Real-Time Event-Driven Load Shedding for Power System Transient Stability Control using Deep Learning Techniques

Jizhe Liu

Master of Philosophy

School of Electrical Engineering and Telecommunications
Faculty of Engineering

April 2022
DECLARATIONS

ORIGINALLITY STATEMENT

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

COPYRIGHT STATEMENT

I hereby grant the University of New South Wales or its agents a non-exclusive licence to archive and to make available (including to members of the public) my thesis or dissertation in whole or part in the University libraries in all forms of media, now or hereafter known. I acknowledge that I retain all intellectual property rights which subsist in my thesis or dissertation, such as copyright and patent rights, subject to applicable law. I also retain the right to use all or part of my thesis or dissertation in future works (such as articles or books). For any substantial portions of copyright material used in this thesis, written permission for use has been obtained, or the copyright material is removed from the final public version of the thesis.

AUTHENTICITY STATEMENT

I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.
UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in the candidate’s thesis in lieu of a Chapter provided:

- The candidate contributed greater than 50% of the content in the publication and are the “primary author”, i.e. they were responsible primarily for the planning, execution and preparation of the work for publication.
- The candidate has obtained approval to include the publication in their thesis in lieu of a Chapter from their Supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis.

The candidate has declared that some of the work described in their thesis has been published and has been documented in the relevant Chapters with acknowledgement.

A short statement on where this work appears in the thesis and how this work is acknowledged within chapters:

My literature review is partially comprised of a review paper that I contributed to and published in the Applied Energy. The results from “Real-time emergency load shedding for power system transient stability control: A risk-averse deep learning method,” in Applied Energy is contained in parts in Chapters 4 and 5. Acknowledgement of the work of the other authors of this paper has been made at the beginning of the chapter as well as in my Acknowledgements section.

Candidate’s Declaration

I declare that I have complied with the Thesis Examination Procedure.
Abstract

With the increasing integration of variational renewable energy and the more active demand side responses, there are more challenges in maintaining secure and reliable power system operation due to the escalated stochasticity and variations in the system. This can be experienced from recent rolling blackouts over the world. Event-driven load shedding (ELS) serves as a fast and effective stability control scheme for power system after a risky disturbance occurs, which can suppress grid oscillation, recover system stability, and prevent cascading failure. Unlike the response-driven control schemes, ELS executes the load shedding action immediately following the disturbance, which aims to control power system stability at an earlier stage with the minimum amount of control cost. The digitalized power systems deploy advanced measurement devices such as phasor measurement units and smart meters, which provides the adequate sensing infrastructure to implement real-time stability assessment and control. However, the conventional approaches for ELS rely on numerical simulations and iterative optimizations which are computationally burdensome and thus slow reactive to the real-time system variations. More recently, artificial intelligence (AI) techniques provide a new way to realize real-time ELS owing to their fast decision making capability. This research identifies the key issues in existing AI-based ELS approaches and proposes a series of novel methodologies based on deep learning techniques to enhance overall ELS performance in practical situations. A deep neural network (DNN) model is first presented to improve the decision-making accuracy on ELS strategy. Moreover, considering the unbalanced control cost induced by an over- and under-estimated ELS amount, a risk-averse learning method for DNN is proposed to increase the likelihood of control success with negligible impairment on control cost. On top of those, a GraphSAGE-based ELS model is proposed to capture and embed the topological structure of power system into deep learning, which further improves the overall control performance of ELS. The proposed methodologies have been tested on New England 39 bus system and Nordic power system. The proposed deep learning methods have shown more exceptional control performance of ELS as compared to the existing methods.
# Table of Contents

Originality Statement  iii  
Acknowledgement  iv  
List of Tables  v  
List of Figures  vi  
List of Abbreviations  viii  

## Chapter 1. Introduction  1  
1.1 Power System Stability  1  
1.2 Power System Stability Control  5  
1.3 Load Shedding in Power System  6  
1.4 Decision-Making on ELS Strategy  7  
1.5 Contributions of This Research  8  

## Chapter 2. ELS Database Generation  11  
2.1 Introduction  11  
2.2 Evolutionary Algorithm  11  
2.3 ELS Data Sampling  12  
2.4 Simulation Results  14  

## Chapter 3. Deep Neural Networks  18  
3.1 Introduction  18  
3.2 Application of DNN in Power System ELS  19  
3.3 Simulation Results  19  

## Chapter 4. Risk-Averse DNN Model  24  
4.1 Introduction  24  
4.2 Overall Framework of Risk- Averse DNN Model  25  
4.3 Transient Stability Margin  26
4.4 EEAC-ELS Relationship 27
4.5 Loss Function Design 28
4.6 Numerical Results 30

Chapter 5. Risk-Averse GraphSAGE for ELS 38

5.1 Introduction 38
5.2 Graphic Neural Networks 39
5.3 GraphSAGE 40
5.4 Association Graph Construction for Power Networks 42
5.6 Simulation Results 43

Chapter 6. Conclusion and Future Works 50

BIBLIOGRAPHY 52
Originality Statement

I hereby certify that the work embodied in this thesis contains published chapters/scholarly work of which I am a joint author. I have included as part of the thesis a written statement, endorsed by my supervisor, attesting to my contribution to the joint publications/scholarly work. My literature review is partially comprised of a review paper that I contributed to and published in the Applied Energy. The results from "Real-time emergency load shedding for power system transient stability control: A risk-averse deep learning method," in Applied Energy is contained in parts in Chapters 4 and 5.

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project’s design and conception or in style, presentation and linguistic expression is acknowledged.

Signed ..................................Jizhe Liu...........................................

Date ................................25/04/2022...........................................

Supervisor:

Signed ...........................Dr. Ke Meng...........................................

Date ................................26/04/2022...........................................
ACKNOWLEDGEMENT

It was incredible and enjoyable to carry out my Mphil research at the University of New South Wales in Australia. I am truly indebted to the following people and organizations who in one way or another contributed to the completion of this thesis. First and foremost, my utmost gratitude goes to my supervisors, Prof. Zhao Yang Dong and Dr. Ke Meng. Without their support and assistance, this thesis would not have been possible. Then, I would like to express gratitude to my co-supervisor, Dr. Yuchen Zhang, for his recognition of my research direction, research guidance, and patient revision on the academic writing. I have imbibed a lot from their excellence. Besides, I would like to extend my appreciation to other researchers, Dr. Ke Meng, Prof. David John Hill, and Dr. Yuchen Zhang. Through the course of interaction with them, I was always imprinted with the source of inspiration in solving the research problems. I am also grateful to my colleagues, Dr. Wang Zhang, Dr. Weicong Kong, Dr. Ziyuan Tong, Dr. Qiwei Zhai, Dr. Cuo Zhang, Dr. Jiadia Yang, Dr. Yongxi Zhang, Dr. Xuekuan Xie, Dr. Junwei Liu, Dr. Shu Wang, Yuan Zhang, Dr. Colin Zhou, Dr. Feng Qiao, and Hesamoddin Marzooghi, for their ideas, thoughts, insights, and feedback during everyday communications and extra-curriculums. I would like to thank my parents, Xida Liu and Ruimei Luo. Their unconditional support is the prerequisite for my research. Moreover, special thanks go to my dear wife, Shuying Wu. She has motivated and assisted me so much along the way. Her love, trust and understanding were in the end what made this thesis possible.
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table I</td>
<td>TESTING RESULTS ON NEW ENGLAND SYSTEM</td>
<td>22</td>
</tr>
<tr>
<td>Table II</td>
<td>TESTING RESULTS ON NORDIC SYSTEM</td>
<td>22</td>
</tr>
<tr>
<td>Table III</td>
<td>AVERAGE COMPUTATION TIME PER SAMPLE (CPU TIME)</td>
<td>22</td>
</tr>
<tr>
<td>Table IV</td>
<td>ACCURACY OF SHEDDING AMOUNT INFERENCE MODELS</td>
<td>32</td>
</tr>
<tr>
<td>Table V</td>
<td>TESTING RESULTS ON NEW ENGLAND SYSTEM</td>
<td>33</td>
</tr>
<tr>
<td>Table VI</td>
<td>TESTING RESULTS ON NORDIC SYSTEM</td>
<td>35</td>
</tr>
<tr>
<td>Table VII</td>
<td>MEAN SHEDDING AMOUNT ERROR ON INDIVIDUAL CANDIDATE BUSES</td>
<td>35</td>
</tr>
<tr>
<td>Table VIII</td>
<td>AVERAGE COMPUTATION TIME PER SAMPLE (CPU TIME)</td>
<td>46</td>
</tr>
<tr>
<td>Table IX</td>
<td>TESTING RESULTS ON NEW ENGLAND SYSTEM</td>
<td>47</td>
</tr>
<tr>
<td>Table X</td>
<td>TESTING RESULTS ON NORDIC SYSTEM</td>
<td>47</td>
</tr>
</tbody>
</table>
# List of Figures

| Fig. 1.1 | Classification of power system stability | 2 |
| Fig. 1.2 | Process of ELS | 6 |
| Fig. 1.3 | Process of underfrequency LS | 7 |
| Fig. 1.4 | The contribution flow chart of this study. The corresponding problems are solved. | 9 |
| Fig. 2.1 | Process of EA | 12 |
| Fig. 2.2 | ELS database generation module. | 14 |
| Fig. 2.3 | Power systems for case study (a) New England 39-bus system, (b) Nordic 41-bus system. | 15 |
| Fig. 2.4 | Variations in objective function value in 5 EA executions. | 17 |
| Fig. 3.1 | The detailed structure of the developed DNN model. | 20 |
| Fig. 4.1 | The overall framework of the proposed Risk-Averse ELS method. | 26 |
| Fig. 4.2 | Optimal load shedding percentages vs. EEAC margins in an experiment. | 28 |
| Fig. 4.3 | The relationship between optimal load shedding amount and EEAC margin subject to (a) NEfault #1, (b) NEfault #2, (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault. | 32 |
| Fig. 4.4 | Distribution of the prediction errors of risk-neutral and risk-averse DNN subject to (a) NEfault #1, (b) NEfault #2, and (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault. | 34 |
| Fig. 4.5 | Variations of MAPE, SR, and ELLD against the change of $\omega$, subject to (a) NEfault #1, (b) NEfault #2, (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault. | 37 |
| Fig. 5.1 | Visual illustration of the GraphSAGE sample and aggregate process. | 41 |
| Fig. 5.2 | Graph Neural Network Applications in Power Systems. | 43 |
| Fig. 5.3 | Variations of MAPE, SR, and ELLD against the change of $\omega$, subject to (a) NEfault #1, (b) NEfault #2, (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault. | 45 |
| Fig. 5.4 | Distribution of the prediction errors of risk-neutral and |
risk-averse GraphSAGE subject to (a) NEfault #1, (b) NEfault #2, and (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault.
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial intelligent</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial neural network</td>
</tr>
<tr>
<td>BP</td>
<td>Backpropagation</td>
</tr>
<tr>
<td>CM</td>
<td>Critical machines</td>
</tr>
<tr>
<td>DNN</td>
<td>Deep neural network</td>
</tr>
<tr>
<td>EA</td>
<td>Evolutionary algorithms</td>
</tr>
<tr>
<td>EAC</td>
<td>Equal area criterion</td>
</tr>
<tr>
<td>EEAC</td>
<td>Extended Equal Area Criterion</td>
</tr>
<tr>
<td>ELM</td>
<td>Extreme learning machine</td>
</tr>
<tr>
<td>ELLD</td>
<td>Expected load loss deviation</td>
</tr>
<tr>
<td>ELS</td>
<td>Event-based load shedding</td>
</tr>
<tr>
<td>FACTS</td>
<td>Flexible AC transmission system</td>
</tr>
<tr>
<td>FIDVR</td>
<td>Fault-induced delayed voltage recovery</td>
</tr>
<tr>
<td>GCN</td>
<td>Graph convolution network</td>
</tr>
<tr>
<td>HVDC</td>
<td>High-voltage direct current</td>
</tr>
<tr>
<td>LS</td>
<td>Load shedding</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long short-term memory</td>
</tr>
<tr>
<td>LTVS</td>
<td>Long term voltage stability</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean absolute error</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean absolute percentage error</td>
</tr>
<tr>
<td>NM</td>
<td>Non-critical machines</td>
</tr>
<tr>
<td>OC</td>
<td>Operating condition</td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary least square</td>
</tr>
<tr>
<td>OMIB</td>
<td>One machine infinite bus</td>
</tr>
<tr>
<td>OPF</td>
<td>Optimal power flow</td>
</tr>
<tr>
<td>PMU</td>
<td>Phasor measurement unit</td>
</tr>
<tr>
<td>RLS</td>
<td>Response-driven load shedding</td>
</tr>
<tr>
<td>RF</td>
<td>Random forest</td>
</tr>
<tr>
<td>RVM</td>
<td>Relevance vector machine</td>
</tr>
<tr>
<td>SGD</td>
<td>Stochastic gradient descent</td>
</tr>
<tr>
<td>SR</td>
<td>Success rate</td>
</tr>
<tr>
<td>STVS</td>
<td>Short term voltage stability</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>TDS</td>
<td>Time-domain simulation</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------</td>
</tr>
<tr>
<td>TSA</td>
<td>Transient security assessment</td>
</tr>
</tbody>
</table>
LIST OF PUBLICATIONS INCLUDED AS PART OF THIS THESIS


