64, 67, 70-72, 75, 76, 81], the parameters include fault dependent variables such as contingency ID and the fault type and location, and fault independent variables taken from the *pre-fault* system such as bus voltage angles, bus voltage angle differences, active power flow, reactive power flow or current flow on transmission lines, and output of generators. In real-time, the fault independent variables are obtained from PMU measurements and sent into the data mining model to screen a pre-selected list of contingencies every 10 to 20 minutes. If the output of the model is predicted to be unstable and the associated contingency has a high probability of occurrence, an appropriate preventive control will be designed and armed. For corrective control [54, 55, 58-60, 62, 64-66, 69, 77], in contrast, the

parameters that describe the contingencies in the off-line database are taken from the *post-fault* system, including generator rotor angles, speeds, and accelerations, voltage magnitudes and angles, and apparent resistance along with its changing rate measured near the electrical centre of the intertie of a system (which will change during loss of synchronism). For on-line application, corrective control will be taken when the system is predicted to be unstable after a fault has actually occurred.

A lot of advantages make the data mining approach the most active direction of research for on-line TSA based on PMU measurements. Although a large number of time domain simulations needs to be done off-line, in real-time application, the monitoring data will be the only input in the Monitoring and Control Centre for the model to make decisions. Furthermore, the decision making process is extremely fast with a high level of confidence and so control actions, if necessary, can be taken early enough.

1.2.1.6 Other Methods

Finally, some completely different approaches for power system TSA have been tried and used.

The analytical tool named trajectory sensitivity was applied in the late 1990's [82-87]. The root theory, which indicates that the sensitivity of a system can be described using ordinary differential equations or differential-algebraic equations, comes from control engineering, and has usually been used for parameter identification in adaptive control systems. In this approach, the sensitivities of variations in the state trajectories (such as rotor angle swing curve) in the post-fault power system are computed according to changing system parameters (such as generator output). It has been discovered that when the system is more stressed, the state trajectories become

more sensitive to parameter variations. Thus these trajectory sensitivities can be used as a measure of system stability. However, this method has only been discussed for off-line system analysis or preventive control.

The curve fitting based trajectory extrapolation method is also used to predict the first swing instability [88-94]. In real time operation, the monitoring data of system parameters, usually three or four data points of generator rotor angles and speeds equally sampled after the clearance of disturbance, are used as input to curve fitting algorithms. The small chunk of resulting trajectory right after the input data points is considered as the prediction of the swing curve. However, the trajectory generated from the curve fitting can, more or less accurately, predict the rotor angle or speed of the generator only within the range from 0.2 s up to 0.5 s, which is both unreliable and insufficient for on-line corrective control application. Further work on this method has not been reported.

Furthermore, reference [95] uses maximal Lyapunov exponent computed from PMU measurements in real time to determine if a post-fault power swing will lead to system instability.

1.2.1.7 Summary

The methodologies that have mostly been studied for power system TSA are summarised in Table 1-1. The purposes for which they can usually be applied and whether on-line PMU measurements are needed as input are outlined in the table.

As previously stated, the task of this research is to predict the system dynamic behaviour in terms of transient stability after the fault clearance in real-time system using on-line monitoring data. It can be seen from the table that the main methods proposed in the literature that match the requirements are Emergency SIME and Data

Mining. Although the Emergency SIME does not involve a large number of off-line simulations, the process that reduces a multi-machine power system into OMIB will only be able to give the impending transient stability status of the post-fault system. Data mining, however, as the advanced technique to extracting valuable information from big data, provides the possibility to extend the object of prediction from transient stability to more detailed information about dynamic behaviour of individual generator or groups of generators.

Therefore, the data mining method is of particular interest in this thesis, as indicated in Table 1-1. The past research on on-line TSA with PMU measurements for corrective control using data mining approach will be reviewed in detail in the next section.

Table 1-1: Summary of the state-of-the-art of methodologies for power system TSA.

Method	*Off-line*	*On-line Preventive Mode*	*On-line Emergency Mode*	*Use PMU Data*
Time Domain Simulation	✓	✓		
Direct	✓	✓		
Hybrid	✓	✓		
Emergency SIME			✓	✓
Data Mining		✓	✓	✓

1.2.2 Past Research on On-line Transient Stability Assessment Using PMU Measurements and Data Mining

1.2.2.1 A Typical Structure of Past Publication

The need for predictive tools to convert real time PMU data into transient stability status and support decisions under emergency conditions has led to widespread

research into the field of data mining. Almost all the past work firstly proposed a data mining technique, developed a framework as in Figure 1-9 in Section 1.2.1.5, and then implemented it on a test network to demonstrate its effectiveness. A typical publication in this area of research usually follows a structure that consists of the following points:

- The technical background of the proposed data mining technique
- The test network
- The offline simulations
- The training of model
- Performance evaluation

Although this thesis only looks into TSA in emergency mode, some references that aim for different studies (such as preventive TSA, small-signal stability assessment and voltage stability assessment) are also used for the discussion. The purposes of these applications are different. However, the overall framework for using data mining to help operational decision making is the same.

1) Data Mining Technique

Over the past 20 years, many different techniques have been proposed for this application. The mostly explored one is Decision Tree (DT). Pioneered by [52, 53] in the late 1980's, DT was firstly proposed to assess power system transient stability using real time PMU measurements for preventive control. Since the 1990's, it has been extensively researched for TSA in both preventive [56, 61, 63, 64, 67, 68, 70, 71, 75] and emergency [54, 55, 58-60, 62, 64-66, 69] mode, and also applied for on-line small-signal stability [96, 97] and voltage stability [96, 98, 99] assessment. The most commonly used tree building algorithm is the Classification And Regression Tree

(CART), as in [55, 58, 59, 61, 63, 66, 69, 71]. The CHi-square Automatic Interaction Detector (CHAID) algorithm is investigated in [60]. Some ensemble DT methods, such as Random Forest (RF) in [71], boosting in [68], and random subspace with boosting in [81] have also been applied to increase the accuracy and robustness of prediction.

Support Vector Machine (SVM) has been proposed in recent years and shown to be promising technique to use for on-line TSA for both preventive [72] and corrective [73, 74] control.

Artificial Neural Network (ANN) has also been frequently investigated. Different types of algorithms have been explored, such as fuzzy neural network in [77], Probabilistic Neural Network (PNN) in [78], adaptive ANN in [80] and recurrent ANN in [79].

Furthermore, some other data mining techniques have been developed for on-line TSA using PMU measurements. In [100], a hybrid intelligent system is proposed which is composed of a pre-processor, an array of neural networks and an interpreter. A new intelligent system that consists of a series of extreme learning machines has been developed in [76].

2) Test Network

Standard IEEE test systems have often been selected in the past research, including the 3-machine 9-bus system [79, 80], the 10-machine 39-bus New England system [54, 55, 68, 72, 73, 77, 81, 100], and the 50-machine 145-bus system [76]. A number of realistic large systems with a high level of complexity and interconnectivity have also been used to demonstrate the effectiveness and robustness of the methodology, such as different sizes of the WECC system in the western United States [60, 63, 66,

68, 69, 96], the Entergy power system which is a part of the Eastern Interconnection in North America [61, 64], the Hydro-Québec grid [65], the Brazilian system [72], the western Danish power system [70, 71, 75], the Brittany region of the French system [99], a dynamic equivalent of the power grid of China [76] and the system for Peninsular Malaysia [78, 80].

3) Off-line Simulations

Generation of Contingencies

To implement the data mining approach on the test network, the first step is to generate a database of contingencies by time domain simulations for the purpose of training. Key to the success of on-line TSA in emergency mode is to cover a wide range of system dynamic signatures (i.e., post-fault responses of the synchronous generators in the system characterized by their rotor angle behaviours) for the database to have enough useful information. In theory, training is the most reliable when this database includes system post-fault behaviour similar to that encountered in real-time. However, as power networks are rapidly evolving into systems with a high level of uncertainty due to a different temporal and spatial mix of distributed generation technologies, increased reliance on renewable energy resources, different operating conditions and topologies of the network and different types and composition of load, it is extremely difficult to capture the full range of variation of system dynamic behaviour using a limited number of simulations. Ideally, a well-trained data mining model should be able to assess transient stability of the system with a required confidence level, without complete information about the system composition and operating state.

In past publications, a contingency in the test system is usually determined by the

system operating condition (characterised by system load level and topology) and the transient disturbance itself (characterised by fault type, location and duration). For standard test systems, to ensure the comprehensiveness of the database, the uncertain factors of the operating conditions are usually sampled randomly or uniformly within a certain range, with the disturbances assumed to be widely distributed in the system. In [74], for instance, disturbances simulated for the New England system are three-phase faults on each bus, and three locations on each transmission line. The fault duration is assumed to be five cycles. The system loading levels are selected to be base level plus 5%, 7% and 10%. For realistic systems, operating conditions and disturbances are selected according to historical record, experiences, or forecasted 24-hour data. For example, for the Salt River Project model which is a part of the WECC system in [63], a series of loading levels that can accurately represent different system snapshots during a representative day, July 21, 2008, is selected. A list of critical N-1 and N-2 disturbances, provided by the system operators, is created. Moreover, in [101] which develops a pattern recognition algorithm to determine the post-contingency dynamic vulnerability regions of power systems in real time, a Monte Carlo approach is used to generate the contingency database for training.

Efficient sampling has been researched for the generation of a training database to maximise information content whilst minimizing computing requirements. A two-stage strategy is developed in [99]. It firstly finds the high information content region in the multidimensional operating parameter state space using a linear sensitivity based method, and then biases the sampling process towards that region using importance sampling. The purpose of training in this work though was to predict voltage stability. Similarly, in [71, 75] (which aim to predict both transient and voltage stability for preventive control) a bisection method is used to approximately

identify a region that contains the security boundary. The boundary is created by polynomial curve fitting of points located between a stable and unstable case. Importance sampling is then applied to select the contingencies that can provide more useful information.

Predictors and Target

Within the database, all the contingencies are recorded using a series of parameters, i.e., *predictor*s, and the class to which they belong, i.e., *target*, so that they can be constructed into the input matrix for training. As introduced in Section 1.2.1.5, for the purpose of TSA in emergency mode, the predictors are taken from the post-fault system. A variety of parameters, including generator rotor angles, speeds, and accelerations [54, 55, 60, 74, 77], voltage magnitudes and angles [66, 73, 74], and apparent resistance along with its changing rate measured near the electrical centre of the intertie of a system (which will change during loss of synchronism) [58, 59, 69], have been used and proved to be effective for different data mining techniques.

When the power system grows in size, the number of predictors increases and so the dimension of the data mining problem becomes extremely high. As some of the originally selected predictors may be redundant and do not contribute to the classification, feature selection has been investigated to reduce and optimise the input parameters to the training algorithm. Within all the publications that utilised DT [52-71, 75, 81, 96-99, 102], one of the most important advantages it offers is that no feature selection technique is needed. The tree building algorithm itself automatically selects the best predictor to test at each node and so the training data set can be best classified. An extensively grown tree is pruned to avoid over-fitting so that the test of predictors, which do not contribute significantly to the accuracy of the tree, is

removed. As a result, only the features helpful for the classification are selected in the resulting tree whilst the others do not get used. With SVM, [72] uses two feature selection techniques, sequential search and genetic algorithms, to reduce the dimensionality of the training space from 224 to 150, 100, and 50, respectively. In [78], both the Correlation Analysis Technique and the Principle Component Analysis are used for feature selection and extraction before the training of PNN. The training time has been extremely reduced. A distance-based algorithm called RELIEF is used in [76] to remove noisy and irrelevant features for the extreme learning machine it uses.

In the extensive past research that focus on the on-line prediction of transient stability for corrective control, the target of prediction is typically whether the system remains stable or goes unstable after the clearance of transient disturbance. During the off-line training process, the post-fault behaviour of a power system is classified only into two classes, stable or unstable.

Construction of Training Database

Using the post-fault system parameters as the predictors for corrective control, two approaches have been utilised to construct the training database. The first one, as in [58-60, 66], treats a *single data point* as an individual object in the training set. For one contingency which is simulated for n cycles, n pairs of predictors-target are created, the targets of which are all assigned as the stability status of this contingency. Therefore a training database constructed from m contingencies includes $m \times n$ objects. Using this approach, the decisions in real-time can be made based on a single point of PMU measurements. If, for example, the decision needs to be made 0.5 s after a fault is cleared from the system, the single data point of predictors sampled at

0.5 s are used as input to the model trained using this method. The other one, as in [55, 73, 74, 77], treats one contingency as one object and uses *cumulative data points* after fault clearance as predictors. A training database constructed from *m* contingencies includes *m* objects, with a larger number of predictors to describe each of them. In real-time application the decisions are made based on a window of PMU measurements. The classifiers are trained with various window lengths to investigate the minimum length of measurements that would give satisfactory prediction results [73, 74]. With a model trained using cumulative data points, the predictors within the window between 0 and 0.5 s are all used as input for the decision to be made 0.5 s after fault clearance.

4) Building the Model

With the training databases constructed from off-line simulations, data mining models are usually built using the proposed technique.

A scheme is proposed in [61, 63, 67, 98] to build the model in two steps before the on-line application. A DT is firstly trained 24-hour ahead using a series of operating conditions representing the projected variation in daily load together with the unit commitment based generation pattern, short term network topology, and the probable disturbances. When it is close-to-real-time, the 24-hour time horizon is divided into periods of equal length (several minutes to tens of minutes). During each period, prospective operating conditions and disturbances for the next period will be used to generate new contingencies to test the existing DT. If its performance is not satisfactory the new contingencies together with the old ones will be used to build a new DT so that the model on-line is periodically updated. This scheme is designed to increase the robustness of the model to variations in uncertain factors such as load

and topology.

Looking at the models themselves, DT is the one that offers the advantage of high human readability, compared to SVM, ANN and some other more compressive methods. It uses a white-box model, whose internal logic and reasoning process is very easy to comprehend. The threshold values of predictors which distinguish the contingencies from one class to the other are identified, and they help system operators to understand how decisions are made. In contrast, classifiers such as SVM and ANN are black-box. The mathematical processes behind the decision making are not easily explainable. The models produced do not naturally provide any useful intuitive reasons about why a particular contingency is classified into one class rather than another.

5) Performance Evaluation

To demonstrate the effectiveness of the models, a number of contingencies which have not been included in the training database are used for testing. A number of factors are usually considered when assessing the performances. The *accuracy* of prediction is clearly important, since incorrect decisions under the emergency conditions (i.e., a large disturbance has actually happened) can cause undesirable consequences: predicting stable contingencies as unstable will trigger unnecessary control whilst predicting unstable contingencies as stable will lead to lateness or failure of control and ultimately the instability of the system. The *speed* of decision making is also a crucial factor. The shorter the time that the model needs to wait (after the clearance of disturbance) to get all the needed measurements is, the longer the time to deploy corrective control will be. In addition, the *robustness* of the model is of great concern since the system conditions become more variable and the level of

uncertainty increases.

Great performances of different types of data mining models have been claimed in past publications for on-line corrective TSA. Their accuracy of prediction is reported to range from about 90% to almost 100%, with the decisions made from four cycles to approximately 1 s after the disturbances are cleared. However, the claimed accuracies are quantified in various, not very consistent, ways. Typical approaches to selecting the test database include: i) Picking up a list of representative individual contingencies which are not used for training [60, 73, 103]; ii) Creating one set of new contingencies using the same way in which the training database is sampled, or generating a larger database in the first place and randomly choosing a portion of it for training and the rest for testing [58, 59, 66, 77]; iii) Creating multiple sets of new contingencies, each of them under the variance of one uncertain factor [55, 74]. For example, in paper [55], the DT models generated for the New England 39-bus test system are tested based on three data sets, containing contingencies with various fault locations and system operating points. Paper [74] generates a SVM model for the same network. The accuracy of transient stability prediction is tested against factors such as type of fault, topology change and type of load, one at a time. Paper [73], on the other hand, used six different contingencies to test its SVM model for the New England test system. The reported accuracies of these models can only demonstrate the robustness of the data mining models to a certain extent, but cannot describe their overall performance in the context of complex and uncertain networks, and the reported accuracy of the models is not comparable to each other. There is still a need for a thorough assessment of the prediction accuracy so that the system operators will know how confident the model is in making the right decision at the right time under emergency conditions.

1.2.2.2 Identification of Generator Grouping in Event of Instability

As previously presented, the targets of the prediction for on-line corrective TSA in most of the past works are binary: whether the post-fault impending system is going to be stable or unstable. When a system goes unstable, however, the generators can be separated into different groups based on the similarity of their dynamic behaviour, i.e., transient responses of generator rotor angles. It is also of interest to the system operators to identify i) The grouping of generators, ii) which groups will lose synchronism and iii) which groups will remain stable. This information can not only assist the selection of appropriate corrective control actions, e.g., generator tripping, load shedding, fast-valving and dynamic braking to prevent instability or to limit the effect of losing some of the generators, but it can also determine the level of stability of stable generators and subsequently apply appropriate damping control. The grouping of generators can also be used as input to controlled islanding schemes which separate the system into smaller islands to avoid cascading outages and system blackouts [104]. Splitting of the system into appropriate islands or dynamically coherent zones can be carried out based on the grouping of generators to maintain the static and dynamic constraints of islands with pre-defined limits and to minimize disruption of customers.

Although great effectiveness of binary classification methods have been shown for on-line transient stability prediction, very little work has been done to further predict the dynamic behaviour of generators in the event of instability. In both [100] (Neural Network based) and [54] (DT based), the method of building parallel classifiers has been proposed to predict the grouping of unstable generators. A series of classifiers needs to be trained to determine whether every pair of generators is in the same synchronized group. Reference [54] also discusses the idea of defining patterns of

grouping in advance and predicts the pattern with a single classifier. The grouping pattern for each contingency was identified by firstly making a similarity matrix, the elements of which depend on the relationship between every pair of generators, and then reconstructing the global behaviour from the matrix. However, not all the patterns in the training database are listed for the test system, and only DT is used as the multiclass classification technique to build the single classifier in [54].

1.2.2.3 Effect of Practical Issues Related to WAMS

In most of the studies reported in the past, it is assumed that the WAMS provide perfect signals required by the data mining models. In a practical environment, however, there are many issues related to the measurements. For example, communication delays will be involved when transferring signals from remote locations to the control centre. Some of the measurements may not be available when needed, due to reasons like randomly increased delays, complete loss of communication links, and unexpected failure of PMUs or PDCs, etc. Furthermore, there might not be enough PMUs installed at the locations where signals are required. The quality of the measurements can also be reduced due to measurement error and noise involved within the data transfer process. All these issues would have an effect on the application of data mining tools.

The issue of missing PMU measurements was addressed in [81], although the application there was for TSA in preventive mode and the predictors for DT were selected as pre-fault quantities. Surrogate split, which is a fully automated mechanism included in CART algorithm [105], is firstly investigated. During the training process, one or more surrogate splits, that can achieve similar splitting results to the primary split, may be generated for every tree node. Therefore, if the value of a predictor is

missing for a test case, the surrogates will be used to decide which child node it should go to for final classification. It is observed, however, that the surrogates splits generated are often parameters measured by the same PMU as the primary split. So the signals of the primary predictor and its corresponding surrogate are always missing at the same time. This fact limits the number of surrogates that can be used. A comprehensive and advanced ensemble DT algorithm which combines random subspace and boosting is then developed to improve the robustness to the missing data. Specifically, multiple small DTs are firstly trained offline using random subspace considering the location information of predictors and the availability information of PMU measurements. Then, the performance of these small DTs is re-checked by using new cases in near real-time. Viable small DTs are identified in case of missing PMU data and boosting is then used to re-weight them when making the decision.

In [81], the impact of measurement error has also been investigated. Since the voltage and current phasor provided by PMU are directly used as predictors, errors are randomly generated using standard Gaussian distribution with maximum TVE not more than 1%. Results show that the misclassification rate of the ensemble DT slightly increases. In [74], in order to test the prediction accuracy of SVM under measurement errors, a random error between 0 and 1% is added to all bus voltage phasors in the testing. The performance is very poor if the SVM is originally trained with perfect signals and can be improved through retraining using signals with errors. Furthermore, although [106] uses the data-driven approach for coherency identification in interconnected power systems instead of transient stability assessment, a White Gaussian Noise (WGN) with Signal-to-Noise Ratio (SNR) of 25 is added to the original data to assess the performance of the Independent Component

Analysis (ICA) and Principal Component Analysis (PCA) method in the presence of noise.

1.2.2.4 Summary of Past Research

Having reviewed the past research on on-line TSA for corrective control using PMU measurements and data mining, DT is found to be the most suitable for this application. Obvious advantages are offered compared to other techniques: it directly finds out critical values of power system parameters which distinguish a contingency between stable and unstable and are easy for humans to comprehend. It also automatically carries out feature selection. DT is therefore adopted as the main classification tool in this thesis. Several areas have been identified which need to be addressed. These are summarised as follows:

- The sensitivity of DT-based accuracy of prediction for transient stability needs to be investigated to different uncertainties in the network. For the corrective control application, the robustness of the DT method has not been tested against a complete list of uncertain factors, including type, location, duration of fault and loading level and topology of the network, one at a time. It is still unclear to which factors the DT method is the most sensitive.
- A framework for the assessment of the accuracy of data mining models for on-line prediction of transient stability is currently lacking. Previous published research in this area has used a variety of, not very consistent, approaches. The system operators need to know the confidence level of the decision that they are making.
- The majority of the existing research classified the transient status of post-fault real-time system into stable and unstable. Very few studies concerning

the patterns of generator grouping in the event of instability have been completed. A methodology which combines the assessment of system stability and the generator grouping is a logical extension of existing work on TSA which is along the way towards application of corrective control.

- Only the issues of measurement error and missing measurements have been discussed in the previously published research. The effect of other practical issues that exist in real life WAMS environment on the on-line identification of power system dynamic signature need to be investigated.

1.3 Aims, Contributions and Thesis Overview

1.3.1 Aims and Objectives

This thesis aims to address many of the unresolved issues which have been identified within the current body of research. The main aim of this research is to develop a robust methodology for on-line identification of dynamic signature, in power systems with an increasing level of uncertainty. Based on incoming system responses from monitoring devices (i.e., PMUs in WAMS), a decision should be reached as soon as possible regarding the stability of the system and the pattern of its dynamic behaviour in the event of instability. In order to achieve these aims, the following research objectives have been defined:

1. To review and summarise existing methodologies for fast detection of power system transient stability based on PMU measurements for the purpose of corrective control, and to identify the appropriate methodologies for future development and application.
2. To implement the selected method for on-line transient stability prediction on suitable test network developed in MATLAB/Simulink.

3. To establish the sensitivity of the accuracy of the selected method for on-line transient stability prediction to a list of power system uncertainties.
4. To develop a generic framework for the assessment of the accuracy of prediction of data mining models, dealing with different levels and types of power system uncertainties.
5. To develop a methodology for real-time prediction of power system dynamic signature, which includes, both transient stability status and the pattern of generator grouping in the event of instability, using PMU measurements and data mining.
6. To assess the robustness of developed methodology to practical issues related to WAMS application.

1.3.2 Main Contributions of this Research

The work within this thesis contributes to a number of areas of power systems stability research, specifically around the data-driven real-time emergency mode Transient Stability Assessment (TSA). The main outcome of this research is the methodology of on-line identification of power system dynamic signature using PMU measurements and data mining techniques. The robustness of the developed methodologies has been evaluated considering a wide range of uncertainties surrounding transient disturbance, operating conditions and WAMS.

The main contributions within this thesis can be summarised as follows (Note: References prefixed with the letter 'B' refer to author's thesis based publications. A full list of international journal and conference publications resulting from this thesis is included in Appendix B):

- The sensitivity of the accuracy of the DT method for on-line corrective

transient stability prediction to a series of power system uncertainties, including fault duration, location, system operating point and pre-fault topology, has been established [B3].

- A generic probabilistic framework for the assessment of the accuracy of a data mining model for on-line prediction of transient stability is developed. The framework allows a consistent comparison of different data mining models based on the accuracy of the prediction [B1]. This is the first original contribution of the thesis.
- A two-step methodology is proposed for PMU-based on-line identification of power system dynamic signature. It firstly predicts transient stability using traditional binary classification and then predicts the pattern of unstable system dynamic behaviour using both clustering and multiclass classification [B2, B5, B6]. This is the second original contribution of the thesis.
- The Hierarchical Clustering has been applied to a database of unstable contingencies to automatically identify characteristic patterns of generator grouping without specifying the number of groups in advance [B2, B5]. This is the third original contribution of the thesis.
- Different multiclass classification techniques, including Decision Tree, Ensemble Decision Tree and Multiclass Support Vector Machine, have been evaluated for the first time for the prediction of generator grouping in the event of instability [B2, B6]. This is the fourth original contribution of the thesis.
- The practical issues related to WAMS implementation have been divided into five categories, including measurement error, communication noise,

wide area signal delays, missing signals and limited number of PMU devices, and the effects of each of them on the performance of data mining methods have been investigated.

- The off-the-shelf ensemble methods for Decision Tree, including bagging, boosting and Random Forest have been critically assessed and compared, with respect to their ability of dealing with missing measurements when predicting power system transient stability [B7].

1.3.3 Thesis Overview

This thesis consists of eight chapters in total. The seven chapters which follow this introductory chapter are outlined below:

Chapter 2 – Power System Modelling and Simulation Tools

This chapter provides the basis of the power system modelling and simulation tools used within this thesis. It firstly presents the mathematical descriptions of power system components. The way in which different types of faults are represented is then explained. Following this, brief descriptions of the tools that are used to perform the dynamic simulation are given, alongside the test network utilised throughout the thesis.

Chapter 3 – Data Mining Techniques

The technical background of all the data mining algorithms used in this thesis is introduced within this chapter, and therefore it is not repeated when the specific algorithms are applied in Chapter 4 to 7. The data mining algorithms considered include DT (both CART and C5.0), Ensemble DT (using bagging, boosting and RF), Multiclass SVM, and Hierarchical Clustering.

Chapter 4 – On-line Identification of Power System Transient Stability Using Decision Tree Method

This chapter firstly outlines the methodology of using DT for on-line prediction of transient stability for corrective control. An implementation of the method is presented on the test network. The CART algorithm is applied to train a DT model using a database of system contingencies, containing generator rotor angles and speeds as predictors. Test datasets are designed to incorporate various uncertainties in the system, including fault duration and location, system operating point and pre-fault system topology. The sensitivity of the accuracy of the DT model is assessed against these uncertainties.

Chapter 5 – Probabilistic Framework for Assessing the Accuracy of Data Mining Tools

A generic framework to assess the accuracy of a data mining model for on-line prediction of transient stability for corrective control is developed within this chapter. The framework performs testing by exhaustively searching a wide range of possible contingencies and then weighs the accuracies according to realistic probability distribution of system uncertainties. DT is used to illustrate the framework, with probabilistic distributions assumed for the test network.

Chapter 6 – On-line Identification of Power System Dynamic Signature

This chapter proposes a novel methodology for PMU-based on-line identification of power system dynamic signature using data mining. The proposed methodology firstly predicts the transient stability status using traditional binary classification and then predicts detailed generator dynamic behaviour in the event of instability. The second step applies unsupervised learning to pre-process the off-line simulated

unstable contingencies, and uses multiclass classifiers to perform prediction. The effect of changes in system operating conditions on the proposed methodology is also discussed.

Chapter 7 –Effect of Practical Issues Related to WAMS

The practical issues related to Wide Area Measurement System application for data provision for real time TSA are categorised at the start of this chapter. The effects of each individual category on the prediction of transient stability are investigated and discussed one after another in the rest of the chapter.

Chapter 8 – Conclusions and Future Work

In this chapter the main conclusions of the research are presented and discussed. Suggestions are also given for the future development and improvement of the presented methodologies, as well as for future work in the general area of real-time TSA.

2 Power System Modelling and Simulation Tools

2.1 Introduction

There are various well documented tools available to engineers for power system modelling and analysis. This chapter describes the fundamental techniques required to complete the transient stability studies (i.e., simulations of power system dynamic response). The mathematical models of all the main system elements are firstly presented, including synchronous generators, excitation systems, power system stabilisers, transformers, transmission lines, system loads and the electrical network. The representation of different types of faults is also presented. Following this, the tools used to perform the dynamic simulation is introduced. Finally, a description of the test network used in this research and its WAMS is presented.

2.2 Modelling Power System Components

This section introduces the models of all main power system components which are used throughout the research to provide the simulation results presented later. The

power system, and all included components are modelled with an orthogonal phase representation, under the assumption that all three phases are balanced [4].

2.2.1 Synchronous Generators

The synchronous generator is the fundamental source of energy within modern electrical power systems. The modelling and analysis of it has been subject of investigations since 1920 [1], and has been covered in a number of books [1, 4, 107]. Accurate modelling of its dynamic performance is extremely important in transient stability studies.

Physically, a synchronous generator consists of two essential elements: the rotor and the stator. The field winding on the rotor carries direct current and produces a magnetic field which induces alternating voltages in the armature windings on the stator. The three-phase windings of armature are distributed 120 degrees apart in space so that the uniform rotation of the magnetic field generates voltages displaced by 120 degrees in time domain. When carrying balanced three-phase currents, the armature will produce a magnetic field in the air-gap rotating at synchronous speed. The field produced by the direct current in the rotor winding, on the other hand, revolves with the rotor. For production of a steady torque, the fields of stator and rotor must rotate at the same speed and therefore the rotor must run at the synchronous speed.

There are two basis rotor structures used, depending on speed. Hydraulic driven generators, which operate at low speeds, usually use salient pole rotors with a high number of magnetic poles. Steam or gas powered generators, the speeds of which are much higher, usually have round rotors with two or four field poles. Separate damper windings or amortisseurs are equipped with salient rotors whilst round rotors do not

have such circuits since the solid steel rotors themselves have equivalent effects.

A symbolic representation of a two-pole round rotor synchronous machine is illustrated in Figure 2-1.

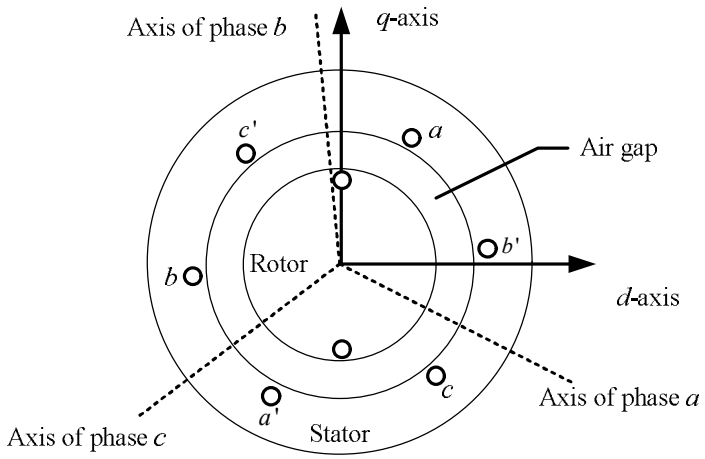


Figure 2-1: Symbolic representation of a two-pole round rotor synchronous machine.

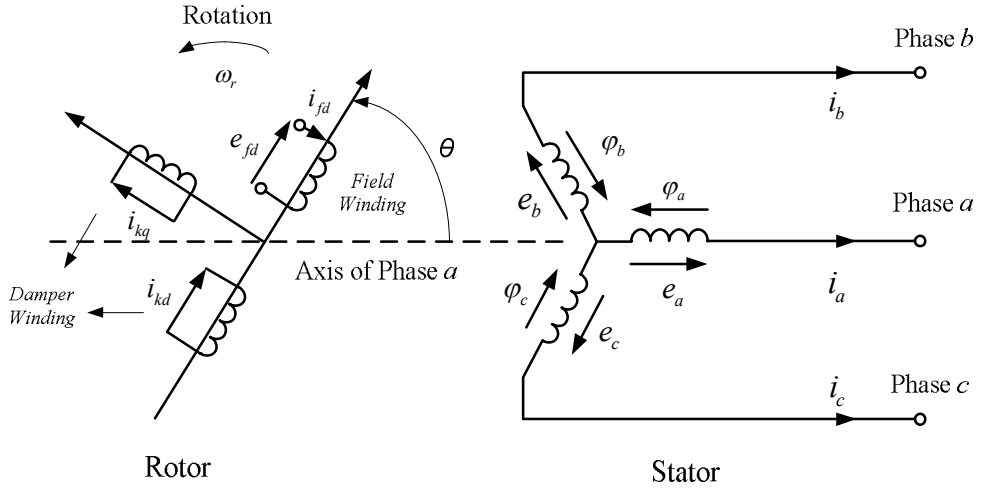


Figure 2-2: Stator and rotor circuits of a synchronous machine.

The circuits involved in a synchronous generator are shown in Figure 2-2, where *a*, *b*, *c* represents stator phase winding. *fd* represents filed winding. *kd* and *kq* are *d*-axis and *q*-axis amortisseur circuit, respectively, whilst *k* is the index of amortisseur circuits. θ is the angle by which *d*-axis leads the centreline of phase *a* winding in electrical radians. ω_r is the rotor angular velocity in electrical rad/s. The electrical

performance of a generator can be developed based on the dynamic equations of the coupled circuits. Park's transformation [1] is used to refer the stator variables to the rotor side with only *d*-axis and *q*-axis.

The mathematical descriptions (i.e., the state equations) with varying state variables and various assumptions for model simplifications for a synchronous generator can be found in [1, 4, 107]. Within this thesis, a sixth order model is used. The first order differential equations for it are given by (2.1) to (2.6).

$$\frac{d}{dt}E_d' = \frac{1}{T_{qo}'}\left[-E_d' + \left(X_q - X_q'\right)\left\{I_q - \frac{X_q' - X_q''}{\left(X_q' - X_{ls}\right)^2}\left(\psi_{2q} + \left(X_q' - X_{ls}\right)I_q + E_d'\right)\right\}\right] \tag{2.1}$$

$$\frac{d}{dt}E_q' = \frac{1}{T_{do}'}\left[-E_q' - \left(X_d - X_d'\right)\left\{I_d - \frac{X_d' - X_d''}{\left(X_d' - X_{ls}\right)^2}\left(\psi_{1d} + \left(X_d' - X_{ls}\right)I_d - E_q'\right)\right\} + E_{fd}\right] \tag{2.2}$$

$$\frac{d}{dt}\psi_{1d} = \frac{1}{T_{do}''}\left[-\psi_{1d} + E_q' - \left(X_d' - X_{ls}\right)I_d\right] \tag{2.3}$$

$$\frac{d}{dt}\psi_{2q} = \frac{1}{T_{qo}''}\left[-\psi_{2q} - E_d' - \left(X_q' - X_{ls}\right)I_q\right] \tag{2.4}$$

$$\frac{d}{dt}\Delta\omega_r = \frac{1}{2H}\left[P_m - P_e - D\Delta\omega_r\right] \tag{2.5}$$

$$\frac{d}{dt}\delta = \Delta\omega_r = \left(\omega_r - \omega_{syn}\right) \tag{2.6}$$

The algebraic equations are given by (2.7) to (2.10), assuming that the generator armature resistance is negligible.

$$E_d = \frac{X_q'' - X_{ls}}{X_q' - X_{ls}}E_d' - \frac{X_q' - X_q''}{X_q' - X_{ls}}\psi_{2q} + X_q''I_q \tag{2.7}$$

$$E_q = \frac{X_d'' - X_{ls}}{X_d' - X_{ls}} E_q' + \frac{X_d' - X_d''}{X_d' - X_{ls}} \psi_{1d} - X_d'' I_d \tag{2.8}$$

$$P_e = E_q I_q + E_d I_d \tag{2.9}$$

$$E_t = \sqrt{E_q^2 + E_d^2} \tag{2.10}$$

The detailed derivation of all the above equations is included in [107]. The electrical quantities and torques are in p.u.; time in seconds; rotor angle in electrical radians and rotor speeds in electrical rad/s. The subscript d and q indicate d- and q- axis. E_{fd} is the field voltage. $E_d^{'}$ and $E_q^{'}$ are the transient stator emfs. ψ_{1d} and ψ_{2q} are the rotor circuit flux linkage. X_d and X_q are the synchronous reactance. $X_d^{'}$ and $X_q^{'}$ are the transient reactance whilst $X_d^{''}$ and $X_q^{''}$ are the sub-transient reactance. $T_{d0}^{'}$ and $T_{q0}^{'}$ are the transient open circuit time constant whilst $T_{d0}^{''}$ and $T_{q0}^{''}$ are the sub-transient open circuit time constant. X_{ls} is the stator leakage reactance. P_m and P_e are the mechanical power and electrical real power, respectively. D is the damping factor. H is the inertia constant. δ is the rotor angle. ω_r and ω_{syn} are the absolute and synchronous speed of rotor and therefore $\Delta\omega_r$ is the rotor relative speed. I_d and I_q are the stator currents. E_d and E_q are the stator voltage and E_t is the combined value of them.

2.2.2 Generator Excitation Systems

The function of an excitation system is to provide direct current to synchronous machine field winding. At the same time, by controlling the field voltage and thereby the field current, the excitation system should contribute to the effective control of voltage and enhancement of system stability [1]. Through the Automatic Voltage

Regulator (AVR), the filed voltage is manipulated in order for the generator stator terminal voltage, E_t , to reach its reference set-point, E_t^{ref} , and to ensure the first-swing stability of the machine. In addition, a Power System Stabiliser (PSS) may also be included to provide an additional input signal to the regulator to damp power system oscillations.

The functional relationship between the synchronous generator, excitation system, and PSS (if included) is illustrated in Figure 2-3.

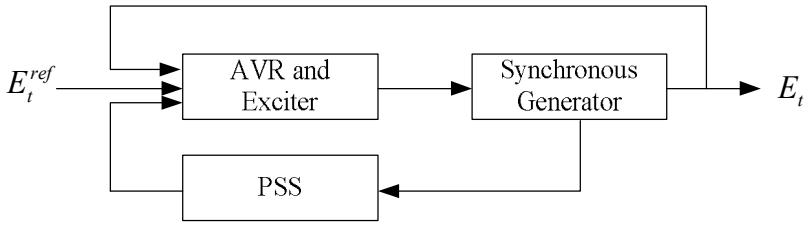


Figure 2-3: Relationship and signals between the synchronous generator, excitation system and power system stabiliser.

Many forms of excitation systems are used in practice. Description of them can be found in [108]. Three different types of excitation system used in this thesis are described.

2.2.2.1 Manual Excitation

Manual excitation is the simplest excitation scheme. No AVR is included in this scheme and the field voltage is maintained at a constant value determined through the synchronous generator parameter initialisation. As a result, the generator terminal voltage may deviate from the desired value if operating conditions change.

2.2.2.2 Static Excitation (IEEE Type ST1A)

Static excitation systems directly supply the field winding of synchronous generators through rectifiers which are fed by either transformers or auxiliary machine windings

[108]. The simplified version of the IEEE Type ST1A static exciter used in this thesis is shown in Figure 2-4. It only consists of the voltage transducer delay block and the exciter block with no time constant. No transient gain reduction block is included. The signal E_{PSS} is a stabilising signal from the PSS if one is used with the exciter. $E_{fd}^{\max}$ and $E_{fd}^{\min}$ are the upper and lower limit of filed voltage E_{fd}, respectively.

