

基于大数据方法的能源互联网 态势感知研究

(申请清华大学工学硕士学位论文)

培 养 单 位 : 自 动 化 系

学 科 : 控 制 科 学 与 工 程

研 究 生 : 赵 兵 兵

指 导 教 师 : 曹 军 威 副 研 究 员

二〇一七年五月

Situation Awareness of Energy Internet based on Big Data Methods

Thesis Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Master of Science

in

Control Science and Engineering

by

Zhao Bingbing

Thesis Supervisor : Associate Professor Cao Junwei

May, 2017

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容。

本人保证遵守上述规定。

（保密的论文在解密后应遵守此规定）

作者签名： _____

导师签名： _____

日 期： _____

日 期： _____

摘要

能源互联网是以可再生清洁能源为主要一次能源，深度融合互联网“开放、平等、分享、互联”思想的新型电网结构。由于风能、太阳能等一系列新型能源具有非常强的随机性、波动性和不确定性，导致能源需求和供给的互相匹配十分困难，也给整个能源互联网的安全和稳定带来挑战。由于能源互联网系统的复杂性和对实时性的要求，传统基于电力系统暂态时域仿真的方法很难满足，因此本文提出利用大数据的方法解决能源互联网态势感知问题。将态势感知分解为三个环节：评估、诊断和决策，结合能源互联网中数据的特点和应用需求，分别提出基于机器学习方法的解决方案。为能源互联网系统的安全稳定评估和控制提供了理论基础和关键技术。

本文的主要研究工作包括以下几个方面：

1)在能源互联网暂态电压稳定性评估和预测方面，重点解决传统基于数据方法的忽略时间特征和空间特征、无法在线运行等缺点。通过合理设置模型特征时段和预测时刻之间的间隔，使得在获得合理预测准确率的同时，获得充足的决策和设备投切时间。同时，通过离线训练和在线预测相结合，实现了暂态电压稳定性的实时评估和系统参数的实时更新。

2)在能源互联网故障诊断方面，主要通过降维方法解决了样本数量少和样本维度高导致的“维度灾难”。利用已知的标签或类别信息引导降维，实现有监督的LLE算法。目标是在选取和创造特征的过程中，有效利用类别信息。本章通过调整LLE算法中的两个重要参数：近邻数 k 和内嵌维度 d 来使得在降维后最小化类内间距，最大化类间间距。最终通过样本的低维空间中的分布来判断故障类型。

3)在能源互联网的决策支持方面，通过数据方法来判断无功补偿装置的位置和补偿的量。通过把决策数据和系统的状态数据融合在一起作为系统的特征，同时训练系统的稳定裕度指标作为输出。通过算例的验证，取得了很好的效果。

4)在系统开发和应用方面，开发了在线电压稳定性分析系统（Real-time Voltage Stability Analysis System, RVSAS）。RVSAS通过组合多种开源大数据分析框架，实现了预测参数的快速训练和学习、在线状态流数据的实时评估与预测和网络模型的实时更新。同时开发了基于仿真软件BPA的数据格式化工具包，为未来电力大数据的研究工作提供了便利。

关键词：能源互联网；态势感知；大数据；故障诊断；无功补偿

Abstract

Energy Internet is a new-type power grid architecture, which takes renewable clean energy as the primary energy source and imitate the design philosophy of internet "Openness, Equal and Sharing". Due to the randomness and uncertainty of renewable energy, the matching between energy supply and energy demand is very difficult, and also pose challenges to safety and stability of Energy Internet. This paper tries to solve the problem of situation awareness of Energy Internet based on big data analysis. After decomposing the situation awareness into 3 steps: assessment, diagnosis and decision-making, we proposed the data-based approach or solution respectively.

The main contributions of this paper includes: 1)For transient voltage stability assessment and prediction of Energy Internet, we propose a new model which set an interval between feature period and prediction point. And in the process of feature selection, we integrate the space features and try to discover the spread rules of faults and find the key nodes in Energy Internet.

2)For fault diagnosis of Energy Internet, we propose a new method based on dimension reduction. Manifold learning is used to reduce redundant information as well as find the intrinsic structure in samples. By the selection of number of neighbours k and embedded dimension d , the label information is utilized to guide the selection of features, which makes LLE from an unsupervised algorithm to a supervised algorithm.

3)For decision-making of Energy Internet, we propose a data-based reactive compensation design method. We get samples by simulating a large number of faults and reactive compensation scheme, and train a classifier to find the best plan.