Chapter 1. Introduction

Power system is a complex dynamic system. For a long time, maintaining the stability of power system has been the basic requirement for safe and uninterrupted power supply to end users. The rising concerns on global climate change have accelerated the energy transformation from fossil fuels to sustainable energy. In such an already-commenced transition, electricity would be the primary energy carrier, with an artifact that future electricity grid will be vastly more complex than the past: increasingly volatile generation as the renewable penetration rises, and increasingly active demand as homes and businesses take the new prosumer roles. However, with the costly and delayed upgrades to the network capability, all these new elements bring about prominent challenges to maintaining the security and stability of the electricity grid operation. A consequence of weakened stability is the recent rolling blackouts, such as the 2016 state-wide blackout in South Australia [1] and 2021 Texas power crisis in United States [2].

This chapter reviews the definition and classification of the stability problems of power system, pinpoints the shortcomings of traditional and state-of-the-art stability control methods, and advocates the contributions of this study.

1.1 Power System Stability

In the operation of power system, various physical disturbances may occur in the course of power generation, transmission, distribution and consumption components and related subsystems. Stability is a prerequisite and as well an important requirement during the operation of a power system. It usually refers to the ability of the system to withstand upcoming disturbance (unexpected events) without interrupting customer service and to keep the whole system intact [3].

However, increasing attention to global climate change has accelerated the shift from fossil fuels to sustainable energy. In such a transformation, electricity will become the main energy carrier, with the fact that the future smart grid will be much more complex than that in the past. With the increase integration of renewable energy penetration, power generation will become increasingly stochastic, which demand will become more active as households and businesses take new prosumer roles. However, due to the high cost and delay of network capability upgrade, all these new elements pose a threat to maintaining the safety and stability of the operation of the power grid. As a result, decline in stability includes recent rolling outages, such as 2016 national outages in South Australia [1] and 2021 Texas power crises in Texas [2].
Concerns about power system stability can be divided into three categories: rotor angle stability, voltage stability, and frequency stability. Within each category, stability issues can be further classified according to the magnitude of the disturbance experienced and/or the length of the time span considered. The power system stability classification is shown in Fig. 1.1.

![Fig. 1.1 Classification of power system stability.](image)

Rotor angular stability. It refers to the ability of a synchronous machine in an interconnected power system to maintain synchronization after disturbances. This devolves on the ability to maintain or restore the balance between the electromagnetic and mechanical torques of all synchronous generators in the system. Generator instability can increase in the form of generator wobble, which can lead to asynchrony with other generators. System stability devolves on the synchronous torque component of each synchronous generator. It is in phase with the rotor angle deviation, damping torque component, and rotational speed deviation. The lack of sufficient synchronizing torque will end in aperiodic or non-oscillating instability, while the lack of damping torque will trigger oscillatory instability. To facilitate analysis and gain insight into the nature of the stability problem, rotor angular stability can be described in terms of the following two subcategories:

1. Small disturbance (or small signal) rotor angular stability is associated with the ability of the power system to maintain synchronization under small disturbances. The perturbations are considered small enough to allow a linearized analysis of the system equations [6], [7], [8]. Small perturbation stability devolves on how the system initially operates. There are two possible forms of instability: I) rotor angle increases through non-oscillating or aperiodic modes due to the lack of synchronizing torque, or II) rotor oscillation amplitude increases due to the lack of sufficient damping torque. The time
of interest in small disturbance stability research ranges from 10 to 20 seconds after disturbance.

(2) Generally speaking, large-disturbance rotor angle stability or transient stability refers to the ability of a power system to maintain synchronization when subjected to severe disturbances (such as transmission line short circuits). The system response involves large deviations in generator rotor angle and is subject to nonlinear power-angle relationships. Transient stability devolves on the initial operation of the system and the severity of the disturbance. Instability is usually aperiodic angular separation caused by insufficient synchronizing torque, manifested as first swing instability. As-a-matter-of-fact, transient instability may not always arise in large power systems during the first mock exam. This may be the result of the superposition of the swing pattern between the slow regions and the local device swing pattern, which ends in a large rotor angle deviation after the first swing [6]. It could also be the result of the first simulated tests of nonlinear effects, which cause instability after the first swing. The time range of interest for transient stability studies is typically 3 to 5 seconds after disturbance. For very large systems with major interregional fluctuations, it may extend to 10-20 seconds.

Voltage stability. It refers to the ability of a power system to keep all busbar voltages stable after being disturbed by given initial operating conditions. It devolves on the ability to maintain/restore the balance between load demand and load supply in the power system. Possible instability in the voltage of some busbars presents with the form of a gradual decrease or increase. One possible result of voltage instability is a loss of load in the area, or a protection system tripping transmission lines and other components, resulting in cascading power failures. Loss of synchronization of some generators can be caused by these outages or operating conditions that violate the field current limit [9]. As shown in Fig. 1.1, voltage stability can be a short-term phenomenon or a long-term phenomenon:

(1) Short-term voltage stability involves the dynamics of fast-acting load components, such as induction motors, electronically controlled loads, and HVDC converters. The study time of interest is on the order of a few seconds, while the analysis demands dealing with the appropriate system differential equations. This is similar to the rotor angular stability analysis. Dynamic modelling of loads is often important. Short circuit faults near the load are more likely to cause short-term voltage stability problems than that of rotor angle stability.

(2) Long term voltage stability issues for slow moving equipment such as tap changer transformers, thermostatically controlled loads and generator current limiters. The study
time of interest can be extended to minutes, requiring long-term simulations to analyse the dynamic performance of the system [11], [12], [13]. Stability usually devolves on the disruption of the device rather than the severity of the initial disturbance. Instability is ascribed to a loss of long-term balance (for example, when a load tries to restore its power beyond the generation capacity of the transmission network, and is connected), the steady state operating point after a disturbance is small disturbance instability, or lack of force to restore balance after a disturbance (eg when the remedy should be adopted too late) [9], [10]. In addition to large disturbances, persistent load increases (such as morning load increases) may also trigger instability events in this subcategory.

Frequency stability. It points to the ability of the power system to maintain a stable frequency after severe system disturbance, resulting in a serious imbalance between power generation and load. This devolves on the ability to maintain/restore the balance between generation and system load while minimising unexpected load losses. Instability, possibly in the form of continuous frequency swings, causing the genset and/or load to trip. Severe system disturbances often end in large changes in frequency, power flow, voltage, and other system variables. It thus causes propagation, control, and protection measures that cannot be modelled in traditional transient stability or voltage stability. During frequency excursions, the characteristic times of activation processes and equipment range from fractions of a second associated with fast responses of equipment (such as low frequency load shedding and generator control and protection) to minutes corresponding to slow responses of equipment, such as prime mover energy supply systems and load voltage regulators. Voltage amplitudes can also change significantly with frequency, especially under islanding conditions where low frequency load shedding can unload the system. High voltages can trigger accidental tripping of the generator due to poorly-designed or uncoordinated field relays or voltage/hertz relays. Overload System voltage that is too low may arise in the impedance relay to malfunction.

1.2 Power System Stability Control

Power systems are armed with control capacities from various resources. As a result, upon an unstable propagation of the system, the appropriate control force can be quickly applied to regain system equilibrium. The power system stability control measures can be divided into two categories. One is to strengthen the grid structure and the other is to protect and control measures to improve the stability of the system. (1) Strengthen the grid structure and improve the stability of the system. The FACTS devices installed in the middle of long transmission lines can quickly adjust the reactive power of the system to maintain the desired voltage level of the line.
important approach to reinforce the network and improve the stability capability of the system. (2) Power system protection and stability control. The control schemes to improve the stability can be summarized into two categories: ① While the system is operating under steady state, one shall take control measures to improve the security of the system so that the system can withstand a credible set of disturbances. This type of control scheme is called preventive control. ② After being subject to a disturbance, one shall take control measures to cease unstable propagation in the system at the emergency stage. This type of control scheme is typically called emergency control, which will be the highlight in this research.

Several main emergency control measures include: (1) Cut off the machine at the sending end and/or the load at the receiving end to improve the stability of the whole system, thus ensuring continuous power supply for most users. (2) Generator excitation system and control: power generation. The machine excitation system, a necessary equipment for the normal operation of the power system, can quickly regulate the voltage at the generator end under fault conditions, while promoting a fast and stable swing of voltage and electromagnetic power. (3) Electric braking and its control device. At the moment of large disturbance, the output electromagnetic power of the generator at the sending end decreases, while the power of the prime mover remains unchanged, resulting in excess power and increasing the power angle between the generator and the system. If no measures are taken, the generator will be out of step. (4) Quick closing valve and its control. In the case of system instability, another measure to reduce power imbalance is to shut off the valve to reduce the input power of the generator.

Based on the above, load shedding is an important control measure to maintain stability of power systems, which can effectively prevent overload and damage on generators, and further avoid system collapse. This research will include load shedding as the control measure to demonstrate the efficacy of the developed methodologies for power system emergency stability control.

1.3 Load Shedding in Power System

As a control action happens urgently, load shedding (LS) can effectively cease potential unstable propagation and protect the power system from instability or collapse. In practice, its strategies include event-driven load shedding (ELS) and response-driven load shedding (RLS).

The ELS strategy is based on a decision table characterized by inputs and outputs, where inputs include quantities representing the current system’s operating condition (OC) and contingencies representing the possible disturbances, and outputs are LS locations and amounts.
Based on the input feature information (i.e. current OC and the encountered disturbance), the output can be immediately indexed from the table and triggered by the corresponding LS action. As a result, there is little time delay after the inception of the disturbance (although communication and relay actions generally lead to a very short delay). ELS is open loop in control logic, it only reacts once to the detected events and does not require further input. The typical process of ELS is shown in Fig. 1.2.