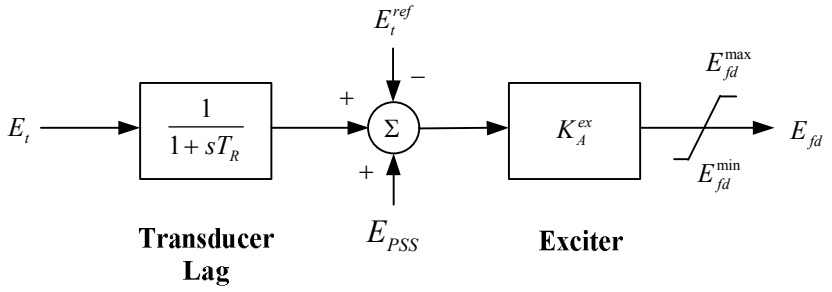


Figure 2-4: Simplified block diagram for the IEEE Type ST1A static exciter.

2.2.2.3 DC Excitation (IEEE Type DC1A)

The excitation systems which use a DC current generator and commutator are referred to as DC exciters. Typically, they respond more slowly than static excitation systems [108]. The block diagram of a simplified version of the IEEE Type DC1A excitation system used in this thesis is illustrated in Figure 2-5. $E_{ex}^{\max}$ and $E_{ex}^{\min}$ are the upper and lower limit of the exciter input voltage, respectively.

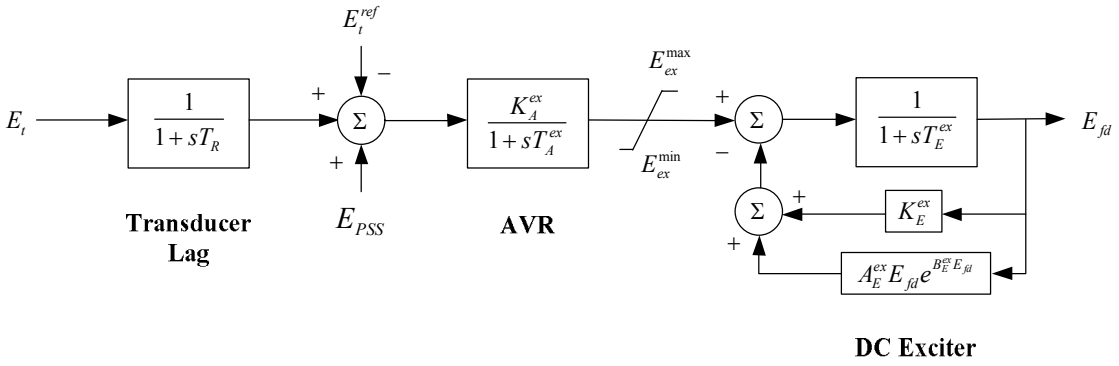


Figure 2-5: Simplified block diagram for the IEEE Type DC1A DC exciter.

2.2.3 Power System Stabilisers

A power system stabiliser adds damping to the generator oscillations by controlling the generator's excitation with supplementary stabilising signal(s). The most commonly and logically used signal to monitor generator rotor oscillations is the rotor speed deviation $\Delta\omega_r$, and it is therefore adopted in this thesis.

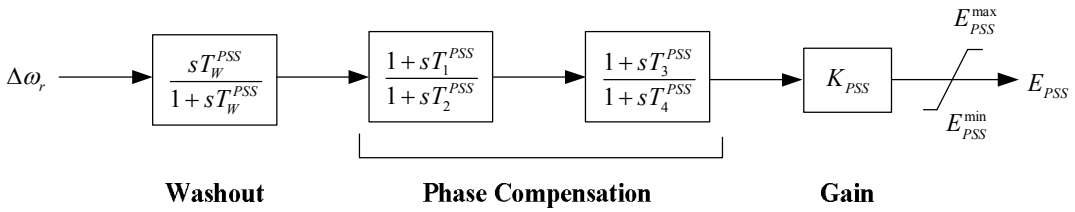


Figure 2-6: Block diagram of a PSS.

The block diagram of PSS is illustrated in Figure 2-6. Since the electrical damping torque component introduced by PSS must be in phase with the rotor speed deviation due to the phase characteristics of the excitation system, the PSS must include suitable phase compensation blocks. A number of phase lead/lag blocks are needed together with a washout filter to ignore steady state changes and a gain to maximise damping. A low-pass filter may also be required to reduce the high frequency output of the PSS so that potential interactions with the torsional mechanical modes of larger steam turbines can be avoided [1]. Within this research, no mechanical system is modelled so that there is no requirement to include the low-pass filter. E_{PSS}^{max} and E_{PSS}^{min} are the upper and lower limit of the output signal.

Although PSS is tuned with respect to the damping of small-signal oscillations (using linearised power system model), it also improves system damping under large-disturbances (if properly tuned). However, the use of PSS does not improve transient stability. After a fault occurs it is the AVR that improves the transient stability by forcing the generator excitation voltage to its upper limit. This serves to maximise

generator load torque and minimise rotor angle swing. Over this crucial during-fault and post-fault period the PSS occasionally acts in opposition to the AVR (to maintain the generation of damping signal in phase with speed deviation), so that when a PSS is employed the field voltage gets off the upper limit value earlier than in the case for the AVR alone. Hence, to avoid reducing transient stability during fault and immediate post-fault period, a PSS signal needs to be suitably limited to ensure the maximum field voltage is continuously maintained until the peak rotor angle swing is reached.

2.2.4 Transmission Lines

Throughout the work presented in this thesis, transmission lines are modelled using the common π circuit with lumped parameters as shown in Figure 2-7 [1]. The length of all lines are assumed to be short enough so that this model is applicable and that more complex π section or distributed parameter representation is not required [109]. V_S and I_S are the voltage and current at the sending end of the line, whilst V_R and I_R are voltage and current at the receiving end. Z represents the series impedance of the line and Y is the total line admittance.

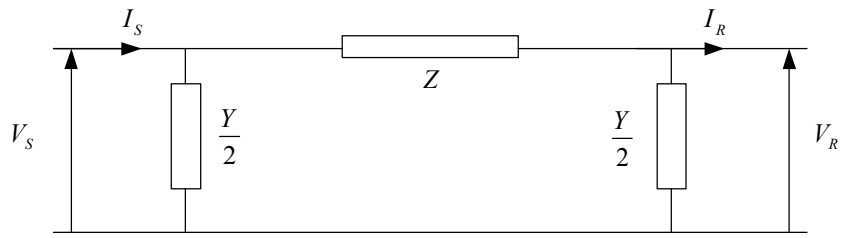


Figure 2-7: Equivalent π circuit of a transmission line [1].

2.2.5 Transformers

An equivalent π circuit of a two-winding transformer is used. As shown in Figure 2-8, $Y_e = 1/Z_e$ in which Z_e is the equivalent leakage impedance of the transformer and

$c = 1/n$ where n is the off-nominal turns ratio of the transformer [1].

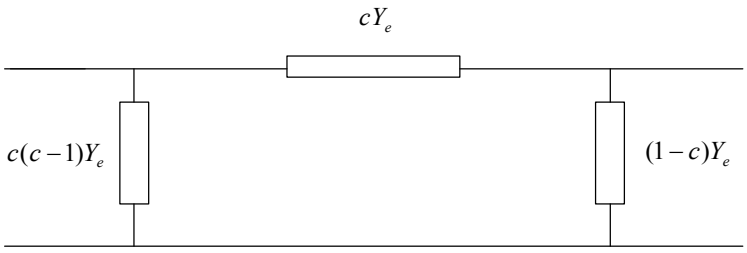


Figure 2-8: Equivalent π circuit of a two winding transformer [1].

2.2.6 Loads

Load characteristics have an important influence on system stability and the way in which power system loads are modelled can significantly affect the results of dynamic simulation [110, 111]. Since voltage magnitude and frequency across the network can be affected by electromechanical oscillations, more detailed models may be required for loads which are sensitive to these changes and so accurate results can be ensured. A summary and examples of load modelling for dynamic performance analysis can be found in [110, 111].

In this thesis, a constant impedance load model is used, represented as shunt admittance Y_i^{load} connected to the i^{th} load bus, as in (2.11) where P_i^{load} and Q_i^{load} are the active and reactive power component of the load and V_i is the bus voltage. Within the time frame of transient stability simulation, the elements of the node admittance matrix of a transmission network are constant except for changes introduced by network switching operations. Load with constant impedance characteristics are included in the node admittance matrix and therefore the simplest to handle. However, the modelling of different types and patterns of loads can be included in the future.

$$Y_i^{load} = \frac{P_i^{load} - jQ_i^{load}}{V_i^2} \tag{2.11}$$

2.2.7 Network

The whole power system network is modelled by combining all transmission lines, transformers and constant impedance loads. The nodal network equation for a network with *N* buses, which describes the relationship between system voltage $\boldsymbol{V}$ and points of current injection $\boldsymbol{I}$ [4], is shown in (2.12),

$$\begin{bmatrix} I_1 \\ \vdots \\ I_i \\ \vdots \\ I_N \end{bmatrix} = \begin{bmatrix} Y_{11} & \cdots & Y_{1i} & \cdots & Y_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_{i1} & \cdots & Y_{ii} & \cdots & Y_{iN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_{N1} & \cdots & Y_{Ni} & \cdots & Y_{NN} \end{bmatrix} \begin{bmatrix} V_1 \\ \vdots \\ V_i \\ \vdots \\ V_N \end{bmatrix} \text{ or } \boldsymbol{I} = \boldsymbol{YV} \tag{2.12}$$

where subscripts *i* and *j* are bus numbers such that Y_{ii} is the self-admittance at bus *i*, and Y_{ij} is the mutual-admittance between bus *i* and *j*.

The network model can be reduced by neglecting all zero-injection buses, and consequently the order of the nodal network equation becomes much smaller [107]. This lowers the computational burden during simulations and power system analysis.

Furthermore, a coordinate transformation is applied between each individual machine model reference frame (*d*-*q*), which rotates at its rotor speed ω_r, and the network common system reference frame (*D*-*Q*), which rotates at synchronous speed ω_{syn}. Figure 2-9 shows the relative position of the two coordinates offset by the rotor angle δ. V_D, V_Q and V_d, V_q are the direct and quadrature component of the generator terminal voltage V_t according to individual machine model reference frame and the network common system reference frame, respectively. Equations (2.13) and (2.14) are used to complete the transformations. Similar transformations can be applied for system current injections.

$$\begin{bmatrix} V_D \\ V_Q \end{bmatrix} = \begin{bmatrix} \cos\delta & -\sin\delta \\ \sin\delta & \cos\delta \end{bmatrix} \begin{bmatrix} V_d \\ V_q \end{bmatrix} \tag{2.13}$$

$$\begin{bmatrix} V_d \\ V_q \end{bmatrix} = \begin{bmatrix} \cos\delta & \sin\delta \\ -\sin\delta & \cos\delta \end{bmatrix} \begin{bmatrix} V_D \\ V_Q \end{bmatrix} \tag{2.14}$$

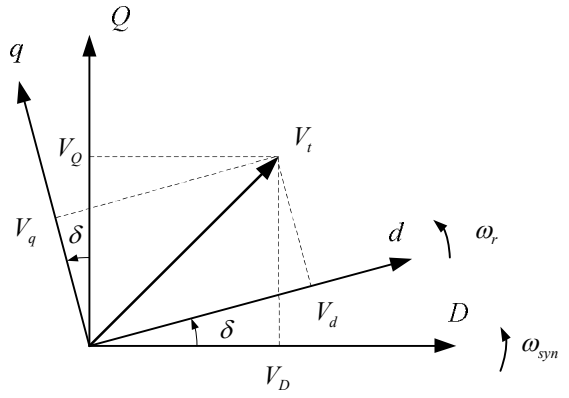


Figure 2-9: Relative position of the individual machine model reference frame (*d-q*) with respect to the network common system reference frame (*D-Q*) [4].

2.3 Modelling Different Types of Faults

With a single phase representation of the network, the use of symmetrical components allows any type of fault, including balanced three-phase, unbalanced single line to ground, line to line, and line to line to ground, to be modelled by a fault shunt impedance Z_F connected between the point of the fault and the neutral. The value of Z_F depends on the type of fault and is given in Table 2-1 [1, 4]. When assessing the non-linear dynamic responses of the network, these are simulated by adding a shunt admittance to the self-admittance Y_{ii} of the faulted bus *i* in the nodal admittance matrix. A value of 10^9 is used within this thesis for the Y_{ii} of three-phase fault.

In Table 2-1, Z_2 and Z_0 are the negative and zero sequence Thevenin equivalent impedance of the network, respectively, as seen from the fault terminal. They are

calculated using a graph theory based algorithm as in [112]. The negative sequence network has the same topology as the positive one whilst the zero-sequence network topology is different and constructed according to the transformer winding connections. In this thesis, only the Y-Y connection of transformer windings with both neutrals grounded is used, the equivalent circuit of which is shown in Figure 2-10. Since both neutrals are grounded, there is a path for the zero sequence current to flow in the primary and secondary, and the transformer is represented as the equivalent leakage impedance per phase.

Table 2-1: Shunt impedance and admittance representing different types of faults.

Fault Type	*Fault Shunt Impedance* (Z_F)	*Fault Shunt Admittance* (Y_{ii})
Three Phase	0	10^9
Single Line to Ground	$Z_2 + Z_0$	$\frac{1}{Z_2 + Z_0}$
Line to Line	Z_2	$\frac{1}{Z_2}$
Line to Line to Ground	$\frac{Z_2 Z_0}{Z_2 + Z_0}$	$\frac{Z_2 + Z_0}{Z_2 Z_0}$

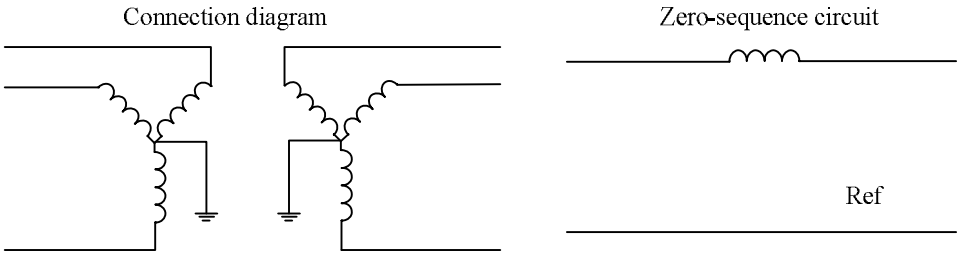


Figure 2-10: Zero-sequence equivalent of Y-Y connected two-winding transformer with both neutral grounded.

2.4 Simulation Tools

Throughout this thesis, all modelling has been completed using MATLAB/Simulink

environment (version 8.2.0.701 R2013b) with direct implementation of the mathematical component descriptions provided within this chapter. For all transient stability simulations, initial load flows are performed using either standard power flow (in Chapter 4) or Optimal Power Flow (OPF) (in Chapter 5 to 7) functions in MATPOWER [113]. Dynamic simulations are carried out with the variable-step type ode23s (stiff/Mod. Rosenbrock) solver in Simulink, which utilises the implicit numerical method based on a second-order modified Rosenbrock formula [114].

2.5 Test Network

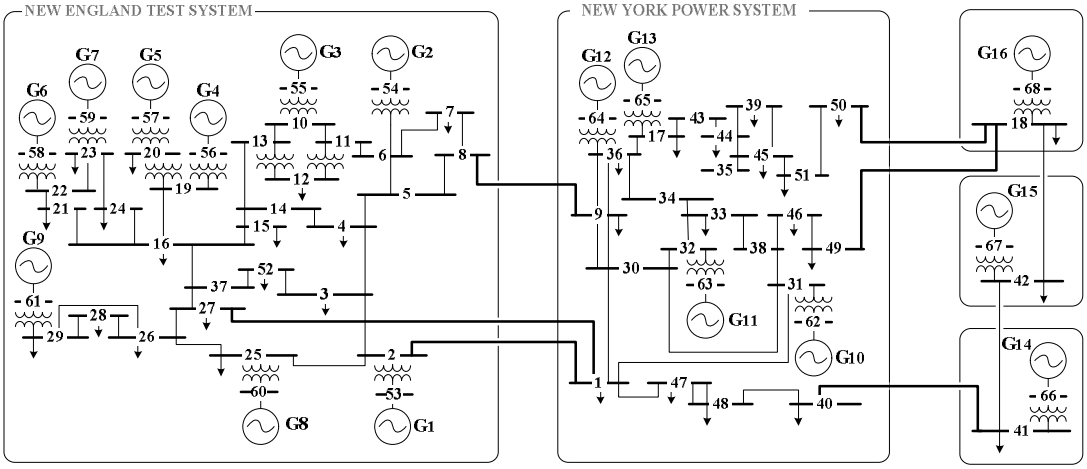


Figure 2-11: NETS-NYPS five-area test network diagram.

A large 16-machine, 68-bus, 86-branch system is utilised throughout this thesis to demonstrate the on-line identification of power system dynamic signature using PMU measurements and data mining. The test system is shown in Figure 2-11 and was originally introduced in [115]. The network represents a reduced order equivalent model of the New England Test System (NETS) and the New York Power System (NYPS). There are five distinct areas in this system: NETS is formed by G1 to G9, NYPS consists of G10 to G13, and three other neighbouring areas are represented by G14, G15 and G16, respectively. The inter-area ties between different areas are highlighted.

All generators are represented by full sixth order models. G1 to G8 are under slow DC excitation (IEEE-DC1A) whilst G9 is equipped with a fast static exciter (IEEE-ST1A) and a PSS [116]. All the remaining generators (G10-G16) use constant manual excitation. Power system loads are modelled as constant impedance.

All system details including line parameters, standard loading, dynamic machine parameters and OPF data are included in Appendix A.

2.5.1 WAMS in the Test Network

Since G13 is modelled as the reference generator with a constant value of rotor angle, it represents an external system which is not of concern for this study. Throughout the thesis, it is assumed that all the other 15 generators in the test network have a PMU and other required measurement devices installed so that generator rotor angles can be directly calibrated in real time from the terminal voltage and rotor position (monitored by optical or magnetic means) without any electrical calculations.

2.6 Summary

This chapter has presented the fundamental modelling and analysis tools which will be used throughout this thesis. Only essential information is given due to wide availability of relevant references books.

The chapter began by presenting the mathematical models of the relevant components of the power systems. These included synchronous generators and their associated controls (excitation systems and PSS), transmission lines, transformers and loads. An explanation was then given as to how different types of faults, including three-phase, single line to ground, line to line, and double line to ground fault, are represented. The simulation tools described after that are used throughout the thesis for all simulations presented in the thesis. Finally, the test network used throughout this

research was introduced, alongside the assumption related to availability and location of PMUs in WAMS.

3 Data Mining Techniques

3.1 Introduction

Data mining is used to describe the computational process of discovering information from large data sets and transforming it into an understandable structure for future uses. The term is a buzzword, and is frequently generalised to any kind of computer decision support system including artificial intelligence, machine learning, statistics, etc. It has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. For instance, firms may commonly apply data mining to business data, to describe, predict, and improve their business performance. In this thesis, the term data mining is used to emphasise that the methodology, for on-line identification of power system dynamic signature based on PMU measurements, focuses on analysing quantities of offline data for predictive purposes.

Two types of data mining techniques will be used in this thesis, namely classification and clustering. Classification falls into the category of supervised learning. The aim of it is to create a predictive model which can map a new object to one of a set of

classes, on the basis of a training dataset containing both objects and related decisions. Clustering, on the other hand, belongs to the category of unsupervised learning. It applies when the objects are not pre-labelled but are to be divided into natural groups. The purpose of clustering techniques is to split a dataset into clusters by maximising some measure of dissimilarity between objects.

This chapter presents the technical background of both classification and clustering. For each of these two, it starts off by introducing the basic concepts, and then describes the specific algorithms that will be used in this thesis, including Decision Tree (DT), Ensemble Decision Tree (EDT), and Support Vector Machine (SVM) for classification, and Hierarchical Clustering (HC) for clustering. The software packages used in this research to implement these algorithms are also introduced.

3.2 Classification

A bank loan officer needs analysis of his or her data to learn which loan applicants are "safe" and which are "risky". A marketing manager at an electronics company needs data analysis to help guess whether a customer with a given profile will buy a new computer. A medical researcher wants to analyse breast cancer data to predict which of the three specific treatments a patient should receive. In the scope of this research, the power system operator needs analysis of the off-line simulated contingency database to predict the transient stability of the power system in real-time. In each of these examples, the data mining task is classification, where a model or classifier is constructed to predict class labels, such as "safe" or "risky" for the loan application data; "yes" or "no" for the marketing data; "treatment A", "treatment B", or "treatment C" for the medical data, or "stable" or "unstable" for the power system dynamic data.

Many classification methods have been proposed by researchers in data mining. The general idea of them will be firstly described in this section. The methods of DT, EDT and SVM will then be introduced. The topic of multiclass classification will also be discussed.

3.2.1 Basic Concepts

The general approach to classification is a two-step process, consisting of a learning or training step where a classification model or classifier is constructed, and a classification step in which the model is used to predict class labels for given input.

In the first step, a classification algorithm builds a model to describe a predetermined training database which is made up of a set of objects and their associated class labels. Within the training database, an object, *X*, is represented by an *n*-dimensional attribute vector $x_1, x_2, \cdots, x_n$. The attributes are typically referred to as "predictors". Each *X* is assumed to belong to a predetermined class. The class label *y* is categorical (or nominal) and usually referred to as the "target" of classification. The process of supervised learning can be viewed as the construction of a model or a function, $y = f(X)$, which can predict the associate class label *y* of a given object *X*.

In the second step, the model is used to classify new objects which are not used during the learning process. For a given test set, the accuracy of classification is the percentage of the test set objects that are correctly classified.

In the bank loan example as previously described (adopted from [117]), within the database of the officer, the objects are the loan applications. The attributes of them (i.e., the predictors) are the age and income of the applicants. The learning process for the officer is to build a model to classify the applicants in the database into the target of the loan decision (i.e., "safe" or "risky").

In this thesis, the objects in the training database are the contingencies simulated for a power system under various operating conditions and disturbances. The attributes of them (i.e., predictors) are the post-fault system parameters such as generator rotor angles and speeds. The class labels (i.e., targets) are either the transient stability status of the post-fault system, or the pattern of the post-fault system behaviour.

The key advantage of classification algorithms that makes them particularly relevant in on-line dynamic security assessment of power systems is their ability to discover, comprehend and generalise relationships between predictors and targets which may not be immediately obvious to a human operative. Furthermore, classification algorithms are also able to identify relationships across a large number of dimensions.

3.2.2 Decision Tree

DT is a type of high dimensional classifier which has been successfully applied to a broad range of tasks in various disciplines such as medicine, financial analysis, manufacturing and production, and molecular biology [117]. A DT is composed of nodes, branches and terminal decisions. Each internal node specifies a test of a predictor, which splits the set of examples within that node into subsets. Each branch descending from a node corresponds to one of the possible outcomes of this test. The leaf nodes with terminal decisions represent the classes to which input examples belong. Essentially, a path from the root to a leaf in a DT is a series of if-then rules, which makes the internal logic and reasoning process behind the model easy to comprehend.

A typical DT which classifies input contingency into two classes is shown in Figure 3-1. It represents the concept of Transient Stability Assessment (TSA), indicating whether the transient stability status of a power system is stable or not. In each of the

internal node, denoted by rectangle, the values of predictors (generator rotor angles in degree in this illustrative example) are compared to their corresponding thresholds. Each of the leaf nodes, denoted by oval, represents a class to which the post-fault power system transient stability status belongs (either stable or unstable). Given a contingency in which the values of rotor angle 1, 2, 4 are 54.5°, 175.8° and 60.1°, respectively, a path indicated with the dashed arrows in Figure 3-1 is traced from the root to the leaf node which holds the stable class.

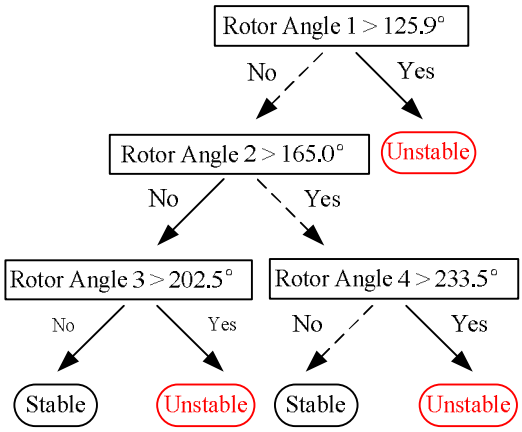


Figure 3-1: A simple decision tree for illustration purposes.

As shown in the example, the representation of acquired knowledge in the tree form is intuitive and easy to assimilate by humans. The learning and classification steps of DT are simple and fast. In general, DT classifiers have good accuracy [117].

3.2.2.1 Tree Building Algorithms

There are many different algorithms to build DTs. They are all similar in the way that the training dataset is recursively split into smaller and smaller groups in a top-down manner, by selecting the most useful predictor to test at each node, until splitting no longer adds value to the predictions [118, 119]. The differences among them are the ways in which they determine the most useful predictor to test (the quantitative measure of predictor selection) at each node.

Figure 3-2 summarises the basic algorithm named **Generate_DT**. The input to the algorithm includes three parameters: *D*, *predictor_list*, and *predictor_selection_method*. *D* is the given data set for training. *predictor_list* is the list of predictors describing the objects within the data set. *predictor_selection_method* specifies a procedure for selecting the predictor that

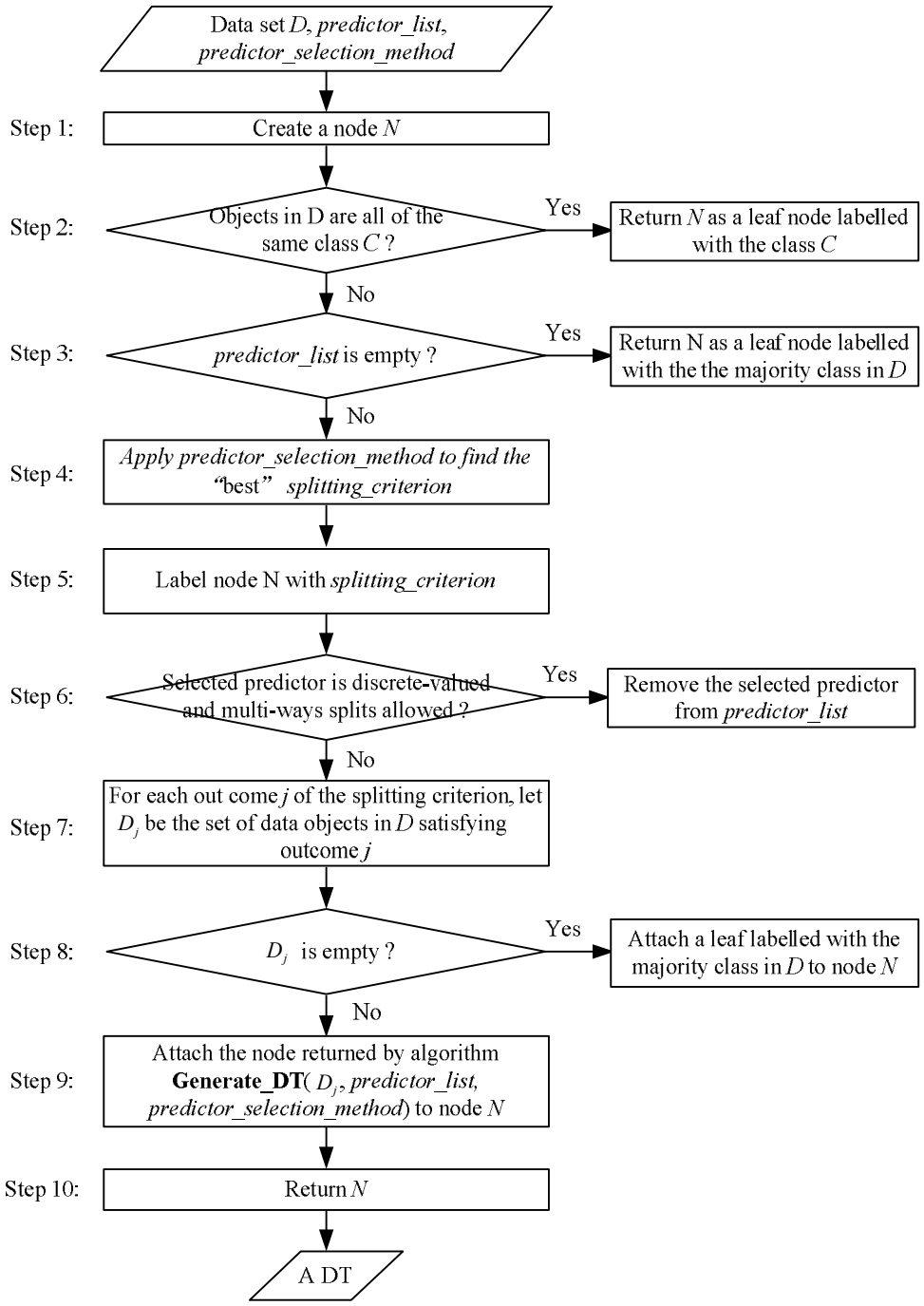


Figure 3-2: Basic algorithm **Generate_DT** for building a decision tree from training objects.

"best" discriminates the given objects according to class. The tree starts as a single node *N*, representing the training objects in *D* (Step 1). If the objects in *D* are all of the same class, the node *N* becomes a leaf and is labelled with the class (Step 2). Note that Step 3 is a terminating condition which will be explained later. Otherwise, the algorithm calls *predictor_selection_method* to determine the splitting criterion (Step 4). The splitting criterion indicates which predictor to test at node *N* and which branches to grow from node *N* with respect to the outcomes of the chosen test. It is determined so that, ideally, the resulting child nodes are as "pure" as possible. After that, the node *N* is labelled with the splitting criterion, which serves as a test at this node (Step 5). A branch is grown from the node for each of the outcomes of the splitting. The objects in *D* are partitioned accordingly (Step 7). There are three possibilities for partitioning objects based on the splitting criterion, as illustrated in Figure 3-3 with examples. Let *P* be the selected predictor which has *v* distinct values, $p_1, p_2, \cdots, p_j, \cdots, p_v$.

- As in Figure 3-3 (a), if *P* is discrete-valued and multi-way splits are allowed (not restricted to binary tree), for instance when the income of the applicants in the bank loan example have three levels: low, medium and high, one branch is grown for each known value, p_j, of *P*. Partition D_j is the subset of objects in *D* having value p_j of *P*. Since all the objects in a given partition have the same *P* value, the predictor *P* does not need to be considered in any future partitioning. It is therefore removed from *predictor_list* (Step 6).
- As in Figure 3-3 (b), if *P* is continuous-valued, for instance when the income is represented by numerical values, the test at node *N* is of the form "*P* > *Threshold*?". Two branches are grown from *N*. The objects are partitioned such

that D_1 holds the subset in which the predictor exceeds its threshold, whilst D_2 holds the rest. In the context of this research, the predictors of generator rotor angles and speeds are also continuous-valued and therefore the partition of contingencies follows this scenario.

- As in Figure 3-3 (c), if P is discrete-valued and a binary tree must be produced, i.e., each node must have exactly two outgoing branches, the test at node N is in the form "$P \in S_P$?", where S_P is the subset of the known value of P. The objects are partitioned such that D_1 holds the subset that satisfies the test whilst D_2 holds the subset that does not satisfy the test.

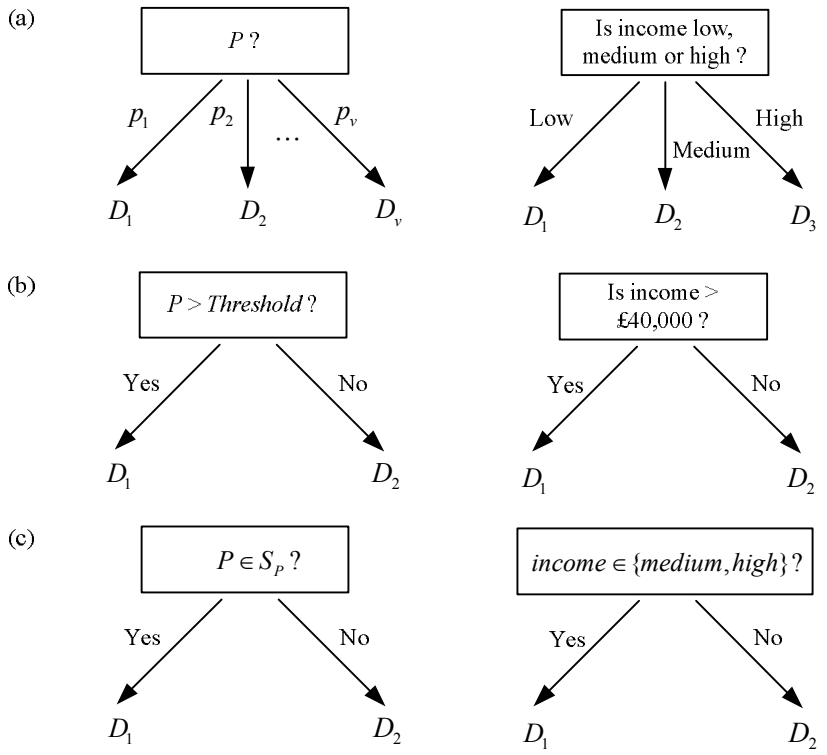


Figure 3-3: Three possibilities for partitioning objects based on the splitting criterion, each with examples.

The algorithm uses the same process recursively to form a DT for the objects at each resulting partition D_j (Step 9). The recursive partition stops only when one of the three terminating conditions is true: i) All the objects in partition D belong to the

same class (Step 2); ii) There are no remaining predictors on which the objects may be further partitioned (Step 3). In this case, node *N* is converted into a leaf and labelled with the most common class in *D*; and iii) There are no objects for a given branch D_j, in which case a leaf is created with the majority class in *D* (Step 8). Finally, the resulting DT model is returned (Step 10).

Commonly used tree building algorithms include Iterative Dichotomiser 3 (ID3), C4.5 (a successor of ID3), C5.0 (a successor of C4.5) and Classification And Regression Tree (CART), etc. Short descriptions of each of those are given below whilst numerical examples can be found in [117].

1) ID 3

ID 3, developed during the late 1970s and early 1980s by J. R. Quinlan [117], uses *information gain* as its predictor selection measure. This measure is based on the information theory, which studies the value or "information content" of messages.

Let node *N* represents the data set *D* in which *m* distinct classes of objects can be found. The expected information needed to classify an object in *D* is given by (3.1), where p_i is the probability that an object in *D* belongs to class C_i (for $i = 1, \cdots, m$) and is estimated by $\left|C_{i,D}\right| / \left|D\right|$. The log function to base 2 is used since the information is encoded in bits. *Info(D)* is the average amount of information needed to identify the class label of an object in *D*, also known as the *entropy* of *D*.

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{3.1}$$

Supposing the objects in *D* are split based on some predictor *P* having *v* distinct values, $p_1, p_2, \cdots, p_v$. If *P* is discrete-valued, it can be used to split *D* directly into *v*

partitions $D_1, D_2, \cdots, D_v$, where D_j contains the objects in D that have the value p_j of P. The information required to classify an object from D based on the partition by P is given in (3.2). The term $|D_j|/|D|$ acts as the weight of the j^{th} partition.

$$Info_P(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j) \tag{3.2}$$

Information gain is defined as the difference between the original information required (i.e., based on just the proportion of classes) and the new requirement (i.e., after partitioning on P), as in (3.3). The information gain therefore is the expected reduction in the information requirement caused by knowing the value of P. The predictor with the highest information gain is selected as the splitting predictor at node N. This predictor minimises the information needed to classify the objects in the resulting partitions and reflects the highest "purity" in these partitions. Such an approach minimises the expected number of tests needed to classify a given object.

$$Gain(P) = Info(D) - Info_P(D) \tag{3.3}$$

If, however, the predictor P is continuous-valued as the generator rotor angles and speeds in this research are, the threshold of P needs to be determined. Typically, the values of P are sorted in increasing order, and the midpoint between each pair of adjacent values is considered as a possible threshold. For every possible threshold for P, $Info_P(D)$ is evaluated. The point with the minimum value of $Info_P(D)$ is selected as the threshold of P.

2) C4.5

The information gain measure is biased toward tests with many outcomes (i.e., it prefers to select predictors having a large number of values). To overcome this bias,

C4.5, a successor of ID 3, uses an extension to information gain known as gain ratio. It applies normalisation to the information gain using "split information" defined as in (3.4), which represents the potential information generated by splitting the training data set D into v partitions. The gain ratio is defined as (3.5). The predictor with the maximum gain ratio is selected as the splitting predictor.

$$SplitInfo_P(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2(\frac{|D_j|}{|D|}) \tag{3.4}$$

$$GainRatio(P) = \frac{Gain(P)}{SplitInfo_P(D)} \tag{3.5}$$

3) C5.0

Although C4.5 has been a benchmark to which newer classification algorithms are often compared, in recent years, its successor C5.0 is available commercially. The tree building algorithm seems to be essentially the same as that used by C4.5. However, its training process is greatly sped up and clearly applies some different techniques, although this has not been described in the open literature [120].

4) CART

Classification And Regression Tree (CART) was developed in 1984 and has been one of the most popular algorithms currently applied in practice [105, 117, 119]. It only constructs binary trees. The criterion used to select predictors is based on "Gini Index" [105, 117, 119].

Using the notation previously described, the Gini Index measures the impurity of D, a data partition or set of training objects, as (3.6), where p_i is the probability that an object in D belongs to class C_i and is estimated by $|C_{i,D}|/|D|$. The sum is computed over m classes.

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2 \tag{3.6}$$

The Gini Index considers a binary split for each predictor. If P is a discrete-valued predictor, to determine the best binary split on P, all the possible subsets that can be formed using known values of P are examined. Each subset, S_P, can be considered as a binary test for predictor P of the form "$P \in S_P$?". A weighted sum of the impurity of each of the two resulting partitions is computed. If a binary split on predictor P splits D into D_1 and D_2, the Gini Index of D, due to that split, is calculated with (3.7). For each predictor, each of the possible binary splits is considered. For continuous-valued predictor, the strategy which is similar to that described earlier for information gain is used to determine the threshold.

$$Gini_P(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \tag{3.7}$$

The reduction in impurity that is incurred by a binary split on a predictor P is given in (3.8). The predictor that maximises the reduction (or, equivalently, has the minimum Gini Index) is selected as the splitting predictor.

$$\Delta Gini(P) = Gini(D) - Gini_P(D) \tag{3.8}$$

In the rest of the thesis, CART and C5.0 (as the most recent version of the algorithm in the series of ID 3, C4.5 and C5.0) will be used for the training of DT.

3.2.2.2 Overfitting

When building the tree, it is usually expected to train the model based on the known dataset to a certain extent so that it will also be able to predict the correct output for unknown input. However, overfitting usually occurs when the model is excessively

complex, such as having too many branches and too many leaf nodes. Although a model like this classifies the existing dataset perfectly, it may have poor prediction performance for unknown inputs. The main reason is the fact that the training dataset may have some random error or noise that cannot describe the underlying relationship between the input and output. Therefore some of the generated nodes and branches may have no causal relation to the target function of the tree.