Compared to ELS, RLS operates in a very different way. It triggers the LS action depending on the response of critical electrical quantities (usually voltage and frequency) after the disturbance occurs. Based on the electrical quantity being monitored, the RLS strategy is divided into undervoltage LS and underfrequency LS. They try to cancel out the voltage/frequency decay after the voltage/frequency has dropped and remained below the pre-set threshold for an unacceptably long period of time. The input is a local voltage/frequency measurement, while the output is a local LS action determined by its settings. These settings include the total number of LS phases, the percentage of load to be unloaded for each phase, and the time delay for each phase. The principle of RLS follows a closed-loop control logic as it demands continuously refreshed response (real-time voltage/frequency measurement) to guide its actions. The logic flow of underfrequency LS is demonstrated in Fig. 1.3.
Since ELS can gain control immediately after detecting the disturbance without significant time delay, its action is set on an earlier stage condition of the system, which enables ELS usually outperforming RLS in controlling cost and efficiency. Strong system ELS from substantial renewable energy sources and more active demand side behaviour are more appropriate. The fast reaction of ELS also well suits solving transient instability problem that could arise before large voltage/frequency excursions [16], [17]. This research touches on developing advanced ELS tools for transient stability control in renewable energy systems.

1.4 Decision-Making on ELS Strategy

The conventional option to determine ELS energy can be drawn into two categories: optimization-based and sensitivity-based approach. The optimization-based approach models ELS as an optimization problem with nonlinear constraints, which can be solved by virtue of state-of-the-art optimization techniques. This approach typically proceeds in an offline-optimization-online-matching fashion: optimal ELS strategies for different contingencies are offline calculated and stored in a lookup table, while the online ELS action is decided as the strategy that best matches the encountered contingency. Although efforts have recently been attached to enhancing optimization efficiency, such as the parallel methods [18], [19] and the constraint relaxation method [20], yet these approaches still suffer from its stagnant reaction to the intense system state variation and require a specific lead time to finalize the ELS strategy. Alternatively, the sensitivity-based approach derives the ELS strategy by adaptively shedding the load on buses with the highest sensitivity to improve stability margin [17], [21]. Although this approach can reduce the computation complexity compared to full optimization-based approach, it impairs the control accuracy, and its developed ELS strategy may fail to guarantee global optimality.
The future energy systems with digital divide propel the comprehensive deployment of advanced sensing facilities, such as phasor measurements and smart metering, to provide wide-area visibility of the electricity grid. In such an environment with massive data, while the system state can be captured in real-time by these devices, it also provides valuable opportunity to decide the ELS strategy in real-time so as to adapt to the intense energy system variations. In recent years, artificial intelligent (AI) have been identified as an effective and efficient data-driven tool for real-time power system stability analysis and other energy applications, such as economic dispatch [22], [23] and infrastructure planning [24]. Its principle is to interpret the nonlinear relationship between operating features and stability assessment/control targets in the offline simulated database. As a result, the intelligent model can be adopted online to make instantaneous decisions depending on the incoming disturbances. In the literature, extensive AI methods, including shallow neural networks [25], [26] and deep learning methods [27], have been developed for real-time stability assessment. Whereas, the approaches for stability control are very limited. In [28], a line load reduction method based on graph convolutional neural network was developed. In [29], a deep reinforcement learning method was developed for under-voltage load shedding. Regarding ELS, extreme learning machine (ELM) algorithm was applied in [30], [31] to maintain post-fault frequency stability. The method was further developed in [32] to alleviate fault-induced delayed voltage recovery (FIDVR).

1.5 Contributions of This Research

Although the reported methods demonstrate some degree of effectiveness in emergency control, they are exposed to the following shortcomings:

1) They are concerned with voltage or frequency stability, but less about transient rotor angle instability that deteriorates system stability more rapidly. It is difficult to solve by response-driven control schemes.

2) Existing data-driven methods determine the load reduction on each load bus by adopting shallow neural networks based on multiple input and single output configuration. However, ELS is essentially a multi-input and multi-output task, involving load reduction strategies at multiple locations for systematic optimization. It also demands integral learning of highly nonlinear relationships. Therefore, machine learning algorithms that can handle more complex problems are required.

3) The prediction error of ELS strategy is bidirectional. In other words, the predicted load reduction can be greater or less than the optimal level. In practice, the costs of excessive load
reduction and conservative load reduction are actually unbalanced. Over-cutting refers to the reduction of more load than what is needed. The cost is only the amount of load loss that exceeds the optimal value. However, under-cutting means that the load reduction is insufficient to restore stability, which can lead to system instability and even cascading events, resulting in higher costs. The unbalanced costs in deciding ELS strategy are excluded in existing methods and their balanced evaluation of prediction accuracy does not really reflect the operating costs involved. This demands a new ELS approach to manage cost imbalances.

The methodologies proposed in this research target on real-time power system stability control, mainly with three key contributions. The contribution flow chart and the corresponding problems solved are given in Fig 1.4.

(1) A deep neural network (DNN) is established as a machine learning model, which is the first attempt to apply deep learning to ELS prediction. Compared with the shallow learning algorithm in the literature, DNN can deeply mine the inherent complex relationships in the data and significantly reduce the ELS prediction error.

Fig. 1.4 The contribution flow chart of this study. The corresponding problems are solved.

(2) In order to alleviate the cost imbalance in the training of deep learning model, a risk-averse learning algorithm with specially-designed loss function is proposed. The loss function
transforms the constrained ELS optimization into an unconstrained problem with post-control stability adjustment. This aims to improve the control success rate of ELS and reduce the overall control cost. This risk-averse learning algorithm intends to avoid insufficient load leading that causes cascading unstable events.

(3) The graph neural network model is introduced into ELS prediction for the first time to embed the networked topology of power grids into machine learning. A graph structure related model suitable for large-scale power system is constructed, which can effectively reduce the control cost and improve the control success rate of ELS in contrast to the fully-connected DNN.
Chapter 2. ELS Database Generation

The data collection process is an important part in developing data-driven models. It directly involves in the training results of the model. The quality of data collected will have a greater impact on the training accuracy and efficiency of the model. This chapter presents the database generation process for data-driven ELS.

2.1 Introduction

ELS database generation is the first step in data-driven application for ELS. The obtained ELS database contains a large number of ELS samples with unstable operation taking the input vector and optimal load shedding strategy as the output vector. The ELS database generation module essentially engages in an operation sampling process and an ELS strategy optimization process to implement the ELS database for each contingency. The obtained ELS database will be adopted to train the AI models for real-time ELS decision-making.

2.2 Evolutionary Algorithm

Evolutionary algorithm (EA) is an "algorithm bunch" [34]. Although it has many changes, it has different gene expression modes, different crossover and mutation operators, and the quotation of special operators, and different regeneration and selection methods. They are inspired by the biological evolution of nature and are a general population-based metaheuristic optimization algorithm. Compared with the traditional optimization algorithms based on calculus and enumeration, evolutionary computing is a mature global optimization method with high robustness and wide applicability. It is self-organized, self-adaptive, and self-learning, which can effectively resolve complex problems that are difficult to solve by traditional optimization algorithms [35].

Evolutionary computation is a robust computational method. It can adapt to different problems in different environments and locate satisfactory and efficient solutions in most cases. Instead of dealing directly with the specific parameters of the problem, it explains a coding scheme for the whole parameter space of the problem. Besides, it searches from a set of initial points rather than search from a single initial point. The search applies information about the value of the objective function, which may not require derivative information of the objective function or special knowledge about specific problems. As a result, evolutionary algorithms are widely used and suitable for highly non-linear problem, easy to modify and parallelizable.
The ELS problem is a high-dimensional, non-linear and non-convex problem regarding to the state of power systems, which has been explained in [4], [5] and [32]. The traditional mathematical calculation methods can be adopted to analyze the optimal LS. Although a certain level of accuracy can be fulfilled, ELS optimization on large-scale power systems is a problem with high-dimension optimization, which is a typical black box problem and cannot be easily solved by generic analytical formulae. In comparison, EA is a heuristic optimization algorithm that relies on the directed evolution of randomly populated samples towards the optimal solution. It can theoretically suit arbitrary nonlinear and non-convex problems including the one for ELS. The specific flow of EA is shown in Fig. 2.1. By setting up the initial population, adjusting the fitness function, choosing the best part of the hyperparameter combination, random crossover and introducing a new hyper-parameter space, we break through the current search limit, which are more conductive to the algorithm to find a better solution. Finally, we get the optimal solution required for ELS.

![Fig. 2.1 Process of EA](image)

### 2.3 ELS Data Sampling

#### 2.3.1 Operation sampling

The integration of renewable energy has injected increasing variations to power system operation. To visualize such variations, renewable generation and load demand are considered independent variables and randomly sampled, from which the operation of the complete system can be derived by optimal power flow. A large number of operations are initially singled out from the high band of load demand, i.e. \[ P_i \times (1 + x\% \times P_i) \] (\( P_i \) is the benchmark load power on the \( i \)-th load bus). By doing so, the sampled operations focus on the situations with exhausted power sources, where instability is more likely to occur in response to large disturbances. As ELS is a control scheme to be activated only under unstable states, such sampling mechanism aims to increase the chance of valid sampling for ELS database. Moreover, the initial samples are sent to time-domain simulations (TDS) to further examine their stability. Only the unstable operations are preserved, which guarantees the validity of all collected samples.
2.3.2 Heuristic ELS optimization

The proposed ELS method is based on supervised learning paradigm where the intelligent model learns from data samples with known outputs to be predicted [22]. For ELS, the prediction output of the intelligent model is the optimal ELS strategy (i.e. the optimal amount of load to be shed at each candidate bus). It is essential to label the sampled operating scenarios and thus construct a complete ELS database for supervised ELS learning. To find the optimal ELS strategy, the following optimization problem is formulated.

\[
\min_L f(L) = \sum_{i=1}^{m} l_i \tag{1-a}
\]

\[
\text{s.t.} \left\{ \begin{array}{l}
0 \leq l_i \leq \bar{l}_i, \ i = 1, 2, ..., m \tag{1-b}
\eta(s, L) > 0 \tag{1-c}
\end{array} \right.
\]

where \(l_i\) is the load shedding amount on the \(i\)th load bus; \(m\) is the number of candidate buses for load shedding; \(f(L)\) is the objective function for ELS optimization. It intends to minimize the total load shedding amount in this research. Nevertheless, other objectives are also applicable depending on the practical needs.

Regarding the constraints, (1-b) refers to the upper bound limit of load shedding percentage as contracted at each candidate bus. (1-c) is the examination of post-control stability, i.e. the stability of the system is regained after applying the ELS strategy. \(\eta(s, L)\) represents the transient stability margin to evaluate the system stability after the control, where \(s\) is the operation and \(L\) is the load shedding action. As shown in Fig. 2.2, \(\eta(s, L)\) is calculated from the post-fault rotor angle trajectories obtained by TDS. A positive margin value signifies the stable condition of the system post to the control action.

In this research, the nonlinear optimization problem in (1-a) - (1-c) is offline solved by evolutionary algorithm [33]. Compared to the deterministic approaches such as trajectory sensitivity analysis method in [32], evolutionary algorithm is a population-based heuristic approach that outfits to solving nonlinear problem and shows better tendency to global optimum for ELS strategy. The iterative process to implement heuristic ELS optimization is presented in Fig. 2.2, which includes initialize ELS population, post-control stability screening, calculate fitness function, terminate condition check, and update population.
2.4 Simulation Results

The ELS database generation method is testified on New England 39-bus system and Nordic 41-bus system for transient stability control. The network structures of the two systems are shown in Fig. 2.3. The New England system has 10 generators, 39 buses and 46 branches. The generators on bus 37 are displaced by wind farms with the same capacity, resulting in renewable penetration of over 20%.