There are two main approaches that can be used to avoid overfitting. One approach is to stop growing the tree earlier, i.e., before it perfectly classifies the training set. The commonly used stopping criteria include maximum tree depth and minimum records in parent and child branch which prevent a split if the number (or percentage) of records in the node to be split is less than the specified value. Another approach, known as pruning, allows full development of the tree and then removal of the bottom-level splits that do not contribute significantly to the accuracy of the tree. Details about pruning methodologies can be found in [105].

3.2.3 Ensemble Decision Tree

Ensemble methods are frequently used in data mining to increase classification accuracy. As shown in Figure 3-4 (adopted from [117]), an ensemble combines a series of k base classifiers, M_1, M_2, …, M_k, with the aim of creating an improved composite model. A given data set D is used to create k training sets, D_1, D_2, $\cdots$, D_k where D_i $(1 \le i \le k)$ is used to build classifier M_i. With the predictors of a new object, the base classifiers each vote by returning a prediction. The ensemble returns a final class according to all the votes. The existing ensemble creation techniques include (but they are not limited to) bagging, boosting, and Random Forest (RF). Each of them is briefly described below and further details can be found in [121].

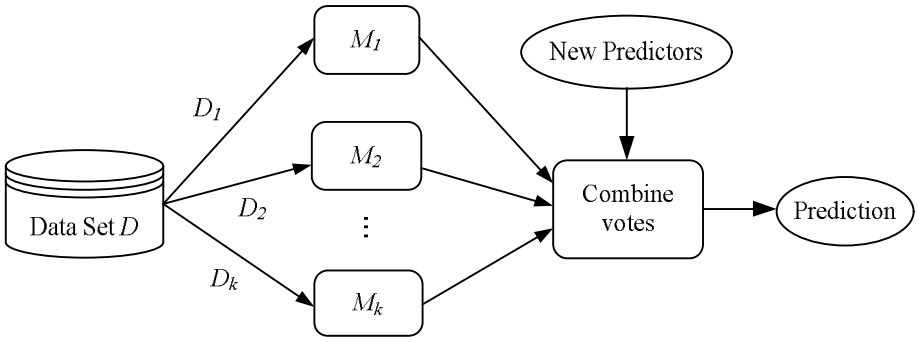


Figure 3-4: Ensemble methods [117].

Within this thesis, the classification technique used to train the base classifiers in ensembles is DT (including both CART and C5.0).

3.2.3.1 Bagging

The basic idea behind bagging can be explained by the following example [3]. When a patient would like to have a diagnosis made based on his or her symptoms, instead of asking one doctor, he or she may choose to ask several. If a certain diagnosis occurs more than any other, the patient may choose it as the final or best diagnosis. That is, the final diagnosis is made based on a majority vote, where each doctor gets an equal vote. Similarly, bagging creates an ensemble of classification models for a classification technique (such as CART and C5.0) where each model gives an equally weighted prediction.

Bagging stands for *bootstrap aggregation*. With the given data set D, of d objects, it works as follows. For iteration i ($i = 1,2,...,\mathrm{k}$), a training set, D_i, of d objects, is sampled with replacement from the original set (i.e., some of the original objects of D may not be included in D_i, whereas others may occur more than once). Each training set is called a *bootstrap* sample. A base classifier, M_i, is learned for each training set, D_i. To classify an unknown object, X, each classifier M_i returns its prediction which counts as one vote. The bagged classifier, M^*, counts the votes and assigns the class

with the most votes to *X*. The algorithm is summarised in Figure 3-5.

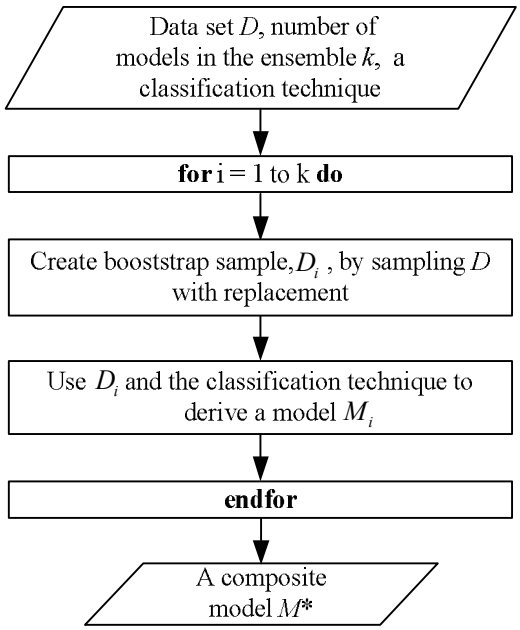


Figure 3-5: Bagging algorithm.

The bagging algorithm often has significantly greater accuracy than a single classifier derived from *D*, the original training data set and is more robust to the effects of noisy data and overfitting.

3.2.3.2 Boosting

As in the previous section, suppose that a patient who has certain symptoms chooses to consult several doctors instead of one. If the patient assigns different weight in terms of value or worth to each doctor's diagnosis based on the accuracies of previous diagnoses they have made, then the final diagnosis is a combination of weighted diagnoses. This represents the essence behind boosting.

Boosting works by building a series of classifiers in a sequence. The basic idea is that when one base classifier is built, it focuses more on the misclassified objects of the previous round. Initially, with the data set *D* which has *d* objects, each object is assigned with an equal weight of $1/\mathrm{d}$. Generating *k* classifiers for the ensemble

requires k rounds through the algorithm. In round i, d objects are sampled from D, with replacement, to form the training set D_i. Each object's chance of being selected is based on its weight. A base classifier model M_i is derived from D_i. Its error is then calculated using D_i as a test set. The weights of the original training objects are then adjusted according to how they are classified. If an object is incorrectly classified, its weight increases. If an object is correctly classified, its weight decreases. These weights will be used to generate the training samples for the base classifier in the next round. In this way, a series of classifiers that complement each other is built. The algorithm is summarised in Figure 3-6.

Within the algorithm, the error rate of model M_i is calculated by (3.9), where w_j is the weight of object X_j in D_i. $err(X_j)$ is the misclassification error of object X_j. If the object is misclassified, $err(X_j)$ is 1; otherwise it is 0. If the performance of classifier M_i is so poor that its error rate exceeds 0.5, this classifier is abandoned.

$$error(M_i) = \sum_{j=1}^{d} w_j \times err(X_j) \tag{3.9}$$

The error rate of M_i affects how the weights of the training objects are updated. If an object in round i was classified correctly, its weight is multiplied by $error(M_i)/(1-error(M_i))$. Once the weights of all the correctly classified objects are updated, the weights for all objects including the misclassified ones are normalised so that their sum remains the same as it was before. The normalisation is done by multiplying it by the sum of the old weights and dividing it by the sum of the new weights. As a result, the weights of the misclassified objects are increased and the weights of the correctly classified ones are decreased.

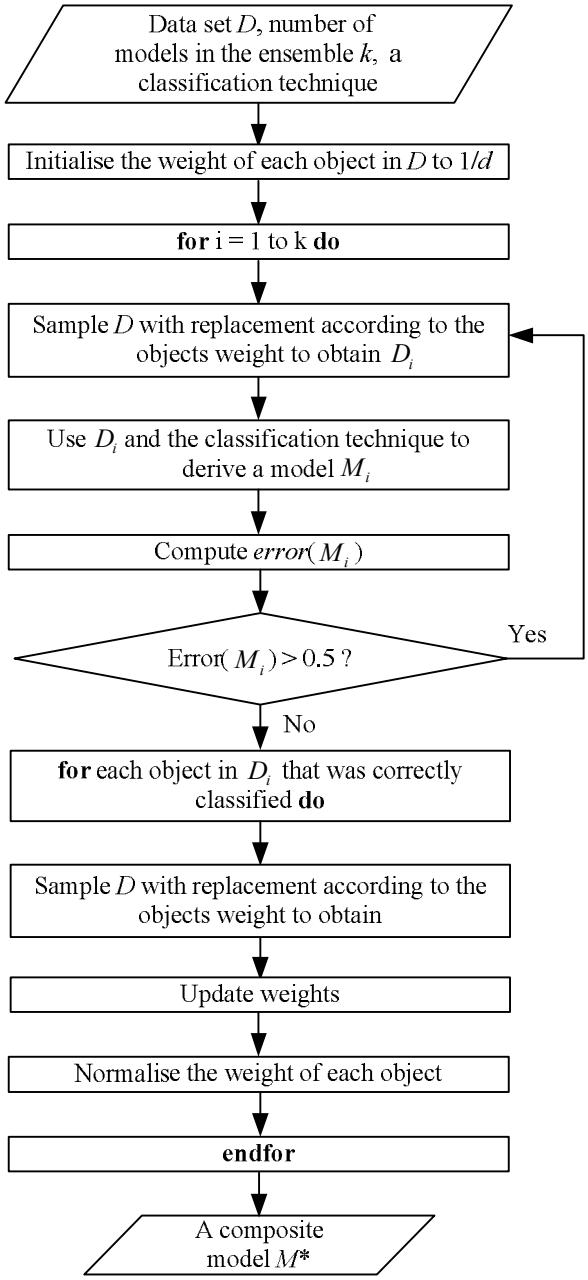


Figure 3-6: Boosting algorithm.

Finally, to classify an unknown object, X, boosting assigns a weight to each classifier's vote based on how well the classifier performed. The weight of the vote of classifier M_i is given in (3.10). The lower its error rate, the more accurate it is, and therefore the higher its weight. For each class, c, the weights of each classifier that assigned c to X is added up. The class with the highest sum is returned as the class prediction for the object X.

$$\log \frac{1-error(M_i)}{error(M_i)} \tag{3.10}$$

Due to the way in which boosting focuses on the misclassified objects, it is more susceptible to the problem of overfitting than bagging. Although both of these two algorithms can significantly improve accuracy in comparison to a single model, boosting tends to achieve greater accuracy [117].

3.2.3.3 Random Forest

Another ensemble method is called Random Forrest (RF). Each of the classifiers in the ensemble is a decision tree and therefore the collection of classifiers is a "forest".

The RF technique applies bagging in tandem with random predictor selection. With a data set D of d objects, the general procedure to generate k decision trees for the ensemble is as follows. For each iteration, i (i = 1, 2, ..., k), the training set D_i consists of d objects sampled with replacement from D, so some examples may occur more than once whilst others may be excluded (i.e., D_i is a *bootstrap* sample of D). To construct a decision tree classifier, M_i, F predictors are randomly selected to determine the split at each node, where F is much smaller than the number of available predictors. The trees are grown with CART algorithm to maximum size and are not pruned. The optimal number of base trees can be determined by observing the out-of-bag error: classifying and voting on each example in D by only those trees which did not use this example for training [22]. During classification, each base tree votes and the most popular class is returned.

RF is comparable in accuracy to boosting. Since it considers much fewer predictors for each split, it is efficient on very large databases and it can be faster than either bagging or boosting.

3.2.4 Support Vector Machine

Support Vector Machine (SVM) was first presented in 1992 and has attracted a great deal of attention lately [117]. It is a method for the classification of both linear and nonlinear data. In a nutshell, SVM works as follows. It uses non-linear mapping to transform the original training data into a higher dimension. Within the new dimension, it searches for the linear optimal separating hyperplane (i.e., a decision boundary separating the objects of one class from another). Due to the ability to model complex nonlinear decision boundaries, SVM is highly accurate, and therefore has been applied to a number of areas, including handwritten digit recognition, object recognition, and speaker identification, etc. [117].

3.2.4.1 The Case When the Data is Linearly Separable

The simplest case in which SVM can be applied is a two-class problem where the training data is linearly separable. Let the data set D be given, in which X_i is one of the objects with associated class label being either +1 or -1 (in this research corresponding to the class *post-fault system will be stable* and *unstable*, respectively). An example based on two predictors, P_1 and P_2, is shown in Figure 3-7 to aid visualisation.

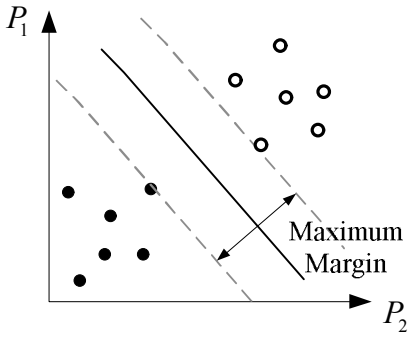


Figure 3-7: The 2-D linearly separable training data set with the best separating hyperplane [117].

Since an infinite number of straight lines can be drawn to separate all the two classes of objects, the best one, which will have minimum classification error on previously unseen objects, needs to be found. Generalising to *n* dimensions, the best hyperplane is needed. As shown in Figure 3-7, the best hyperplane is the one with the largest margin. It can be written as (3.11), where $\boldsymbol{W}$ is a weight vector $w_1, w_2, \cdots, w_n$, $\boldsymbol{X}$ is the predictor vector of a training object $X_1, X_2, \cdots, X_n$, *n* is the number of predictors and *b* is a scalar, often referred to as a bias. Any object that lies above the separating hyperplane satisfies (3.12) and any object that lies below the separating hyperplane satisfies (3.13). Any training objects that fall on the sides of the margin are called support vectors.

$$\boldsymbol{W} \cdot \boldsymbol{X} + b = 0 \tag{3.11}$$

$$\boldsymbol{W} \cdot \boldsymbol{X} + b > 0 \tag{3.12}$$

$$\boldsymbol{W} \cdot \boldsymbol{X} + b < 0 \tag{3.13}$$

The best hyperplane can be found through a constrained quadratic optimisation problem; it involves complex mathematics, beyond the scope of this thesis. Further details can be found in [117]. When a trained SVM classifies a test object, the decision depends on whether the data point lies above or below this hyperplane.

3.2.4.2 The Case When the Data is Linearly Inseparable

When the data set is not linearly separable, as shown in Figure 3-8, i.e., no straight line can be found to separate the classes, the original data set firstly needs to be transformed into a high dimensional space using a non-linear mapping.

In the new higher dimensional space, when searching for the hyperplane, the training objects appear only in the form of dot products, $\Phi(X_i) \cdot \Phi(X_j)$, where $\Phi(X)$ is the

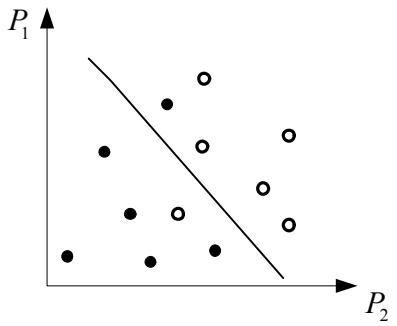


Figure 3-8: The 2-D linearly inseparable training data set [117].

non-linear mapping function applied to transform the training objects. Mathematically, this is equivalent to applying a *kernel function*, $K(X_i, X_j)$, to the original input data, which is represented by (3.14). As a result, everywhere that $\Phi(X_i) \cdot \Phi(X_j)$ appears in the training algorithm it can be replaced with $K(X_i, X_j)$. By using this substitution, the mapping can be completely avoided.

$$K(X_i, X_j) = \Phi(X_i) \cdot \Phi(X_j) \tag{3.14}$$

The commonly used kernel functions include Polynomial, Gaussian Radial Basis Function (RBF), and Sigmoid. There are no rules for choosing the best one. In this thesis, RBF is selected since it gives the most accurate SVM based on the data in hand. In the RBF kernel, as in (3.15), parameter σ is the width of the Gaussian distribution.

$$K(X_i, X_j) = e^{-\|X_i - X_j\|^2 / 2\sigma^2} \tag{3.15}$$

During the process of finding the best separable hyperplane, a parameter C is involved, which represents the trade-off between maximising the margin and minimising the training error.

The choice of values for C and σ will affect the classification performance of the SVM. To obtain a high level of accuracy, the optimal values of these two parameters

can be found using a grid-search.

3.2.5 Multiclass Classification

Among the existing classification algorithms, some of them, such as DT, naturally permit the use of more than two classes. An example of DT for multiclass classification is shown in Figure 3-9.

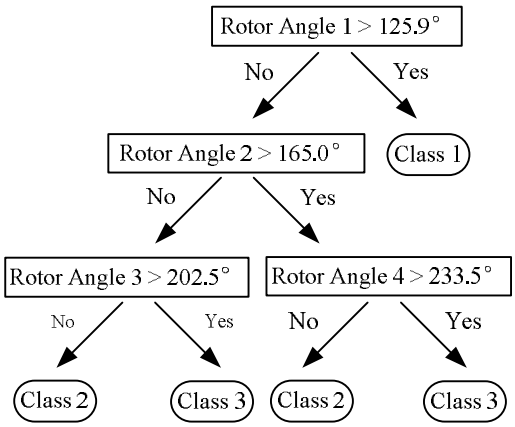


Figure 3-9: A multiclass decision tree for illustration purpose.

Other algorithms, like SVM, are by nature binary. Several methods have been proposed to effectively extend them for multiclass classification. They fall into two different approaches: one is to combine a series of binary classifiers whilst the other is to consider all data in one optimisation formulation directly. Reference [122] compares several of these methods and indicates that the "one-against-one" is the most suitable for practical use. It builds a binary classifier between every pair of classes so that for a training data set with m classes, $m(m-1)/2$ binary classifiers are constructed. Each of them is trained using the objects of the two classes it should discriminate. To classify an unknown object, each classifier votes and the object is assigned the class with the maximum number of votes.

For example, for a given data set D in which the training objects have class labels of three types. D_1, D_2 and D_3 are the subsets of objects with class label C_1, C_2 and C_3,

respectively. Three binary classifiers are trained, using D_1 and D_2 to distinguish C_1 and C_2, D_1 and D_3 to distinguish C_1 and C_3, and D_1 and D_3 to distinguish C_1 and C_3. If the output of the three classifiers for an unknown object is C_1, C_3 and C_3, respectively, the final decision of this object is class label C_3.

This "one-against-one" approach is adopted in this thesis to extend SVM to Multiclass Support Vector Machine (MSVM).

3.3 Clustering

When the Customer Services Director at an electronics company needs to organise all the customers into groups and assign a different manager to each group, strategically, he or she would like the customers in each group to be as similar as possible in terms of their business patterns, to develop customer relationships that specifically target each group. In this research, when the system operators need to divide the generators in a post-fault power system into groups, they want the generators in each group to be as similar as possible based on their dynamic behaviour so that they can be controlled together. In these examples, the data mining task is clustering, where the class label of each customer or generator is unknown.

The basic concepts of clustering are firstly given in this section. Hierarchical Clustering (HC) which will be used in this research is then introduced in detail.

3.3.1 Basic Concepts

Clustering is the process of partitioning a set of objects into subsets (or groups). Each subset is a cluster. The objects within a cluster have high similarity, but are dissimilar to objects in other clusters. The similarity between groups of objects can be customised depending on the target problem. In this context, different clustering algorithms may generate different clusters on the same data set. The partitioning is

not performed by humans, but by the algorithm automatically. Therefore, clustering is useful in that it can lead to the discovery of previously unknown groups within the data. Compared to classification, it is a form of *learning by observation*, rather than *learning by example*, and therefore known as unsupervised learning.

As a data mining function, clustering can be used as a standalone tool to obtain insight into the distribution of data and to observe the characteristics of each cluster. Alternatively, it can serve as a pre-processing step for other algorithms, such as classification, which would then operate on the result of cluster analysis. In this thesis, clustering is applied for the latter purpose.

There are a large number of clustering algorithms in the literature, such as *k*-means, *k*-medoids, Hierarchical Clustering (HC), probabilistic model-based clustering, etc. [117, 123]. Most of these algorithms use the number of clusters as the starting point. Hierarchical Clustering, however, groups data over a variety of levels by creating a *dendrogram*, which represents the results of grouping not by a single set of clusters but a multilevel hierarchy where clusters at one level are merged to form clusters at the next level. The number of clusters is not needed in advance. Since the number of groups of generators varies in different contingencies in power systems and needs to be found out automatically, HC is selected in this thesis.

3.3.2 Hierarchical Clustering

HC can either be *agglomerative* or *divisive*. The agglomerative method uses a bottom-up strategy. It starts by letting each object in the data set form its own cluster. The single cluster becomes the hierarchy's root. The clustering procedure firstly finds the two clusters that are closest to each other according to some similarity measure, and combines the two to form one cluster. Through a linkage criterion, two clusters

are merged at per iteration until all objects are linked into a hierarchical cluster tree. The divisive method, on the other hand, employs a top-down strategy. It starts by putting all objects in one cluster, which is the hierarchy's root. It then divides the root cluster into several smaller sub-clusters, and recursively partitions those clusters into smaller ones, until each cluster at the lowest level is coherent enough (i.e., either containing only one object or the objects within a cluster are sufficiently similar to each other). Figure 3-10 illustrates the application of HC on data objects ***a***, ***b***, ***c***, ***d***, ***e***. The arrows indicate the process for the agglomerative and divisive methods.

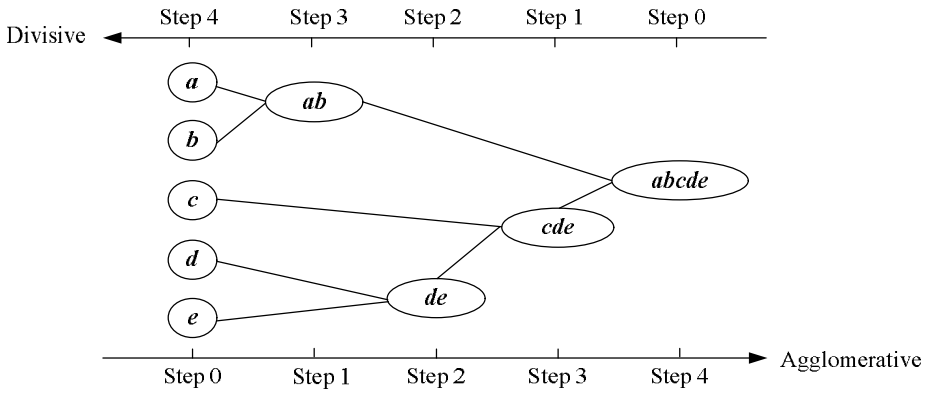


Figure 3-10: Agglomerative and divisive hierarchical clustering on data objects ***a***, ***b***, ***c***, ***d***, ***e*** [117].

With the divisive method the challenge is how to partition a large cluster into several smaller ones. There are $2^{n-1}-1$ possible ways where n is the number of objects. When n is large, it is computationally prohibitive to examine all possibilities. Heuristics is usually used in such cases however it can lead to inaccurate results. Therefore, to ensure accuracy and computational efficiency, the agglomerative method is selected in this thesis.

A tree structure called a *dendrogram* is commonly used to represent the process of HC. It shows how objects are grouped together step-by-step. Figure 3-11 is a dendrogram for the agglomerative process presented in Figure 3-10. A vertical axis is

used to show the similarity scale between clusters. The most appropriate clusters can be determined by cutting off the dendrogram.

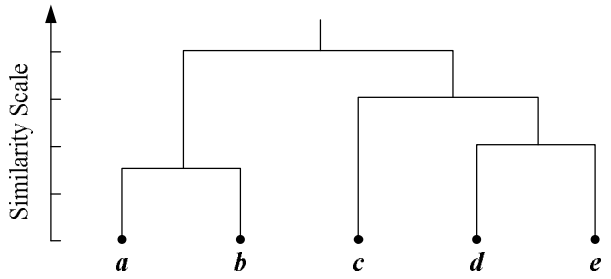


Figure 3-11: Dendrogram representation for agglomerative hierarchical clustering of data objects ***a***, ***b***, ***c***, ***d***, ***e***.

3.3.2.1 Similarity Measure

There is a variety of similarity measures that can be used in HC to quantify how close two objects are, such as Euclidean distance, Squared Euclidean distance, Manhattan distance, and Maximum distance. When p_k and q_k are the value of object *p* and *q* on the k^{th} dimension, these distances are defined as follows.

$$Euclidean\ distance = \sqrt{\sum_k (p_k - q_k)^2} \tag{3.16}$$

$$Squared\ Euclidean\ distance = \sum_k (p_k - q_k)^2 \tag{3.17}$$

$$Manhattan\ distance = \sum_k |p_k - q_k| \tag{3.18}$$

$$Maximum\ distance = \max_k |p_k - q_k| \tag{3.19}$$

Among these measures, Euclidean distance is the most widely used and generally gives satisfactory clustering results. It is therefore adopted in this thesis.

3.3.2.2 Linkage Criterion

Within the process of HC, the linkage criterion determines the distance between two

clusters, where each cluster is generally a set of objects. Commonly used linkage criteria include single-linkage, complete-linkage, centroid-linkage, and average-linkage. The distance between clusters for each of them is defined by (3.20) to (3.23), where $|p-q|$ is the distance between two objects, p and q; m_i and m_j are the mean for cluster C_i and C_j; and n_i and n_j are the number of objects in C_i and C_j.

$$dist_{single-linkage}(C_i, C_j) = \min_{p \in C_i, q \in C_j} \{|p-q|\} \tag{3.20}$$

$$dist_{complete-linkage}(C_i, C_j) = \max_{p \in C_i, q \in C_j} \{|p-q|\} \tag{3.21}$$

$$dist_{centroid-linkage}(C_i, C_j) = |m_i - m_j| \tag{3.22}$$

$$dist_{average-linkage}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{p \in C_i, q \in C_j} |p-q| \tag{3.23}$$

As given in the equations above, the single-linkage criterion defines the distance between clusters as the distance between two individual objects, one from each cluster, that are closest to each other. Complete-linkage criterion is the opposite, in which the distance between clusters is determined as the distance between two objects, one from each cluster, that are farthest away from each other. In the centroid-criterion, clusters are represented by their mean values, and the distance between clusters is defined as that between these two means. In the average-criterion, the average of the distances between all pairs of individual objects, one from each cluster, is taken as the distance between two clusters.

In this thesis, the linkage criterion selected is the complete-linkage. This is because the two nearest clusters of generators will not be linked together if the difference between any two of the generator rotor angles, one from each cluster, exceeds a

threshold.

3.3.2.3 A numerical Example

A numerical example of the operation of agglomerative HC using Euclidean distance and complete-linkage is given below. The method is applied to five (arbitrarily chosen) two-dimensional data points, as shown in Figure 3-12. The value of each point on the x-axis and y-axis are listed in the first and second row of matrix *M*, respectively.

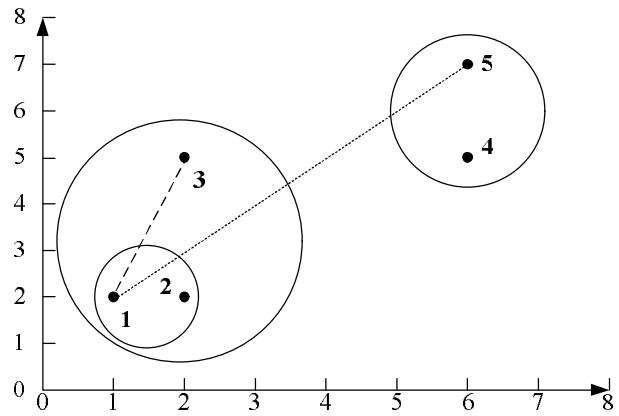


Figure 3-12: Two-dimensional plot for the five arbitrarily chosen data points.

$$M = \begin{bmatrix} 1 & 2 & 2 & 6 & 6 \\ 2 & 2 & 5 & 5 & 7 \end{bmatrix} \tag{3.24}$$

The similarity matrix of the data points is as follows:

$$D_1 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1 & 3.16 & 5.83 & 7.07 \\ 1 & 0 & 3 & 5 & 6.4 \\ 3.16 & 3 & 0 & 4 & 4.47 \\ 5.83 & 5 & 4 & 0 & 2 \\ 7.07 & 6.4 & 4.47 & 2 & 0 \end{bmatrix} \end{array} \tag{3.25}$$

The smallest element in the matrix is that for object 1 and 2, consequently these are joined to form a two-member cluster. Distances between this new cluster and the other three individuals are obtained as:

$$d_{(12)3} = \max\left[d_{13}, d_{23}\right] = d_{13} = 3.16 \tag{3.26}$$

$$d_{(12)4} = \max\left[d_{14}, d_{24}\right] = d_{14} = 5.83 \tag{3.27}$$

$$d_{(12)5} = \max\left[d_{15}, d_{25}\right] = d_{15} = 7.07 \tag{3.28}$$

A new similarity matrix can now be constructed whose elements are inter-object and cluster-object distances.

$$D_2 = \begin{array}{c} \\ (12) \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccc} (12) & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 3.16 & 5.83 & 7.07 \\ 3.16 & 0 & 4 & 4.47 \\ 5.83 & 4 & 0 & 2 \\ 7.07 & 4.47 & 2 & 0 \end{bmatrix} \end{array} \tag{3.29}$$

The smallest element in D_2 is that for object 4 and 5, so these now form a second two-member cluster. A new set of distances is found as:

$$d_{(12)3} = \max\left[d_{13}, d_{23}\right] = d_{13} = 3.16 \text{ (as before)} \tag{3.30}$$

$$d_{(12)(45)} = \max\left[d_{14}, d_{15}, d_{24}, d_{25}\right] = d_{15} = 7.07 \tag{3.31}$$

$$d_{(45)3} = \max\left[d_{34}, d_{35}\right] = d_{35} = 4.47 \tag{3.32}$$

These can be arranged in a matrix D_3.

$$D_3 = \begin{array}{c} \\ (12) \\ 3 \\ (45) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & 3 & (45) \end{array} \\ \begin{bmatrix} 0 & 3.16 & 7.07 \\ 3.16 & 0 & 4.47 \\ 7.07 & 4.47 & 0 \end{bmatrix} \end{array} \tag{3.33}$$

The smallest element now is $d_{(12)3}$ (indicated as the dashed line in Figure 3-12) so object 3 is added to the cluster containing objects 1 and 2. Finally, the cluster containing objects 1, 2, 3 and 4, 5 are combined into a single cluster. The distance between these two clusters (shown as the dotted line in Figure 3-12) is calculated as

(3.34).

$$d_{(123)(45)} = \max\left[d_{14}, d_{15}, d_{24}, d_{25}, d_{34}, d_{35}\right] = d_{15} = 7.07 \tag{3.34}$$

The corresponding dendrogram is shown in Figure 3-13.

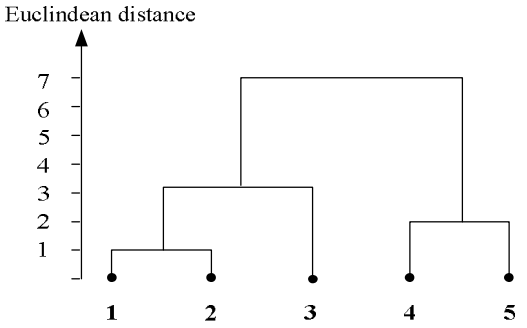


Figure 3-13: Dendrogram illustrating the clustering process in the numerical example.

3.4 Data Mining Software

Various software packages are available (either commercially or for free) to help customers develop data mining models and bring predictive intelligence to decisions. They have different capabilities, ease or difficulty of use, and user interface. Several of them have been used in past research related to on-line dynamic security assessment, such as SAS Enterprise Minder [60], Salford Predictive Modeler [58, 59, 66], WEKA [99] and ORANGE [99].

The three packages used in this research are IBM SPSS Modeler (version 14.2) [124], MATLAB Statistics and Machine Learning Toolbox (version 8.2.0.701 R2013b) [125], and LIBSVM (version 2.8) [126].

3.4.1 IBM SPSS Modeler

IBM SPSS Modeler is a data mining software application built by IBM. The workbench has an easy-to-use interface and allows users to quickly develop predictive models without programming. Working with IBM SPSS Modeler is a

three-step process of working with data, including uploading the data onto the software, running the data through a series of manipulations, and sending the data into a destination (either a model or a type of data output). The results can easily be directed to other databases.

A variety of modelling methods is offered by IBM SPSS Modeler. For DT, apart from the CART algorithm, C5.0 is also available which provides a high level of classification accuracy, especially with the powerful boosting function.

3.4.2 MATLAB Statistics and Machine Learning Toolbox

The Statistics and Machine Learning Toolbox in MATLAB provides functions to describe, analyse, and model data. Only the CART algorithm is available to build DT, with the ensemble methods of boosting, bagging and RF. The advantage of using MATLAB is that the model trained can be called in a loop when it needs to be tested multiple times (for instance, for a large number of scenarios of missing predictors). Hierarchical Clustering algorithm is provided in the toolbox for unsupervised learning.

3.4.3 LIBSVM

LIBSVM is a widely used and open source library for support vector classification, regression and distribution estimation. It supports multiclass classification with the "one-against-one" algorithm. Developed at the National Taiwan University, LIBSVM is written in C++ and a simple interface is provided to allow users to easily link it with their own programs in MATLAB. Functions are available to find the optimal values of parameters (C and σ as introduced in Section 3.2.4.2) using a grid-search, which leads to high accuracy of classification.

For each data mining model developed within this thesis, the software package used

for implementation will be outlined along with the presentation of the training process.

3.5 Summary

The technical background of all the data mining algorithms that will be used in this thesis was introduced in this chapter. Classification is the process of supervised learning, and is going to be the core technique that implements the on-line identification of power system dynamic signature. The algorithms described include CART and C5.0 for DT, bagging, boosting, and RF to create EDT, and SVM. All DT methods allow both binary and multiclass applications whilst SVM has to be extended into Multiclass SVM (MSVM) through a one-against-one strategy for classification tasks with more than two categories. On the other hand, clustering is unsupervised learning, and is going to be used as the pre-processing step of classification. The agglomerative Hierarchical Clustering was presented. Finally, the software packages that are used to implement the data mining techniques in this thesis were introduced.

In both Chapter 4 and Chapter 5, only CART algorithm of DT will be used for the purpose of binary classification. In Chapter 6, C5.0 will be used for binary classification whilst CART, C5.0, boosted C5.0, RF and MSVM will be compared for multiclass application. Hierarchical Clustering will also be utilised in this chapter. Bagged CART, boosted CART and RF will be compared in Chapter 7 in terms of the ability to deal with missing input data. Further details on the implementation of these algorithms and their performance evaluation can be found in later chapters.

4 On-line Identification of Power System Transient Stability Using Decision Tree Method

4.1 Introduction

As introduced in Chapter 1, in the context of increased epistemic and aleatory uncertainties, network complexity and variable power transfers, future power systems will operate closer to their stability limits. To enable safer system operation, corrective control and stabilisation is becoming a potentially viable option. Extensive research efforts have been consequently devoted to fast and reliable on-line identification of power system transient stability using PMU measurements and data mining. A review of this work has been presented in Section 1.2.2.

In the framework of wide-area measurements in power systems, data mining has been investigated as the predictive tool to convert PMU data into transient stability

information and support decisions under emergency conditions. A data mining model is generally trained using a large number of off-line simulated post-fault system responses. When applied in real-time, it makes use of PMU data to "warn" operators and system-level controllers about impending transient stability issues, and to trigger appropriate control actions so that a loss of stability can be avoided or the consequences of it minimised. The aim of developing such a model is to enable safer system operation closer to its stability limit, in the context of increasingly stressed power networks with ever increasing model, parameter and operational uncertainties.

This chapter demonstrates the implementation of the Decision Tree (DT) method on the 16-machine, 68-bus NETS-NYPS test network. The study investigates the sensitivity of the prediction accuracy of the DT to a number of system uncertainties, including the duration and location of fault, system operating point and pre-fault network topology.

4.2 Methodology

Figure 4-1 provides an overall description of the methodology for on-line prediction of transient stability for corrective control using PMU measurements and DT. The shaded portion in the flowchart shows the off-line stream which is the training of DT. Predictors are usually selected as post-disturbance system parameters, such as generator rotor angles, speeds, and accelerations [54, 55, 60, 74, 77], voltage magnitudes and angles [66, 73, 74], and apparent resistance along with its changing rate measured near the electrical centre of the intertie of a system (which will change during loss of synchronism) [58, 59, 69]. The target of training is the system stability status after the disturbance is cleared. The library of system dynamic signatures is generated through time domain simulation [20]. The underlying relationships

between predictors and targets are automatically learned by the DT during training so that the model can fit into the on-line stream (presented by the unshaded portion of the flowchart).

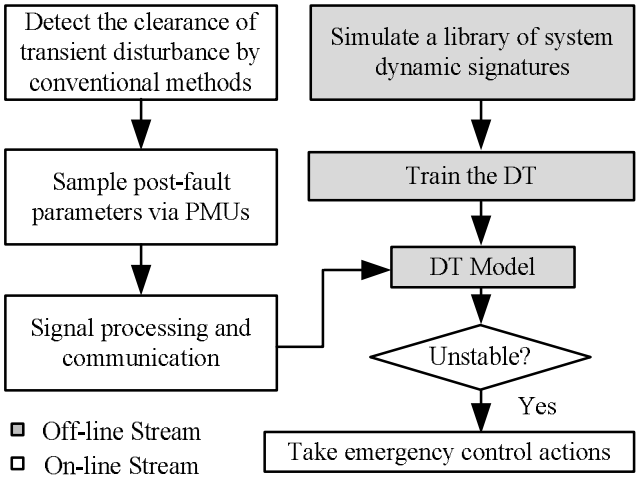


Figure 4-1: DT approach for on-line prediction of transient stability.

In real-time application, after it is detected that a fault has just been cleared from the system (by conventional methods such as monitoring voltages, frequency or the status of circuit breakers in the network [21]), samples of predictors will be taken and constructed as the input to the DT model. The patterns of input are expected to be recognised and intelligent decisions are expected to be made. As mentioned in Chapter 1, incorrect decisions under emergency conditions can cause undesirable consequences: predicting stable cases as unstable will trigger unnecessary control actions whilst predicting unstable cases as stable will lead to lateness or failure of control and ultimately the instability of the system.

4.3 Applications on Test Network

The NETS-NYPS five-area system as described in Chapter 2 is used as the test network. The pre-fault system initial conditions are performed using standard power flow.

4.3.1 Training of Decision Tree

4.3.1.1 Predictors and Targets

Generator rotor angles and speeds are selected as predictors in this chapter since their variation after disturbance best describes the transient behaviour of the power system and they are the most straightforward to use when assessing the degree to which the system is disturbed. Since it is assumed that each of the 15 generator buses (excluding G13 which is the reference with constant rotor angle) has a PMU and other required measurement devices installed, the rotor angles can be directly provided without any electrical calculation. It is assumed in this study that rotor speeds are calculated as the rate of change of rotor angle after the signals are transferred into the Monitoring and Control Centre.

In total, 30 predictors are chosen for the test system including 15 rotor angles (of G1 to G12 and G14 to G16) and 15 rotor speeds. The target of training is whether the system will be stable or unstable after the fault is cleared.

4.3.1.2 Generation of Training Set

To generate the training database, three-phase self-clearing faults (one at a time) with various fault location (from bus 1 to bus 68) and fault clearing time were simulated in the test network. The fault clearing time was varied between 0.05 s and 0.29 s, with an increment of 0.01 s in order to generate a mix of stable and unstable system conditions. In total, 1700 different faults were simulated. The post-fault system behaviour (rotor angle responses) was recorded for 6 s following each fault, since in most of the unstable contingencies at least one of the generators in the test system experiences the loss of synchronism within this period of time. The transient stability of the system is assessed based on the recorded generator rotor angle responses.

It is generally difficult to distinguish between stable and unstable responses within a short period of time after the fault. The criterion used to detect the instability is usually a matter of operational practice [5]. In this chapter, as in some past work [55, 62], it is defined as follows: If the difference between any two of the generator rotor angles exceeds 360 degrees within the 6 s of simulation the system is considered to be unstable, otherwise it is stable. Using this criterion, 1238 (72.8%) out of 1700 simulated cases were classified as stable whilst 462 (27.2%) were classified as unstable. An example of the swing curves is given for a stable and an unstable case in Figure 4-2 and Figure 4-3, respectively. All the rotor angles oscillate in the stable case with decreasing amplitude whilst two generators start to lose synchronism after about 3 s in the unstable case.

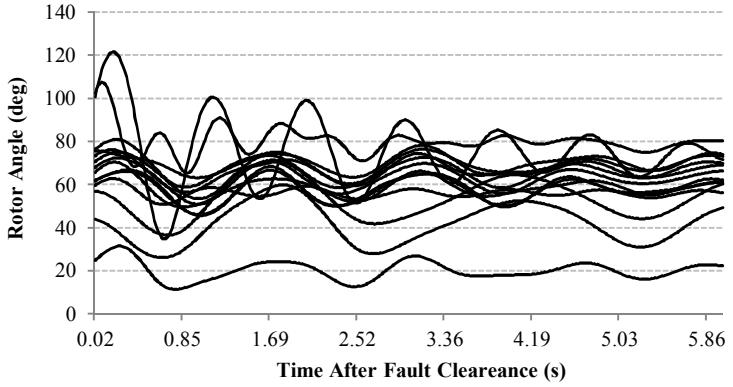


Figure 4-2: Example of post-fault generator rotor angle swings in a stable case.

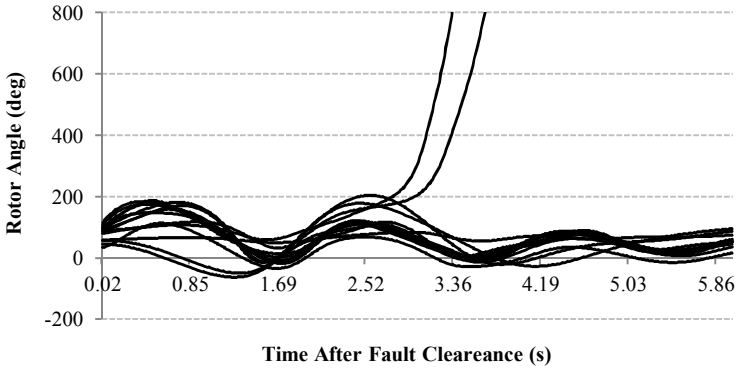


Figure 4-3: Example of post-fault generator rotor angle swings in an unstable case.

In the simulations of system dynamic behaviour, the post-fault generator rotor angles are computed by solving a set of differential-algebraic equations which describes the power system. Therefore, the computation results are discrete collections of numerical data on rotor angle values along with the corresponding time stamps. The "time series" function in MATLAB is used to construct the data collections into curves and then resample them with the user defined sampling rate.

In each simulation, all 15 rotor angles are sampled every cycle (i.e., 0.0167 s), which is the same as the required PMU reporting rate [10]. The rotor speed for each generator is then calculated using a backward difference approximation [60, 103], as given by (4.1),

$$\omega(t) = \frac{[\delta(t) - \delta(t - \Delta t)]}{\Delta t} \tag{4.1}$$

where $\omega(t)$ and $\delta(t)$ are the rotor speed and angle of the generator at time t, respectively. Δt is the sampling time interval, i.e., 0.0167 s for the 60 Hz system.

Although the post-fault system behaviour was simulated for 6 s, only the first 5.01 s of rotor angles and speeds data are converted into training data [62]. Therefore 300 collections of predictors (sets containing information on rotor angles and speeds) are created for one simulation. Each collection is assigned a target variable which holds a binary value "1" (unstable) or "0" (stable). The targets of all 300 collections from unstable simulations are set as unstable and vice versa, in order to allow the DT building algorithm to find the threshold value of predictors as early as possible [66]. For the 1700 faults, 510,000 predictor-target pairs are constructed in total. Each of them is used as an individual example for training. The information about their time stamps however, cannot be seen by the DT.

4.3.1.3 Building the Tree

The DT in this chapter is built using the CART algorithm in IBM SPSS Modeler [124]. As the software allows users to specify parameters such as maximum tree depth, minimum records in parent and child branch (either in percentage or absolute value) and tree pruning rule, which will affect the presentation of the resulting tree, multiple DTs are built so that the one with best performance on new test data can be chosen. The problem of overfitting can be avoided by abandoning those trees with too many nodes or splits. The final DT chosen takes 5 min and 50 s to build (on a PC with 2.66 GHz quad core CPU and 3.25 GB RAM) and has 11 levels below the root node. It was found that 12 out of the 30 predictors (rotor angles and speeds of 12 generators), listed in Table 4-1, contribute to making a decision. The test at each node sets a threshold value for that particular predictor, which essentially distinguishes stable and unstable simulations far earlier than the time when the loss of stability actually happens.

Table 4-1: Predictors contributing to the final decision tree.

Predictors (Generator Bus #)	
Rotor Angle	*Rotor Speed*
2, 4, 9, 10, 11, 12, 14	3, 7, 8, 11, 16

4.3.2 Sensitivity of Prediction Accuracy to Different Uncertainties

4.3.2.1 Generation of Test Sets

The chosen set of uncertain factors in the network consists of fault duration, fault location, the system operating point and the pre-fault system topology. Different sets of test data, previously unseen by the DT, are used to investigate the sensitivity of the prediction accuracy to each factor. In all simulations, only one three-phase self-

clearing fault at a time is applied in the system and the post-fault system is simulated for 5 s. The asymmetrical faults were not considered at this stage of research.

1) ***For fault duration***: The system operating point and pre-fault topology is unchanged, i.e., loading and generation is as provided in [115, 127] and all lines are in service. The location of the fault is from bus 1 to bus 68. The fault duration was varied randomly between 0.01 s and 0.29 s and 25 different fault durations were selected. Selected fault durations were different from the ones used in the training set. A test set containing 1700 different faults is designed in this way.
2) ***For fault location***: The system operating point and pre-fault topology is unchanged. The fault durations are the same as those in the training set. The fault location varies among 6 points on each transmission line (at 20%, 30%, 40%, 50%, 60%, and 70% of the length). As there are 66 lines without transformer in the test system, a test set consisting of 9900 faults is designed.
3) ***For system operating point***: The faults with the same duration and location as in the training set are repeated for 7 different loading conditions (loading levels of 0.5, 0.9, 0.95, 1.05, 1.1, 1.15, and 1.2 of base load), with the pre-fault system topology unchanged. (In the UK, the extremely low level of 0.5 p.u. usually occurs from 3 am to 5 am during summer days [128].) The number of faults simulated for each loading level is 1700.
4) ***For system pre-fault topology***: 8 different cases are investigated independently. In each case, one of the following transmission lines (indicated as red dashed lines in Figure 4-4) is removed from service: $L_{14\text{-}15}$, $L_{4\text{-}14}$, $L_{1\text{-}30}$, $L_{3\text{-}52}$, $L_{9\text{-}30}$, $L_{25\text{-}26}$, $L_{40\text{-}41}$, and $L_{40\text{-}48}$. The first and last two lines are selected for disconnection since they will alter the system's pre-fault power flow and

voltage profile very little ($L_{14\text{-}15}$ and $L_{4\text{-}14}$) and significantly ($L_{40\text{-}41}$ and $L_{40\text{-}48}$). The rest of the cases are evenly chosen between these extremes. For each case, a set of 1700 faults is repeated as at the training stage, with the system operating point unchanged. 8 test sets are created in total.

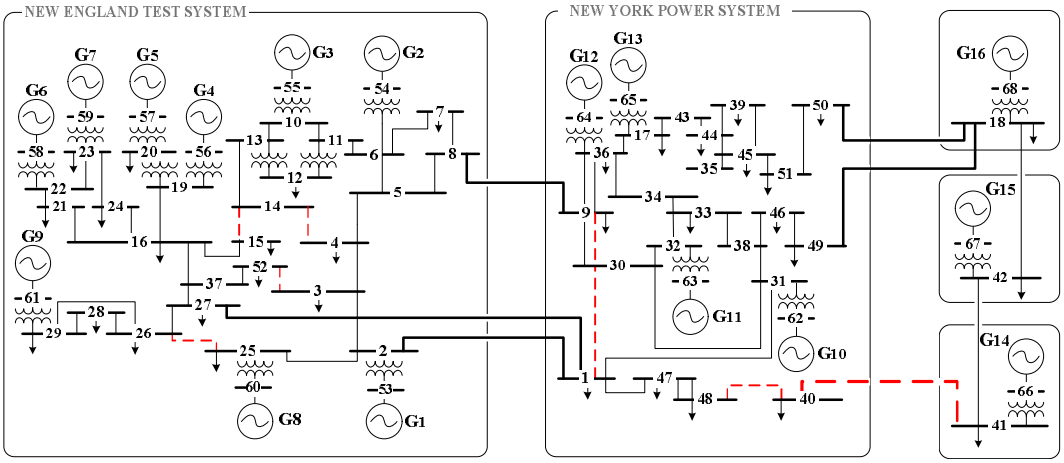


Figure 4-4: NETS-NYPS five-area test network with removed transmission line indicated (lines were removed one at a time in 8 independent cases).

All the above sets of test data are sent as inputs into the DT model in IBM SPSS Modeler. Since the speed of processing is extremely fast the outputs to 510,000 predictor collections were obtained within 10 s on the same PC as previously described.

4.1.1.1 Results and Discussions

In real-time application, the PMU measurements from the network are transmitted into the control centre every cycle. After it has been identified that a fault is cleared (e.g., by monitoring voltages, frequency or the status of circuit breakers in the network) a time series of predictor collections are sent into the DT as input. A decision regarding whether the system will be stable or unstable is made for each collection so that the output from DT is also a time series. As a result, in this chapter, for a single set of test data the accuracy of prediction is evaluated at every instance

when decisions are made based on values of predictors at that time. For example, for the test set designed for uncertainty of fault duration, which contains 1700 different faults, 92% prediction accuracy at 0.2 s indicates that the transient stability of the test system is predicted correctly for 1564 faults using the predictors sampled *at* 0.2 s after the clearing of the fault.

The evolutions of prediction accuracy for different test sets, according to the time after which the fault is cleared, are shown in Figure 4-5 to Figure 4-8. Each figure demonstrates the sensitivity of DT performance to one of the four assumed uncertainties, by comparing the results for test sets to the classification accuracy curve for the training database.

Generally speaking, as shown in all these figures, the level of DT prediction accuracy increases with time. After the fault is cleared from the system, the longer the DT waits, the more confident it will be in making the decision of system transient stability. Take the training set as an example, using the predictors sampled about 0.5 s after the fault is cleared, the stability is classified correctly in 95% of the cases. Using predictors sampled approximately 3 s after fault is cleared, close to 100% of the cases are correctly classified.

Looking first at Figure 4-5 and Figure 4-6, the shape of prediction accuracy curve is very weakly affected by the uncertainty of fault duration and location in the test sets. This indicates that the DT model generated previously is robust to unseen inputs that involve these two uncertain factors.

In Figure 4-7, each prediction accuracy curve is obtained from the test set designed for a separate system Loading Condition (LC). Numbers in the legend indicate the level of loading and generation. When system loading is as low as 0.5 p.u., the system

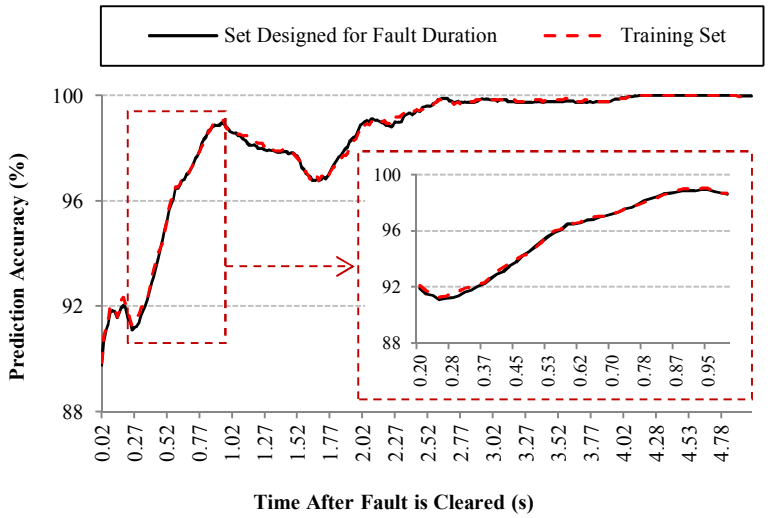


Figure 4-5: Sensitivity of decision tree based prediction accuracy to different fault durations.

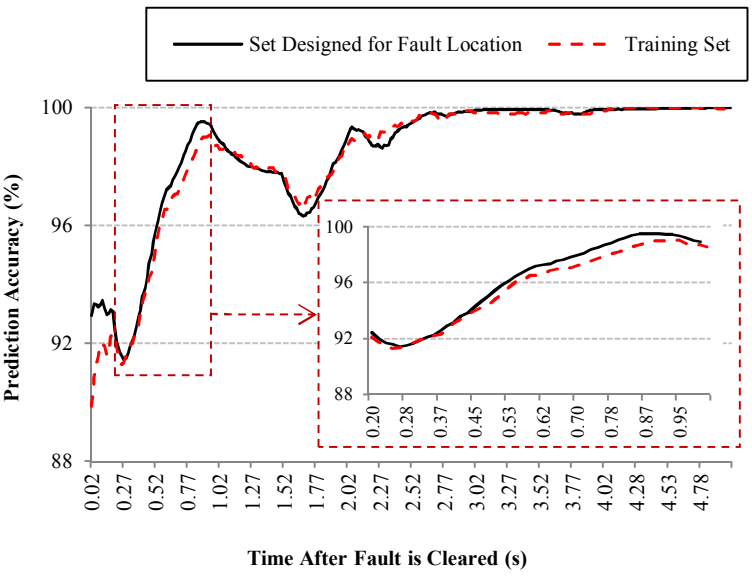


Figure 4-6: Sensitivity of decision tree based prediction accuracy to different fault locations.

is much less stressed than before and very far from its stability limit. The accuracy of prediction is higher than 95% when predictors are sampled immediately after the fault is cleared, and reaches 100% when decisions are made after about 0.3 s. When the LC is close to 1 (from 0.9 to 1.05), the shape of the curve changes. However, the curves corresponding to these LC can be considered as similar. When the system is highly loaded (from 1.1 to 1.2) it becomes more stressed and, the accuracy of prediction becomes lower. The accuracy does not increase to over 98% until 4 s after fault is

cleared.

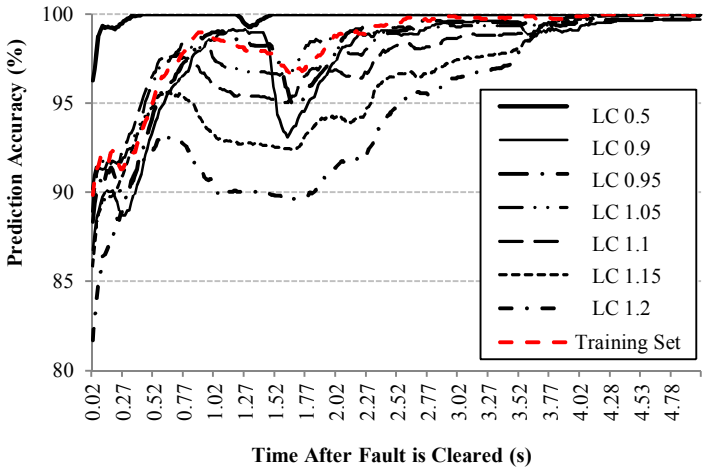


Figure 4-7: Sensitivity of decision tree based prediction accuracy to different system operation points.

The prediction accuracy curves of the 8 test sets designed for assessing the effect of uncertainty in pre-fault system topology are separately shown in Figure 4-8 (a) and (b). The legends indicate the line taken out from service in different cases. In Figure 4-8 (a), the removal of line $L_{14\text{-}15}$, $L_{4\text{-}14}$, $L_{3\text{-}52}$, or $L_{25\text{-}26}$, makes the pre-fault system slightly more stressed. The shapes of corresponding curves stay close to the training set's curve. The removal of line $L_{1\text{-}30}$, or $L_{9\text{-}30}$, however, pushes the pre-fault system into a much more stressed operating condition. Compared to previous cases, the prediction accuracy of these two test sets is lower, especially when the time exceeds 2 s.

Figure 4-8 (b) demonstrates that the performance of the DT can be much worse compared to previous cases if one of the key transmission lines in the system is removed. When the line $L_{40\text{-}41}$, or $L_{40\text{-}48}$, is taken out, the pre-fault system becomes extremely stressed. The prediction accuracy in this case dropped between 70% and 75% within the first 2 s after fault is cleared and does not go over 95% until after 4 s.

As a whole, the DT based prediction accuracy for system transient stability is much

more sensitive to the uncertainty in system loading and pre-fault topology than to that of the fault duration and location. The more stressed the pre-fault system is, the poorer the prediction performance. The reason can be explained through closer inspection of simulations in different test sets. In the set constructed from the system with a high stress level, a larger portion of the faults will cause the system to be marginally stable (or unstable). The post-fault generator rotor angles and speeds either oscillate for a very long time or go out of step towards the end of simulation (after few seconds), and therefore it is very difficult for the DT to predict whether the system will be stable or unstable.

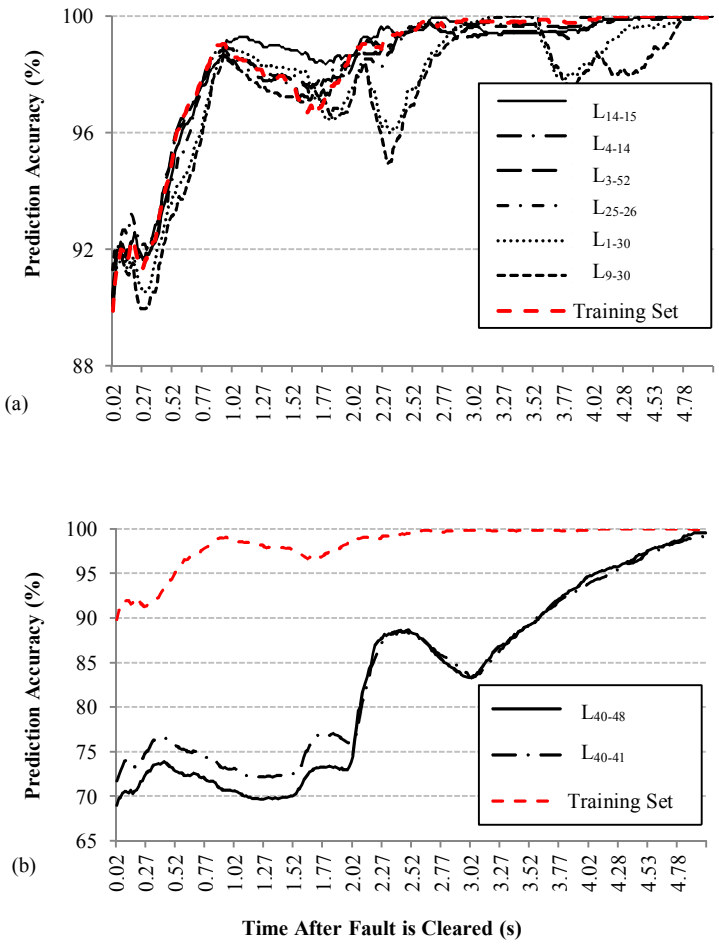


Figure 4-8: Sensitivity of decision tree based prediction accuracy to different pre-fault system topology: (a) An arbitrary line in the system is removed from service. (b) Key line in the system is removed from service.

Furthermore, Figure 4-9 was obtained by plotting all curves from Figure 4-5 to Figure 4-8 in the same figure. The shaded area indicates the range in which the prediction accuracy of DT will most probably be, no matter which type of uncertainty is involved in the inputs. The curves above and below the area demonstrate the extremely good and poor performance of DT when the system is under extremely unfavourable operating conditions and/or topology.

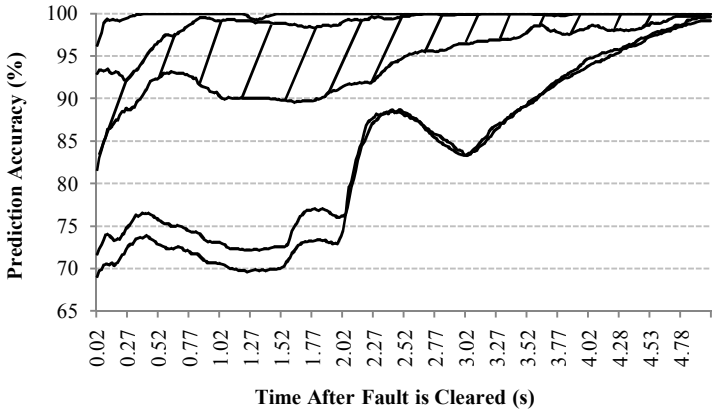


Figure 4-9: Most likely range of decision tree based prediction accuracy to different uncertainties.

4.4 Summary

In this chapter, the application of DT method for PMU based online identification of transient stability was illustrated. The prediction accuracy of the DT model for unseen inputs was evaluated every time when the predictors are sampled during the post-fault period. The sensitivity of prediction accuracy has been investigated considering four types of uncertainties in the system.

It has been shown that the accuracy of prediction is more sensitive to the uncertainty in system operating point and pre-fault network topology, than to fault duration and fault location. It should be pointed out though that only three-phase fault has been used at this stage of research as they are the most critical for system stability.

The developed DT is typically able to predict system post-fault behaviour with

accuracy between 88% and 92% as fast as 0.25 s after the fault is cleared and with accuracy between 90% and 99% about 1 s after the fault. These initial results suggest that DT is a useful tool for online assessment of system transient stability and hence applicable to corrective control approaches.

Although the test results provide useful information about the prediction accuracy of the trained DT, a more thorough assessment of the model robustness is required for a wide range of feasible operating conditions and disturbances. This will be addressed in the following chapter.

5 Probabilistic Framework for Assessing the Accuracy of Data Mining Tools

5.1 Introduction

In the previous chapter, the accuracy of DT for on-line prediction of transient stability was tested according to a list of power system uncertainties. Other frequently used ways of constructing test databases in the literature include selecting several individual representative contingencies and sampling one set of contingencies according to the distributions of all uncertain factors or past experiences. The system operators (or system level controllers), however, need to be confident regarding the accuracy of prediction when making control decision under emergency conditions.

This chapter presents a generic probabilistic framework to assess the accuracy of a data mining model, which is the *first original contribution* of this research. The framework allows a thorough and realistic evaluation of the reliability of a data mining model in that it makes the right decision at the right time. It also, and more

importantly, allows a consistent comparison of different data mining models based on the accuracy of the prediction.

The probabilistic approach is applied in the testing process. The algorithm performs an exhaustive search of possible contingencies and then weighs the accuracies according to the realistic probability distribution of uncertain system factors. The assessment results highlight the effect of different uncertain factors on the accuracy of prediction of system transient stability, and identify the confidence level of a data mining model in real time.

The data mining technique used as an example in this chapter is also DT. Any other data mining techniques could have been used to illustrate the methodology without any loss of generality.

5.2 Proposed Methodology

Many basic initiating factors that can affect transient stability are probabilistic in nature, including the type and location of the fault, operating times associated with the fault clearing equipment, and the system operating conditions at the time of fault inception such as loading levels, generation capacities and network topology [129-131]. A probabilistic approach is therefore the most suitable way to establish a thorough evaluation of the prediction accuracy of a data mining model.

To simplify the description and illustration of the proposed methodology only four probabilistic factors are used in the procedure of testing. They are: i) fault clearing time; ii) fault location; iii) fault type; iv) system loading level. The inclusion of more factors for practical implementations would be straightforward and would not affect the methodology, though the numerical results would be different.

For a certain type of fault, at one possible location and a certain load level, the fault

clearing time is varied in a probabilistic manner to generate a set of contingencies. The probabilistic distribution of fault clearing time is defined according to the protection system in the network [129, 130, 132]. The accuracy of prediction A_{ijk} at a location i for a fault type j and at load level k can be found based on the test results of this set.

At the same time, system faults can be categorised into single line to ground (LG), line to line (LL), line to line to ground (LLG) and three phase (LLL) faults. The faults can be located at different buses or on various transmission lines (including both close-in and mid-line region [130]). The system load varies during the day and through the year. The probability distribution of the above three factors for a practical power system can be found in historical data. Since the type and location of fault and the system conditions at the time of the fault occurrence are all independent, according to the probability theory, the probability, P_{ijk}, of prediction accuracy being A_{ijk} is given by (5.1), where P_i, P_j and P_k are the probability of having a fault at location i, of type j, and at load level k, respectively. Therefore, the probabilistic accuracy of prediction (A) at all possible locations (I) for all fault types (J) and all system load levels (K) can be calculated from (5.2).

$$P_{ijk} = P_i \times P_j \times P_k \tag{5.1}$$

$$A = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} A_{ijk} \times P_{ijk} \tag{5.2}$$

The expression for accuracy of prediction (5.2) can be easily extended to include other uncertain factors including network topology changes, generation patterns, load models and parameters, presence and output of renewable generation, etc. This extension would result in (5.1) and (5.2) being modified to include further

probabilities on the right hand side of (5.1) and summations on the right hand side of (5.2). The extension would yield different quantitative results for A and more computational effort to get them; however, the methodology to obtain the result would remain the same.

5.3 Probabilistic Distribution Used in the Test System

The probabilities associated with the clearing time, type and location of fault, and system loading level are expected to be obtained from the available system data. In the absence of such data for the NETS-NYPS test system, the distributions are assumed as follows.

5.3.1 Fault Clearing Time

In this study, a discrete uniform distribution of fault clearing time, ranging from 0.05 s to 0.29 s with an increment of 0.02 s, is assumed for any location in the network. This assumption allows a reasonable mix of stable and unstable post fault system conditions to be generated. However, in practical networks more complex distributions based on reliability modelling of the protection system, such as normal distribution, could be used [129, 130, 132].

5.3.2 Fault Location

For an existing power system the probabilities of fault location can usually be obtained from the historical transmission outage statistics. It is also often the case that during the study period, some buses and lines did not experience transient faults at all [130]. Since the data for the test network used in this study is not available, the probabilities listed in Table 5-1 are arbitrary. It can be seen from the table that 25 different locations are chosen for this study. Five of them are generator buses whilst the remaining 20 locations are on the transmission lines (in the middle of each line for

simplicity). Furthermore, as highlighted in the table, Line 16-19 is the most probable location where faults can happen (the probability is 0.1382) whilst Bus 54, Line 1-31, Line 33-34 and Line 39-44 all have the lowest probability of 0.0046.

Table 5-1: Fault location probabilities.

No.	*Location*	*Probability*	*No.*	*Location*	*Probability*
1	Bus 54	0.0046	14	Line 1-30	0.0369
2	Bus 57	0.0138	15	Line 30-31	0.0414
3	Bus 61	0.0099	16	Line 1-31	0.0046
4	Bus 62	0.0276	17	Line 32-33	0.0690
5	Bus 63	0.0185	18	Line 33-34	0.0046
6	Line 17-36	0.0369	19	Line 16-37	0.0967
7	Line 16-19	0.1382	20	Line 39-44	0.0046
8	Line 16-21	0.0876	21	Line 37-52	0.0277
9	Line 22-23	0.0230	22	Line 3-4	0.0834
10	Line 16-24	0.0276	23	Line 10-11	0.0311
11	Line 26-27	0.0415	24	Line 15-16	0.0089
12	Line 26-28	0.0507	25	Line 1-27	0.01
13	Line 28-29	0.1012			

5.3.3 Fault Type

Table 5-2: Fault type probabilities.

No.	*Fault Type*	*Probability*
1	Single Line to Ground	0.7
2	Line to Line	0.15
3	Line to Line to Ground	0.1
4	Three Phase	0.05

Although less severe in terms of consequences that they leave on power systems,

most of the faults in realistic power systems are unbalanced faults. The most severe, three-phase faults, in contrast, have relatively low probabilities of occurrence. The distribution of the fault type used in this study is adopted from [133] and it is shown in Table 5-2.

5.3.4 System Loading Level

As shown in previous work, a multi-step loading model derived from a Load Duration Curve (LDC) can be used to describe the variations in system loading [129, 134, 135]. Therefore, a six-step approximate LDC as illustrated in Figure 5-1 is assumed for the test network. Different loading levels (as percentage of maximum) and their probabilities, as detailed in Table III, are all arbitrary.

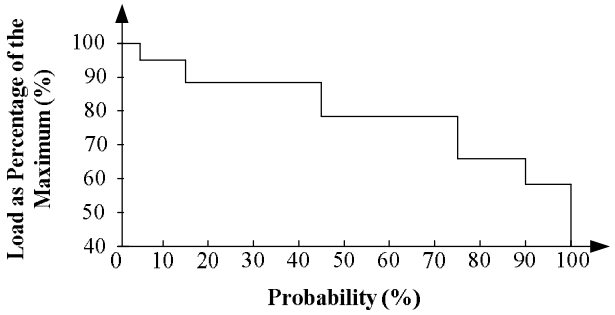


Figure 5-1: Multi-step load model for the test system.

Table 5-3: System load probabilities.

Load Step	*Load Level as Percentage of Maximum*	*Loading Factor Based on Given Load*	*Probability*
1	100%	1.173	0.05
2	95%	1.114	0.1
3	88%	1.03	0.3
4	79%	0.927	0.3
5	66%	0.774	0.15
6	58%	0.681	0.1

To determine the maximum load of the test network, all nominal loads as given in [115, 127], either real or reactive, are initially multiplied by the same Loading Factor (LF). The OPF is then run repeatedly with a gradual increase of the LF. The maximum load can be found just before the non-converge of OPF occurs. As a result, the maximum load level (100%) corresponds to the LF 1.173. The LFs for the other five levels are calculated and listed in Table 5-3.

5.4 Training of Decision Tree

5.4.1 Predictors and Target

As in the previous chapter, 30 predictors are chosen including 15 rotor angles (of G1 to G12 and G14 to G16) and the 15 rotor speeds. The target of training is the system stability status after the fault is cleared.

5.4.2 Generation of Training Set

To generate the training database, only three-phase self-clearing faults (one at a time) with varying fault location (from bus 1 to bus 68) and fault clearing time (from 0.05 s to 0.29 s, with an increment of 0.02 s) were simulated in the test network, at the given nominal load level (LF equals 1). The pre-fault system initial conditions are obtained using OPF. In total, 884 different faults were simulated and the post-fault system behaviour (rotor angle responses) was recorded for 6 s following each fault.

The transient stability of the system is assessed again according to the criterion that if the difference between any two of the generator rotor angles exceeds 360 degrees within the 6 s of simulation the system is considered to be unstable, otherwise it is stable. As a result, 746 simulations (84.4%) among the 884 were classified as stable whilst 138 simulations (15.6%) were unstable.

In each simulation, all 15 rotor angles are sampled every cycle (i.e., 0.0167 s, based

on 60Hz system) [10]. The rotor speed for each generator is calculated using a backward difference approximation [60, 103], as given by (4.1).

The first 5.01 s of rotor angles and speeds are converted into training data [62] and so for one simulation 300 vectors of predictors are created. Each vector has 30 elements, and is further assigned a target variable which has a binary value "1" (unstable) or "0" (stable). The targets of all the 300 vectors from one unstable simulation are set as unstable and vice versa. For the 884 faults, 265,200 predictor-target pairs are constructed in total. Each of them is used as an individual example for training.

5.4.3 Building the Tree

The DT in this chapter is also built using the CART algorithm in IBM SPSS Modeler, in 4 min and 7 s (on a PC with 2.66 GHz quad core CPU and 3.25 GB RAM) and has 11 levels below the root node. Only 12 (mix of speeds and rotor angles of 11 generators) out of the original 30 predictors are used to make a decision, due to the feature selection of the training process. They are listed in Table 5-4.

Table 5-4: Predictors contributing to the final decision tree.

Predictors (Generator Bus #)	
Rotor Angle	*Rotor Speed*
3, 5, 7, 9, 10, 11, 12, 16	2, 4, 8, 9

5.5 Probabilistic Evaluation of Prediction Accuracy

The methodology for the probabilistic evaluation of prediction accuracy outlined in Section 5.2 is applied to the DT model, using the probability distributions assumed in Section 5.3. The procedure is described in detail as follows.

For each of the four types of fault, at each of the 25 locations and at each of the six system loading levels, a testing set containing 13 self-clearing faults is simulated by

varying the fault clearing time from 0.05 s to 0.29 s with an increment of 0.02 s. Both symmetrical and asymmetrical faults are modelled as presented in Section 2.3. The post-fault system is simulated for 6 s for each fault. In total, there are 600 testing sets including 7800 different faults which are not used during the training of DT. The system stability status, after each fault, is assessed using the criterion previously described, and then later compared to the output of DT.

After predictor vectors are constructed from the first 5.01 s of data from the above simulations (in the same way as for training set), they are sent as inputs into the DT model in IBM SPSS Modeler. For each simulation (including 300 predictor vectors) the outputs can be obtained within approximately 6 ms execution time on the same PC as previously described.

As presented in the last chapter, for a single testing set, the accuracy of prediction is evaluated at every instance when decisions are made based on the values of predictors at that time. For example, for the testing set which contains 13 three-phase faults at bus 2 when system LF is 1.173, prediction accuracy of 76.9% at 0.1 s indicates that: in 11 of the faults, the post-fault transient stability status of the test system (stable/unstable) is predicted correctly using the predictors sampled at 0.1 s after fault clearing. Therefore, A_{ijk} for each of these 600 testing sets is actually an array with 300 elements which represent the prediction accuracies from 0.0167 s to 5.01 s (with an increment of 0.0167 s) after the fault is cleared from the system.

After calculating A_{ijk} for all 600 testing sets, the probability of having a fault at location i, of type j, and at load level k can be applied. There are three types of prediction accuracy which can firstly be worked out. They are: i) The probabilistic accuracy at 25 different locations for all types of faults and at all possible load levels;

ii) The probabilistic accuracy for four different types of faults at all possible locations and all possible load levels; iii) The probabilistic accuracy at six different load levels for all types of fault and at all possible locations. They can be calculated using (5.3) to (5.5) and the results can indicate to what extent each of them affects the accuracy of prediction.

$$A_i = \sum_{j=1}^{4}\sum_{k=1}^{8} A_{ijk} \times P_{ijk} \tag{5.3}$$

$$A_j = \sum_{i=1}^{25}\sum_{k=1}^{6} A_{ijk} \times P_{ijk} \tag{5.4}$$

$$A_k = \sum_{i=1}^{25}\sum_{j=1}^{4} A_{ijk} \times P_{ijk} \tag{5.5}$$

5.6 Results and Discussion

The evolutions of different types of probabilistic accuracy of prediction according to the time after the fault is cleared are presented in Figure 5-2 to Figure 5-5. The data in these figures corresponds to the accuracy of prediction by the DT algorithm made at different times following the fault clearance. The algorithm detects an event and waits for a pre-determined time to use the output from one time sample to make a prediction. As shown in all these figures, the level of DT prediction accuracy increases with time. After the clearance of the fault, the longer the DT waits, the more confident it will be in making the decision about system transient stability. However, the sooner the prediction is made the longer the time available to take control actions to avoid loss of stability will be; therefore these figures can help balance the trade-off between the decision making time and the acceptable level of accuracy of prediction.

Looking first at Figure 5-2, the 25 overlapped curves show that for any fault type and at any system load level, the probabilistic accuracy of prediction of the DT varies

depending on the fault location. However, it can also be seen that based on the probability distribution assumed for this test system, the effect of the fault location on the probabilistic accuracy of prediction is very small. All of the 25 accuracies go above 98% after about 0.2 s and the differences between their values are low (within 2%).

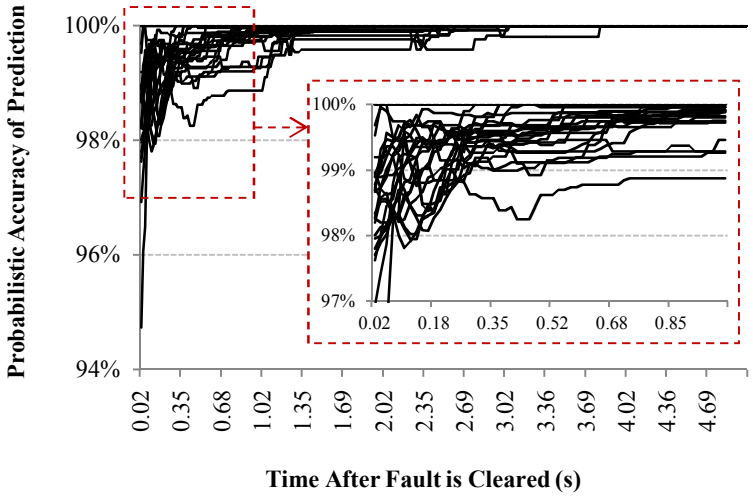


Figure 5-2: Probabilistic accuracy of prediction at 25 different possible locations.

Figure 5-3 presents the curves of probabilistic accuracy of prediction for the four different types of fault, at any possible location and any possible system loading level. In fact, unbalance faults in the test system cause fewer stability problems since the system itself is designed to withstand the "worst case" scenario. The number (or percentage) of unstable simulations increases when the fault type varies from single line to ground to three phase fault, as listed in Table 5-5. Therefore, the straight line at the top of Figure 5-3 which has the value of 100% throughout the 5 s, demonstrates the high reliability of the DT model to predict the most frequent type of fault, single line to ground fault, correctly, i.e., to predict stable simulations as stable. For line to line and double line to ground fault, which cause comparatively more unstable simulations, the probabilistic accuracies are approximately 98% when predictors are

sampled immediately after the fault is cleared, and reaches more than 99% after about 0.5 s. For the most severe but least frequent type of fault, i.e., the three-phase fault, the probabilistic accuracy of prediction becomes lower. It is only around 90% immediately after the clearance of the fault and does not increase to above 98% until 0.8 s following the fault clearance.