In comparison, the Nordic system is powerful and more complicated, covering the actual power grids in Finland, Sweden, Denmark, and Norway, which includes 23 generators (including several synchronous condenser), 41 buses, and 69 branches, with generators G1, G10, and G17b replaced by wind farms. The renewable penetration goes over 30%.
Fig. 2.3. Power systems for case study (a) New England 39-bus system, (b) Nordic 41-bus system.
As illustrated in Section 3, the database is completed by the following factors: operations, nominated contingencies, and available load shedding resources. TDS is run on PSS/E platform to derive the EEAC margin for each operation. The TDS step size is set to 0.01s and the maximum simulation time is 10 seconds. Evolutionary algorithm (EA) is considered as the heuristic algorithm to find the optimal ELS strategy for each generated sample.

2.4.1 Pre-fault operations

Based on the operation sampling method proposed in Section 3.1, 3766 unstable operations are modelled for New England system by applying OPF and TDS, while 3426 unstable operations are generated for Nordic system. For both systems, the load levels are uniformly singled out from 80% to 120% of rated power. Whereas the wind power is singled out from 0% to 100% of wind farm capacity. Each operation includes the following features: the active and reactive power of the generator, the voltage amplitude and angle, the active and reactive loads, and the current on the transmission line. In total, there are 314 operating features for New England system and 407 for Nordic system.

2.4.2 Nominated contingencies

In this study, two severe three-phase faults are nominated as contingencies for each test system to verify the proposed method. One database is generated on each nominated fault. A third database is also generated for each test system by mixing the data of the two faults. In this way, the ELS performance on multiple faults can be investigated.

For New England system, the fault location is bus 17 (noted as NEfault # 1) and bus 25 (noted as NEfault # 2), with fault duration 0.2 s. NEFault #2 is cleared by tripping the line between buses 25 and 26.

For Nordic system, the fault location is bus 4031 (noted as NDfault # 1) and bus 4044 (noted as NDfault # 2), with fault duration 0.2 s. NDfault #1 is cleared by tripping the line between buses 4022 and 4031, while NDfault #2 is cleared by tripping the line between buses 4041 and 4044.

2.4.3 Available load shedding resources

Based on the preliminary sensitivity analysis in [22], buses 4, 8, 20 and 39 of New England system and buses 1044, 4043, 4051 and 4072 of Nordic system are nominated as candidate buses for load shedding. Moreover, in practice, it is inappropriate in ELS to shed all the loads on a bus. In this study, the maximum load shedding percentage at all candidate buses is maintained to 90% of the load.
2.4.4 ELS optimization

Each data sample is labelled by its optimal ELS strategy computed using EA algorithm. The population size of EA is set to 50; the mutation rate of each generation is 0.3; while the maximum number of iterations is 100. The iterative updating is stopped when the optimal solution of five consecutive generations remains unchanged.

As a heuristic algorithm, EA may end in a local optimum in some optimization problems [44], [45]. To verify it, a sample is randomly nominated from the database and EA algorithm is rolled out on this sample for 100 times with randomized initial populations. It reveals that the 100 EA executions all converge to the same optimal result. Five of the executions are randomly singled out to illustrate the convergence result, and the variations in the best objective function value at each iteration are displayed in Fig. 2.4. All the executions converging to the same optimal objective function value verifies that EA is able to ensure global optimization when solving ELS problem.

Fig. 2.4. Variations in objective function value in 5 EA executions.
Chapter 3. Deep Neural Networks

The existing data-driven methods applied the shallow neural network based on multi input single output configuration to determine the load reduction of each load bus respectively. However, ELS is essentially a multi-input and multi-output task, which involves the load shedding strategy of system optimization. It is necessary to integrate the highly nonlinear relationship. Therefore, the deep learning model with multi-layer network is better for this research.

3.1 Introduction

DNN is a multi-layer neural network model. It contains three types of layers: an input layer, hidden layers, and an output layer. A typical DNN structure can be represented as:

\[ h_0 = s \]

\[ h_i = \sigma(a_i h_{i-1} + b_i), i = 1, 2, ... K \]

\[ L = \sigma'(a_o h_K - b_o) \]

where \( h_0 \) is the network input vector (i.e. the real-time operation \( s \)) received at the input layer; \( h_i \) is the output vector at the \( i \)th hidden layer; \( h_K \) is the output vector at the output layer; and \( L \) is the predicted ELS strategy represented by a vector of the percentages of load to be shed at each candidate bus.

The core of DNN is the back-propagation algorithm, or BP algorithm for short. The back-propagation algorithm mainly includes two links (excitation propagation and weight update) repetition and loop iteration until the network’s response to the input reaches the predetermined target range.

The learning process of BP algorithm includes forward propagation process and backward propagation process. In the former, the input information is processed layer by layer and transmitted to the output layer through the input layer and hidden layer. If the expected output value is unavailable in the output layer, the sum of squares of the output and expected error is taken as the objective function and passed to backpropagation. The partial derivative of the objective function to the weight of each neuron is calculated layer by layer. The gradient of the objective function to the weight vector is formed as the basis for modifying the weight. The learning of the network is done in the process of modifying the weights. When the error reaches the expected value, the network learning ends [44].
3.2 Application of DNN in Power System ELS

There is a high-dimensional non-linear relationship between the power system operation and optimal ELS strategy. The existing data-driven ELS methods in [30]-[32] adopt shallow neural networks with a single hidden layer. These models are defined in relatively simple model structures that may oversimplify the complicated data relationship in the ELS problem, which impairs the accuracy of ELS strategy prediction. By contrast, DNN model can compress the inherent relationship into a multi-hidden-layer structure, which shows better ability to fit complex nonlinear mapping [45] and thus tends to provide more accurate ELS prediction results.

The overall aim of DNN training is to find the best values of the learnable parameters (i.e. weights $\alpha$ and biases $\beta$) at each layer. By adjusting these parameters, the training process can be regarded as minimizing the mean value of the loss function given the training samples:

$$\min_{\theta} \frac{1}{N_{\text{train}}} \sum_{k=1}^{N_{\text{train}}} \Gamma_k \quad (5)$$

where $\theta$ represents the parameters set including the weights and biases across all layers of DNN. $N_{\text{train}}$ is the number of training samples, $\Gamma_k$ is the loss function value of the $k$th training sample.

An efficient training technique, namely stochastic gradient descent (SGD) [46], is used in this research to train DNN. The mechanism of SGD is to select a subset of the summation function at each step, which can significantly reduce the computation burden in each iteration. The training is carried out by the back propagation model and the parameters are updated by SGD:

$$\Delta \theta(t+1) = \Delta \theta(t) + \rho \frac{\partial L_{\text{total}}}{\partial \theta} \quad (6)$$

where $\rho$ is the learning rate.

3.3 Simulation Results

The performance of DNN is tested using the database generated in Section 2.3, on both New England and Nordic systems. The whole database is divided into three subsets (i.e. training set, validation set, and testing set). The training set is adopted to develop DNN model. The validation set is adopted to tune the hyper-parameters involved. The testing set is adopted to test and verify the performance of proposed method. In this study, the training set, validation set and testing set are randomly singled out from the full ELS database in a ratio of 3:1:1.
3.3.1 DNN configuration

DNN serves as the core machine learning algorithm for the proposed method. Choosing appropriate hyper-parameters, such as number of hidden layers and neurons in each layer, is important for DNN to fulfill its full learning strength and avoid the unnecessary training burden.

In this study, the DNN model is developed on Python platform using PyTorch 1.6. It utilizes SGD method for DNN training. During training, epoch is set to 200 and batch size is 64. The learning rate is 0.001. Other parameters involved in the training process, including $\omega$, are adjusted based on the validation results. As a result, we use $\omega = 0.0001$ for DNN training. The DNN hyper-parameters are also tuned according to the validated results. The final structure of the developed DNN is stereoscopically shown in Fig. 3.1. The DNN has 3 hidden layers, with 128, 64, and 32 neurons in each layer.

![Diagram of DNN model](image)

Fig. 3.1. The detailed structure of the developed DNN model.

3.3.2 Performance Metrics

We use three metrics to evaluate the ELS performance of DNN. They are mean absolute percentage error (MAPE) and success rate (SR). The details of these metrics are presented below.
The MAPE is a widely-used to measure the regression error of intelligent models. For ELS prediction, MAPE refers to the mean deviation between the predicted and the true optimal load shedding percentage on each candidate bus, which is formulated as

$$MAPE = \frac{1}{N_t} \sum_{i=1}^{N_t} \left( \frac{1}{N} \sum_{j=1}^{N} |\hat{y}_i - y_i| / y_i \right)$$

(7)

where $N_t$ is the total number of tested samples.

Besides the ELS prediction error measured by MAPE, the chance to achieve post-control stability, namely SR, is also essential to evaluate the ELS performance. Here, we apply the predicted ELS action on each tested sample and run further TDS to calculate the EEAC margin. The sign of EEAC margin indicates the post-control stability. SR is the percentage of tested samples whose transient stability is successfully regained after applying the predicted ELS action, which is formulated as

$$SR = \frac{\sum_{i=1}^{N_t} e_i}{N_t}$$

(8)

where $e_i$ is a binary value to note the success or failure of the ELS action. According to the post-control stability constraint in (3-c), if $\eta_{EEAC} > 0$, $e_i$ is set to 1; otherwise, $e_i$ is set to 0.

3.3.3 Comparative Studies

The index comparison research described in the previous section is carried out to validate the effectiveness of the proposed DNN model in ELS method. MAPE and SR indicators are calculated based on their ELS results. They are stated in Table I and table II. The proposed ELS method adopts deep learning technology to enable more accurate ELS estimation. Therefore, the two tables compare the proposed DNN model with other deep learning algorithms and the most advanced shallow learning algorithms. The MAPE, SR of different algorithms on the two test systems are presented in Table I and table II. Shallow algorithms include limit learning machine (ELM) [32], random forest (RF), support vector machine (SVR) and relevance vector machine (RVM) [47]. The results in the two tables outline that deep network has greater advantages in the complex system of power system comparing with shallow network. Each index is far better than other artificial intelligence models.
Computation efficiency

3.3.4 Compared to the traditional optimization-based ELS method, the core of proposed intelligent ELS method is its real-time computation speed to adapt to the variations in operation. The average computation time to predict the ELS action for each tested sample is presented in Table III. The proposed DNN model can deliver the results in milli-second level. It is a significant elevation in computation efficiency compared to the 15- and 20-minute computation time of optimization-based method. The fast online computation speed of DNN

Table I
TESTING RESULTS ON NEW ENGLAND SYSTEM

<table>
<thead>
<tr>
<th>Model</th>
<th>NEfault #1</th>
<th></th>
<th>NEfault #2</th>
<th></th>
<th>Mixed Fault</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>SR</td>
<td>MAPE</td>
<td>SR</td>
<td>MAPE</td>
<td>SR</td>
</tr>
<tr>
<td>DNN</td>
<td>3.85%</td>
<td>84.19%</td>
<td>3.36%</td>
<td>85.17%</td>
<td>4.05%</td>
<td>83.49%</td>
</tr>
<tr>
<td>ELM</td>
<td>8.58%</td>
<td>79.16%</td>
<td>8.35%</td>
<td>80.14%</td>
<td>9.19%</td>
<td>78.14%</td>
</tr>
<tr>
<td>RF</td>
<td>5.61%</td>
<td>82.91%</td>
<td>4.93%</td>
<td>84.01%</td>
<td>5.02%</td>
<td>81.16%</td>
</tr>
<tr>
<td>SVM</td>
<td>7.37%</td>
<td>80.49%</td>
<td>7.23%</td>
<td>81.06%</td>
<td>8.87%</td>
<td>79.13%</td>
</tr>
<tr>
<td>RVM</td>
<td>6.42%</td>
<td>81.34%</td>
<td>5.98%</td>
<td>82.94%</td>
<td>6.25%</td>
<td>80.06%</td>
</tr>
</tbody>
</table>