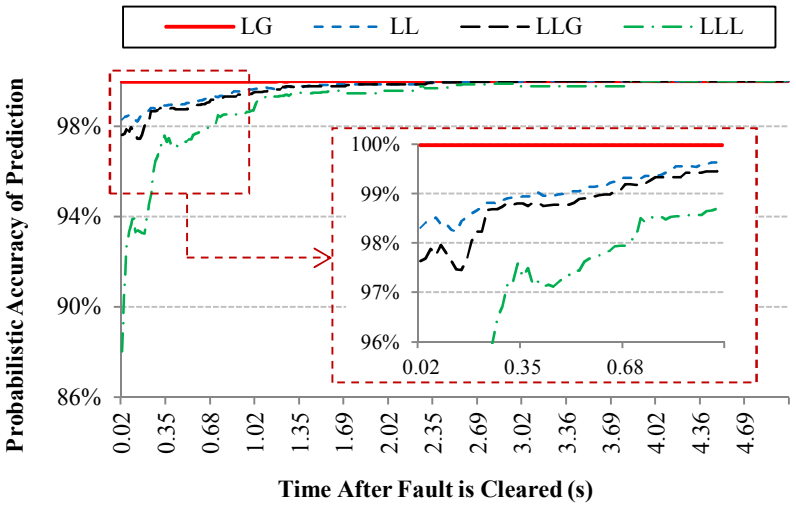


Figure 5-3: Probabilistic accuracy of prediction for 4 different types of fault.

Table 5-5: Simulation results based on fault types.

Fault Type	*Stable*	*Unstable*
LG	1949 (99.95%)	1 (0.05%)
LL	1818 (93.23%)	132 (6.77%)
LLG	1778 (91.18%)	172 (8.82%)
LLL	1420 (72.18%)	530 (27.18%)

In Figure 5-4, the accuracy of prediction at six different system loading levels for any type of fault and at any possible location is shown. When the LF is lower, the system operates at a less stressed level and is further away from its stability limit. A smaller number (or percentage) of unstable simulations are generated, as detailed in Table 5-6. The curves in Figure 5-4 for LF from 0.681 to 1.03 are all above 99% immediately

after the fault is cleared, which demonstrates the DT model's high reliability to predict the system at normal loading level. When LF is 1.114, the probabilistic accuracy also exceeds 98% after approximately 0.2 s. When the system is under extremely unfavourable operating conditions and the load level is at its maximum value, approximately 35% of the faults simulated (of all types of fault and at all possible locations) are unstable. The prediction accuracy of DT can only achieve 92% accuracy immediately following the fault clearance, and does not go above 98% until 0.9 s.

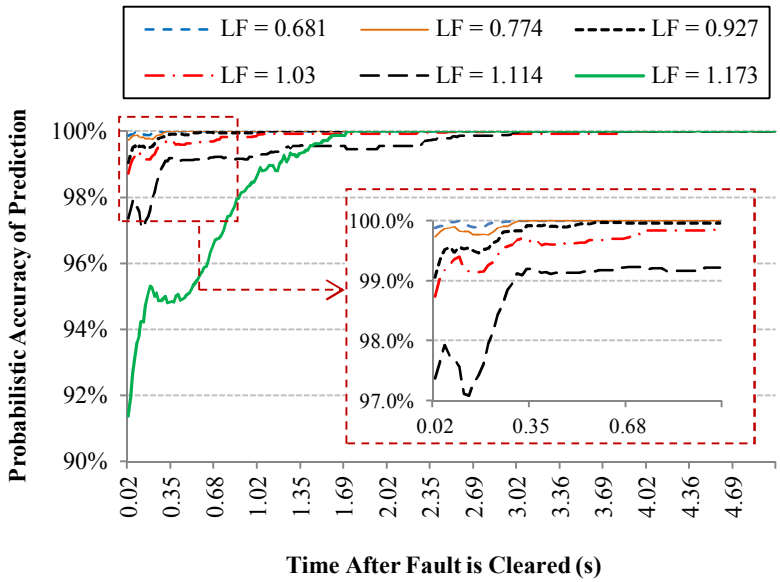


Figure 5-4: Probabilistic accuracy of prediction at 6 different loading levels.

Table 5-6: Simulation results based on system loading factor.

LF	Stable	Unstable
0.681	1289 (99.15%)	11 (0.85%)
0.774	1282 (98.62%)	18 (1.38%)
0.927	1253 (96.38%)	47 (3.62%)
1.03	1197 (92.08%)	103 (7.92%)
1.114	1100 (84.62%)	200 (15.38%)
1.173	844 (64.92%)	456 (35.08%)

Finally, Figure 5-5 shows the overall accuracy of prediction of the DT model developed in Section 5.4.3. The curve in Figure 5-5 is obtained by applying different weights on the curves in Figure 5-2, Figure 5-3 or Figure 5-4, according to the probabilities listed in Table 5-1, Table 5-2, or Table 5-3 respectively. It can be seen that this DT model is generally very reliable for making the right decision under emergency conditions of the test system, considering all four probabilistic factors. For any type of fault at any possible location and any system load level, the DT predicts the system stability status with 99% accuracy within 0.2 s after the clearance of fault, and close to 100% accuracy after about 2.5 s.

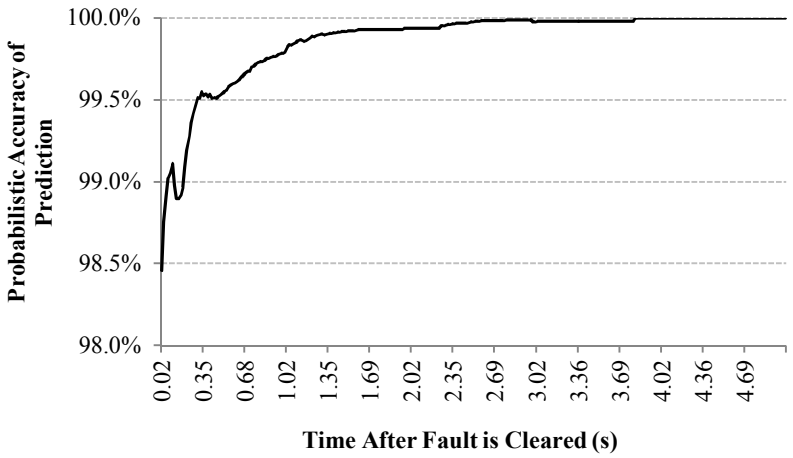


Figure 5-5: Overall probabilistic accuracy of prediction.

5.7 Summary

This chapter presented a probabilistic framework for the assessment of a data mining model's accuracy of prediction for on-line transient stability. This has been applied to the 16-machine, 68-bus NETS-NYPS test network. Although the DT based methodology was used to illustrate the proposed approach, the probabilistic framework for the assessment itself is data mining technique independent.

The training of DT was only based on three-phase faults at different buses and at the given nominal load level of the test system. However, for assessing the probabilistic

accuracy of prediction, the probabilistic nature in fault type, fault location, fault clearing time and system load level are all taken into consideration. Although probability distributions of these factors in the test network were assumed arbitrarily, historical data of existing power systems can be used for practical applications if available.

Finally, the results were presented as the evolution of probabilistic accuracy of prediction according to the time after the fault clearance. It has been shown that the model developed in this study is generally very reliable.

Both the DTs in this and the previous chapter only predict power system transient stability to assist with corrective control. The next chapter, however, will go a step further in predicting the dynamic behaviour of generators in the event of instability.

6 On-line Identification of Power System Dynamic Signature

6.1 Introduction

As in much of the past research on on-line transient stability prediction using PMU measurements and data mining, the DT in both Chapter 4 and 5 predict whether the post-fault power system remains stable or goes unstable. The post-fault behaviour of the system is classified only into two classes during the off-line supervised training process.

This chapter proposes a two-stage methodology for the problem of PMU-based on-line identification of power system dynamic signature. After a transient disturbance is cleared in real-time, the first stage is to predict the transient stability status using traditional binary classification. If the system is determined to be unstable, then the second stage predicts detailed generator dynamic behaviour. A novel methodology is developed in this chapter to implement the second stage. It is also based on data

mining, but firstly applies unsupervised learning to pre-process the off-line simulated database of unstable contingencies, and then trains multiclass classifiers through supervised learning. A variety of multiclass classification techniques, including DT, EDT and multiclass SVM (MSVM), are compared in order to determine the most suitable one for the task in hand. The proposed two-stage methodology is demonstrated on the test network.

This chapter presents the *second, third, and fourth original contribution* of this thesis.

6.2 Proposed Methodology

6.2.1 Generating a Library of System Dynamic Responses

To predict the dynamic signature of a system in real time, the first step is to produce a library of system dynamic responses (characterised by post-disturbance rotor angle swings of individual generators) by performing extensive off-line contingency simulations. The probabilistic nature of phenomena considered is taken into account to ensure comprehensiveness of the library. The uncertainties associated with system topology, loading levels, generation capacities [129, 130], type and location of disturbance and fault clearing times need to be modelled. The library generation procedure is shown by the flow chart of Figure 6-1.

The sampling of uncertain factors can be uniform or random according to their individual probability distributions which are obtained from historical or forecasted 24-hour data [99]. In the contingency analysis, the length of time domain simulation after the fault clearance is dependent on the dynamic behaviour of the system. It should be long enough so that: i) The grouping of generators are fully developed and can be clearly distinguished in most of the contingencies; ii) Cases in which some of the generators experience loss of synchronism towards the end of the simulation are

avoided.

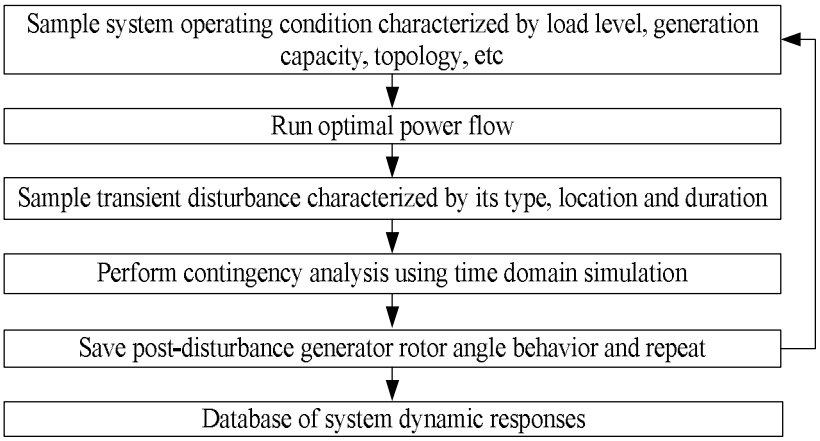


Figure 6-1: Procedure for generating a library of system dynamic response.

6.2.2 Developing a Binary Classifier for Transient Stability Status

A binary training database is firstly constructed from all contingencies in the library. Generator rotor angles are used as predictors and system transient stability status is the target of prediction. The criterion to distinguish between stable and unstable contingencies is the difference between rotor angles of any two generators in the system [55, 62]. If this difference exceeds 360 degrees within a certain simulation time the system is labelled as unstable, otherwise it is stable. A classification technique, C5.0 algorithm of DT, is then applied to train a binary classifier to be used in the first stage of on-line application.

6.2.3 Developing a Multiclass Classifier for Unstable Dynamic Behaviour

In the second stage of the proposed methodology, only the unstable contingencies from the library are further processed. The characterization of unstable dynamic behaviour involves two off-line steps:

Step 1. The patterns of unstable dynamic behaviour are identified within the database using an unsupervised learning approach. Each contingency is labelled with the

category of pattern. The purpose of this step is to find prototype examples in the training database so that the on-line prediction task becomes a supervised classification problem.

Step 2. Following Step 1, multiclass classifiers are trained through supervised learning. The predictors are again generator rotor angles whilst the target is the pattern of unstable dynamic behaviour, instead of transient stability status.

Further details about each of the above steps are given below.

6.2.3.1 Step 1: Data Pre-processing

Finding the patterns of post-disturbance rotor angle swings in a large number of simulations is very difficult for the following reasons. In each individual contingency, the post-disturbance rotor angle swings of all generators in the system need to be grouped using a clustering algorithm. The number of groups, however, varies for different contingencies and is not known in advance. To group generators for all contingencies automatically, a uniform standard that specifies under what conditions two generators should be grouped together needs to be defined. A variety of clustering algorithms have been applied in the past to identify the coherency of generators using trajectory of rotor angles or speeds, including Fuzzy C-means, Principle Component Analysis, Independent Component Analysis, Support Vector Clustering, Hierarchical Clustering (HC), etc. [106, 136-138]. In all these studies, however, the number of groups of generators exhibiting similar behaviour, or other types of parameters, need to be decided in advance for specific contingencies.

The methodology proposed in this chapter uses HC, as a pre-processing step of the supervised classification problem, to identify characteristic patterns of unstable dynamic behaviour of a power system from the database of unstable post-disturbance

system responses. The number of groups therefore is not specified in advance as was the case in past published work where clustering algorithms were used to identify the coherency of generators.

The illustrative, arbitrarily drown, post-disturbance generator rotor angle swings in one contingency are shown in Figure 6-2. They are saved as a matrix of time series in the generated database, as shown by (6.1),

$$X = \begin{bmatrix} \delta_1^{t_1} & \delta_1^{t_2} & \cdots & \delta_1^{t_n} \\ \delta_2^{t_1} & \delta_2^{t_2} & \cdots & \delta_2^{t_n} \\ \vdots & \vdots & \delta_j^{t_j} & \vdots \\ \delta_m^{t_1} & \delta_m^{t_2} & \cdots & \delta_m^{t_n} \end{bmatrix} \tag{6.1}$$

where t_n is the length of simulation time whilst m is the total number of generators. The time interval between every two data samples is selected as one cycle.

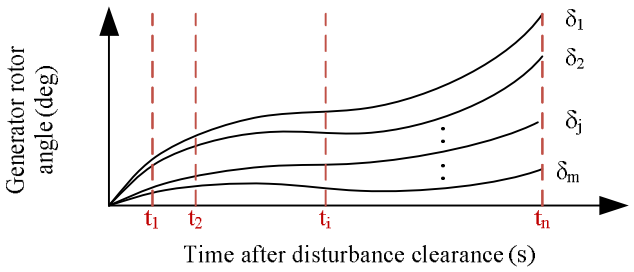


Figure 6-2: Illustration of post-disturbance generator rotor angle swings.

Two different approaches are proposed to apply HC to group the generators.

1) Approach I

In this approach, only the single sample of rotor angles at the last instant of simulation is used to identify the grouping of generators. The objects of Hierarchical Clustering therefore become m data points ranging from (t_n, $\delta_1^{t_n}$), (t_n, $\delta_2^{t_n}$) to (t_n, $\delta_m^{t_n}$). Euclidean distance [123] is used to measure the similarity between pairs of points. Complete-linkage is chosen as the linkage criterion, i.e., the distance between

clusters is determined as the distance between two elements, one from each cluster, that are farthest away from each other [123]. The unit of height in the resulting dendrogram is degree.

As 360 degrees is usually the threshold of rotor angle difference that determines the system transient stability [55, 62], it is used here as the cut-off value to form the clusters from the dendrogram of each contingency. In Figure 6-3, for instance, the first five generators are divided into three separate clusters. G4 and G5 are placed in cluster 1, G1 and G3 in cluster 2, and G2 in cluster 3. In this way, the difference between any two rotor angles in each cluster is less than 360 degrees at the end of the simulation.

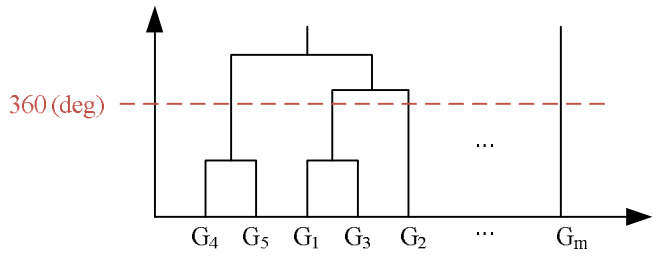


Figure 6-3: Example dendrogram constructed from data points at the end of simulation.

Applying the above approach to generators in every unstable contingency, a dendrogram can be created for each of them. By using a uniform threshold to cut all the trees, the grouping of generators, which can vary from one contingency to another, can be found for all contingencies. In this way, the number of clusters/groups does not need to be specified for each loop in advance. Instead it is identified automatically. The way in which the dendrogram is cut off represents the key novelty of the applied Hierarchical Clustering.

The procedure of Approach I for identifying patterns of generator dynamic behaviour for all unstable contingencies is presented in Figure 6-4. As indicated by the dashed

line frame, a further step can be taken on the result of Hierarchical Clustering to identify the groups of generators that lose synchronism. Whether to take this step or not depends on the aim of the study, i.e., what one wants to predict on-line for the purpose of corrective control. Within a library of responses, multiple (a limited number) patterns of unstable dynamic behaviour can usually be identified. A label/category can be assigned to each contingency according to which pattern it results in. This label will be used as the target for subsequent supervised learning.

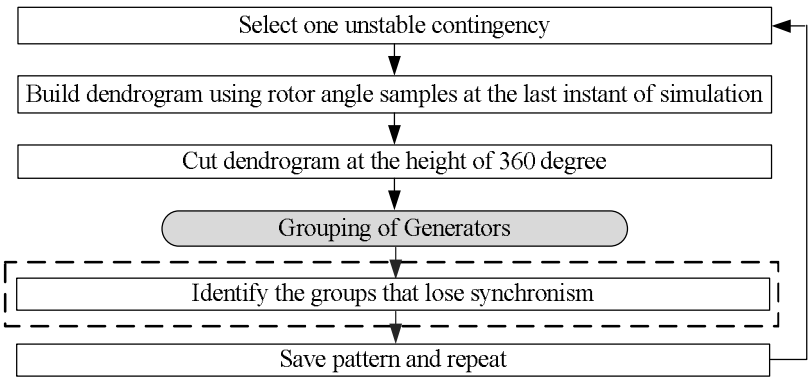


Figure 6-4: Procedure for identifying patterns of generator dynamic behaviour for all unstable contingencies: Approach I.

Although this approach is simple and effective, it has some shortcomings. It only utilises the information of rotor angles at the last instant, and does not take into account the complete rotor angle responses. When the number of Monte Carlo simulations is large, there can be a few cases in which two separate curves happen to join together around the end of the simulation and will be separated again if the simulation time is extended. In these cases, the grouping results obtained through this approach may not be optimal.

2) Approach II

The second approach proposed here considers the complete simulated swing curves of the generator rotor angles instead of one individual sample. The objects of

Hierarchical Clustering change to *m* *n*-dimensional vectors represented by the rows of matrix *X* in (6.1). Euclidean distance [123] between two vectors for generator *j* and *k* is shown in (6.2) whilst the complete-linkage criterion is used again.

$$d(\delta_j, \delta_k) = \sqrt{\sum_{i=1}^{n} (\delta_j^{t_i} - \delta_k^{t_i})^2} \tag{6.2}$$

The threshold value used to cut the dendrogram, however, cannot easily be defined. To find this value, all the stable contingencies (identified using the criterion mentioned before) within the database are collected first. The maximum Euclidean distance between any of the two rotor angle curves $d_{\max}$ is found in each of the stable contingencies. The threshold value is then determined based on the distribution of $d_{\max}$ for all stable contingencies. The procedure of this approach is summarised in Figure 6-5.

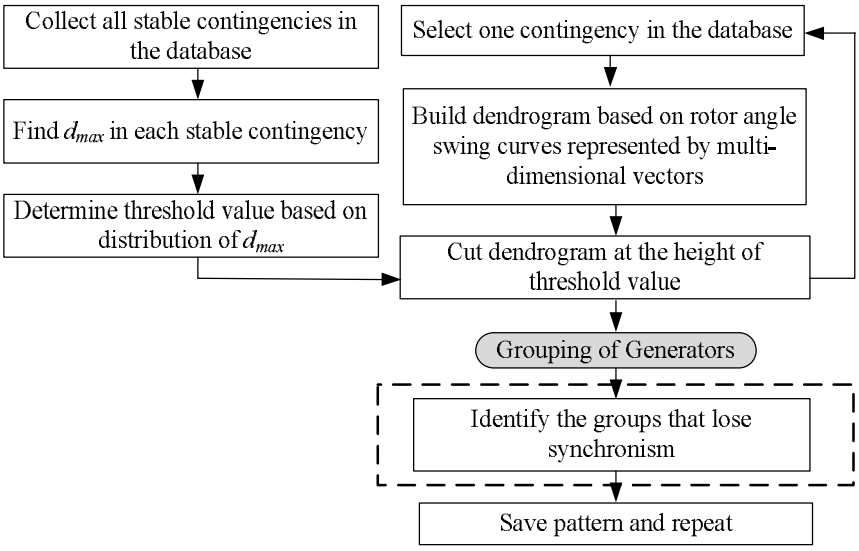


Figure 6-5: Procedure of identifying generator grouping patterns for all unstable contingencies: Approach II.

6.2.3.2 Step 2: Multiclass Classification

In order to predict which pattern the post-disturbance system will fall into using a short period of real-time rotor angle samples, a multiclass classifier needs to be

trained. As introduced in Section 3.2.5, different classification algorithms exist, some of which naturally permit the use of more than two classes. Others are by nature binary, but can be developed into multi-class [117]. The most popular algorithms that have been successfully applied in various disciplines, including DT, EDT and MSVM, are investigated and compared in this chapter.

6.2.4 Identifying Power System Dynamic Signature in Real-Time

Finally, in real-time application, after it has been detected that a transient disturbance has just been cleared (using conventional methods such as monitoring voltages, frequency or status of circuit breakers), generator rotor angles are recorded and sent to the binary classifier to predict the system transient stability status. If the system is identified as unstable, predictors are sent into the multiclass classifier. The result predicts the behaviour of rotor angles of the unstable generators at the time corresponding to the duration of the off-line simulation (i.e., t_n). If the system is defined as stable, on the other hand, appropriate damping control could be deployed if required, i.e., if the system oscillations are deemed to be poorly damped.

A summary of the complete methodology for online identification of power system dynamic signature is presented in Figure 6-6.

6.3 Test System Uncertainties

The NETS-NYPS five-area system is again used as the test network. The pre-fault system initial conditions are performed using OPF.

For the system operating conditions, the load variation is considered as the only uncertain factor in this chapter. It follows a normal distribution with nominal mean values as given in [115] and standard deviation of 3.33% (10% at 3σ). All the loads are modelled as completely dependent by scaling them with the same loading factor

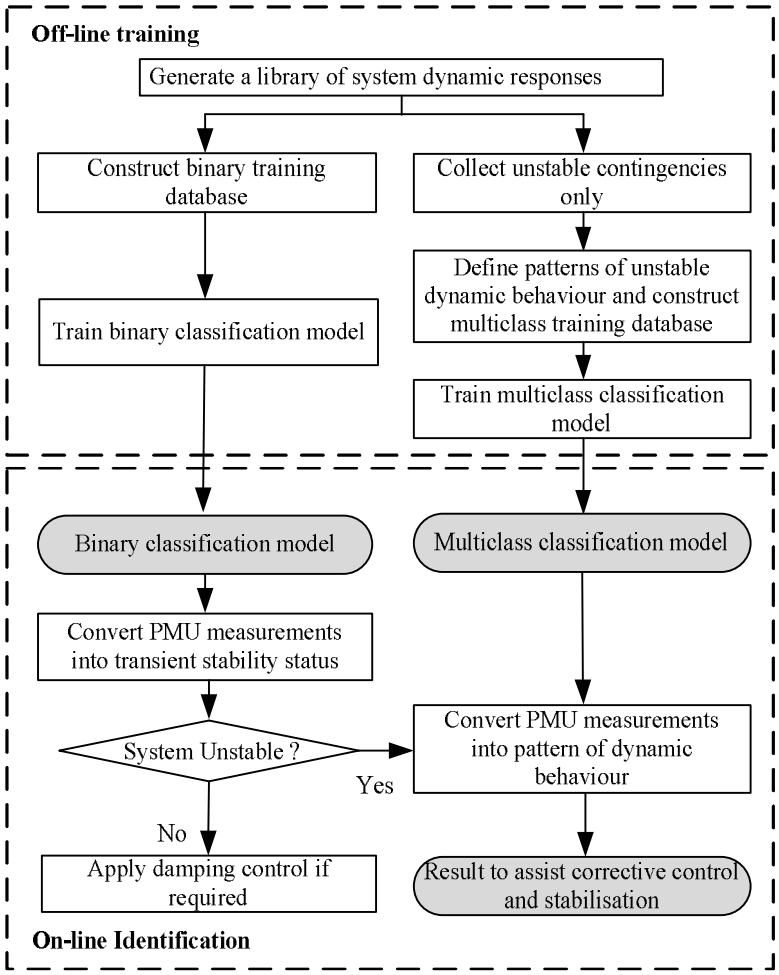


Figure 6-6: Flow chart summary of the proposed methodology.

at a time. The generation output at each loading level is determined using OPF that minimises the generation cost based on traditional quadratic cost functions for each generator. Topology change is not considered at this stage of research. Only three-phase faults are used at this stage of research as the most critical for system stability. It is assumed that all faults occur on transmission lines, and not on buses or within transformers. The faulted line is selected uniformly among the 66 available lines. The fault position along the length of the line is randomly selected following a uniform distribution. The duration of the fault is modelled as a normal distribution as in [130] with a mean value of 13 cycles and a 0.667 cycles standard deviation (2 cycles at 3σ). Although this is a long fault clearing time for high voltage systems and only realistic

when circuit breakers fail or when delayed tripping is involved [1], it is selected to generate a reasonable mix of stable and unstable conditions in the test system. The faults are assumed to be cleared without tripping any line. Clearly, many other different contingencies and cases could be simulated; however, these are deemed sufficient for the illustration of the proposed methodology.

6.4 Applications, Results and Discussions

The methodology described in Section 6.2 has been applied to the test system. The performance of it is evaluated based on two criteria. The first is how fast the prediction can be made after the fault clearance. The sooner it is, the longer the time is available for corrective control. The second is how accurate the prediction is. The misclassification of dynamic signature (post fault behaviour of the system) can lead to inappropriate control actions and in turn to instability or even system blackout.

6.4.1 Generation of Training Database

To generate the training database, 5000 contingencies are simulated using a Monte Carlo approach. The uncertain factors for each contingency, including system loading level, fault location and fault duration, are randomly sampled according to their corresponding distributions as presented in Section 6.3. In many of these cases a few generators lost synchronism after 10 to 15 s. Therefore, the test system is simulated for 20 s after the fault clearance. The time interval between samples of the rotor angle swings is 0.0167 s for the 60 Hz system (sampling rate of 60Hz). Since there are 15 generators (excluding the slack G13 with constant rotor angle), the matrix X in (6.1) for each contingency has 15 rows and 1197 columns.

In the MATLAB/Simulink environment, simulating 1000 contingencies (approximately 90% of which are stable whilst 10% are unstable) for the test system

for 20 s after the fault clearance takes about 14 hours of computation time on a PC with 2.66-GHz quad core CPU and 3.25 GB of RAM. For practical power systems, the time domain simulation can be executed by software such as DIgSILENT/Power Factory and Transient Security Assessment Tool (TSAT) as in [71] and [61], respectively.

6.4.2 Binary Classification

Table 6-1: Training databases with different length of post-fault rotor angle responses.

Training Database No.	*Length of rotor angle data in cycles (p)*	*Length of rotor angle data in second (t_p)*	*No. of Predictors (15*p)*
1	10	0.167	150
2	20	0.334	300
3	30	0.501	450
4	40	0.668	600
5	50	0.835	750
6	60	1.002	900

All 5000 contingencies are used in the training for transient stability status. Although the post-fault system is simulated for 20s, only the first p cycles (i.e., t_p s) of each generator rotor angle are used as predictors. Training databases have been created with varying values of p (varying lengths of rotor angle responses), detailed in Table 6-1. This has been completed to assess the influence of the length of the training data on the prediction accuracy. According to the criterion specified in Section 6.2.2, 438 (8.76%) contingencies are identified as unstable whilst the remaining 4562 (91.24%) are stable. Binary targets with the value of "1" (unstable) or "0" (stable) are then assigned to every contingency. As a result, the binary training database can be represented by the matrix Y, as shown in (6.3), which has 5000 rows and $(15 \times p + 1)$

columns. In (6.3), $\delta_{i,j}^{t_k}$ is the rotor angle of generator j in the i^{th} contingency at time t_k whilst T_i is the stability status of the i^{th} contingency.

$$Y=\begin{bmatrix} \delta_{1,1}^{t_1} & \delta_{1,1}^{t_2} & \cdots & \delta_{1,1}^{t_n}, & \delta_{1,2}^{t_1} & \delta_{1,2}^{t_2} & \cdots & \delta_{1,2}^{t_n}, & \cdots, & \delta_{1,15}^{t_1} & \delta_{1,15}^{t_2} & \cdots & \delta_{1,15}^{t_n}, & T_1 \\ \delta_{2,1}^{t_1} & \delta_{2,1}^{t_2} & \cdots & \delta_{2,1}^{t_n}, & \delta_{2,1}^{t_n}, & \delta_{2,2}^{t_1} & \cdots & \delta_{2,2}^{t_n}, & \cdots, & \delta_{2,15}^{t_1} & \delta_{2,15}^{t_2} & \cdots & \delta_{2,15}^{t_n}, & T_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \delta_{5000,1}^{t_1} & \delta_{5000,1}^{t_2} & \cdots & \delta_{5000,1}^{t_n}, & \delta_{5000,2}^{t_1} & \delta_{5000,2}^{t_2} & \cdots & \delta_{5000,2}^{t_n}, & \cdots, & \delta_{5000,15}^{t_1} & \delta_{5000,15}^{t_2} & \cdots & \delta_{5000,15}^{t_n}, & T_{5000} \end{bmatrix} \quad (6.3)$$

6.4.3 Multiclass Classification

Only the 438 unstable contingencies are used for further analysis of unstable dynamic behaviour. Training databases have been created again according to Table 6-1. Having completed the procedure of data pre-processing in both of the two approaches outlined in Section 6.2.3.1, two types of nominal targets can be defined depending on whether the step in the dashed line frame in Figure 6-4 and Figure 6-5 is taken and assigned to each contingency. They are:

- Target I: Pattern which indicates generator grouping only.
- Target II: Pattern which indicates both, generator grouping and groups that lose synchronism.

6.4.3.1 Dynamic Signatures Identified from Approach I

Using Approach I, 12 different patterns are identified for target I, shown in Table 6-2. Pattern No. 1, with G10 and G11 becoming unstable, occurs most frequently. This is the result of many factors such as the generator controls, generator parameters, and system topology. As mentioned before, G10 to G16 have constant excitation while the other generators are equipped with AVRs. Furthermore, the inertia constants of G10 and G11 are relatively small compared to the inertia constants of G12 to G16. Therefore, G10 and G11 are the most likely to become unstable following system disturbances.

Table 6-2: Patterns of unstable dynamic behaviour in training database – Target I.

Pattern #	*No. of Contingencies*	*Grouping of Generators*
1	243	(G1-G9, G12, G14-G16) (G10) (G11)
2	110	(G1-G8, G10-G12, G14-G16) (G9)
3	2	(G1, G3-G9) (G2) (G10) (G11) (G12) (G14-G16)
4	18	(G1-G9) (G10) (G11) (G12) (G14-G15) (G16)
5	45	(G1-G9) (G10) (G11) (G12) (G14-G16)
6	1	(G1, G2, G4-G9, G12, G14-16) (G3) (G10) (G11)
7	11	(G1-G9, G14, G15) (G10) (G11) (G12) (G16)
8	4	(G1, G2, G4-G16) (G3)
9	1	(G1, G4-G9) (G2, G3) (G10) (G11) (G12) (G14, G15) (G16)
10	1	(G1-G8) (G9) (G10) (G11) (G12) (G14-G16)
11	1	(G1-G10, G12-G16) (G11)
12	1	(G1-G9, G11-G16) (G10)

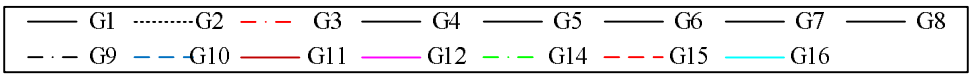
For target II, the groups of generators that lose synchronism are identified. The criterion is whether the rotor angle difference between any generator in a group and the reference generator (G13) is greater than 360 degrees within the simulation time. The number of patterns increases to 16 as detailed in Table 6-3. This indicates that for some patterns in Table 6-2, although the generator groupings of all contingencies are the same, the groups that lose synchronism can actually be different.

Examples are given in Figure 6-7 to show how the post-fault rotor angle responses look like for the patterns listed in Table 6-3.

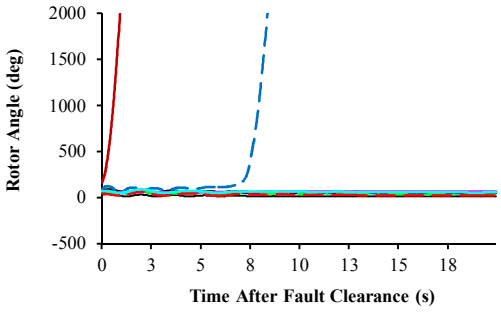
In this database, the situation where two separate curves join together at the end of the simulation, as mentioned before, does not occur due to long enough simulation after fault clearance (i.e., 20 s).

Table 6-3: Patterns of unstable dynamic behaviour in training database – Target II.

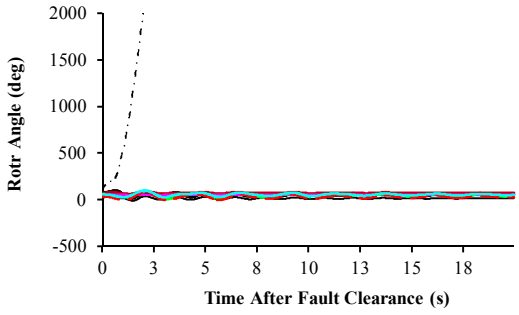
Pattern #	*No. of Contin-gencies*	*Grouping of Generators with Unsynchronised Groups Highlighted*
1	243	(G1-G9, G12, G14-G16) (G10) (G11)
2	110	(G1-G8, G10-G12, G14-G16) (G9)
3	1	(G1, G3-G9) (G2) (G10) (G11) (G12) (G14-G16)
4	1	(G14-G16) (G1, G3-G9) (G2) (G10) (G11) (G12)
5	9	(G1-G9) (G10) (G11) (G12) (G14-G15) (G16)
6	9	(G1-G9) (G10) (G11) (G12) (G14-G15) (G16)
7	9	(G1-G9) (G10) (G11) (G12) (G14-G16)
8	31	(G14-G16) (G1-G9) (G10) (G11) (G12)
9	5	(G1-G9) (G10) (G11) (G12) (G14-G16)
10	1	(G1, G2, G4-G9,G12, G14-16) (G3) (G10) (G11)
11	11	(G1-G9, G14, G15) (G10) (G11) (G12) (G16)
12	4	(G1, G2, G4-G16) (G3)
13	1	(G1, G4-G9) (G2, G3) (G10) (G11) (G12) (G14, G15) (G16)
14	1	(G14-G16) (G1-G8) (G9) (G10) (G11) (G12)
15	1	(G1-G10, G12-G16) (G11)
16	1	(G1-G9, G11-G16) (G10)



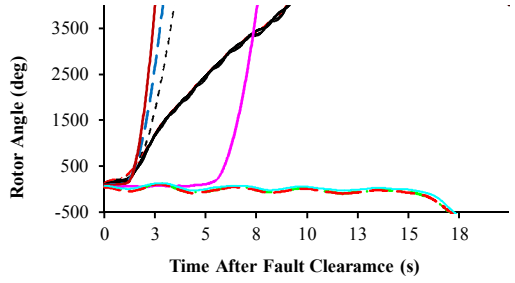
(a) Pattern 1



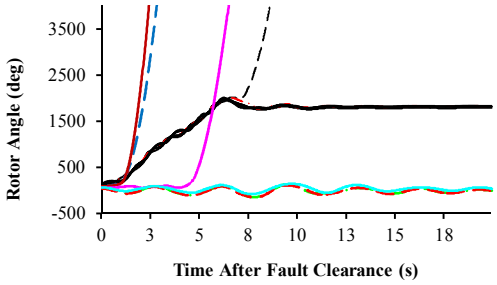
(b) Pattern 2



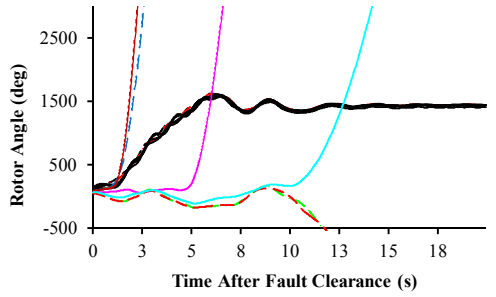
(c) Pattern 3



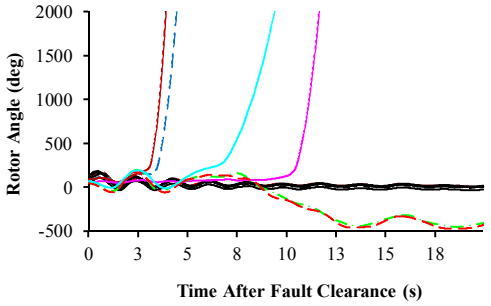
(d) Pattern 4



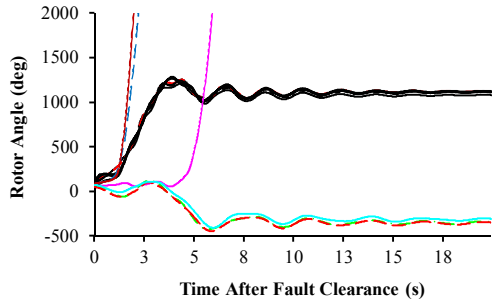
(e) Pattern 5



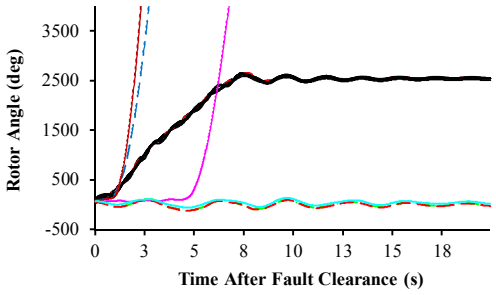
(f) Pattern 6



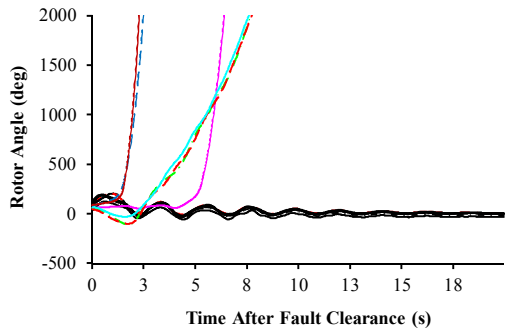
(g) Pattern 7



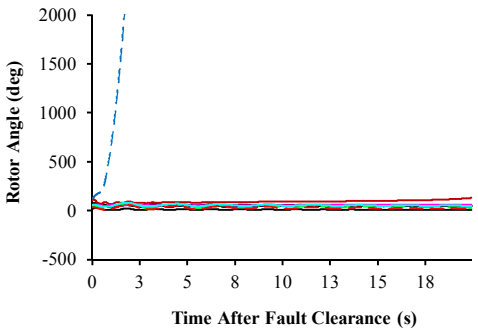
(h) Pattern 8



(i) Pattern 9



(j) Pattern 10



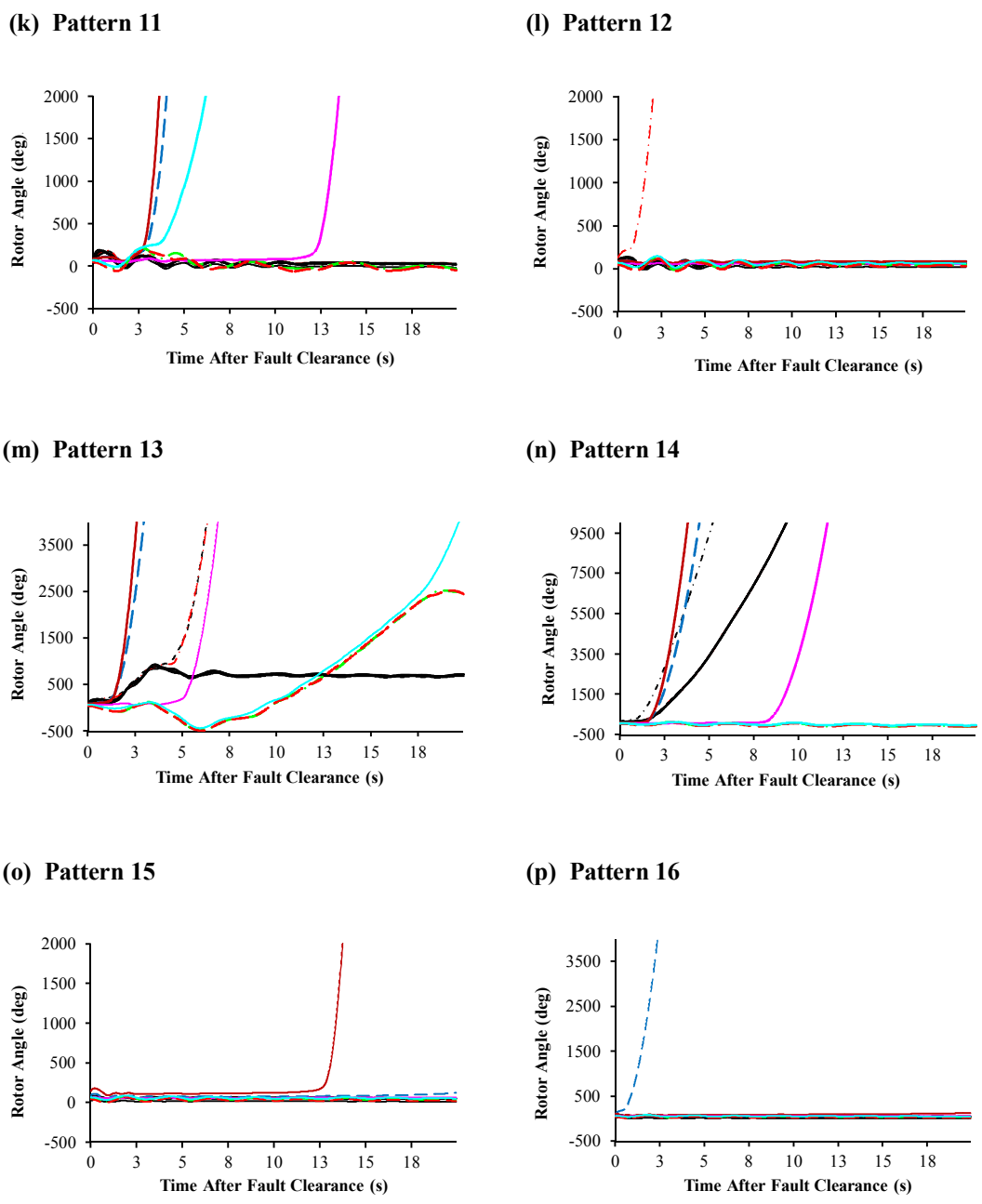


Figure 6-7: Examples of post-fault rotor angle behaviour for the 16 patterns listed in Table 6-3.