Table II
TESTING RESULTS ON NORDIC SYSTEM

<table>
<thead>
<tr>
<th>Model</th>
<th>Fault#1</th>
<th></th>
<th>Fault#2</th>
<th></th>
<th>Mixed Fault</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>SR</td>
<td>MAPE</td>
<td>SR</td>
<td>MAPE</td>
<td>SR</td>
</tr>
<tr>
<td>DNN</td>
<td>4.33%</td>
<td>80.16%</td>
<td>4.23%</td>
<td>79.10%</td>
<td>4.36%</td>
<td>81.22%</td>
</tr>
<tr>
<td>ELM</td>
<td>9.52%</td>
<td>76.15%</td>
<td>9.67%</td>
<td>75.15%</td>
<td>9.64%</td>
<td>77.26%</td>
</tr>
<tr>
<td>RF</td>
<td>5.39%</td>
<td>81.12%</td>
<td>5.89%</td>
<td>80.19%</td>
<td>6.14%</td>
<td>82.14%</td>
</tr>
<tr>
<td>SVM</td>
<td>10.12%</td>
<td>76.24%</td>
<td>9.12%</td>
<td>74.26%</td>
<td>9.94%</td>
<td>74.18%</td>
</tr>
<tr>
<td>RVM</td>
<td>9.61%</td>
<td>78.12%</td>
<td>8.94%</td>
<td>77.42%</td>
<td>9.67%</td>
<td>78.37%</td>
</tr>
</tbody>
</table>

Table III
AVERAGE COMPUTATION TIME PER SAMPLE (CPU TIME)

<table>
<thead>
<tr>
<th></th>
<th>New England System</th>
<th>Nordic System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data-driven Method</td>
<td>0.248 ms</td>
<td>0.262 ms</td>
</tr>
<tr>
<td>Optimization-based Method</td>
<td>15.24 min</td>
<td>20.13 min</td>
</tr>
</tbody>
</table>
realizes the real-time ELS prediction, whereas the long computation time of optimization-based method disables it to rapidly respond to the operation variation.
Chapter 4. Risk-Averse DNN Model for ELS

Predicting ELS strategy using traditional deep learning model can generate errors in both directions. In other words, the predicted load reduction can be greater or less than the optimal level. As-a-matter-of-fact, he costs of excessive load reduction and conservative load reduction are actually unbalanced. Over-cutting refers to the reduction of more load than is needed, while the cost is only the amount of load loss that exceeds the optimal value. However, under-cutting means that the load reduction is not sufficient to restore stability, which can lead to system instability and even catastrophic events, resulting in significantly higher costs. However, how to usher the training direction to ensure orientated training results is one of the cores of this study.

4.1 Introduction

The proposed ELS method exploits DNN as the core engine to deep mine the relationship between operation and optimal ELS strategy from database. It then provides real-time prediction of online ELS actions. Although it is a data-driven approach, the overall aim is basically to solve the optimization problem (1-a)- (1-c) in real-time to meet its objectives and constraints. The ultimate goal is to minimize the total load shedding amount while ensuring the post-control stability.

The standard DNN is typically trained to minimize the generated error between its estimated results and the true output labels. It specifies that the DNN is exposed to bi-directional error with balanced costs, i.e. the predication results have equal chances to be higher or lower than the true values. Eventually, the ELS control costs generated by the errors in the two directions are actually imbalanced. An ELS strategy with over-estimated shedding amount only increases the control cost incurred by the amount exceeding the optimal value. However, an ELS strategy with under-estimated shedding amount can end in more loads to be shed by the activation of subsequent RLS schemes or even lead to catastrophic blackout events. This means huge costs will be generated and must be avoided. Following such logic, since the standard learning mechanism of DNN does not take the constraints of the post-control stability into consideration, it is not suitable for ELS prediction task.

Considering above inadequacies, this chapter proposes a deep risk-averse ELS model with regulated deep learning process to elevate the chance of post-control stability. It presents the framework of risk-averse ELS method, elaborates the loss function design and outlines the corresponding DNN training method.
4.2 Overall Framework of Risk-Averse DNN Model

The proposed ELS method aims to develop an intelligent model that can make real-time decisions on the optimal ELS strategy. Its objective is to minimize load shedding cost while ensuring the post-control stability of the system [49]. The overall framework is presented in Fig. 4.1 where the whole method is implemented in a data-driven manner, including a ELS model development phase and a real-time ELS decision-making phase. ELS model development refers to an offline process where the risk-averse ELS model is developed through an ELS database generation module, a load shedding inference modelling module, and a risk-averse training module.

As explained in Chapter 2, the purpose of ELS database generation is to create a comprehensive database that contains numerous unstable samples with essential ELS information for intelligent model training. For each sample, its pre-fault operation is singled out from the feasible operating space defined by the grid standard, with the optimal ELS strategy derived in a heuristic optimization process. An effective ELS action demands the system stability to be regained in response to the ELS control [32]. However, this post-control stability information is unknown prior to taking the load shedding action.

The load shedding inference modelling is to locate the best model that maps the relationship between the transient stability margin and the total load shedding amount, from which the post-control stability can be speculated. Such load shedding inference model is adopted to convert the transient stability assessment results into an effective regulation for optimal ELS decision-making. It thus serves as a preliminary step to get ready the essential ingredients for ELS model training.

Risk-averse training is the key module for developing risk-averse ELS model. The post-control stability inferred from transient stability margin is incorporated to penalize the unfavored deep learning direction. It can help reduce the transient instability risk should the ELS action be taken.

During the real-time ELS decision-making phase, the fully developed ELS model is applied online to decide the optimal ELS strategy instantaneously based on the real-time system measurement and the encountered contingencies.
4.3 Transient stability margin

In the proposed shedding amount inference model, an appropriate transient stability margin that can reliably measure the severity of system instability serves as a basis to infer the minimum load shedding amount. In the literature, the criterion for transient stability is usually designed based on the maximum post-fault rotor angle deviation bounded by a pre-defined threshold, e.g., $360^\circ$ [32], [38], [39]. Such threshold-based criterion is always system and/or operation dependent, which may fail to truly reflect the stability status of the system [40].

Considering above inadequacies, this research leverages extended equal area criterion (EEAC) [38], [40] theory to define accurate and unified assessment on transient stability. The principle of EEAC is to first drive a full TDS to obtain post-fault multi-machine trajectories (here, complex system dynamic models can be used, and any network topology change will be considered). Then, it converts the multiplex transient trajectories into a one machine infinite bus (OMIB) trajectory from which the transient stability severity can be measured by equal area criterion (EAC). The details of EEAC are referred to [37], [38], [39] and [40], and some significant applications of EEAC are available in [41], [43].

Based on the EEAC theory, a transient stability margin has been established to quantitatively evaluate the stability severity of power system. It is widely shown that such EEAC margin has a quasi-linear relationship with many key power system parameters, such as generation output, fault clearing time, etc [35]. The EEAC margin is calculated as follows. Based on the post-fault trajectories, EEAC separates synchronous machines into two exclusive groups. The machines responsible for loss of synchronism are grouped as critical machines (CM), while other machines
are grouped as non-critical machines (NM). Then the OMIB trajectory is calculated as follows [42]:

\[
\delta(t) = \left(\sum_{i \in \mathcal{C}} M_i\right)^{-1} \times \sum_{i \in \mathcal{C}} M_i \delta_i(t) - \left(\sum_{j \in \mathcal{N}} M_j\right)^{-1} \times \sum_{j \in \mathcal{N}} M_j \delta_j(t) \tag{9}
\]

\[
P_m(t) = M \cdot \left[\left(\sum_{i \in \mathcal{C}} M_i\right)^{-1} \cdot \sum_{i \in \mathcal{C}} P_{mi}(t) - \left(\sum_{j \in \mathcal{N}} M_j\right)^{-1} \cdot \sum_{j \in \mathcal{N}} P_{mj}(t)\right] \tag{10}
\]

\[
P_e(t) = M \cdot \left[\left(\sum_{i \in \mathcal{C}} M_i\right)^{-1} \cdot \sum_{i \in \mathcal{C}} P_{ei}(t) - \left(\sum_{j \in \mathcal{N}} M_j\right)^{-1} \cdot \sum_{j \in \mathcal{N}} P_{ej}(t)\right] \tag{11}
\]

\[
M = \left(\sum_{i \in \mathcal{C}} M_i \cdot \sum_{j \in \mathcal{N}} M_j\right) \cdot \left(\sum_{i \in \mathcal{C}} M_i + \sum_{j \in \mathcal{N}} M_j\right)^{-1} \tag{12}
\]

where \(C\) and \(N\) denotes the set of CMs and NMs, respectively; \(\delta\) denotes the rotor angle of the OMIB; \(P_m\) and \(P_e\) denote the mechanical and electrical power of the OMIB, respectively.

By observing the performance of mechanical and electrical power of the OMIB in the \(P-\delta\) plane, the stability criterion is defined as follows:

\[
\eta_{EEAC} = A_{dec} - A_{acc} \tag{13}
\]

where \(\eta_{EEAC}\) is the EEAC margin; \(A_{dec}\) and \(A_{acc}\) refers to the decelerating (\(P_m(t) > P_e(t)\)) and accelerating (\(P_m(t) < P_e(t)\)) area, respectively. By definition, the range of EEAC margin is \([-100, 100]\), where positive and negative \(\eta_{EEAC}\) imply transiently stable and unstable situations, respectively.

In some occasions, there is no enclosed accelerating or decelerating areas in the OMIB \(P-\delta\) plane. In these situations, \(\eta_{EEAC}\) is set to -100 for the “extremely unstable” cases and 100 for the “extremely stable” cases.

### 4.4 EEAC-ELS Relationship

Provided that EEAC shows quasi-linear relationship with many key power system parameters [42], we can also estimate that the relationship between EEAC margin and the optimal load shedding percentage tends to be linear. An experiment on a benchmark power system is conducted to investigate this relationship. In this experiment, 500 ELS samples on a severe contingency are created following the database generation process as specified in Chapter 2 section 3. The optimal load shedding percentages and the EEAC margins of these samples are plotted in Fig. 4.2.
It is evident that the overall load shedding percentage in the optimal ELS solution shows a quasi-linear relationship with the EEAC margin of these unstable cases. We use the least square method [48] to get an approximate straight line to fit their relationship, from which the minimum load shedding amount can be derived:

\[
P_{\text{opt}} \propto \eta_{\text{EEAC}} \Rightarrow P_{\text{opt}} = -k_0 \times \eta_{\text{EEAC}} + b \Rightarrow LS_{\text{min}} = P_{\text{opt}} \times \sum_{i=1}^{N} P_{L,i}
\]

(14)

where \(P_{\text{opt}}\) is the minimum load shedding percentage; \(\eta_{\text{EEAC}}\) is the EEAC margin based on the power system operation and the experienced contingency; \(N\) is the total number of candidate buses for load shedding; \(P_{L,i}\) is the active load power at candidate bus \(i\); \(LS_{\text{min}}\) is the minimum load shedding amount to regain system stability; \(k_0\) is a positive constant coefficient; and \(b\) is a constant intercept. Since zero load shedding is needed when \(\eta_{\text{EEAC}} = 0\), \(b\) should theoretically be 0. \(k_0\) is estimated using ordinary least square (OLS) method on the full database to minimize the linearization error.