6.4.3.2 Dynamic Signatures Identified from Approach II

The distribution of d_{max} which is the maximum Euclidean distance between any of the two rotor angle curves in one stable contingency is shown in Figure 6-8. It can be seen that the highest value of d_{max} among the 4562 stable cases is less than 2800. Since the database may not contain any marginally stable cases, 3000 is selected as the threshold distance to cut the dendrogram. The dynamic signatures identified in

this approach (for both targets) are exactly the same as those in the previous approach, as shown in Table 6-2 and Table 6-3. It is important to emphasise that the threshold value here is not suitable for different applications. It depends on the dynamic behaviour of the test system and the length of post-fault simulation.

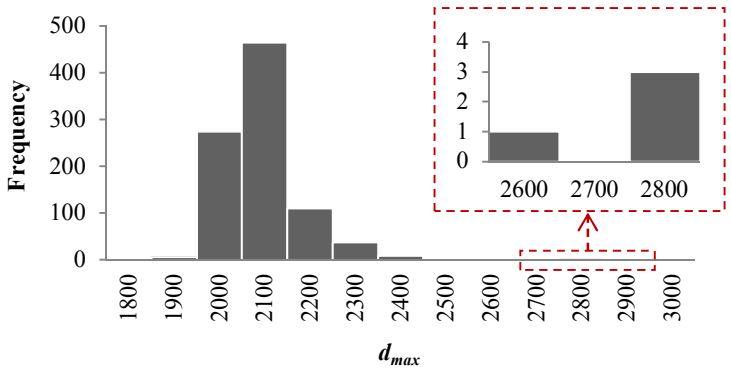


Figure 6-8: Frequency distribution of d_{max} in 4562 stable contingencies.

Nominal targets are then assigned to every contingency, the value of which range from 1 to 12 for target I and from 1 to 16 for target II. The multi-class training database can be represented by the matrix Z, as shown in (6.4), which has 438 rows and $(15\times p+1)$ columns. In (6.4), $\delta_{i,j}^{t_k}$ is the rotor angle of generator j in the i^{th} contingency at time t_k whilst T_i is the pattern number of the i^{th} contingency.

$$Z=\begin{bmatrix} \delta_{1,1}^{t_1} & \delta_{1,1}^{t_2} & \cdots & \delta_{1,1}^{t_n}, & \delta_{1,2}^{t_1} & \delta_{1,2}^{t_2} & \cdots & \delta_{1,2}^{t_n}, & \cdots\cdots, & \delta_{1,15}^{t_1} & \delta_{1,15}^{t_2} & \cdots & \delta_{1,15}^{t_n}, & T_1 \\ \delta_{2,1}^{t_1} & \delta_{2,1}^{t_2} & \cdots & \delta_{2,1}^{t_n}, & \delta_{2,1}^{t_n}, & \delta_{2,2}^{t_1} & \cdots & \delta_{2,2}^{t_n}, & \cdots\cdots, & \delta_{2,15}^{t_1} & \delta_{2,15}^{t_2} & \cdots & \delta_{2,15}^{t_n}, & T_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \delta_{438,1}^{t_1} & \delta_{438,1}^{t_2} & \cdots & \delta_{438,1}^{t_n}, & \delta_{438,2}^{t_1} & \delta_{438,2}^{t_2} & \cdots & \delta_{438,2}^{t_n}, & \cdots\cdots, & \delta_{438,15}^{t_1} & \delta_{438,15}^{t_2} & \cdots & \delta_{438,15}^{t_n}, & T_{438} \end{bmatrix} \quad (6.4)$$

6.4.4 Building the Models

All the DTs except RFs used in this study are built with IBM SPSS Modeler. 10 basic trees are built for each boosted model. RFs are built using MATLAB Statistics and Machine Learning Toolbox. 5 predictors are selected at random for each decision split and 50 basic trees are generated for each RF.

LIBSVM [126] is used to develop all the Multiclass Support Vector Machine

(MSVM) models. As described in Section 3.2.2, the kernel function selected work is Radial Basis Function (RBF). The two parameters in RBF that need to be set for a given training set are C and σ. The optimal values of them are firstly found using a grid-search and a 5-fold cross-validation process, and then used to train the whole training database.

6.4.5 Generation of Testing Databases

A further 2000 contingencies which were not included in the training database are simulated for testing, using the Monte Carlo approach with the same probability distribution. 167 (8.35%) of these contingencies are unstable whilst 1833 (91.65%) are stable. Furthermore, the patterns of dynamic behaviour that each unstable contingency belongs to are labelled so that they can be compared with the classifiers' predictions. Within the 167 unstable contingencies, 165 fall into the patterns that are included in the training database, as detailed in Table 6-4 and Table 6-5 for target I (i.e., Pattern which indicates generator grouping only) and target II (i.e., Pattern which indicates both, generator grouping and groups that lose synchronism), respectively. The remaining 2 which do not occur during training are listed in Table 6-6.

Table 6-4: Patterns of unstable dynamic behaviour in testing database included in training – Target I*.

Pattern #	1	2	3	4	5	7	8
No. of Contingencies	92	40	1	9	20	2	1

* Pattern which indicates generator grouping only.

Table 6-5: Patterns of unstable dynamic behaviour in testing database included in training – Target II*.

Pattern #	1	2	4	5	6	7	8	9	11	12
No. of Contingencies	92	40	1	2	7	3	14	3	2	1

*Pattern which indicates both, generator grouping and groups that lose synchronism.

Table 6-6: Two patterns of unstable dynamic behaviour in testing database excluded in training.

No. of Contingencies	*Grouping of Generators with Unsynchronised Groups Highlighted*
1	(G1, G4-G9) (G2, G3) (G10) (G11) (G12) (G14-G16)
1	(G1-G8) (G9) (G10) (G11) (G12) (G14, G15) (G16)

6.4.6 Results and Discussions

6.4.6.1 Transient Stability Status

For each of the 6 binary training databases (of varying length), a separate DT is trained using C5.0. The 2000 test data are sent into the DTs as input. The speed of processing is extremely fast so that in real time application more time can be saved deploying control actions. The computation time required to obtain the output of each contingency is in the order of 10^{-5} seconds using the same PC as previously described.

Table 6-7: Classification accuracy for transient stability status.

Length of rotor angle (cycle)	*Overall Accuracy*	*Stable (classified as stable)*	*Unstable (classified as unstable)*
10	98.7% (1974/2000)	98.96% (1814/1833)	95.81% (160/167)
20	98.85% (1977/2000)	99.02% (1815/1833)	97.01% (162/167)
30	99.25% (1985/2000)	99.45% (1823/1833)	97.01% (162/167)
40	99.55% (1991/2000)	99.73% (1828/1883)	97.6% (163/167)
50	99.75% (1995/2000)	99.84% (1830/1883)	98.8% (165/167)
60	99.75% (1995/2000)	99.84% (1830/1883)	98.8% (165/167)

The results are summarised in Table 6-7. Separate accuracies are listed for stable and unstable contingencies. The plots in Figure 6-9 show that the accuracy of prediction increases with the amount of data used. The longer it waits to collect the predictors after the fault clearance, the more confident it will be in making the right decision

about the system transient stability status. The binary classification method is demonstrated to be very effective. At 10 cycles (0. 167 s) after the fault clearance, close to 99% of the stable contingencies and more than 95% of the unstable ones are correctly identified. At 60 cycles (1 s) after the fault, only 3 out of 1883 stable cases and 2 out of 167 unstable cases are misclassified.

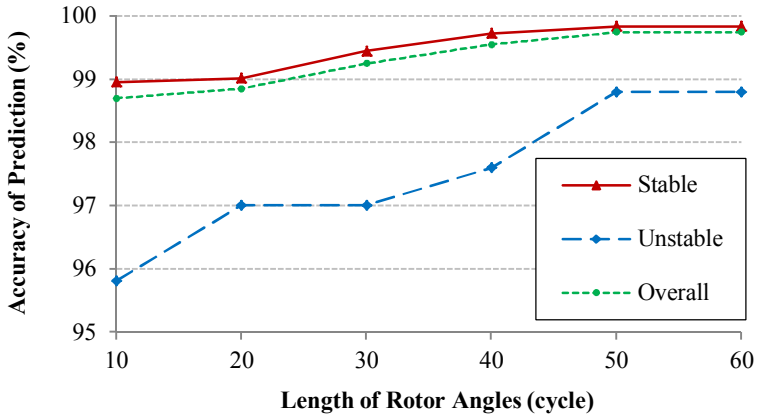


Figure 6-9: Variation of transient stability status prediction accuracy with length of post-fault rotor angle responses.

6.4.6.2 Unstable Dynamic Behaviour – Target I

For each of the 6 multiclass training databases with the target of generator grouping (i.e., target I), a separate classifier is trained with the five classification techniques: DT (including CART and C5.0), EDT (including Boosting and RF), and MSVM. This results in a total of 30 multiclass classifiers. In real-time application, all the contingencies that are determined to be unstable in the first stage will be sent for further processing into the second stage. These include the unstable cases that are correctly classified and the stable cases that are misclassified. In this work, only the correctly classified unstable cases in the binary classification are used as input to corresponding multiclass classifiers for the purpose of further analysis. For instance, using 10 cycles of rotor angles, 160 out of the 167 unstable cases are correctly identified by the binary classifier (according to Table 6-7), and therefore send into the

second stage. With the CART model 134 out of these 160 cases are classified into the correct pattern of generator grouping, as indicated by (134/160) in Table 6-8. For the ensemble models, although it takes slightly longer to give the output compared to an individual model, the order of computation time for a single contingency is still 10^{-4} seconds. The results are shown in Table 6-8 and Figure 6-10.

Table 6-8: Classification accuracy for unstable dynamic behaviour – Target I*.

Length of rotor angle (cycle)	*CART*	*C5.0*	*Boosted C5.0*	*RF*	*MSVM*
10	83.75% (134/160)	87.50% (140/160)	**89.38% (143/160)**	**89.38% (143/160)**	88.13% (141/160)
20	83.95% (136/162)	88.27% (143/162)	**90.12% (146/162)**	**90.12% (146/162)**	88.89% (144/162)
30	88.27% (143/162)	89.51% (145/162)	**91.36% (148/162)**	90.12% (146/162)	89.51% (145/162)
40	87.12% (142/163)	90.18% (147/163)	**91.41% (149/163)**	90.18% (147/163)	88.96% (145/163)
50	87.88% (145/165)	90.91% (150/165)	**91.52% (151/165)**	**91.52% (151/165)**	89.70% (148/165)
60	88.48% (146/165)	91.52% (151/165)	**92.12% (152/165)**	91.52% (151/165)	90.91% (150/165)

* Pattern which indicates generator grouping only.

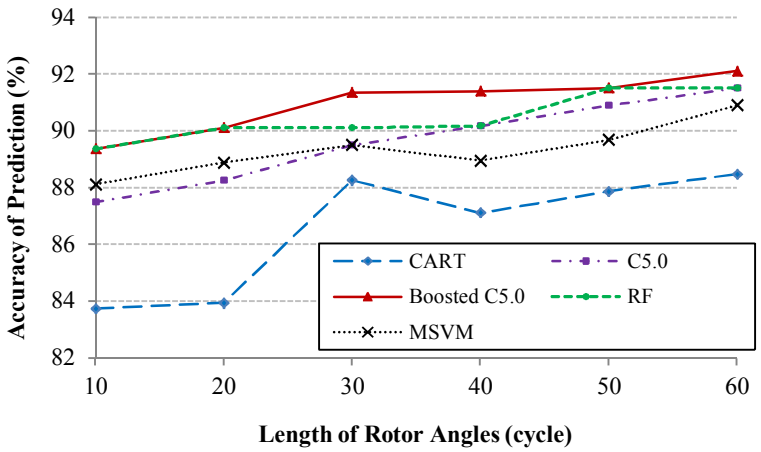


Figure 6-10: Variation of unstable dynamic behaviour prediction accuracy (target I) with length of post-fault rotor angle responses.

Looking first at the trends of curves along the x-axis in Figure 6-10, it can be seen

that the classification accuracy for patterns of generator grouping also increases with the amount of rotor angle data used. The comparison of different curves indicates that among all the multiclass classification methods applied, CART performs the worst since the accuracy of it is only 83.75% with 10 cycles of post-fault rotor angles, and does not go over 88% until 60 cycles of rotor angles data are used. C5.0 and MSVM perform better. Their accuracies are around 88% at 10 cycles after the fault clearance and over 90% at 60 cycles after the fault clearance. The performance of the two EDT methods, Boosted C5.0 and RF, is comparable and the best in this application. Their accuracies are over 89% at 10 cycles after the fault. After 20 cycles, Boosting C5.0 performs slightly better than RF.

6.4.6.3 Unstable Dynamic Behaviour – Target II

Table 6-9: Classification accuracy for unstable dynamic behaviour – Target II*.

Length of rotor angle (cycle)					
10	*20*	*30*	*40*	*50*	*60*
86.88% (139/160)	87.04% (141/162)	87.65% (142/162)	87.12% (142/163)	87.88% (145/165)	88.48% (146/165)

*Pattern which indicates both, generator grouping and groups that lose synchronism.

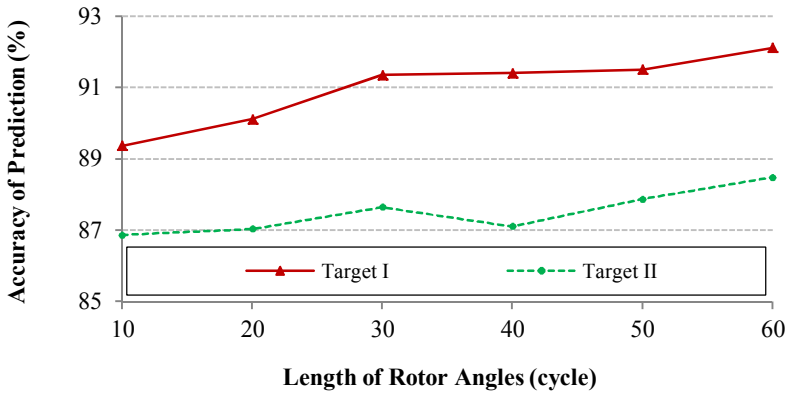


Figure 6-11: Comparison of unstable dynamic behaviour prediction accuracy for target I and II using Boosted C5.0.

For the training databases with the target that indicates both generator grouping and unsynchronised groups (i.e., target II), only Boosted C5.0 is applied because of its

best performance in the previous cases. The testing results are listed in Table 6-9. Figure 6-11 compares the accuracy of Boosted C5.0 when classifying the unstable contingencies for the two different targets. The accuracy is lower for target II as expected but still about 87% at 10 cycles after the fault clearance. The reason for the lower accuracy is discussed in the next subsection.

6.4.6.4 Discussions: The Multiclass Unbalance Problems

It is clear from the above results that the accuracy of on-line identification of dynamic behaviour for unstable contingencies is not as good as that of identifying transient stability status. This is because the multiple classes increase the data complexity and negatively affect the classification performance. More importantly, the multiclass training data is highly unbalanced. Many of the classes are extremely underrepresented compared to others due to the low probability of such patterns. As shown in Table 6-2, for target I, the first two classes have 353 cases in total whilst the remaining 85 cases are distributed in 10 different classes. There are 5 classes with only one case. This makes the conventional classification methods less effective since they tend to achieve high accuracy by always predicting the majority classes, and therefore the minority classes are often misclassified. The data distribution is even more skewed for target II as shown in Table 6-3, making the task of multiclass classification more difficult.

In fact, the class unbalance problem is very common in the research area of data mining [139]. A number of solutions have been proposed at the data and algorithm levels. They either change the distribution of data in the training set by over-sampling or under-sampling, or modify existing learning algorithms, e.g., apply cost-sensitive learning where the costs of errors, per class, are not equal. However, these methods work relatively well only for two-class unbalance problems, and have been shown to

be less effective or even cause a negative effect in dealing with multiclass tasks [139]. The most common ensemble methods, which are also popular and important solutions, have been investigated in this application. The results indicate that they do help to increase the accuracy but not significantly. More complicated and effective solutions remain an area for future research.

6.5 Effect of Changes in Operating Conditions

In the application of the proposed methodology, the load variation is the only uncertain factor involved in the system operating conditions. It follows a normal distribution with a given mean value, which can be considered as modelling of the short-term forecast error. Changes in system loading level and pre-fault topology, which are common in the day-to-day operation of power systems, are not taken into consideration. However, as the previous studies in Section 4.3.2 demonstrate, these two factors significantly affect the accuracy of prediction of the DT method for transient stability status, and should be carefully considered when designing the training database.

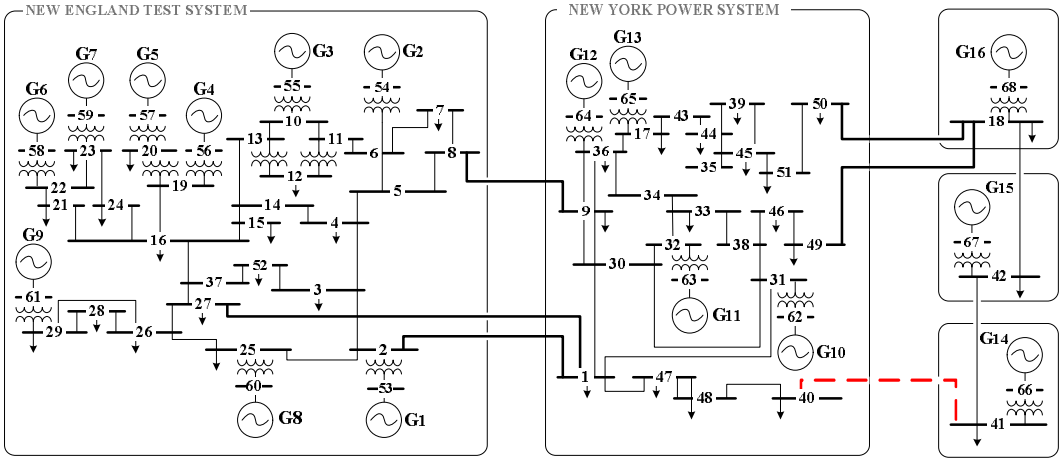


Figure 6-12: NETS-NYPS five-area test network with removed transmission line indicated.

To initially investigate the effect of pre-fault system topology change on the unstable

dynamic behaviour of power systems, a case study is carried out in which $L_{40\text{-}41}$ in the test network (the transmission line connected bus 40 and 41 as indicated in Figure 6-12) is removed from service. 1000 contingencies are simulated according to the same probability distributions of uncertain factors provided in Section 6.3, within which 185 are unstable. Table 6-10 shows the 13 patterns of unstable dynamic behaviour identified using the Hierarchal Clustering approach described in Section 6.2.3.1. These patterns only indicate generator groupings without the groups that lose synchronism.

Table 6-10: Patterns of unstable dynamic behaviour in the 185 unstable contingencies generated with pre-fault system topology change.

Pattern #	*No. of Contingencies*	*Grouping of Generators*
1	2	(G1-G10, G12-G16) (G11)
2	2	(G1-G9, G11-G16) (G10)
3	88	(G1-G9, G12, G14-G16) (G10) (G11)
4	3	(G1-G9, G12) (G10) (G11) (G14-G16)
5	10	(G1-G9, G12) (G10) (G11) (G14-G15) (G16)
6	2	(G1-G9, G14, G15) (G10) (G11) (G12) (G16)
7	1	(G1-G9) (G10) (G11) (G12, G14-G15) (G16)
8	13	(G1-G9) (G10) (G11) (G12) (G14-G16)
9	33	(G1-G9) (G10) (G11) (G12) (G14-G15) (G16)
10	17	(G1-G8, G10-G12, G14-G16) (G9)
11	5	(G1-G8, G12, G14-G16) (G9) (G10) (G11)
12	8	(G1-G8) (G9) (G10) (G11) (G12) (G14-G15) (G16)
13	1	(G1, G4-G9) (G2) (G3) (G10) (G11) (G12) (G14-16)

Comparing Table 6-10 to Table 6-2 which shows the patterns used for previous training, it can be seen that when the pre-fault system topology is changed, 6 new

types (as highlighted) of unstable dynamic behaviour occur, containing 28 unstable contingencies. After constructing all the 185 unstable contingencies into a test database, they are sent into the Boosted C5.0 multiclass classifier built in Section 6.4.4 with 60 cycles of rotor angles as predictors. The accuracy of prediction is only 63.78%. Although preliminary, this result indicates that the topology change has a significant impact on the effectiveness of the proposed methodology.

Since the target of this research is to identify system dynamic signature in real-time for corrective control, it is not feasible to train a model which is able to deal with contingencies under variations of topology and loading level over a very long time (say a year or even a month). The feasible way is to divide the time into intervals and train different models based on the most probable topologies and loading levels within each interval. As mentioned in the review of past research in Chapter 1, in [61, 63, 67, 98], a scheme is proposed to handle the changes in load and topology. In this scheme, a DT is firstly built offline for a 24-hour horizon for the system, using the prospective operating conditions in the next 24 hours which can be obtained from a short-term load forecast and network topologies. Then the time horizon is divided into periods of equal length (typically several minutes to tens of minutes). Prospective operating conditions in the next period can be predicted. If they are close to the conditions which have already been used during the previous training, the DT will remain unchanged until the next period. If new conditions appear, the DT will be updated by using new contingencies to re-train the tree.

Another framework may solve the problem of uncertainties involved in system operating conditions. Based on the historical information of a network, a set of data mining model pairs (one binary for the prediction of transient stability and one multiclass for the prediction of unstable dynamic behaviour) can be pre-trained

offline for a finite number of representative operating conditions (characterised by load, topology and some other probable uncertain factors such as patterns of renewable generation), so that each pair of models can be mapped to a particular hour of a day at a particular time of a year. In real-time application, according to the short-term forecast of system operating conditions, the most appropriate pair of models can be selected from the set. A lot of future research needs to be carried out to implement this initial idea. With this overall picture of research, the methodology proposed in this chapter can be considered as the way of building one individual model pair for a particular representative operating condition.

6.6 Summary

This chapter proposed a two-stage methodology for on-line identification of power system dynamic signature using PMU measurements and data mining.

In the available literature, the transient stability assessment for corrective control usually focuses on transient stability status without dealing with the dynamic behaviour of generators in the event of instability. The methodology presented in this chapter addresses this gap. It takes the traditional binary classification to identify system transient stability in the first stage, and establishes a novel methodology to identify the unstable dynamic behaviour in the second stage. The method firstly applies Hierarchical Clustering in order to pre-define patterns of unstable dynamic behaviour, within a database of post-disturbance system responses obtained by the Monte Carlo simulation. Two different approaches are proposed to cut off the dendrogram so that generators can be grouped based on the similarity of their rotor angle behaviour for a large number of contingencies automatically. The results are used to label the training data for on-line prediction. Different multiclass

classification techniques, including DT, EDT and MSVM, are then applied to identify characterised unstable responses.

Finally, the effect of changes in operating conditions on the proposed methodology has been initially investigated. A case study of pre-fault system topology change is presented to demonstrate significant change in the system dynamic behaviour and consequently significant reduction of the prediction accuracy of the multiclass classification. A potential analysis framework that may overcome the issue of reduced accuracy due to uncertainties associated with system operating conditions (e.g., loading level and topology) is briefly discussed and proposed for future study.

The two-stage methodology presented in this chapter can help with the selection of corrective control actions in real-time and more importantly, inform decision making regarding the most effective controlled islanding scheme.

7 Effect of Practical Issues Related to WAMS

7.1 Introduction

Until now, this thesis has shown that data mining approaches can be used to predict power systems transient stability and dynamic signature in real-time based on incoming monitoring data. Within the previous three chapters, it is assumed that signals from the WAMS are perfect, which is not realistic in a practical environment. In the past research, the effect of missing measurements [81] and errors in the signals [74, 81, 106] are investigated to demonstrate the robustness of the proposed data-driven methodologies.

This chapter, for the first time in the literature, divides the practical issues related to WAMS into five categories: including measurement error, communication noise, wide area signal delays, missing measurements, and limited number of PMUs. Each of these five will be discussed, although some investigations are preliminary. The identification of power system dynamic signature consists of two sequential steps as proposed in Chapter 6; however, all the cases studies in this chapter only deal with

the prediction of the system transient stability status.

7.2 Effect of Measurement Error and Communication Noise

Measurement error is the difference between a measured value of quantity and its true value. According to the IEEE C37.118.1 - 2011 standard, the Total Vector Error (TVE) of the phasor measured by PMU should be less than 1% [10]. For mechanical parameters such as generator rotor angles, other required devices are involved to measure the rotor position (using *Rotor Position Measurement Method*) and the limit of error is not provided in [10]. Communication noise, by contrast, occurs during the process of transmitting signals from the remote locations to the Monitoring and Control Centre (MCC).

To investigate the effect of measurement error and communication noise present in the practical signals on the performance of the DT method, the DT model which has been generated in Chapter 5 for the NETS-NYPS test network is used in this section. A large number of faults which are not included in the training database are simulated for testing. Six different system loading levels are assumed to model the variations in the demand of the test network as in Figure 5-1. The loading factors based on a given load are chosen to be 1.173, 1.114, 1.03, 0.927, 0.774 and 0.681. 25 different fault locations are arbitrarily chosen as listed in Table 5-1, including 5 generator buses and the middle points of 20 transmission lines. At each loading level and each location, four types of self-clearing fault including single line to ground, line to line, double line to ground and three phase fault are applied with a clearing time ranging from 0.05 s to 0.29 s (with an increment of 0.02 s). The post-fault system is simulated for 5.01 s for each fault. In total, 7800 faults are simulated and constructed into the original test database in the same way as in Section 5.4.2. Following 6965 (89.3%) of

these faults, the system remains stable whilst it becomes unstable for the remaining 835 (10.3%) of faults.

Since the generator rotor angles and speeds are selected as predictors, the measurement errors can come from both the terminal voltages and the rotor positions. Noise will be involved in the data transfer process. As a result, White Gaussian Noise (WGN), which is a linear addition of white noise with a constant spectral density and a Gaussian distribution of amplitude, is added to the original simulated test database with various Signal-to-Noise Ratio (SNR) [106]. Three cases are designed with the SNR of 50, 40, and 30 dB. The smaller the SNR is, the higher the level of background noise. Figure 7-1 illustrates an example of rotor angle signal with WGN.

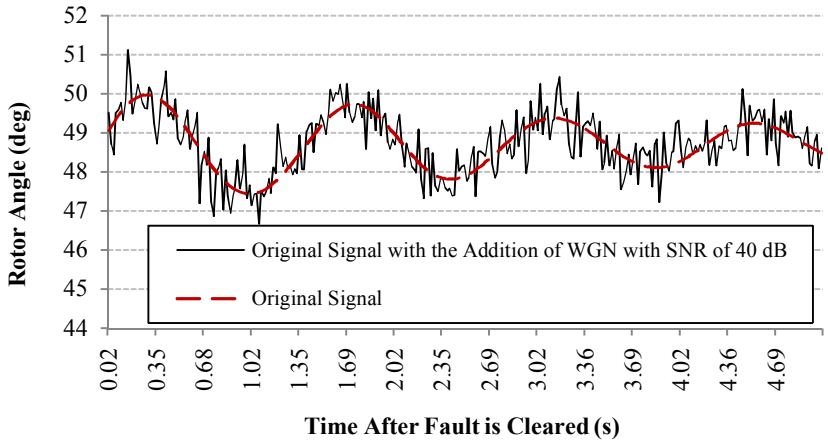


Figure 7-1: Example of generator rotor angle signal with the addition of WGN with SNR of 40 dB.

The result for each of these three cases alongside the original test database is presented in Figure 7-2. The accuracy of prediction is evaluated at every instance when decisions are made based on the values of predictors at that time, i.e., every 0.0167 s after the fault is cleared from the system. It can be seen that the accuracy drops slightly when the SNR of WGN is 50 or 40 dB, compared to that of the original test database. When the SNR goes down to 30 dB, however, the accuracy experiences a significant reduction. It oscillates approximately between 86% and 92% over the

5.01 s after the fault clearance.

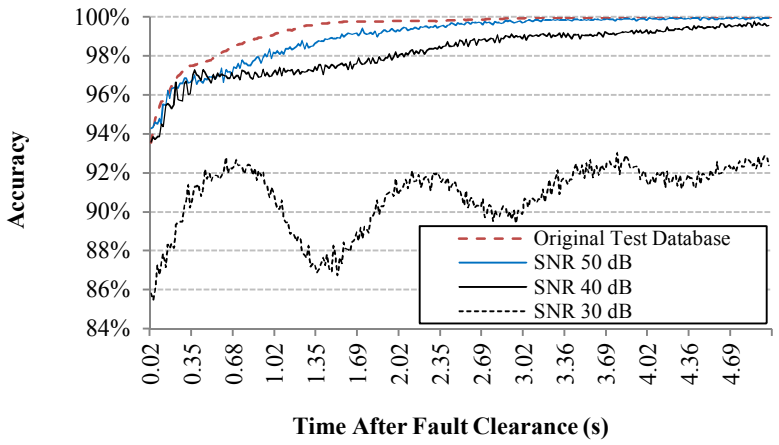


Figure 7-2: Accuracy of prediction for signals with the addition of WGN with various SNR.

7.3 Effect of Wide Area Signal Delays

The devices in the WAMS, including PMUs and PDCs, are usually geographically widespread in practical power systems. Therefore measurement signals from remote locations will often be sent through pre-existing satellite communication links as dedicated hard-wired links may prove prohibitively expensive. As such, they could potentially be subject to delay of several hundreds of milliseconds. If faster communication channels (for example, fibre optic links) are available then there should be shorter associated delays.

Due to the physical distances from various PMUs to the MCC, the measurements within the WAMS sampled at the same instant synchronised through the GPS system would arrive at the MCC after different time of delays. In Figure 7-3, for example, rotor angle signals from three different PMUs (along with their rotor position measurement devices) are shown according to the time when they arrive at the MCC, using the instant when a transient disturbance is identified to be cleared from the system as the starting point. The three signals are delayed by t_1, $t_1 + t_2$ and $t_1 + t_2 + t_3$, respectively. If the rotor angles of length t are needed for the data mining model (e.g.,

DT) to make a decision, it has to wait $t_1 + t_2 + t_3 + t$ for all the measurements to be collected (i.e., until the last signal arrives). Therefore the effect of wide area signal delays (if the delays are not randomly increased or even completely lost) is that the decision would be made until the farthest signal has been collected.

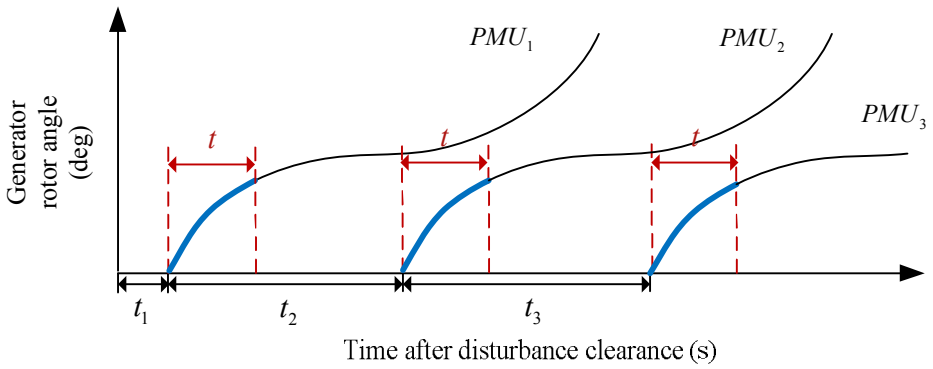


Figure 7-3: Illustration of effect of wide area signal delays.

7.4 Effect of Missing Measurements

Following the discussions above, randomly increased delays or even complete loss of signals can be further involved in practical WAMS, especially when satellite links are used. Therefore, in real-time application, some of the input data needed by the data mining model can be missing when the decision has to be made. Moreover, the unexpected failure of the PMUs or PDCs can also result in the unavailability of measurement signals. As a result, it is necessary to investigate the impact of the availability of real-time signals on the application of data mining tools.

7.4.1 Handling Missing Measurements Using Surrogate Split Method in Decision Tree

Handling missing data in the predictors during the prediction process has been a topic of great interest in the field of data mining. Various techniques have been studied along with their effect on the prediction accuracy of classification models [140]. For

example, imputation is a type of method by which an estimation of missing values is used [140]. In the on-line prediction of power system dynamic signature, one possible way of estimating the missing signals could be by treating them as targets and using the remaining PMU measurements as predictors of a regression model. However, due to the various possibilities of losing different combination of signals, a large number of regression models need to be built offline, which is time and effort consuming. Although [81] has claimed that the surrogate split method provided by CART algorithm is not effective enough due to the fact that the surrogates generated are usually parameters measured by the same PMU as the primary predictors, it will be explored again in this thesis.

7.4.1.1 The Surrogate Split Method

As introduced in Chapter 3, CART is one of the most commonly used DT building algorithms [141]. It constructs binary trees by splitting a node into exactly two "child" nodes in a top-down manner, beginning with the root node that contains the whole training data. At each node, the Gini Index is used as the criterion to determine the most useful split to best classify the data, i.e., make the resulting child node the "purest". Supposing that $X \le a$ is the best split at a node, its surrogate split is a split based on another predictor Y, e.g., $Y \le b$, such that the most similar splitting results can be achieved. During the training process, one or more surrogate splits may be generated for every tree node. Therefore if the value of X is missing for a test case, the best surrogate split will be used to decide which child node it should go to for final classification. If the best surrogate is also missing, the second best is used, and so on.

With the surrogate split method described above, the problem of missing data in the predictors can be handled during the on-line prediction of transient stability.

7.4.1.2 Applications, Results and Discussions

1) Training of Decision Tree and Generation of Test Database

In this section, the same DT model and the original test database as in the previous section are utilised. The primary predictor and its two best surrogates of each node in the model are listed in Table 7-1. These nodes are ranked according to the improvement scores of their primary predictors [119]. The greater the improvement is, the greater the reduction in impurity (measured by the Gini Index) between the parent and child nodes if that predictor is used. In other words, a high improvement generally indicates a useful split for this DT. The numbers of rotor angles and speeds in the table are the numbers of generators from which these signals are measured.

Table 7-1: Surrogates in the DT.

Node	*Primary Predictor*	*First Surrogate*	*Second Surrogate*
1	Angle 11	Speed 11	Angle 1
2	Angle 9	Speed 9	Angle 1
3	Angle 10	Speed 10	Angle 15
4	Angle 3	Speed 3	Angle 2
5	Angle 16	Speed 16	-
6	Angle 5	Speed 5	Angle 4
7	Speed 8	Angle 8	-
8	Angle 12	Speed 12	-
9	Angle 7	Speed 7	Angle 6
10	Speed 2	Angle 2	-
11	Speed 4	Angle 4	-
12	Speed 9	Angle 9	Angle 14

It can be seen from the table that the best surrogate for every primary predictor of rotor angle is its corresponding speed and vice versa. Since the rotor speed of a

generator is calculated from its rotor angle using a backward difference approximation, these two predictors certainly have the highest correlation, and are always unavailable at the same time. Therefore, the first surrogate listed in the table cannot be used in on-line applications. Furthermore, there are several nodes in the DT which do not have a second surrogate. If the primary predictors of these nodes are missing, the majority rule is used so that the child node with a larger number of training data would be chosen.

2) Effect of Missing Measurements

Although it is quite obvious that the accuracy of the prediction of transient stability would degrade when surrogate split is used in the DT, and the more useful the missing predictor is, the greater the impact on the performance of the tree, case studies are carried out to quantify the extent of this impact. Five cases are considered for illustrative purposes, as listed in Table 7-2, despite the fact that the total number of missing signals scenarios is very large.

Table 7-2: Cases of missing measurements.

Case No.	*The Generator from which Signal is Missing*
1	None
2	G4
3	G9
4	G11
5	G4, G2, G7, G12, G8, G5, G16, G3, G10, G9, G11

In Case 1, all the measurements needed by the DT are available so the original test database is used directly. For Cases 2 to 4, it is assumed that the PMU measurement is missing from only one generator. The test database for each of these cases is

obtained by replacing all the values of rotor angle and speed of that generator in the original test database by empty strings. The signal from G4 is located at the bottom node of the DT whilst the signals from G11 and G9 are located at the top two nodes, respectively. To investigate the worst case scenario, in Case 5, it is assumed that all the primary predictors in the DT (along with their first surrogate as shown in Table 7-1) are missing.

The test database for each of the five cases is sent as input into the DT model. The evolutions of the accuracy of the prediction, according to the time after the clearance of the fault for the five cases, are presented in Figure 7-4. It can be observed that there is no significant difference between the performance of the DT for the test set in Case 2 and Case 1. The accuracy of the prediction in both of these two cases reaches 96% when the assessment is made about 0.13 s after the fault is cleared, and reaches almost 100% if the assessment is made approximately 2.7 s after the fault clearance. This is not surprising since the predictor at the bottom node only contributes very slightly to the performance of the tree. In Case 3 and Case 4 where the most useful predictors are missing, however, the accuracy does not get to 96% until about 0.3 s and 0.6 s, respectively, and can only reach 99% and 98% for Case 3 and Case 4, respectively. Furthermore, the test result for Case 5 demonstrates that in the worst case, the accuracy of prediction cannot get to 96% until about 0.8 s and can only reach 96.7%.

Through closer inspection of the test results, it is found out that the reduction of the prediction accuracy when measurements are missing is mainly caused by the decrease of the accuracy for classifying the unstable simulations. Figure 7-5 shows the accuracy of the prediction of unstable simulations only. As indicated by the arrows, at 0.35 s after the clearance of the fault, about 80% of the unstable simulations are

correctly predicted in Case 1 and 2. However, only 70%, 50% and 40% of the unstable simulations are correctly predicted in Case 3, 4 and 5, respectively. Until approximately 2.7 s, all unstable simulations have been identified in Case 1 and 2 whilst only 90%, 80% and 70% of them are found in the remaining three cases. Therefore, although the reduction of the overall prediction accuracy does not seem to be huge even in the worst case scenario, Figure 7-5 indicates that the performance of the surrogate split method on handling missing signals of the DT for the on-line prediction of transient stability is in need of further improvement.

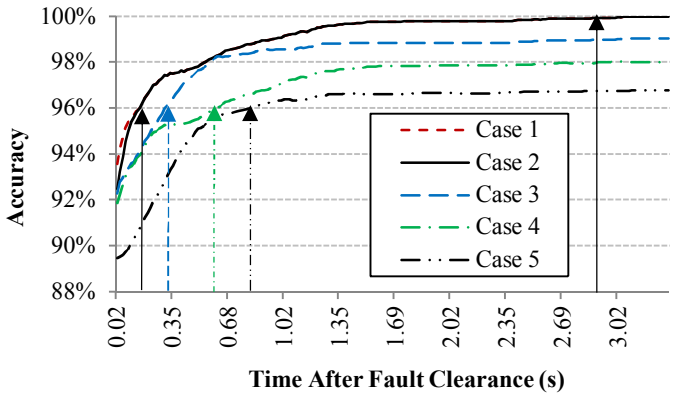


Figure 7-4: Accuracy of prediction for five cases of missing measurements.

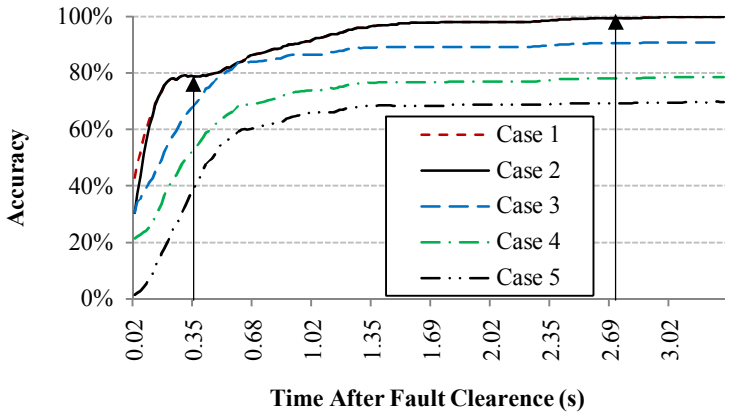


Figure 7-5: Accuracy of prediction of unstable simulations for five cases of missing measurements.

7.4.2 Comparison of Ensemble Decision Tree Method Considering Availability of PMU Measurements

This section applies a series of commonly used EDT methods, including bagging,

boosting, and Random Forest (RF) which are directly available in most of the data mining software packages, for the on-line identification of power system transient stability with missing PMU measurements. Although [81] develops a comprehensive and advanced ensemble algorithm which combines random subspace and boosting together to improve the DT's robustness to missing data, the aim of this section is to critically assess and compare these off-the-shelf and easy to use ensemble techniques for the first time, so that the most effective ones can be suggested.

7.4.2.1 Methodology for Performance Evaluation

1) Availability of PMU Measurements

As introduced in Section 1.4, the WAMS utilises the highly precise synchronous clock system GPS to build a unified time-space ordinate for the whole system. It consists of PMUs, PDCs, MCC, as well as the high-speed data communication networks [142]. It usually has a hierarchical structure as shown in Figure 1-7. The availability of the signal from an individual PMU can be calculated using (7.1),

$$A = A^{PMU} \times A^{Link_{PMU-PDC}} \times A^{PDC} \times A^{Link_{PDC-MCC}} \tag{7.1}$$

where A^{PMU} is the availability of the PMU, A^{PDC} is the availability of the PDC that this PMU is connected to, $A^{Link_{PMU-PDC}}$ and $A^{Link_{PDC-MCC}}$ are the availability of the communication link from PMU to the PDC and from PDC to the MCC, respectively.

2) Probabilistic Performance Evaluation

To predict the post-fault power system transient stability status in real-time, a library of system dynamic responses is firstly generated by performing extensive off-line contingency simulations. The probabilistic nature of power systems is taken into consideration to ensure comprehensiveness of the training database. The uncertainties surrounding initiating factors, including type and location of disturbance, the

operating time associated with fault clearing equipment, and system operating conditions at the time of disturbance inception such as system topology, loading levels, and generation capacities [129, 130], can be modelled using a numerical Monte Carlo approach according to their independent probability distribution obtained from realistic historical data or forecasted 24-hour data [99]. A training database is then constructed from the library, with which an ensemble model is trained by applying one of the EDT algorithms.

In real-time application, after it has been detected that a transient disturbance has just been cleared from the system, the on-line PMU measurements, some of which may be missing, will be sent into the Monitoring and Control Centre to make a decision. For the purpose of performance evaluation, a number of contingencies which have not been used in training are further generated and constructed into a testing database.

For a system which has N PMUs installed, assuming that signals from all of them are used as predictors in the training process, there are 2^N possible failure scenarios in total. This test is repeated for all scenarios in a loop. For the j^{th} scenario represented by $S(j)$, the testing database is sent into the EDT model M with all the predictors from signals (directly measured or derived from actually measured signals) of the failed PMUs missing, i.e., the values of these predictors are removed from every contingency. The misclassification error in this scenario is the percentage of the incorrectly classified contingencies within the test database and is represented by $e(M|S(j))$. When all the PMUs are unavailable, $e(M|S(j))$ is set to be 1. After finishing all the scenarios, an *overall probabilistic misclassification error* is calculated by (7.2) [81] to indicate the performance of the model M to identify post-fault transient stability status with missing PMU measurements.

$$e(M) = \sum_{j=1}^{2^N} Prob(S(j)) \times e(M \mid S(j)) \tag{7.2}$$

In the above equation *Prob(S(j))* is the probability of *S(j)* to happen. Assuming that all the PMU measurements in the network have the same availability (i.e., the value of A for all PMU measurements are the same), with *n* PMUs available and *(N-n)* failed, *Prob(S(j))* is calculated using (7.3).

$$Prob(S(j)) = A^n \times (1-A)^{N-n} \tag{7.3}$$

The process of evaluating the performance of an ensemble DT for on-line transient stability prediction probabilistically considering the availability of PMU measurements can be summarised by Figure 7-6.

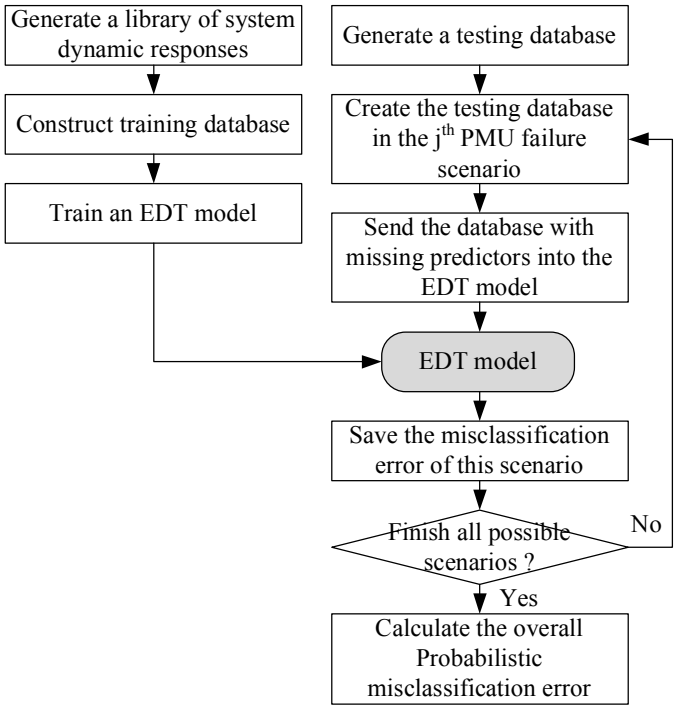


Figure 7-6: Flow chart summary of the process to evaluate the performance of an ensemble DT algorithm for on-line transient stability prediction considering the availability of PMU measurements.

7.4.2.2 Availability of PMU Measurements in the Test Network

As previously assumed, each of the 15 generators in the NETS-NYPS test network has a PMU and other required measurement devices installed so that generator rotor

angles can be directly calibrated in real time from the terminal voltage and rotor position with good accuracy. Based on the reliability modelling of PMUs using fuzzy sets in reference [81, 143], the availability A^{PMU} is in the range of [0.979975, 0.998920], and is assumed to be the same for all the PMUs in a system. All communication links from the PMUs to PDC have the same availability $A^{Link_{PMU-PDC}} = 0.999$. Moreover, the availability of all the PDCs and the communication links from the PDC to the MCC is assumed to be 1. By using (7.1), the availability of the signal from an individual PMU is $A = A^{PMU} \times 0.999 \times 1 \times 1 = 0.999 A^{PMU}$. Therefore $A \in$ [0.97897, 0.997921].

7.4.2.3 Applications, Results and Discussions

In this study, the three types of EDT, the theoretical background of which are all introduced in Section 3.2.3, are trained for the test network. Their performances with missing PMU measurements are evaluated during testing.

1) Training of Ensemble Decision Tree and Generation of Test Database

The 5000 contingencies simulated to generate the training database in Chapter 6 are used in this section for training. The first 30 cycles (i.e., 0.501 s in the 60 Hz system) of each generator rotor angle response (excluding G13) are used as predictors. The target is the transient stability status, i.e., stable or unstable. As a result, the training matrix has 5000 rows and 451 columns (15*30 = 450 columns of predictors and 1 column of target).

For each of the three ensemble methods (i.e., bagging, boosting and RF), an ensemble model is trained using CART method (with surrogate on) to build base classifiers. A single CART with surrogate is also generated as a benchmark. Since an extremely large number of signal failure scenarios needs to be tested in a loop, these four

classification models are built using MATLAB Statistics and Machine Learning Tool Box [125].

Furthermore, the 2000 contingencies which were used in Chapter 6 as the test database are utilised here for testing. Since there are 15 PMUs, the total number of possible failure scenarios is $2^{15} = 32768$. For each of the four classification models, the test is repeated for all scenarios, by sending the 2000 contingencies as input with a certain number of predictors removed. The overall probabilistic misclassification error of each classification model is calculated as presented in Section 7.4.2.1.

2) Results and Discussions

As previously described, the availability of measurements from all 15 PMUs are the same, and in the range of [0.97897, 0.997921]. The tests of classification models are performed for various values of A, and the results are illustrated in Figure 7-7. It is obvious that the ensemble models are much more robust to the missing PMU measurements compared to a single DT with surrogate. The overall probabilistic misclassification errors of the three ensemble models are almost straight lines with values under 2%. The error of the single CART model, however, is always significantly higher than 2% and only drops to about 3% when A is 0.995. The gap between the errors increases as A decreases. The error of the single CART model is about 14% when A is 0.975.

Looking into the three ensemble models, as shown in Figure 7-8, although their performances are comparable, RF is the best one with the error only slightly higher than 1% even when the value of A decreases to 0.975.

Figure 7-9 and Figure 7-10 show the performance of all four models for classifying stable and unstable cases, respectively. The overall probabilistic misclassification

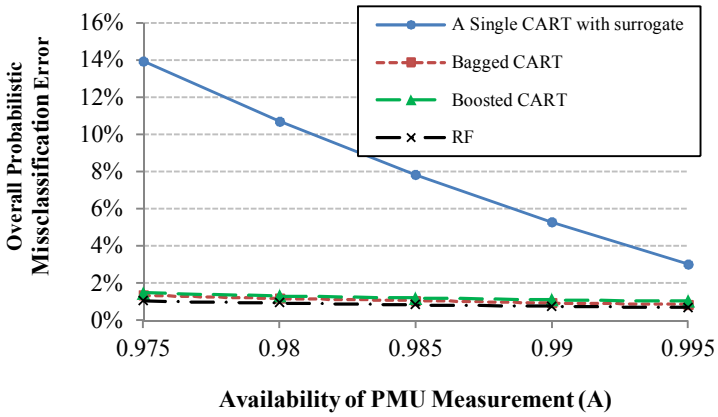


Figure 7-7: Overall performance of all the four classification models with various values of A.

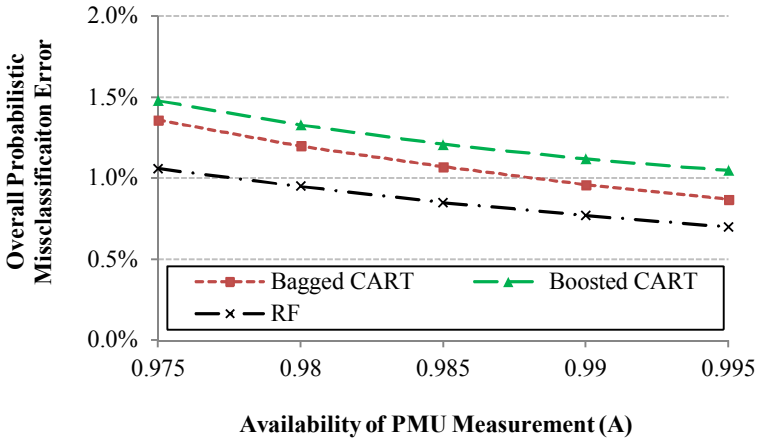


Figure 7-8: Overall performance of the three ensemble models with various values of A.

errors are calculated with equation (7.2) as well, but the error in one scenario is the percentage of the stable contingency classified as unstable in Figure 7-9, and the percentage of the unstable contingency classified as stable in Figure 7-10. It can be seen that the overall error has been reduced by using any of the ensemble methods for both, stable and unstable cases, although there is a smaller reduction in error for unstable cases especially with a low value of A. RF is again shown to be the most effective one among the three.

From Figure 7-7 to Figure 7-10 it can also be seen that the larger the availability of PMU measurements is, the more confident the models (all of them) are when making the predictions.

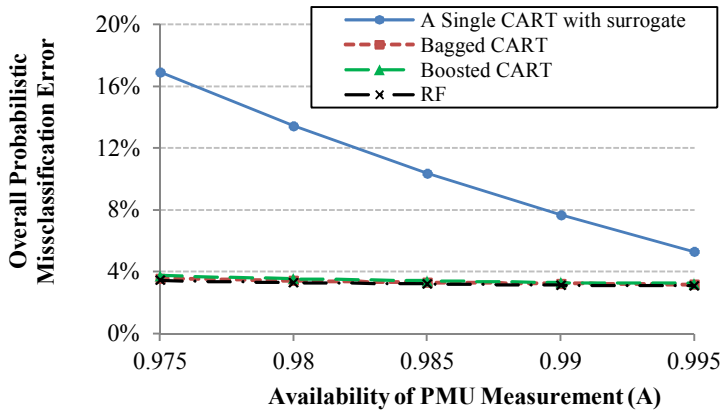


Figure 7-9: Performance of all the four classification models for classifying stable cases with various values of A.

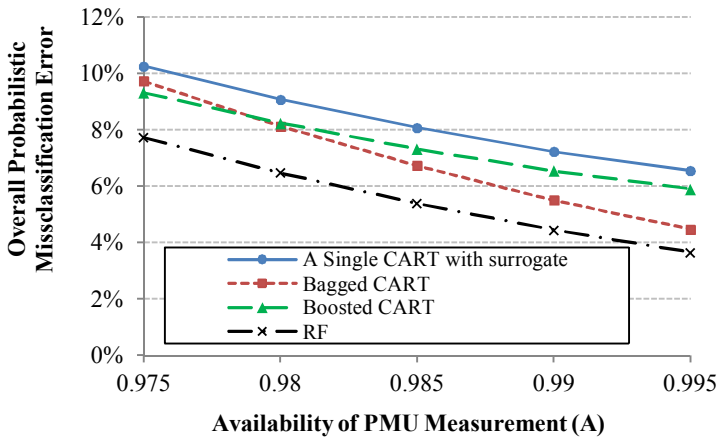


Figure 7-10: Performance of all the four classification models for classifying unstable cases with various values of A.

7.5 Effect of Limited Number of PMUs

The final practical issue that will be discussed is a limited number of PMU devices. As stated in Section 2.5.1, it is assumed that all the 15 generators (excluding the reference G13) in the test network are equipped with a PMU and other required measurement devices for generator rotor angle measurements throughout the thesis. In practical power systems, there might not be enough PMUs installed at the locations where feasible signals can be provided. Although thorough research is required, a very simple case study is provided here to reveal the possible problems to be studied.

Supposing that there are actually only four PMUs (with rotor position measurement devices) in the NETS-NYPS system, as shown in Figure 7-11, each of which is installed at the terminal bus of G3, G8, G12, and G15, respectively. The four generators are arbitrarily, randomly, selected, for initial investigation. Under such conditions, the question may arise as whether to train the data mining model with all 15 rotor angles available in the off-line simulations, or to train it with the four available rotor angles only. Using the 5000 contingencies in the training database in Chapter 6 again, two different RFs are built for binary target (i.e., transient stability status): one (RF_1) with 30 cycles of all 15 generator rotor angles (excluding G13) as predictors whilst the other (RF_2) with 30 cycles of the four available generator rotor angles. When sending the 2000 contingencies into these two RFs for testing, surely only four inputs are given. RF_1 treats the measurements it needs for the remaining 11 rotor angles as missing.

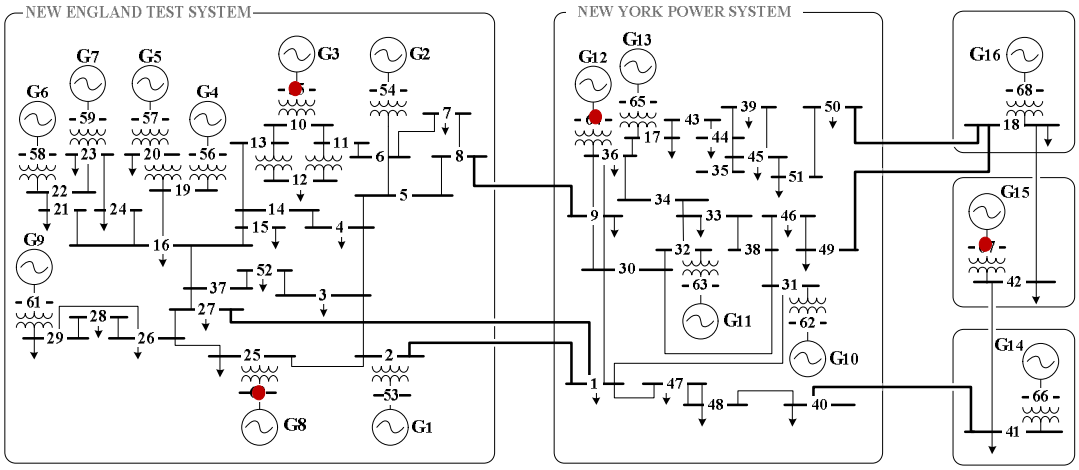


Figure 7-11: NETS-NYPS five-area test network with only four PMUs with rotor position measurement devices.

Table 7-3 lists the test results of the two RFs trained with different numbers of signals. If all 15 rotor angles are used (i.e., RF_1), with only the four PMUs (with rotor position measurement devices) in the network, all the 2000 contingencies in the test database

are classified as stable, i.e., no unstable contingencies are identified. However, if only the rotor angles from the four generators equipped with PMUs are used for training in the first place (i.e., RF_2), 98.8% stable contingencies are correctly predicted as stable and 78.44% unstable contingencies are correctly predicted as unstable.

Table 7-3: Accuracy of prediction of the two RFs trained with different numbers of rotor angle signals.

	Overall Accuracy	*Stable (Classified as Stable)*	*Unstable (Classified as Unstable)*
RF_1	91.65% (1833/2000)	100% (1833/1833)	0% (0/167)
RF_2	97.5% (1950/2000)	98.8% (1811/1833)	78.44% (139/167)

Although the study of a limited number of PMUs in the WAMS is extremely preliminary in the above example, it is shown that this practical issue will have a significant effect on the data mining methods discussed in this thesis for the on-line identification of power system dynamic signature. It is of vital importance to look into this issue more thoroughly in future study.

7.6 Summary

This chapter looked into the effect of practical issues related to WAMS on the data mining methodologies for on-line identification of power systems transient stability. Five categories of issues were discussed, including measurement error, communication noise, wide area signal delays, missing measurements, and a limited number of PMUs.

The presence of measurement error and communication noise in the real-time rotor angle signals were firstly modelled by adding WGN with various levels of SNR to the original signals. The degradation of the performance of the DT method was quantified. The accuracy of the prediction drops significantly when the SNR

decreases to 30 dB.

Secondly, the effect of wide area signal delays due to the physical distances from various PMUs to the MCC was presented. As long as the delays do not randomly increase or are not completely lost, the decision of data mining models would be made until the farthest signal has been collected. Then the problem of missing measurements was handled with surrogate split included in the CART algorithm. However, only four case studies of missing signals were carried out without investigating all the possible scenarios. The results showed that the ability of DT to identify the impending loss of stability would significantly decrease when surrogates were used and it was suggested that more effective methods be explored.

Following this, three off-the-shelf and commonly used ensemble methods for DT, including bagging, boosting and RF were investigated in order to deal with the missing measurements. A single DT with surrogate split was used as the benchmark. When evaluating their performance, all possible scenarios of missing PMU measurements were tested. A probabilistic classification error was calculated for each classification model according to the availability of PMU signals. The results showed that all three tested ensemble algorithms can significantly improve the robustness of DT decision making when dealing with the problem of missing PMU measurements, and, therefore, each of them can be used for these types of studies. Even though their performances are quite similar, the RF method was found to be slightly better than the other two.

Finally, the issue of the limited number of PMUs installed in practical power systems was preliminarily discussed. Although the case study, in which only 4 PMUs were randomly installed in the test network, was very simple, it reveals the issues to be

addressed in future study to improve the accuracy of estimation with fewer PMUs.

8 Conclusions and Future Work

8.1 Conclusions

This thesis has developed a robust methodology for on-line identification of power system dynamic signature based on incoming system responses from PMUs in WAMS. Within the methodology, data mining techniques have been used to convert real-time monitoring data into transient stability of power systems and the pattern of system dynamic behaviour in the event of instability.

The methodology is both important and relevant for system operators and controllers. The future power system may operate closer to its stability limit, and the changing types and patterns of load and generation are resulting in highly variable operating conditions. Corrective control and stabilisation is becoming a potentially viable option to enable safer system operation. Additionally, the number of WAMS projects and PMUs is rising, which will significantly improve the system situational awareness. The combination of all these factors means that it is of vital importance to exploit a new and efficient transient stability assessment tool in order to use real-time

PMU data to support decisions for corrective control actions.

An overview of the state-of-the-art of transient stability assessment revealed that a series of methodologies have been used for both off-line and on-line preventive assessment. Several methods have been previously proposed for on-line corrective use, among which data mining was identified to be the most promising. The past research surrounding the data mining approach for on-line transient stability assessment using PMU measurements was therefore thoroughly reviewed. It was found out that Decision Tree (DT) is the most suitable technique for this application.

The DT method has been implemented on the 16-machine, 68-bus NETS-NYPS test network. The idea was to train a DT model based on a database of off-line simulated contingencies, and use this model in real-time to convert PMU measurements into transient stability status after a fault is identified to be cleared. The database of contingences has been generated for the network with only three-phase faults of different lengths and at various locations. The training process was implemented using the CART algorithm, with generator rotor angles and speeds during the post-fault period as predictors. After that, the sensitivity of the prediction accuracy of the model was investigated according to four types of uncertainties in the network, including fault duration and location, system operating point and pre-fault system topology. It was shown that the developed DT model is typically able to predict system post-fault behaviour with accuracy between 88% and 92% as fast as 0.25 s after the fault clearance and with accuracy between 90% and 99% about 1 s after the fault clearance. The accuracy of prediction is more sensitive to the uncertainty in the system operating point and pre-fault network topology, than to fault duration and fault location. The implementation demonstrated that DT is a useful tool for online assessment of system transient stability and hence applicable to corrective control

approaches.

The review of past research revealed that a variety of, not very consistent, approaches has been used to assess the accuracy of data mining models for on-line prediction of transient stability. Frequently used ways of constructing test databases included selecting several individual representative contingencies, sampling one set of contingencies based on the distributions of all system uncertain factors or past experiences, and creating several sets of contingencies according to a list of uncertainties. The *first original contribution* of this thesis is a generic probabilistic framework for the assessment of the prediction accuracy of data mining models. This new framework performs an exhaustive search of possible contingencies in the testing process and weighs the accuracies according to the realistic probability distribution of uncertain system factors. The assessment results can not only inform the system operators and controllers of the confidence level of a data mining model in real-time, but also highlight the effect of different uncertainties on the accuracy of prediction. The implementation of this framework was also illustrated on the NETS-NYPS test network, using DT as the example of data mining technique. The performance of the DT model developed in this study was tested using a wide variety of disturbances with probabilistically modelled locations, durations, types of fault and the system loading levels. The accuracy of prediction is approximately 98.5% immediately following the fault clearance and can increase to almost 100% if the prediction is made 2.5 s after the fault clearance.

Although great effectiveness has been shown in the past research for the binary classification methods to classify the transient status of post-fault real-time system into stable and unstable, very little work has been done to further predict the pattern of generator grouping in the event of instability. This grouping is of interest to the

system operators since it can assist better selection of corrective control actions, and inform decision making regarding the most effective controlled islanding scheme. The *second original contribution* of this thesis is a methodology for on-line identification of power system dynamic signature. The methodology is a two-stage scheme, which takes the traditional binary classification to identify system transient stability in the first stage, and establishes a novel method to identify the unstable system dynamic behaviour in the second stage using both clustering and multiclass classification. The application of Hierarchical Clustering to a database of unstable contingencies to pre-define the patterns of generator grouping is the *third original contribution* of this thesis. Different multiclass classification techniques, including DT, Ensemble DT and Multiclass Support Vector Machine, were evaluated for the on-line prediction in the event of instability. Boosted C5.0 was identified to be the most suitable for this application. This represents the *fourth original contribution* of this thesis.

Implementation of the two-stage methodology on the NETS-NYPS test system showed that an overall prediction accuracy of over 99% was achieved in terms of transient stability status at 0.5 s after the fault clearance, using the DT trained by C5.0 algorithm. Furthermore, with the best-performing Boosted C5.0 (compared to CART, C5.0, Random Forest and Multiclass SVM), over 91% of the correctly identified unstable contingencies were then successively classified, 0.5 s after the fault clearance, into the right pattern of generator dynamic behaviour, and about 88% into the right pattern of generator dynamic behaviour with the indicated unsynchronised generator groups.

The final study in this thesis looked into the effect of practical issues related to WAMS on the data-driven methodologies developed for on-line prediction of system

dynamic signature. Five categories of practical issues were discussed, including measurement error, communication noise, wide area signal delays, missing measurements, and limited number of PMUs. Although some of the investigations were preliminary, it was found out that the performance of the DT method on on-line transient stability prediction was degraded when adding a White Gaussian Noise to the original signals to model the presence of measurement error and communication noise. The decision made by the DT model would be delayed until the farthest signal reached the Monitoring and Control Centre. In addition, the surrogate split method included in the CART algorithm was identified as not good enough for dealing with missing signals. All of the three off-the-shelf ensemble methods, including bagging, boosting and Random Forest, could significantly improve the robustness of DT, although Random Forest performed slightly better than the other two. Finally, a simple case study in which the number of PMUs installed in the test system is limited revealed that this issue has a significant effect on the data mining methods.

8.2 Future Work

The work presented within this thesis has fulfilled all of the research aims which were initially defined. Nevertheless, there are a number of areas where this research could be extended in order to further develop the ideas and methodologies which have been established.

- The data mining approach presented in this thesis requires a large database of contingency. To generate this database, both pre-fault system operating conditions and the disturbance are sampled uniformly or randomly according to their probability distribution. The original review of past research revealed some efficient sampling approaches used for the generation of training

database for on-line stability assessment in preventive mode. It would be desirable to identify or develop some of the existing or new efficient sampling techniques which can be applied in the training process of both the binary and multiclass classification for on-line identification of power system dynamic signature. These techniques can help to maximise the information content in the database whilst minimise the computing requirements.

- Two approaches of constructing the training database from the post-fault system parameters have been found in the review of past research. The first approach treats a single data point as an individual object in the training set, and is applied in this thesis in Chapter 4, Chapter 5, Section 7.2, and Section 7.4.1. The second approach, on the other hand, treats cumulative data points in one contingency as one object, and is used in Chapter 6, Section 7.4.2, and Section 7.5. These two approaches should be compared to each other, by using the same training database on the test network and the same DT algorithm to build two models, and by using the same testing database to evaluate their relative performance (including both accuracy and speed of decision making, robustness to system uncertainties, and advantages and disadvantages in practical power system implementation). A study like this may be able to suggest a better way of training to the system operators.
- Multiclass classification has been proposed in this thesis to predict the pattern of generator grouping in the event of instability. The result on the test network was not as good as that of transient stability status prediction. As well as the natural increasing difficulty in the task of multiclass classification, the highly unbalanced training data were also responsible for making the conventional classifiers less effective. Potential solutions, such as changing the training

data distribution and modifying existing learning algorithms, should be investigated to deal with the unbalance problem in multiclass classification, so that the accuracy of predicting the pattern of generator grouping can be further improved.

- Throughout this thesis, the developed methodology for on-line identification of power system dynamic signature has been implemented on a system in which the power sources are all synchronous generations. However, a growing number of conventional controlled power plants is being replaced by stochastic and intermittent renewable generation sources such as wind and solar, resulting in more uncertainties in operating conditions. It is, therefore, of vital importance to investigate the impact of non-conventional generations on the dynamic behaviour of the system, and in turn the data mining approach proposed in this thesis for identifying system dynamic signature.
- Apart from generation, other conditions such as changing load types and patterns, and system topology will also contribute to the increasing level of uncertainty. Since the target of this thesis is to identify the system dynamic signature in real-time, it may not be feasible to train a data mining model which is able to deal with contingencies under the combination of all these uncertain conditions. It would be desirable to cluster the various operating conditions of a power system within a period of time (for instance a year) into a finite number of patterns. Such work could allow a set of data mining models to be pre-trained off-line, so that the most appropriate one can be selected in real-time based on a short-term forecast of system operating conditions.
- The preliminary case study in this thesis has shown that a limited number of

PMUs have a significant effect on the data mining methods. In practice, such an issue is very likely to occur due to the design and implementation of WAMS. Further and thorough investigation on this may reveal the level of information content that the signal from each generator has, provide a minimum number of PMUs required for the data mining methods to achieve a certain level of accuracy, and suggest the optimal locations for PMU installation from the transient stability assessment point of view.

- Furthermore, in the two-stage methodology that has been proposed in this thesis for on-line identification of power system dynamic signature, a binary classifier is used to predict the transient stability status and a multiclass classifier is used to predict the generator grouping when the system loses stability. The method could be extended to predict the generators dynamic behaviour when the system remains stable. Such information can determine the level of stability and help the selection of appropriate damping controls. Similar to the generator grouping in the event of instability, patterns of generator coherency could also be pre-defined off-line through some clustering techniques. Another multiclass classifier could be trained for stable cases and used in parallel with the one for unstable cases in real-time application.
- A visualisation tool to communicate the power system dynamic signature identified in real-time to the system operators could also be developed. A user-friendly graphical interface could be designed to facilitate the visualisation and to gain a clear understanding of system transient stability and dynamic behaviour.
- Finally, the future work following this thesis should integrate the

methodologies developed here with other strands of research which have been undertaken in the Self* Network Operation and Control (SNOC) stream of the Grand Challenge Autonomic Power Systems (APS) Project funded by the Engineering and Physical Sciences Research Council (EPSRC). A promising approach is to use the pattern of generator dynamic behaviour in the event of transient instability as one of the criteria to define the "dynamic zones of control".

References

[1] P. Kundur, *Power System Stability and Control*. London: McGraw Hill, 1994.

[2] R. Preece, "A Probabilistic Approach to Improving the Stability of Meshed Power Networks with Embedded HVDC Lines," Ph.D. Thesis, the School of Electrical and Electronic Engineering, The University of Manchester, U.K., 2013.

[3] J. J. Grainger and W. D. Stevenson, *Power System Analysis*. London: McGraw-Hill, 1994.

[4] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power System Dynamics and Stability*: John Wiley & Sons, 1997.

[5] M. Pavella, D. Ernst, and D. Ruiz-Vega, *Transient Stability of Power Systems A Unified Approach to Assessment and Control*: KAP, 2000.

[6] A. G. Phadke, "Synchronized phasor measurements-a historical overview," in *IEEE Transmission and Distribution Conference and Exhibition 2002: Asia Pacific*, 2002.

[7] V. Centeno, J. D. L. Ree, A. G. Phadke, G. Michel, J. Murphy, and R. Burnett, "Adaptive out-of-step relaying using phasor measurement techniques," in *IEEE Computer Applications in Power,* vol. 6, pp. 12-17, 1993.

[8] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. New York: Springer, 2008.

[9] F. Aminifar, S. Bagheri-Shouraki, M. Fotuhi-Firuzabad, and M. Shahidehpour, "Reliability Modeling of PMUs Using Fuzzy Sets," *IEEE Trans on Power Delivery,* vol. 25, pp. 2384-2391, 2010.

[10] "IEEE Standard for Synchrophasor Measurements for Power Systems," *IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005),* pp. 1-61.

[11] X. Xiaorong, X. Yaozhong, X. Jinyu, W. Jingtao, and H. Yingduo, "WAMS applications in Chinese power systems," *Power and Energy Magazine, IEEE,* vol. 4, pp. 54-63, 2006.

[12] J. W. Stahlhut, T. J. Browne, G. T. Heydt, and V. Vittal, "Latency viewed as a stochastic process and its impact on wide area power system control signals," *IEEE Trans on Power Systems,* vol. 23, pp. 84-91, 2008.

[13] M. Chenine, Z. Kun, and L. Nordstrom, "Survey on priorities and communication requirements for PMU-based applications in the Nordic Region," in *IEEE PowerTech*, Bucharest, Romania, 2009.

[14] V. Madani, A. Vaccaro, D. Villacci, and R. L. King, "Satellite based communication network for large scale power system applications," in *Bulk Power System Dynamics and Control - VII. Revitalizing Operational Reliability, iREP Symposium*, 2007.

[15] I. Kamwa, J. Beland, G. Trudel, R. Grondin, C. Lafond, and D. McNabb, "Wide-area monitoring and control at Hydro-Quebec: past, present and future," in *IEEE Power Engineering Society General Meeting*, Montreal, Canada, 2006.

[16] T. Rauhala, K. Saarinen, M. Latvala, M. Laasonen, and M. Uusitalo, "Applications of phasor measurement units and wide-area measurement system in Finland," in *IEEE PowerTech*, Trondheim, Norway, 2011.

[17] I. C. Decker, D. Dotta, M. N. Agostini, S. L. Zimath, and A. S. de Silva, "Performance of a synchronized phasor measurements system in the Brazilian power system," in *IEEE Power Engineering Society General Meeting*, Montreal, Canada, 2006.

[18] E. M. Martinez, "Wide area measurement & control system in Mexico," in *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, Nanjing, China, 2008.

[19] P. M. Ashton, G. A. Taylor, M. R. Irving, A. M. Carter, and M. E. Bradley, "Prospective Wide Area Monitoring of the Great Britain Transmission System using Phasor Measurement Units," in *IEEE Power and Energy Society General Meeting*, 2012, pp. 1-8.

[20] J. Rasmussen and P. Jorgensen, "Synchronized phasor measurements of a power system event in eastern Denmark," *IEEE Trans on Power Systems,* vol. 21, pp. 278-284, 2006.

[21] A. B. Leirbukt, J. O. Gjerde, P. Korba, K. Uhlen, L. K. Vormedal, and L. Warland, "Wide area monitoring experiences in Norway," in *IEEE PES Power Systems Conference and Exposition* Atlanta, USA, 2006.

[22] W. Sattinger, "Application of PMU measurements in Europe TSO approach and experience," in *IEEE PowerTech*, Trondheim, Norway, 2011.

[23] Y. Ota, T. Hashiguchi, H. Ukai, M. Sonoda, Y. Miwa, and A. Takeuchi, "Monitoring of interconnected power system parameters using PMU based WAMS," in *IEEE PowerTech* Lausanne, Switzerland, 2007.

[24] M. La Scala, G. Sblendorio, and R. Sbrizzai, "Parallel-in-time implementation of transient stability simulations on a transputer network," *IEEE Trans on Power Systems,* vol. 9, pp. 1117-1125, 1994.

[25] W. Jun Qiang, A. Bose, H. Jin An, A. Valette, and F. Lafrance, "Parallel implementation of power system transient stability analysis," *IEEE Trans on Power Systems,* vol. 10, pp. 1226-1233, 1995.