### 4.5 Loss Function Design

In the proposed method, a specially-designed loss function is adopted to train the DNN, which serves as the key innovation towards the risk-averse feature of DNN. Based on the optimization (1-a) - (1-c), there is a tradeoff between the applied load shedding amount and the
post-control stability. On the one hand, higher load shedding amount more likely regains the stability of the system but increases the control cost. On the other hand, the operators prefer to apply lower load shedding amount to reduce the control cost but takes higher risk on post-control instability. The loss function is designed to balance such tradeoff in DNN training, which can minimize the overall load shedding amount with regulation on post-control stability.

The loss function to train the risk-averse DNN model is formulated as follows

$$\Gamma = \lambda_{\text{reg}} + \omega \cdot \lambda_{\text{pen}}$$

(15)

where $\Gamma$ consists of two terms: $\lambda_{\text{reg}}$ and $\lambda_{\text{pen}}$. $\lambda_{\text{reg}}$ is the regular term represented by the ELS prediction error. $\lambda_{\text{pen}}$ is a penalty term to penalize the training outputs that under-estimate the optimal load shedding amount. It helps regulate the DNN to meet the post-control stability constraint. The incorporation of regular term and penalty term can transform the constrained optimization problem in (1-a)-(1-c) into an unconstrained one that can be solved in typical DNN training process [37]. $\omega$ is a positive weight value to balance the effects of the two terms.

Same as in the classic design of loss function, the regular term $\lambda_{\text{reg}}$ refers to the mean squared error (MSE) between the predicted ELS strategy and the true optimal solution, which is calculated as follows:

$$\lambda_{\text{reg}} = \frac{1}{N} \sum_{i=1}^{N} (\bar{l}_i - l_i)^2$$

(16)

where $\bar{l}_i$ and $l_i$ respectively represent the predicted and the optimal load shedding percentage at the candidate bus $i$.

The penalty term plays the role to penalize the violation of post-control stability constraint. Considering the difficulties in directly obtaining the stability information, we speculate the post-control stability by comparing the total load shedding amount in the predicted ELS strategy and the minimum load shedding amount inferred from EEAC margin. If the total load shedding amount in the predicted ELS strategy is less than the minimum value, the load shedding amount is under-estimated. As a result, the control force is insufficient to regain transient stability, simplifying the violation of post-control stability constraint. Following such logic, the penalty term is designed as follows.

$$\lambda_{\text{pen}} = P\left(\sum_{i=1}^{N} \bar{l}_i - LS_{\text{min}}\right)$$

(17)

where $P(x) = \begin{cases} x^2, & \text{if } x < 0 \\ 0, & \text{if } x \geq 0 \end{cases}$

(18)
where and $LS_{min}$ is the minimum load shedding amount derived in (7). $x < 0$ means the predicted total load shedding amount is lower than the minimum value to support transient stability, confirming the violation of post-control stability constraint. In this situation, the penalty factor is required to reconstruct the loss function for DNN training, which can guide the training direction and make the DNN parameters more inclined to the stable operation of the power system. Persisting such guidance throughout the iterative training process can avoid the recurrence of control failure as much as possible. It finally ends in a DNN model that is self-regulated by the post-control stability constraint.

The overall aim of DNN training is to sort out the best values of the learnable parameters (i.e. weights $\alpha$ and biases $\beta$) at each layer. By adjusting these parameters, the training process can be regarded as minimizing the average value of the loss function given the training samples:

$$\min_{\theta} \frac{1}{N_{\text{train}}} \sum_{k=1}^{N_{\text{train}}} \Gamma_k$$

(19)

where $\theta$ represents the parameters set including the weights and biases across all layers of DNN, $N_{\text{train}}$ is the number of training samples, $\Gamma_k$ is the loss function value of the $k$th training sample.

An efficient training technique, namely stochastic gradient descent (SGD) [40], is adopted in this chapter to train DNN. The mechanism of SGD is to nominate a subset of the summation function at each step, which can significantly reduce the computation burden in each iteration. The training is carried out by the back propagation model and the parameters are updated by SGD:

$$\Delta \theta(t+1) = \Delta \theta(t) + \rho \frac{\partial L_{\text{total}}}{\partial \theta}$$

(20)

where $\rho$ is the learning rate.

4.6 Numerical Results

4.6.1 Shedding amount inference modelling

The EEAC-ELS relationships manifested in the three database are plotted in Fig. 3.2 where the quasi-linear relationships are clearly interpreted. Based on the OLS result, linearized functions are fit for each power network. Each contingency -- $k_0$ for NEfault #1, NEfault #2, and mixed fault of New England system are respective 1785, 1796 and 1800; while $k_0$ for NDfault #1, NDfault #2, and mixed fault of Nordic system are respectively 1791, 1779 and 1804. The percentage errors of OLS are 2.76% for New England system and 2.16% for Nordic system.
The accuracy of the shedding amount inference model is evaluated on the ELS strategies for testing samples. These ELS strategies are generated by classic DNNs. The samples with overall ELS percentage falling below and above the fitted line in Fig. 4.3 are recognized as load under-cutting and over-cutting events, respectively. For the load under-cutting events, we obtain their true post-control stability conditions by running TDS. The accuracy of the shedding amount inference model in detecting ELS strategies that end in post-control unstable events is collected in Table IV. It is worth noting that this test is performed on the ELS solutions from well-trained DNNs. Since the DNNs have already been trained towards the risk-averse direction, their ELS outputs should be very close to the decision boundary of load under-/over-cutting. Even in this situation, the shedding amount inference model can still achieve high accuracy (between 90.18% and 93.07%). This encapsulates the model accuracy would be much higher on the arbitrary ELS strategies to be screened during the DNN training process.
Fig. 4.3. The relationship between optimal load shedding amount and EEAC margin, subject to (a) NEfault #1, (b) NEfault #2, (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault.

Table IV
ACCURACY OF SHEDDING AMOUNT INFERENCEx MODELS

<table>
<thead>
<tr>
<th></th>
<th>Fault #1</th>
<th>Fault #2</th>
<th>Mixed Fault</th>
</tr>
</thead>
<tbody>
<tr>
<td>New England System</td>
<td>91.06%</td>
<td>90.18%</td>
<td>92.43%</td>
</tr>
<tr>
<td>Nordic System</td>
<td>92.04%</td>
<td>91.98%</td>
<td>93.07%</td>
</tr>
</tbody>
</table>

4.6.2 Performance Metrics for Risk-Averse ELS

Considering the co-effect of MAPE and SR in evaluating the overall performance of risk-averse ELS, a new metrics, namely expected load loss deviation (ELLD), is proposed to provide a more objective account of evaluation. ELLD is a risk-based metrics quantifying the deviation between the expected load loss resulted from the predicted and the optimal ELS actions. The load loss incurred by control failure is also accounted. ELLD is formulated as

\[
ELLD = SR \times (\bar{LS} - LS_{opt}) + (1 - SR) \times (90\% \times \sum_{i=1}^{N} p_{Li} - LS_{opt})
\]  

(21)

where \( \bar{LS} \) and \( LS_{opt} \) respectively represents the total load shedding amount by the predicted ELS strategy and the optimal ELS strategy. \( SR \) and \( 1 - SR \) serve as the weight values in ELLD to represent the probability of post-control stable and unstable conditions, respectively.
According to (21), the load loss deviations for post-control stable or unstable scenarios are calculated as follows. If the power system becomes stable in response to the ELS action, the load loss deviation is numbered as the absolute error in total load shedding amount since no further load loss is incurred on top of the shed amount. Besides, if the ELS action fails to regain stability, we expect a consequence where 90% (i.e. the maximum load percentage to be shed) of the load on all candidate buses are lost. This is decided on our preliminary test results showing that 90% load reduction can ensure the transient stability for all the tested samples. In this situation, the load loss deviation is calculated as the difference between 90% load reduction and the optimal load shedding amount.

4.6.3 Comparative Studies

Two comparative studies are conducted to demonstrate the effectiveness of the proposed risk-averse ELS method. The former is to compare the performance of risk-averse DNN and risk-neutral DNN (i.e. classic DNN with $\omega = 0$). The prediction error distributions of the two DNNs are shown in Fig. 4.4. The MAPE, SR, and ELLD metrics are computed from their ELS results and presented in Table V and VI. In Fig. 3.3, the proposed risk-averse learning mechanism successfully shifts the prediction errors towards the positive side, meaning the predicted ELS amount tends to be higher than the optimal amount to avoid control failure. In Table II and III, compared to the risk-neutral DNN, the SR is significantly improved by the risk-averse DNN, with a negligible increase in MAPE. As a result, the ELLD is decreased by nearly half for all faults and systems. It means the proposed ELS method that aims to avoid control failure can achieve a better trade-off balance between the load shedding amount and the transient instability risk, which broadly reduces the ELS control cost. It shall be noted that the load loss calculation in ELLD is built on an optimistic situation where an immediate 90% load reduction on candidate buses can cease the unstable propagation. Nonetheless, in worse cases where cascading failure or even blackout events are underway, the load loss would be significantly higher and thus the proposed ELS method would provide more significant benefits to the operator.

<table>
<thead>
<tr>
<th>Model</th>
<th>NEfault #1</th>
<th>NEfault #2</th>
<th>Mixed Fault</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>SR</td>
<td>ELLD</td>
</tr>
<tr>
<td>Risk-Averse DNN</td>
<td>3.94%</td>
<td>93.68%</td>
<td>98.3</td>
</tr>
<tr>
<td>Risk-Neutral DNN</td>
<td>3.85%</td>
<td>84.19%</td>
<td>174.4</td>
</tr>
</tbody>
</table>
Fig. 4.4. Distribution of the prediction errors of risk-neutral and risk-averse DNN, subject to (a) NEfault #1, (b) NEfault #2, and (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault.
To compare the results for the two test systems (i.e. compare Table 5, Table 6), it can be outlined that the proposed method demonstrates more significant enhancement in SR on Nordic system than on New England system. This indicates the suitability of the proposed risk-averse deep learning method on large and complex power systems. It also verifies the practicability and effectiveness of the proposed ELS method on actual power grids.

4.6.4 Testing results on load shedding locations

Besides the overall accuracy on load shedding amount, the prediction accuracy on individual load shedding locations is also significant in implementing economical and effective ELS strategy. The mean error of load shedding amount on individual candidate buses is shown in Table 7 where uniform prediction error is shown across the different load shedding buses. The result in Table 7 indicates the uniform performance of risk-averse DNN on individual outputs. This test is applied to the mixed fault dataset to verify the performance for different fault scenarios.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fault#1</th>
<th>Fault#2</th>
<th>Mixed Fault</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>SR</td>
<td>ELLD</td>
</tr>
<tr>
<td>Risk-Averse DNN</td>
<td>5.31%</td>
<td>94.12%</td>
<td>125.7</td>
</tr>
<tr>
<td>Risk-Neutral DNN</td>
<td>4.33%</td>
<td>80.16%</td>
<td>201.5</td>
</tr>
</tbody>
</table>

To compare the results for the two test systems (i.e. compare Table 5, Table 6), it can be outlined that the proposed method demonstrates more significant enhancement in SR on Nordic system than on New England system. This indicates the suitability of the proposed risk-averse deep learning method on large and complex power systems. It also verifies the practicability and effectiveness of the proposed ELS method on actual power grids.