[26] I. C. Decker, D. M. Falcao, and E. Kaszkurewicz, "Parallel implementation of a power system dynamic simulation methodology using the conjugate gradient method," *IEEE Trans on Power Systems,* vol. 7, pp. 458-465, 1992.

[27] V. Jalili-Marandi and V. Dinavahi, "Instantaneous relaxation-based real-time transient stability simulation," *IEEE Trans on Power Systems,* vol. 24, pp. 1327-1336, 2009.

[28] E. Vaahedi, Y. Mansour, and E. K. Tse, "A general purpose method for on-line dynamic security assessment," *IEEE Trans on Power Systems,* vol. 13, pp. 243-249, 1998.

[29] G. Aloisio, M. A. Bochicchio, M. La Scala, and R. Sbrizzai, "A distributed computing approach for real-time transient stability analysis," *IEEE Trans on Power Systems,* vol. 12, pp. 981-987, 1997.

[30] P. C. Magnusson, "The transient-energy method of calculating stability," *Transactions of the American Institute of Electrical Engineers,* vol. 66, pp. 747-755, 1947.

[31] P. D. Aylett, "The energy-integral criterion of transient stability limits of power systems," *Proceedings of the IEE - Part C: Monographs,* vol. 105, pp. 527-536, 1958.

[32] A. H. El-Abiad and K. Nagappan, "Transient stability regions of multimachine power systems," *IEEE Trans on Power Apparatus and Systems,* vol. PAS-85, pp. 169-179, 1966.

[33] F. Da-Zhong, T. S. Chung, Z. Yao, and S. Wennan, "Transient stability limit conditions analysis using a corrected transient energy function approach," *IEEE Trans on Power Systems,* vol. 15, pp. 804-810, 2000.

[34] Y. Z. Sun, Y. H. Song, and X. Li, "Novel energy-based Lyapunov function for controlled power systems," *IEEE Power Engineering Review,* vol. 20, pp. 55-57, 2000.

[35] M. Anghel, F. Milano, and A. Papachristodoulou, "Algorithmic construction of Lyapunov functions for power system stability analysis," *IEEE Trans on Circuits and Systems I: Regular Papers,* vol. 60, pp. 2533-2546, 2013.

[36] C. Jing, V. Vittal, G. C. Ejebe, and G. D. Irisarri, "Incorporation of HVDC and SVC models in the Northern State Power Co. (NSP) network; for on-line implementation of direct transient stability assessment," *IEEE Trans on Power Systems* vol. 10, pp. 898-906, 1995.

[37] V. Vittal, N. Bhatia, A. A. A. Fouad, G. A. Maria, and H. M. Z. El-Din, "Incorporation of nonlinear load models in the transient energy function method," *IEEE Trans on Power Systems,* vol. 4, pp. 1031-1036, 1989.

[38] A. A. Fouad, V. Vittal, Y. X. Ni, H. R. Pota, K. Nodehi, H. M. Zein-Eldin, E. Vasahedi, and J. Kim, "Direct transient stability assessment with excitation control," *IEEE Trans on Power Systems,* vol. 4, pp. 75-82, 1989.

[39] R. Mihalic and U. Gabrijel, "A structure-preserving energy function for a static series synchronous compensator," *IEEE Trans on Power Systems,* vol. 19, pp. 1501-1507, 2004.

[40] PSERC, "On-Line Transient Stability Assessment Scoping Study," February 2005.

[41] Y. Xue, T. Van Cutsem, and M. Ribbens-Pavella, "Real-time analytic sensitivity method for transient security assessment and preventive control," *IEE Proceedings C, Generation, Transmission and Distribution,* vol. 135, pp. 107-117, 1988.

[42] Y. Xue, T. Van Cutsem, and M. Ribbens-Pavella, "A simple direct method for fast transient stability assessment of large power systems," *IEEE Trans on Power Systems,* vol. 3, pp. 400-412, 1988.

[43] Y. Xue, T. Van Custem, and M. Ribbens-Pavella, "Extended equal area criterion justifications, generalizations, applications," *IEEE Trans on Power Systems,* vol. 4, pp. 44-52, 1989.

[44] G. A. Maria, C. Tang, and J. Kim, "Hybrid transient stability analysis," *IEEE Trans on Power Systems,* vol. 5, pp. 384-393, 1990.

[45] D. Z. Fang, T. S. Chung, and A. K. David, "Improved techniques for hybrid method in fast-transient stability assessment," *IEE Proceedings - Generation, Transmission and Distribution,* vol. 144, pp. 107-112, 1997.

[46] M. Pavella, "Generalized one-machine equivalents in transient stability studies," *IEEE Power Engineering Review,* vol. 18, pp. 50-52, 1998.

[47] D. Ernst, D. Ruiz-Vega, M. Pavella, P. Hirsch, and D. Sobajic, "A unified approach to transient stability contingency filtering, ranking and assessment," in *IEEE Power Engineering Society Summer Meeting*, Vancouver, Canada, 2001.

[48] B. Lee, S. H. Kwon, J. Lee, H. K. Nam, J. B. Choo, and D. H. Jeon, "Fast contingency screening for online transient stability monitoring and assessment of the KEPCO system," *IEE Proceedings - Generation, Transmission and Distribution,* vol. 150, pp. 399-404, 2003.

[49] D. Ruiz-Vega and M. Pavella, "A comprehensive approach to transient stability control. I. Near optimal preventive control," *IEEE Trans on Power Systems,* vol. 18, pp. 1446-1453, 2003.

[50] Y. Zhang, L. Wehenkel, P. Rousseaux, and M. Pavella, "SIME: A hybrid approach to fast transient stability assessment and contingency selection," *International Journal of Electrical Power & Energy Systems,* vol. 19, pp. 195-208, 1997.

[51] D. Ruiz-Vega and M. Pavella, "A comprehensive approach to transient stability control. II. Open loop emergency control," *IEEE Trans on Power Systems,* vol. 18, pp. 1454-1460, 2003.

[52] L. Wehenkel, T. V. Cutsem, and M. Ribbens-Pavella, "An artificial intelligence framework for on-line transient stability assessment of power systems," *IEEE Trans on Power Systems,* vol. 4, pp. 789-800, 1989.

[53] L. Wehenkel and M. Pavella, "Decision trees and transient stability of electric power systems," *International Federation of Automatic Control,* vol. 27, pp. 115-134, 1991.

[54] S. Kretsinger, S. Rovnyak, D. Brown, and J. Thorp, "Parallel decision trees for predicing groups of unstable generators from synchronized phasor measurements," in *Precise Measurements in Power Systems Conference*, Washington, D.C., USA, 1993.

[55] S. Rovnyak, S. Kretsinger, J. Thorp, and D. Brown, "Decision trees for real-time transient stability prediction," *IEEE Trans on Power Systems,* vol. 9, pp. 1417-1426, 1994.

[56] L. Wehenkel, M. Pavella, E. Euxibie, and B. Heilbronn, "Decision tree based transient stability method a case study," *IEEE Trans on Power Systems* vol. 9, pp. 459-469, 1994.

[57] S. Rovnyak, L. Chih-Wen, J. Lu, W. Ma, and J. Thorp, "Predicting future behavior of transient events rapidly enough to evaluate remedial control options in real-time," *IEEE Trans on Power Systems,* vol. 10, pp. 1195-1203, 1995.

[58] S. M. Rovnyak, C. W. Taylor, and Y. Sheng, "Decision trees using apparent resistance to detect impending loss of synchronism," *IEEE Trans on Power Delivery,* vol. 15, pp. 1157-1162, 2000.

[59] K. Mei and S. M. Rovnyak, "Response-based decision trees to trigger one-shot stabilizing control," *IEEE Trans on Power Systems,* vol. 19, pp. 531-537, 2004.

[60] N. Senroy, G. T. Heydt, and V. Vittal, "Decision tree assisted controlled islanding," *IEEE Trans on Power Systems,* vol. 21, pp. 1790-1797, 2006.

[61] K. Sun, S. Likhate, V. Vittal, V. S. Kolluri, and S. Mandal, "An online dynamic security assessment scheme using phasor measurements and decision trees," *IEEE Trans on Power Systems,* vol. 22, pp. 1935-1943, 2007.

[62] S. M. Rovnyak, "Discussion of "Decision tree assisted controlled islanding"," *IEEE Trans on Power Systems,* vol. 22, pp. 2292-2293, 2007.

[63] R. Diao, V. Vittal, and N. Logic, "Design of a real-time security assessment tool for situational awareness enhancement in modern power systems," *IEEE Trans on Power Systems,* vol. 25, pp. 957-965, 2010.

[64] I. Genc, R. Diao, V. Vittal, S. Kolluri, and S. Mandal, "Decision tree-based preventive and corrective control applications for dynamic security enhancement in power systems," *IEEE Trans on Power Systems,* vol. 25, pp. 1611-1619, 2010.

[65] I. Kamwa, S. R. Samantaray, and G. Joos, "Development of rule-based classifiers for rapid stability assessment of wide-area post disturbance records," in *IEEE Power and Energy Society General Meeting*, Minneapolis, USA, 2010.

[66] Q. Gao and S. M. Rovnyak, "Decision trees using synchronized phasor measurements for wide-area response-based Control," *IEEE Trans on Power Systems,* vol. 26, pp. 855-861, 2011.

[67] V. Vittal, "Application of phasor measurements for dynamic security assessment using decision trees," in *IEEE Power and Energy Society General Meeting*, San Diego, USA, 2012.

[68] M. He, J. Zhang, and V. Vittal, "Robust online dynamic security assessment using adaptive ensemble decision-tree learning," *IEEE Trans on Power Systems,* vol. 28, pp. 4089-4098, 2013.

[69] A. P. Meiyan Li, Arun G. Phadke, James S. Thorp, "Transient stability prediction based on apparent impedance trajectory recorded by PMUs," *International Journal of Electrical Power & Energy Systems,* vol. 54, pp. 498–504, 2013.

[70] Z. H. Rather, L. Chengxi, C. Zhe, C. L. Bak, and P. Thogersen, "Dynamic security assessment of Danish power system based on decision trees: Today and tomorrow," in *IEEE PowerTech*, Grenoble, France, 2013.

[71] L. Chengxi, S. Kai, Z. H. Rather, C. Zhe, C. L. Bak, P. Thogersen, and P. Lund, "A systematic approach for dynamic security assessment and the corresponding preventive control scheme based on decision trees," *IEEE Trans on Power Systems,* vol. 29, pp. 717-730, 2014.

[72] L. S. Moulin, A. P. A. da Silva, M. A. El-Sharkawi, and R. J. Marks, II, "Support vector machines for transient stability analysis of large-scale power systems," *IEEE Trans on Power Systems,* vol. 19, pp. 818-825, 2004.

[73] A. D. Rajapakse, F. Gomez, K. Nanayakkara, P. A. Crossley, and V. V. Terzija, "Rotor angle instability prediction using post-disturbance voltage trajectories," *IEEE Trans on Power Systems,* vol. 25, pp. 947-956, 2010.

[74] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, "Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements," *IEEE Trans on Power Systems,* vol. 26, pp. 1474-1483, 2011.

[75] L. Chengxi, Z. H. Rather, C. Zhe, C. L. Bak, and P. Thogersen, "Importance sampling based decision trees for security assessment and the corresponding preventive control schemes: The Danish case study," in *IEEE PowerTech* Grenoble, France, 2013.

[76] X. Yan, D. Zhao Yang, Z. Jun Hua, Z. Pei, and W. Kit Po, "A reliable intelligent system for real-time dynamic security assessment of power systems," *IEEE Trans on Power Systems,* vol. 27, pp. 1253-1263, 2012.

[77] C.-W. Liu, M. Su, S.-S. Tsay, and Y.-J. Wang, "Application of a novel fuzzy neural network to real-time transient stability swings prediction based on synchronized phasor measurements," *IEEE Trans on Power Systems,* vol. 14, pp. 685-692, 1999.

[78] N. I. A. Wahab, A. Mohamed, and A. Hussain, "Fast transient stability assessment of large power system using probabilistic neural network with feature reduction techniques," *Expert Systems with Applications,* vol. 38, pp. 11112-11119, 2011.

[79] A. G. Bahbah and A. A. Girgis, "New method for generators' angles and angular velocities prediction for transient stability assessment of multimachine power systems using recurrent artificial neural network," *IEEE Trans on Power Systems,* vol. 19, pp. 1015-1022, 2004.

[80] A. N. Al-Masri, M. Z. A. Ab Kadir, H. Hizam, and N. Mariun, "A novel implementation for generator rotor angle stability prediction using an adaptive artificial neural network application for dynamic security assessment," *IEEE Trans on Power Systems,* vol. 28, pp. 2516-2525, 2013.

[81] M. He, V. Vittal, and J. Zhang, "Online dynamic security assessment with missing PMU measurements: a data mining approach," *IEEE Trans on Power Systems,* vol. 28, pp. 1969-1977, 2013.

[82] M. J. Laufenberg and M. A. Pai, "A new approach to dynamic security assessment using trajectory sensitivities," *IEEE Trans on Power Systems,* vol. 13, pp. 953-958, 1998.

[83] M. J. Laufenberg and M. A. Pai, "Sensitivity theory in power systems: application in dynamic security analysis," in *IEEE International Conference on Control Applications*, Dearborn, USA, 1996, pp. 738-743.

[84] T. B. Nguyen and M. A. Pai, "Dynamic security-constrained rescheduling of power systems using trajectory sensitivities," *IEEE Trans on Power Systems,* vol. 18, pp. 848-854, 2003.

[85] I. A. Hiskens and M. A. Pai, "Trajectory sensitivity analysis of hybrid systems," *IEEE Trans on Circuits and Systems I: Fundamental Theory and Applications,* vol. 47, pp. 204-220, 2000.

[86] I. A. Hiskens and M. Akke, "Analysis of the Nordel power grid disturbance of January 1, 1997 using trajectory sensitivities," *IEEE Trans on Power Systems,* vol. 14, pp. 987-994, 1999.

[87] I. A. Hiskens and M. A. Pai, "Power system applications of trajectory sensitivities," in *IEEE Power Engineering Society Winter Meeting*, 2002.

[88] M. Anjia, "Ball catching: the inspiration to power system stability control, a fast algorithm for the generator's disturbed trajectory prediction," in *IEEE Power Engineering Society General Meeting*, Tampa, USA, 2007.

[89] M. H. Haque and A. H. M. A. Rahim, "Determination of first swing stability limit of multimachine power systems through Taylor series expansions," *IEE Proceedings C, Generation, Transmission and Distribution,* vol. 136, pp. 373-380, 1989.

[90] M. Takahashi, K. Matsuzawa, M. Sato, K. Omata, R. Tsukui, T. Nakamura, and S. Mizuguchi, "Fast generation shedding equipment based on the observation of swings of generators," *IEEE Trans on Power Systems,* vol. 3, pp. 439-446, 1988.

[91] A. A. Daoud, G. G. Karady, and R. A. Amin, "A new fast-learning algorithm for predicting power system stability," in *IEEE Power Engineering Society Winter Meeting*, Columbus, USA, 2001.

[92] M. A. Mohamed, G. G. Karady, and A. M.Yousef, "New strategy agents to improve power system transient stability," presented at the World Academy of Science, Engineering and Technology, 2005.

[93] G. G. Karady, A. A. Daoud, and M. A. Mohamed, "On-line transient stability enhancement using multi-agent technique," in *IEEE Power Engineering Society Winter Meeting* 2002.

[94] K. Matsuzawa, K. Yanagihashi, J. Tsukita, M. Sato, T. Nakamura, and A. Takeuchi, "Stabilizing control, system preventing loss of synchronism from extension and its actual operating experience," *IEEE Trans on Power Systems,* vol. 10, pp. 1606-1613, 1995.

[95] J. Yan, C.-C. Liu, and U. Vaidya, "PMU-based monitoring of rotor angle dynamics," *IEEE Trans on Power Systems* vol. 26, pp. 2125-2133, 2011.

[96] C. Zheng, V. Malbasa, and M. Kezunovic, "Regression tree for stability margin prediction using synchrophasor measurements," *IEEE Trans on Power Systems,* vol. 28, pp. 1978-1987, 2013.

[97] S. P. Teeuwsen, I. Erlich, M. A. El-Sharkawi, and U. Bachmann, "Genetic algorithm and decision tree-based oscillatory stability assessment," *IEEE Trans on Power Systems,* vol. 21, pp. 746-753, 2006.

[98] R. Diao, K. Sun, V. Vittal, R. J. O'Keefe, M. R. Richardson, N. Bhatt, D. Stradford, and S. K. Sarawgi, "Decision tree-based online voltage security assessment using PMU measurements," *IEEE Trans on Power Systems,* vol. 24, pp. 832-839, 2009.

[99] V. Krishnan, J. D. McCalley, S. Henry, and S. Issad, "Efficient database generation for decision tree based power system security assessment," *IEEE Trans on Power Systems,* vol. 26, pp. 2319-2327, 2011.

[100] N. Amjady and S. F. Majedi, "Transient stability prediction by a hybrid intelligent system," *IEEE Trans on Power Systems,* vol. 22, pp. 1275-1283, 2007.

[101] J. C. Cepeda, J. L. Rueda, I. Erlich, and D. G. Colome, "Recognition of post-contingency dynamic vulnerability regions: Towards smart grids," in *IEEE Power and Energy Society General Meeting*, San Diego, USA, 2012.

[102] C. Liu, Z. H. Rather, Z. Chen, and C. L. Bak, "An overview of decision tree applied to power systems," *International Journal of Smart Grid and Clean Energy,* vol. 2, pp. 414-419, 2013.

[103] N. Senroy, "Emergency state stability control of power systems through intelligent islanding," Ph.D. dissertation, Arizona State Univ., Tempe, USA, May 2006.

[104] M. Jin, T. S. Sidhu, and K. Sun, "A new system splitting scheme based on the unified stability control framework," *IEEE Trans on Power Systems,* vol. 22, pp. 433-441, 2007.

[105] T. M. Mitchell, *Machine Learning*: McGraw-Hill, 1997.

[106] M. A. M. Ariff and B. C. Pal, "Coherency identification in interconnected power system - an independent component analysis approach," *IEEE Trans on Power Systems,* vol. 28, pp. 1747-1755, 2013.

[107] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*: Prentice Hall, Inc., 1998.

[108] "IEEE recommended practice for excitation system models for power system stability studies," *IEEE Std 421.5-2005 (Revision of IEEE Std 421.5-1992),* 2006.

[109] P. M. Anderson and A. A. Found, *Power System Control and Stability*. New York: IEEE Press, 1993.

[110] "Load representation for dynamic performance analysis," *IEEE Trans on Power Systems,* vol. 8, pp. 472-482, 1993.

[111] C. Concordia and S. Ihara, "Load representation in power system stability studies," *IEEE Trans on Power Apparatus and Systems,* vol. PAS-101, pp. 969-977, 1982.

[112] H. Saadat, *Power System Analysis*: McGrawHill, 2002.

[113] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans on Power Systems* vol. 26, pp. 12-19, 2011.

[114] The MathWorks, Inc., *Simulink User's Guide R2014b*, 2014.

[115] G. Rogers, *Power System Oscillations*. Norwell: Kluwer Academic Publishers, 2000.

[116] R. Preece, J. V. Milanovic, A. M. Almutairi, and O. Marjanovic, "Probabilistic evaluation of damping controller in networks with multiple VSC-HVDC lines," *IEEE Trans on Power Systems,* vol. 28, pp. 367-376, 2013.

[117] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*: Morgan Kaufmann, 2012.

[118] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans on Systems, Man and Cybernetics,* vol. 21, pp. 660-674, 1991.

[119] IBM Cooporation, *IBM SPSS Modeler 14.2 Modeling Nodes*, 2011.

[120] I. H. Witten and E. Frank, *Data Mining: practical machine learning tools and techniques*: Morgan Kaufmann, 2005.

[121] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *IEEE Trans on Pattern Analysis and Machine Intelligence,* vol. 29, pp. 173-180, 2007.

[122] C.-W. Hsu and L. C.-J., "A comparison of methods for multiclass support vector machines," *IEEE Trans on Neural Networks,* vol. 13, pp. 415-425, 2002.

[123] B. S. Everitt, S. Landau, and M. Leese, *Cluster Analysis*. London: Arnold 2001.

[124] [Online] http://www.01.ibm.com/software/analytics/spss/products/modeler/features.

[125] The MathWorks, Inc., *Statistics Toolbox User's Guide R2014b*, 2014.

[126] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans on Intelligent Systems and Technology,* vol. 2, pp. 27:1--27:27, 2011.

[127] B. Pal and B. Chaudhuri, *Robust Control in Power Systems*. New York: Springer Science & Business Media, Inc., 2005.

[128] C. Pout, F. MacKenzie, and E. Olloqui, "The impact of changing energy use patterns in buildings on peak electricity demand in the UK," Department of Energy & Climate Change, 2008.

[129] E. Vaahedi, W. Li, T. Chia, and H. Dommel, "Large scale probabilistic transient stability assessment using BC Hydro's on-line tool," *IEEE Trans on Power Systems,* vol. 15, pp. 661-667, 2000.

[130] R. Billinton and P. R. S. Kuruganty, "Probabilistic assessment of transient stability in a practical multimachine system," *IEEE Trans on Power Apparatus and Systems* vol. PAS-100, pp. 3634-3641, 1981.

[131] R. Billinton, P. R. S. Kuruganty, and M. F. Carvalho, "An approximate method for probabilistic assessment of transient stability," *IEEE Trans on Reliability,* vol. R-28, pp. 255-258, 1979.

[132] P. R. S. Kuruganty and R. Billinton, "Protection system modelling in a probabilistic assessment of transient stability," *IEEE Trans on Power Apparatus and Systems,* vol. PAS-100, pp. 2163-2170, 1981.

[133] IEEE Committee Report, "Single phase tripping and auto reclosing of transmission lines," *IEEE Trans on Power Delivery,* vol. 7, pp. 182-192, 1992.

[134] Y.-Y. Hsu and C.-L. Chang, "Probabilistic transient stability studies using the conditional probability approach," *IEEE Trans on Power Systems,* vol. 3, pp. 1565-1572, 1988.

[135] A. A. Alabduljabbar, "Optimal placement of FACTS devices for the improvement of techno-economic performance of power netwrks," Ph.D. thesis, the School of Electrical and Electronic Engineering, The University of Manchester, U.K., 2008.

[136] I. Kamwa, A. K. Pradhan, and G. Joos, "Automatic segmentation of large power systems into fuzzy coherent areas for dynamic vulnerability assessment," *IEEE Trans on Power Systems,* vol. 22, pp. 1974-1985, 2007.

[137] K. K. Anaparthi, B. Chaudhuri, N. F. Thornhill, and B. C. Pal, "Coherency identification in power systems through principal component analysis," *IEEE Trans on Power Systems,* vol. 20, pp. 1658-1660, 2005.

[138] R. Agrawal and D. Thukaram, "Identification of coherent synchronous generators in a multi-machine power system using support vector clustering," in *International Conference on Power and Energy Systems*, 2011.

[139] S. Wang and X. Yao, "Multiclass imbalance problems: analysis and potential solutions," *IEEE Trans on Systems, Man, and Cybernetics,* vol. 42, pp. 1119-1130, 2012.

[140] M. Saar-Tsechansky and F. Provost, "Handling missing values when applying classification models," *The Journal of Machine Learning Research,* vol. 8, pp. 1623-1657, 2007.

[141] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. London, UK: Chapman and Hall, 1984.

[142] Y. Wang, W. Li, and J. Lu, "Reliability analysis of wide-area measurement system," *IEEE Trans on Power Delivery,* vol. 25, pp. 1483-1491, 2010.

[143] F. Aminifar, S. Bagheri-Shouraki, M. Fotuhi-Firuzabad, and M. Shahidehpour, "Reliability modeling of PMUs using fuzzy sets," *IEEE Trans on Power Delivery,* vol. 25, pp. 2384-2391, 2010.

Appendix A: Network Data

This appendix provides the data required in order to perform dynamic studies on the NETS-NYPS test network used throughout this thesis. A system base of 100 MVA is used. Full system details, generator and exciter parameters are adopted from [127] with PSS settings for G9 sourced from [115].

A.1 Line Impedances

The line impedance data for the network is presented in Table A-1, including transformer off-nominal turns ratio (ONR) where applicable.

Table A-1: Line data for the NETS-NYPS test network.

From Bus	*To Bus*	*R(pu)*	*X(pu)*	*B(pu)*	*ONR*	*From Bus*	*To Bus*	*R(pu)*	*X(pu)*	*B(pu)*	*ONR*
2	53	0	0.0181	0	1.025	33	34	0.0011	0.0157	0.202	–
6	54	0	0.025	0	1.07	35	34	0.0001	0.0074	0	0.946
10	55	0	0.02	0	1.07	34	36	0.0033	0.0111	1.45	–
19	56	0.0007	0.0142	0	1.07	9	36	0.0022	0.0196	0.34	–
20	57	0.0009	0.018	0	1.009	9	36	0.0022	0.0196	0.34	–
22	58	0	0.0143	0	1.025	16	37	0.0007	0.0089	0.1342	–
23	59	0.0005	0.0272	0	1	31	38	0.0011	0.0147	0.247	–
25	60	0.0006	0.0232	0	1.025	33	38	0.0036	0.0444	0.693	–
29	61	0.0008	0.0156	0	1.025	41	40	0.006	0.084	3.15	–

From Bus	*To Bus*	*R(pu)*	*X(pu)*	*B(pu)*	*ONR*	*From Bus*	*To Bus*	*R(pu)*	*X(pu)*	*B(pu)*	*ONR*
31	62	0	0.026	0	1.04	48	40	0.002	0.022	1.28	–
32	63	0	0.013	0	1.04	42	41	0.004	0.06	2.25	–
36	64	0	0.0075	0	1.04	18	42	0.004	0.06	2.25	–
17	65	0	0.0033	0	1.04	17	43	0.0005	0.0276	0	–
41	66	0	0.0015	0	1	39	44	0	0.0411	0	–
42	67	0	0.0015	0	1	43	44	0.0001	0.0011	0	–
18	68	0	0.003	0	1	35	45	0.0007	0.0175	1.39	–
36	17	0.0005	0.0045	0.32	–	39	45	0	0.0839	0	–
49	18	0.0076	0.1141	1.16	–	44	45	0.0025	0.073	0	–
16	19	0.0016	0.0195	0.304	–	38	46	0.0022	0.0284	0.43	–
19	20	0.0007	0.0138	0	1.06	1	47	0.0013	0.0188	1.31	–
16	21	0.0008	0.0135	0.2548	–	47	48	0.0025	0.0268	0.4	–
21	22	0.0008	0.014	0.2565	–	47	48	0.0025	0.0268	0.4	–
22	23	0.0006	0.0096	0.1846	–	46	49	0.0018	0.0274	0.27	–
23	24	0.0022	0.035	0.361	–	45	51	0.0004	0.0105	0.72	–
16	24	0.0003	0.0059	0.068	–	50	51	0.0009	0.0221	1.62	–
2	25	0.007	0.0086	0.146	–	37	52	0.0007	0.0082	0.1319	–
25	26	0.0032	0.0323	0.531	–	3	52	0.0011	0.0133	0.2138	–
37	27	0.0013	0.0173	0.3216	–	1	2	0.0035	0.0411	0.6987	–
26	27	0.0014	0.0147	0.2396	–	2	3	0.0013	0.0151	0.2572	–
26	28	0.0043	0.0474	0.7802	–	3	4	0.0013	0.0213	0.2214	–
26	29	0.0057	0.0625	1.029	–	4	5	0.0008	0.0128	0.1342	–
28	29	0.0014	0.0151	0.249	–	5	6	0.0002	0.0026	0.0434	–
1	30	0.0008	0.0074	0.48	–	6	7	0.0006	0.0092	0.113	–
9	30	0.0019	0.0183	0.29	–	5	8	0.0008	0.0112	0.1476	–
9	30	0.0019	0.0183	0.29	–	7	8	0.0004	0.0046	0.078	–
30	31	0.0013	0.0187	0.333	–	8	9	0.0023	0.0363	0.3804	–
1	31	0.0016	0.0163	0.25	–	6	11	0.0007	0.0082	0.1389	–
30	32	0.0024	0.0288	0.488	–	10	11	0.0004	0.0043	0.0729	–
32	33	0.0008	0.0099	0.168	–	12	11	0.0016	0.0435	0	1.06
4	14	0.0008	0.0129	0.1382	–	10	13	0.0004	0.0043	0.0729	–
13	14	0.0009	0.0101	0.1723	–	12	13	0.0016	0.0435	0	1.06
14	15	0.0018	0.0217	0.366	–	1	27	0.032	0.32	0.41	–
15	16	0.0009	0.0094	0.171	–	50	18	0.0012	0.0288	2.06	–

A.2 Standard Power Flow Data

The data required to complete standard power flow is included in Table A-2. Bus 65 is the slack.

Table A-2: Standard power flow data for the NETS-NYPS test network.

Bus	*V (pu)*	*θ (°)*	P_G *(MW)*	P_L *(MW)*	Q_L *(MVar)*	*Bus*	*V (pu)*	*θ (°)*	P_G *(MW)*	P_L *(MW)*	Q_L *(MVar)*
1	–	–	–	252.7	118.56	44	–	–	–	267.55	4.84
3	–	–	–	322	2	45	–	–	–	208	21
4	–	–	–	200	73.6	46	–	–	–	150.7	28.5
7	–	–	–	234	84	47	–	–	–	203.12	32.59
8	–	–	–	208.8	70.8	48	–	–	–	241.2	2.2
9	–	–	–	104	125	49	–	–	–	164	29
12	–	–	–	9	88	50	–	–	–	100	-147
15	–	–	–	320	153	51	–	–	–	337	-122
16	–	–	–	329	32	52	–	–	–	158	30
17	–	–	–	6000	300	53	1.045	–	250	–	–
18	–	–	–	2470	123	54	0.98	–	545	–	–
20	–	–	–	680	103	55	0.983	–	650	–	–
21	–	–	–	274	115	56	0.997	–	632	–	–
23	–	–	–	248	85	57	1.011	–	505	–	–
24	–	–	–	309	-92	58	1.05	–	700	–	–
25	–	–	–	224	47	59	1.063	–	560	–	–
26	–	–	–	139	17	60	1.03	–	540	–	–
27	–	–	–	281	76	61	1.025	–	800	–	–
28	–	–	–	206	28	62	1.01	–	500	–	–
29	–	–	–	284	27	63	1	–	1000	–	–
33	–	–	–	112	0	64	1.0156	–	1350	–	–
36	–	–	–	102	-19.46	65	1.011	0	–	–	–
39	–	–	–	267	12.6	66	1	–	1785	–	–
40	–	–	–	65.63	23.53	67	1	–	1000	–	–
41	–	–	–	1000	250	68	1	–	4000	–	–
42	–	–	–	1150	250						

A.3 Optimal Power Flow Data

From Chapter 5 to 7, an optimal power flow solution is incorporated within the test system. The optimisation minimises the total cost of generation for the given loading scenario, where each generator is subject to the standard cost function (A.1).

$$Cost = c_0 + c_1 P_e + c_2 P_e^2 \ \$/\text{hour} \tag{A.1}$$

The coefficient values for each generator are given in Table A-3 and taken from [2]. Also included in the table are the constraints on active and reactive power for each generating unit. Furthermore, all bus voltages are constrained to between 0.9 and 1.1 p.u.. The thermal constraints of transmission lines are not considered.

Table A-3: Standard power flow data for the NETS-NYPS test network.

Generator	c_0	c_1	c_2	P^{max} *(MW)*	P^{min} *(MW)*	Q^{max} *(MVar)*	Q^{min} *(MVar)*
G1	0	6.9	0.0193	450	20	148.5	-148.5
G2	0	3.7	0.0111	745	80	245.85	-245.85
G3	0	2.8	0.0104	850	40	280.5	-280.5
G4	0	4.7	0.0088	832	40	274.56	-274.56
G5	0	2.8	0.0128	705	50	232.65	-232.65
G6	0	3.7	0.0094	700	45	297	-297
G7	0	4.8	0.0099	760	50	250.8	-250.8
G8	0	3.6	0.0113	740	75	244.2	-244.2
G9	0	3.7	0.0071	1000	70	330	-330
G10	0	3.9	0.0090	700	35	231	-231
G11	0	4.0	0.0050	1200	60	396	-396
G12	0	2.9	0.0040	1550	80	511.5	-511.5
G13	0	2.5	0.0019	3213	40	1060.29	-1060.29
G14	0	3.3	0.0033	1985	20	655.05	-655.05
G15	0	3.8	0.0050	1200	20	396	-396
G16	0	9.5	0.0014	4200	50	1386	-1386

A.4 Generator Dynamic Data

The generator dynamic data is given in Table A-4 and Table A-5, scaled to the given machine base.

Table A-4: Generator dynamic data for the NETS-NYPS test network (1).

Gen	*Bus*	*Rating (MVA)*	X_{lk} *(pu)*	X_d *(pu)*	X'_d *(pu)*	X''_d *(pu)*	T'_{d0} *(s)*	T''_{d0} *(s)*
G1	53	100	0.0125	0.1	0.031	0.025	10.2	0.05
G2	54	100	0.035	0.295	0.0697	0.05	6.56	0.05
G3	55	100	0.0304	0.2495	0.0531	0.045	5.7	0.05
G4	56	100	0.0295	0.262	0.0436	0.035	5.69	0.05
G5	57	100	0.027	0.33	0.066	0.05	5.4	0.05
G6	58	100	0.0224	0.254	0.05	0.04	7.3	0.05
G7	59	100	0.0322	0.295	0.049	0.04	5.66	0.05
G8	60	100	0.028	0.29	0.057	0.045	6.7	0.05
G9	61	100	0.0298	0.2106	0.057	0.045	4.79	0.05
G10	62	100	0.0199	0.169	0.0457	0.04	9.37	0.05
G11	63	100	0.0103	0.128	0.018	0.012	4.1	0.05
G12	64	100	0.022	0.101	0.031	0.025	7.4	0.05
G13	65	200	0.003	0.0296	0.0055	0.004	5.9	0.05
G14	66	100	0.0017	0.018	0.00285	0.0023	4.1	0.05
G15	67	100	0.0017	0.018	0.00285	0.0023	4.1	0.05
G16	68	200	0.0041	0.0356	0.0071	0.0055	7.8	0.05

Table A-5: Generator dynamic data for the NETS-NYPS test network (2).

Gen	*Bus*	*Rating (MVA)*	X_q *(pu)*	X'_q *(pu)*	X''_q *(pu)*	T'_{q0} *(s)*	T''_{q0} *(s)*	*H (s)*	*D*
G1	53	100	0.069	0.028	0.025	1.5	0.035	42	4
G2	54	100	0.282	0.06	0.05	1.5	0.035	30.2	9.75
G3	55	100	0.237	0.05	0.045	1.5	0.035	35.8	10
G4	56	100	0.258	0.04	0.035	1.5	0.035	28.6	10
G5	57	100	0.31	0.06	0.05	0.44	0.035	26	3
G6	58	100	0.241	0.045	0.04	0.4	0.035	34.8	10
G7	59	100	0.292	0.045	0.04	1.5	0.035	26.4	8
G8	60	100	0.28	0.05	0.045	0.41	0.035	24.3	9

Gen	*Bus*	*Rating (MVA)*	X_q *(pu)*	$X_q^{'}$ *(pu)*	$X_q^{''}$ *(pu)*	$T_{q0}^{'}$ *(s)*	$T_{q0}^{''}$ *(s)*	*H (s)*	*D*
G9	61	100	0.205	0.05	0.045	1.96	0.035	34.5	14
G10	62	100	0.115	0.045	0.04	1.5	0.035	31	5.56
G11	63	100	0.123	0.015	0.012	1.5	0.035	28.2	13.6
G12	64	100	0.095	0.028	0.025	1.5	0.035	92.3	13.5
G13	65	200	0.0286	0.005	0.004	1.5	0.035	248	33
G14	66	100	0.0173	0.0025	0.0023	1.5	0.035	300	100
G15	67	100	0.0173	0.0025	0.0023	1.5	0.035	300	100
G16	68	200	0.0334	0.006	0.0055	1.5	0.035	225	50

Generators G1-G8 all use type DC1A exciters, with the following parameters:

$T_R = 0.01$, $K_A^{ex} = 40$, $T_A^{ex} = 0.02$, $E_{ex}^{\min} = -10$, $E_{ex}^{\max} = -10$, $T_E^{ex} = 0.785$, $K_E^{ex} = 1$,

$A_E^{ex} = 0.07$, $B_E^{ex} = 0.91$.

Generator G9 uses a type ST1A exciter, with the following parameters:

$T_R = 0.01$, $K_A^{ex} = 200$, $E_{fd}^{\min} = -5$, $E_{fd}^{\max} = 5$.

Generator G9 is also equipped with a PSS with the following settings:

$T_W^{PSS} = 10$, $T_1^{PSS} = 0.05$, $T_2^{PSS} = 0.01$, $T_3^{PSS} = 0.05$, $T_4^{PSS} = 0.02$, $K_{PSS} = 10$,

$E_{PSS}^{\min} = -0.5$, $E_{PSS}^{\max} = 0.5$.

Appendix B: List of Author's Thesis Based Publications

B.1 International Journal Publications

[B1] **T. Guo** and J. V. Milanović, "Probabilistic Framework for Assessing the Accuracy of Data Mining Tool for Online Prediction of Transient Stability", *IEEE Transactions on Power Systems,* vol. 29, pp. 377-385, 2014.

[B2] **T. Guo** and J. V. Milanović, "On-line Identification of Power System Dynamic Signature Using PMU Measurements and Data Mining", *IEEE Transactions on Power Systems*, TPWRS-00940-2014.R2 (accepted for publication on 05/07/2015).

B.2 International Conference Publications

[B3] **T. Guo** and J. V. Milanović, "On-line prediction of transient stability using decision tree method - Sensitivity of accuracy of prediction to different uncertainties", *IEEE PowerTech 2013*, Grenoble, France, 16-20 June, 2013.

[B4] **T. Guo** and J. V. Milanović, "The Effect of Quality and Availability of Measurment Signals on Accuracy of On-line Prediction of Transient Stability Using Decision Tree Method", *4th IEEE PES ISGT Europe*, Copenhagen, Denmark, 6-9 October, 2013.

[B5] **T. Guo** and J. V. Milanović, "Identification of Power System Dynamic Signature Using Hierarchical Clustering", *IEEE Power and Energy Society General Meeting*, Washington, DC Metro Area, USA, 27-31 July 2014.

[B6] **T. Guo**, J. He, Z. Li, and J. V. Milanović, "Evaluation of Classification Methods for On-line Identification of Power System Dynamic Signature," *18th Power Systems Computation Conference*, Wroclaw, Poland, 18-22 August, 2014.

[B7] **T. Guo**, P. Papadopoulos, P. Mohammed, and J. V. Milanović, "Comparison of Ensemble Decision Tree Methods for On-line Identification of Power System Dynamic Signature Considering Availability of PMU Measurements", *IEEE PowerTech 2015*, Eindhoven, Netherlands, 29 June-2 July, 2015.