### Table VII MEAN SHEDDING AMOUNT ERROR ON INDIVIDUAL CANDIDATE BUSES.

<table>
<thead>
<tr>
<th>Bus 4</th>
<th>Bus 8</th>
<th>Bus 20</th>
<th>Bus 39</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.40 MW</td>
<td>25.74 MW</td>
<td>28.36 MW</td>
<td>28.74 MW</td>
</tr>
</tbody>
</table>

New England System

Nordic System
<table>
<thead>
<tr>
<th>Bus 1044</th>
<th>Bus4043</th>
<th>Bus4051</th>
<th>Bus 4072</th>
</tr>
</thead>
<tbody>
<tr>
<td>47.44 MW</td>
<td>42.39 MW</td>
<td>40.08 MW</td>
<td>48.60 MW</td>
</tr>
</tbody>
</table>

Furthermore, a new performance index, namely, bus inactivation accuracy (BIA), is adopted to measure the recognition accuracy on the inactive load shedding buses for each test ELS event.

\[
BIA = \frac{1}{N_t \times N_{0,j}} \sum_{j=1}^{N_t} \sum_{i=1}^{N_{0,j}} S_{i,j}
\]

(22)

where \( N_{0,j} \) represents the number of the inactive buses in the \( j \)th ELS event. This inactive bus refers to the buses resulting in zero load shedding amount in the true optimal ELS solution. \( S_{i,j} \) is a binary value representing the activation status of the \( i \)th candidate bus in the \( j \)th ELS event. With the aim to pick up the predicted inactive buses, \( S_{i,j} \) is respectively given 0 and 1 for active and inactive buses in the predicted ELS strategy. In this test, the buses with predicted load shedding percentage less than 1% are recognized as inactive. We have verified that the control effectiveness will not be affected without applying such small amount in ELS. For the mixed fault dataset, the BIA for New England and Nordic systems results are respectively 97.1% and 96.6%. It thus demonstrates the high precision of the proposed method in deciding load shedding locations.

4.6.5 Sensitivity analysis

In the proposed loss function for risk-averse DNN training, the weight \( \omega \) is the key parameter to balance the regular term and the penalty term. A sensitivity analysis is conducted to visualize the impact of changing this weight \( \omega \) on ELS performance, from which the final choice of \( \omega \) values can be justified. Hereby, we exponentially increase \( \omega \) from \( 10^{-4} \) to \( 10^{-1} \) in a factor of 10. The performance metrics of the risk-averse DNN model is shown in Fig. 4.5 where MAPE and SR both increase along the increase of \( \omega \). It can be elevated that the MAPE for all faults is insensitive when \( \omega \leq 0.0001 \), but experiences dramatical increase thereafter. Ironically, SR shows the opposite pattern with significant jump for \( \omega \leq 0.0001 \) and then steady increase after that. Based on such changes of MAPE and SR, the ELLD reaches its minimum at \( \omega = 0.0001 \). Therefore, 0.0001 was nominated as the optimal value for \( \omega \), which is expected to end in the minimum load loss incurred by the ELS control.
Fig. 4.5. Variations of MAPE, SR, and ELLD against the change of $\omega$, subject to (a) NEfault #1, (b) NEfault #2, (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault
Chapter 5. Risk-Averse GraphSAGE for ELS

A power system network can be mathematically laid out as an irregular graph [50]. The DNN in previous chapters only provided a straightforward fully-connected mapping between power system OCs and their corresponding control strategies. The graph nature of the power network has not been well modelled, resulting in the contained topological information to be largely lost. Because of the lack of spatial information, the way to link the nodes with deviated electrical features are unknown to the resulting model. Moreover, in actual large power networks, the topology can be very complicated. It is difficult to model the graph topology in a precisely mathematical way, which demands efficient methods to integrate power network topology into the deep learning models. Effective modelling power network topology into graph structure that suits data-driven analytics is an important but challenging task for real-time ELS.

5.1 Introduction

In some existing articles, many proposed options attempt to use rule diagrams or images to represent dynamic state variables [51], [52], [53] in power systems according to predefined rules. However, the actual effect is limited, and the topology information cannot be completely collected. If there is a variation in topology induced by maintenance or system faults, these approaches are not adaptable and completely retraining of the predictor is needed, which is inefficient in practical applications [54].

In the literature, graph neural networks (GNN) have been explored as a promising class of algorithms to integrate the network topology of power systems in machine learning. Tailored for applications with graph-structured data, GNNs are able to exploit the inherent structure of graph data by virtue of relational inductive biases in deep learning architectures, thereby providing compelling ways to learn, reason, and generalize from graph data abilities. Moreover, approaches based on graph convolution networks (GCN) have been developed by integrating graph structure into convolution operation [55], [56]. GCN applies convolution filters directly to graph nodes and their neighbors [55]. Its principle is to extract high-dimensional information related to nodes into vector form through dimensionality reduction. For instance, the K neighborhood nodes of each node are extracted and normalized to a convoluted perception field. In this process, the selection process of K neighbors per node is independent, which is not spatially variant. For a large power network with unevenly distributed information, the main challenge for graph learning is to define a local operator that can handle communities of varying sizes. In the meantime, it aims to maintain the weight-sharing nature of DNNs. Though GNNs has achieved great success in other disciplines [57], [58] and [59], there is little effective work in power system ELS. In this chapter, GNN is
first introduced as a deep learning concept for ELS. Then, the GraphSAGE algorithm is introduced. Besides, a new ELS method based on GraphSAGE is further proposed to tackle the drawbacks of existing GNNs in large systems. Then, it combines GraphSAGE with the risk-averse method in the previous chapter to improve the overall control performance.

### 5.2 Graphic Neural Networks

Graphical formats describe structural information by modeling a set of objects and their relationships. Objects and relationships are represented as nodes and edges, respectively. In solving power system problems, the buses are typically modelled as nodes while the transmission lines connecting the buses are modelled as edges. GNNs provide a way to automatically capture and encode the rich topological information in the power flow data. In comparison, the conventional machine learning algorithms could work with graph data. Yet, it works typically in a cumbersome process with hand-engineered features.

A key feature of GNNs is node embedding, which is to learn the embedding representation of nodes in a graph by combining topological structure and node attribute information. This process intends to encode nodes into vectors in lower dimension, as well as maintain the key position and relationship information in the original format. However, classic GNN algorithms, such as GCN, demands embedding of learning nodes in a definite graph. It cannot be directly generalized to nodes that have not been observed in the training process, i.e. a direct push learning. Although it can capture the global information of a graph, all nodes need to be included in the training of node embedding in a transudative learning mode, which is not an efficient way to get embedding for a new node.

To compensate for the shortcomings of GCN, GraphSAGE algorithm is used in this research for power system ELS. GraphSAGE uses neighbor sampling and feature aggregation to make a linear transformation between the node's own attributes and the sampled neighbor node's attributes. It then concatenates them and perform a linear transformation to get the feature representation of the target node. The advantages of GraphSAGE are (1) it effectively alleviates the problem of GCN memory explosion by neighbor sampling and is suitable for large-scale problems; (2) Transudative learning is transformed into inductive learning to avoid the need for retraining over each node and support incremental features; (3) By introducing neighbor sampling, the direct-push node representation will be transformed into a generalized representation of nodes corresponding to a variety of local structures. These advantages motivate the exploration of GraphSAGE-based method in this research to be well suited in large systems such as large-scale power systems and preserve graph information in neural network models.
5.3 GraphSAGE

GraphSAGE is a representative inductive learning algorithm in GNNs. It is a deep learning method based on node embedding, which mainly includes the construction of association graph, forward propagation stage and backward propagation stage.

5.3.1 Construction of Association Graph

Association graph \( G = (D, E) \), where \( D = \{d_1, d_2, \ldots, d_n\} \in \mathbb{R}^n \) represents the nodes and \( E \) represents the edges. \( A = \{a_{11}, a_{12}, \ldots, a_{ij}, \ldots, a_{nn}\} \in \mathbb{R}^{n \times n} \), namely adjacency matrix, is usually constructed to mathematically model the edges, where \( a_{ij} = 1 \) if there is an association between nodes \( d_i \) and \( d_j \), and otherwise \( a_{ij} = 0 \).

The quality of the association graph affects the performance of GraphSAGE. When GraphSAGE is applied for other typical applications, such as social network prediction, protein molecular structure prediction, and road traffic information prediction, it is easy to obtain association graphs. For the molecular structure of a protein, the diagram reflects the connections between protein molecules, which are inherent characteristics of protein molecules. In the field of traffic prediction, graphs represent the connections between each traffic intersection, which is a physical representation. In the literature, a common method is to explain the positional relationship of nodes in Euclidean space through Euclidean distance and establish connections between closer nodes.

5.3.2 Forward Propagation Stage

GraphSAGE is an inductive learning framework that can efficiently generate embeddings for unknown nodes by virtue of the attribute information of nodes. The main idea is to aggregate adjacent nodes by learning a function to generate the embedding vector of the target node.

As shown in Fig. 5.1, GraphSAGE generates the target node (red) embedding and provides predictions for downstream tasks following the procedure below:

1. Random sampling of k-hop neighbors. For each node, GraphSAGE randomly selects a certain number of neighbor nodes instead of all neighbors for aggregation, which can effectively reduce the computational complexity (number of 1-hop neighbors = 3, numbers of 2-hop neighbor sampling = 5).
(2) Generate target node embedding. Taking 2-hop neighbor aggregation as an example, this step starts by aggregating 2-hop neighbor features to generate one-hop neighbor embedding. It then aggregates one-hop neighbor embedding to generate the target node embedding. This step finally embeds the two-hop neighbor information into the target node.

(3) Predict the label of the target node by using embedding as input to the fully-connected layers.

Fig. 5.1. Visual illustration of the GraphSAGE sample and aggregation process.

The above node embedding method demands K aggregators to aggregate K times. At each aggregation, the features of each node embedding obtained from the previous layer are aggregated. It obtains the features of the current layer assuming that the nodes in the previous layer have their own attributes. The aggregation process runs for k times until the features are aggregated at the target node in the ending graphSAGE layer.

Many aggregation functions are available to implement forward aggregation, such as mean aggregator, GCN aggregator, Pooling aggregator and LSTM aggregator. In this research, mean aggregator is adopted. Mean aggregator stitches the k-1 vector of the target node with the k-1 vector of the neighbour node. It then averages each dimension of the vector, making a nonlinear transformation to produce the k-level representation of the target node. This process can be presented as the following equation:

$$
\mathbf{h}_v^k \leftarrow \sigma \left( \mathbf{W} \cdot \text{MEAN} \left( \left\{ \mathbf{h}_v^{k-1} \right\} \cup \left\{ \mathbf{h}_u^{k-1}, \forall u \in N(v) \right\} \right) \right)
$$

(23)
where $\mathbf{h}_k^v$ represents the target node vector at $k$ layer; $\mathbf{h}_{k-1}^v$ represents the target node vector at $k-1$ layer; $\mathbf{h}_{k-1}^u$ represents the neighbor nodes at $k-1$ layer. The $k-1$ layer vectors of the target node and its adjacent nodes are spliced together. The average value of each dimension of the vector is calculated, and the result is non-linearly transformed to generate the $k$-layer representation vector of the target node.

5.3.3 Backward Propagation Stage

For supervised learning, GraphSAGE is trained through a back-propagation process identical to other neural networks. The loss function for training can be chosen based on the specific task (usually mean squared error is used as the loss function for regression tasks and cross entropy is used for classification loss function for the task). This research is a regression task in supervised learning, whose loss function is designed on mean square error.

5.4 Association graph construction for power networks

Based on GraphSAGE theory, an association graph must be appropriately constructed to translate the graph information into effective inputs to the neural network. In general, the power network can be deemed as an association graph with buses as nodes and transmission lines as edges. However, in this situation, it would be adverse to define the adjacency matrix that can fully embed the operation information (including both bus voltages and power flows between buses) of the system into a single graph. Therefore, in this research, two association graphs namely bus-based graph and line-based graph are constructed as a solution to separately model the bus connections and line connections. Bus-based graph aims to embed the bus voltage information into power network topology, while line-based graph aims to embed the line current (including net current injections by generators and loads) information. These two association graphs can be presented as follows:

1. Bus-based graph: $G_B = (D_B, E_B)$, where the nodes $D_B = \{v_1, v_2, \ldots, v_{n_B}, \theta_1, \theta_2, \ldots, \theta_{n_B}\} \in \mathbb{R}^{2 \times n_B}$ contains the voltage magnitudes and angles on all buses and $E_B$ represents the edges. The adjacency matrix $A_B$ is defined as follows:

\[
a_{ij} = \begin{cases} 
1, & \text{if buses of $i$th and $j$th quantities are physically connected} \\
0, & \text{if buses of $i$th and $j$th quantities are not physically connected}
\end{cases}
\]  

(24)

2. Line-based graph: $G_L = (D_L, E_L)$, where the nodes $D_L = \{i_1, i_2, \ldots, i_{n_G+n_D+n_L}\} \in \mathbb{R}^{n_G+n_D+n_L}$ contains the current injected by all generators and loads and the current...
flowing on all transmission lines. $E_L$ represents the edges. The adjacency matrix $A_L$ is defined as follows

$$a_{ij} = \begin{cases} 1, & \text{if lines of } i\text{th and } j\text{th quantities are physically adjacent} \\ 0, & \text{if lines of } i\text{th and } j\text{th quantities are not physically adjacent} \end{cases}$$ (25)

Fig. 5.2. Graph Neural Network Applications in Power Systems.

The information of the power system includes bus voltages and power flows, as well as the fault characteristics. Each feature is pre-processed by its type in GraphSAGE. The bus voltages are processed in a graph to aggregate and embed the voltage information of the connected nodes into each node. The tidal currents are organized in another graph to model the adjacency of the transmission lines. The fault information includes the fault location and duration. As shown in Fig 5., the node voltage information, line current information and fault information are pre-processed in the GraphSAGE layer. Then the various types of information from different graphs are jointly concatenated as the input into the fully-connected layers of the neural network which delivers the optimal ELS strategy as the final output.

Moreover, by adding the penalty term to the loss function as presented in Chapter 4, the GraphSAGE training can be guided to the flavoured direction to enable the risk-averse learning capability. By doing so, the overall control performance of the resultant ELS strategy is expected to improve.
5.5 Simulation Results

GraphSAGE is the core deep learning algorithm for the method in this chapter. It is tested and validated using the databases developed in Chapter 2. PyTorch 1.7.1 is adopted to develop the GraphSAGE model on Python platform, and SGD method is adopted to train the model.

5.5.1 Hyperparameter Selection

The performance of GraphSAGE model is closely associated with the hyper-parameters for model fitting. For instance, when the number of layers is insufficient, the learning parameters fail to complete the fitting results of the model. However, when there are too many layers, the processing will take up excessive computing resources in the training process and will affect the experimental results due to overfitting. Choosing appropriate hyper-parameters, such as the number of layers in the graph structure, the number of hidden layers, and the number of neurons in each layer, is decisive for the model to make full use of its learning ability and avoid unnecessary training burden, thus avoiding waste of computing resources and improving the accuracy of the model.

In this chapter, a method of cyclic measurement is adopted to determine the applied hyper-parameters of GraphSAGE model. The model starts with an initial setting, and the optimal hyperparameter setting is determined by increasing the number of layers at each cyclic step. A partial training dataset is adopted to train the model for each cycle, while a partial testing dataset is adopted to validate the model so that the optimal model size for different power systems can be determined through iterations.

As a result, the final structure of the developed GraphSAGE is stereoscopically presented in Fig. 5.2. The GraphSAGE has two layers of graph structure collection, 3 hidden layers, with 128, 64, and 32 neurons in each layer. During training, epoch is maintained to 200 and batch size is 64. The learning rate is 0.001.

The weight of penalty factor also has a great impact on the proposed risk-averse model. As shown in Chapter 4, sensitivity analysis is conducted to visualize the influence of changing the weight $\omega$ on ELS performance. Then the optimal $\omega$ can be decided. Fig. 5.3 shows the sensitivity analysis results, where $\omega$ is multiplied by 10 from $10^{-4}$ to $10^{-1}$, and the ELS performance (i.e. MAPE, SR, and ELLD) are displayed correspondingly. The optimal $\omega$ is determined on the comprehensive performance index ELLD. It can be concluded that 0.0001 is chosen as the optimal weight value of penalty factor.
Fig. 5.3. Variations of MAPE, SR, and ELLD against the change of $\omega$, subject to (a) NEfault #1, (b) NEfault #2, (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault.

5.5.2 Model Efficiency Comparison

Compared with the DNN-based ELS method, the ELS method based on graph neural network is the same now that its main computing speed can adapt to the changes in working state.
Table VIII outlines the average calculation time of predicting ELS action for a test sample using different methods. The DNN method, the GCN method and the GraphSAGE method are compared. As shown in Table VIII, though the graph-based methods increase the computation time as compared the DNN method due to the additional process of graph information collection, yet they all complete the prediction within a millisecond, which is well sufficient for real-time implementation to adapt to the intense variations in the system.

<table>
<thead>
<tr>
<th>Method</th>
<th>New England System</th>
<th>Nordic System</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNN Method</td>
<td>0.248 ms</td>
<td>0.262 ms</td>
</tr>
<tr>
<td>GCN Method</td>
<td>0.286 ms</td>
<td>0.331 ms</td>
</tr>
<tr>
<td>Proposed GraphSAGE Method</td>
<td>0.359 ms</td>
<td>0.407 ms</td>
</tr>
</tbody>
</table>

5.5.3 ELS Performance Comparison

The performance of risk-averse GraphSAGE model is verified by respectively comparing with the risk-neutral GraphSAGE model, risk-neutral DNN model, risk-averse DNN model and classical graph neural network, on the systems of New England 39 bus and Nordic 41 bus as shown in Fig 2.4. The applied fault is the same as that in Chapter 2. The GCN model [60] collects the node voltage on each bus and the full connection layer as the state of the experimental model. In GraphSAGE model, the two-layer graph collection layer is adopted to collect the node voltage on each bus and the power flow information on the transmission line, which is transmitted to the two-layer full connection layer through combination.

The effectiveness of the proposed graph-based neural network for risk aversion is verified by a comparative study of three performance indices—MAPE, SR, and ELLD. The resulting indices of different methods are illustrated in tables IX and X. It can be noticed that, with appropriate modelling of network topology, the risk-averse-GraphSAGE model significantly improves SR and MAPE metrics compared with risk-neutral DNN, risk-averse DNN, and traditional GCN models. Therefore, the ELLD of the risk-averse-GraphSAGE model is superior to other models for all systems and contingencies. It thus implies that ELS methods designed to avoid control failures can achieve a better balance between load shedding amount and transient instability risk, thereby widely reducing overall control costs.
5.5.4 Comparison of Risk Models

The mechanism to achieve risk aversion with the proposed GraphSAGE method is demonstrated in Fig. 5.4. It shows the distribution of prediction errors generated by the two models. In that, the proposed risk-averse learning mechanism successfully shifted prediction errors to a positive side, which means that the predicted ELS amount tends to be higher than the optimal volume to avoid control failure.
Fig. 5.4. Distribution of the prediction errors of risk-neutral and risk-averse GraphSAGE, subject to (a) NEfault #1, (b) NEfault #2, and (c) New England mixed fault, (d) NDfault #1, (e) NDfault #2, and (f) Nordic mixed fault.
5.5.5 Discussion on Experimental Results

Based on Chapter 4 and 5, we can summarize the feasibility of the proposed methods in the decision-making of ELS subject to a power system contingency. Four points can be made from the experimental results:

(1) By using the deep learning techniques, both the DNN and the GraphSAGE are far faster than the traditional TDS and optimization-based method. The neural network models can make an ELS decision within tens of milliseconds, which can be considered as an instantaneous response. These data-driven approaches greatly improve the computational efficiency, showing the potential to achieve real-time ELS decision-making.

(2) By adding the penalty term to the loss function and adjusting the associated weights, the cascading impact of power system instability caused by the unfavoured direction of prediction error can be effectively alleviated without significantly impairing accuracy. The executability of risk aversion is fully proved in two models (traditional DNN model and GraphSAGE model). This risk-averse method can be adopted to deal with the problem of unreliability in some models that need directional training.

(3) The successful application of GNN in power system verifies that embedding graph structure in deep learning is important and necessary for further improving ELS performance. Power network does not conform to the complex model of Euclidean geometry. In this situation, GNNs is better than traditional DNN in fitting ability. It also shows significant improvement by using GraphSAGE in power system ELS. Its ability to distributionally retrieve node information and the combination with fully connected layers also strengthen the ability of graph information collection and embedding.

(4) The proposed method has been tested in power grid systems with different renewable energy ratios and different sizes. These tests, to some extent, can verify the scalability of the method. The method only needs the real-time operating condition of the system to decide the emergency control action.
Chapter 6. Conclusion and Future Works

This research proposes a series of deep learning methods for real-time ELS. In the proposed methods, a shedding amount inference model is built to infer the minimum load shedding amount from the EEAC margin, from which the post-control stability condition can be efficiently speculated without the need of TDS. A deep learning model is trained by a new loss function that is designed to minimize the ELS prediction error while penalize the cases with insufficient load shedding amount. It also enables the deep learning model more reluctant to control failure risks. In addition, a GraphSAGE-based ELS method is proposed to capture and embed the network information into graph features to be learnt by deep learning models, which helps reflect the true topological relationship between power system features and reduce the prediction error on optimal ELS strategies.

The proposed ELS methods are validated on two power systems with considerable renewable penetration. Based on the testing results, the proposed methods show the following advantages:

1) Integrating risk-averse learning into the deep learning model can significantly improve the success rate of ELS control. The increase of load shedding amount is negligible, which can greatly save the control cost when considering the cascading impact of control failure. This shows that the proposed risk-averse method for deep learning provides a more economical and reliable ELS solution for transient stability control and meets the practical needs of emergency control.

2) Compared to the existing shallow learning algorithms for intelligent ELS, the proposed deep learning method significantly reduce the ELS prediction error. It thus improves the control effectiveness in terms of both economy and security.

3) As an intelligent approach, the proposed method is highly efficient in solving ELS problem compared to the conventional optimization-based method, which gains real-time capability to make ELS decisions. This ability is essential in power grids with intense variations from renewable energy sources and load-side demands.

4) Graph neural networks are conducted in power systems to effectively solve this complex system that does not satisfy Euclidean geometry. GraphSAGE effectively solves the drawback of slow data retrieval in large networks. Large-scale information can be downsized into small-scale information and then stitched to increase computational efficiency and model accuracy.

The future works can be carried out in three areas:
1) It shall speculate on the post-control stability. Probabilistic earning could be integrated in the future for uncertainty evaluation and provide more robust judgement on stability constraints.

2) Some improvements can be made on the training algorithms, in order to effectively improve the deep learning efficiency and enhance the convergence of the models.

3) Developing more practical economic models would be helpful to precisely evaluate ELS costs, based on which the tradeoff between prediction error and control success rate would be more precisely modelled and controlled.
BIBLIOGRAPHY


53


[50] H. Cai, D. J. Hill, ”Short-term Voltage Stability Assessment with Spatial-Temporal Graph Convolutional Networks.”, IEEE Transactions on Power Systems, Apr. 2022,