4)For the system development and application, we design Real-time Voltage Stability Analysis System(RVSAS). RVSAS realize the fast training and learning of system parameters, realtime assessment of status data stream and real-time update of system parameters. At the same time, a toolkit for data formatting in BPA is developed, which can help the future scholars the get data samples in Energy Internet quickly and easily.

Key words: Energy Internet;Situation Awareness;Big Data;Fault Diagnosis;Reactive Compensation

目 录

第1章 引言	1
1.1 研究背景	1
1.1.1 能源互联网	1
1.1.2 研究综述	3
1.1.3 典型能源互联网系统	4
1.1.4 态势感知	7
1.1.5 基于大数据的能源互联网的态势感知	9
1.2 论文研究内容与本人贡献	9
1.3 章节安排	10
第2章 能源互联网暂态电压稳定性预测	11
2.1 本章引言	11
2.2 暂态电压稳定性预测	12
2.2.1 传统暂态稳定预测方法	12
2.2.2 基于数据的暂态稳定预测和评估	13
2.2.3 存在的问题	14
2.3 电压稳定性预测模型	15
2.4 预测结果与分析	18
2.4.1 样本生成	18
2.4.2 算法流程	19
2.4.3 算法结果与分析	24
2.5 在线电压稳定性分析系统	26
2.5.1 系统架构	27
2.5.2 离线学习与参数更新: Spark	30
2.5.3 在线稳定性预测: Storm	30
2.5.4 参数传递: Kafka	31
2.6 本章小结	32
第3章 能源互联网故障诊断	33
3.1 本章引言	33
3.2 流形学习	35
3.2.1 降维	35

3.2.2	流形学习	36
3.2.3	LLE	37
3.2.4	ISOMAP	38
3.3	基于有监督LLE算法的能源互联网故障诊断	38
3.4	参数选择：近邻数k和内嵌维度d	43
3.5	仿真样本	46
3.6	故障诊断结果与分析	48
3.7	本章小结	51
第4章	能源互联网无功补偿	54
4.1	本章引言	54
4.2	无功补偿	54
4.3	基于大数据无功补偿优化设计	56
4.3.1	从AlphaGo获得的启示	56
4.3.2	模型设计	57
4.4	样本生成和训练	58
4.5	算例分析	61
4.6	面向BPA的数据处理工具包	67
4.7	本章小结	69
第5章	总结与展望	71
5.1	论文工作总结	71
5.2	未来研究工作展望	72
	参考文献	73
	致 谢	78
	声 明	79
	个人简历、在学期间发表的学术论文与研究成果	80

主要符号对照表

SCADA	数据采集与监视控制系统
PMU	同步相量测量装置
WAMS	广域监测系统
SVG	静止无功发生器
PCA	主成分分析方法
LLE	局部线性嵌入
Storm	流式计算处理框架
Kafka	分布式发布订阅消息系统
Spark	开源集群计算环境
HDFS	分布式文件系统
RVSAS	在线电压稳定性分析系统

第1章 引言

1.1 研究背景

1.1.1 能源互联网

随着经济与工业的快速发展，对能源的需求量快速增长。传统以石油、煤炭等化石资源的大量开采使用，带来了环境污染、资源稀缺等一系列问题，因此，开发多样化的清洁可再生能源是解决这一系列问题的唯一出路。所谓能源互联网，是在现有电力系统架构的基础上，以可再生清洁能源为主要一次能源，充分利用互联网设计思想和大数据技术的分布式能源平台。能源互联网充分利用信息技术和电力电子技术，将大量分布式能量采集装置和能量储存装置互联起来，实现能量和信息的双向流动和对等交换。最终保证区域能源可靠供应，实现区域能源协调供给。

能源互联网与之前的传统电网之间最大的不同在于借鉴了互联网的思维和利用了互联网的技术。互联网极大丰富了人类获取和利用信息的渠道，悄然改变了人们沟通和交流的方式。同时，它也逐步演化成为一种思维模式，颠覆了诸多传统产业的生产经营模式。

能源互联网从互联网中借鉴的思想主要是开放、对等、分享、互联。传统电网的设计理念是中心式的，通过不断的提高网络覆盖、电压等级、机组容量等来获得发展，也加强了自然垄断。互联网通过摆脱了电话网络对中心交换的依赖，分布式的架构使其获得了自我服务、自我维护、自我更新的能力，双向的信息流动使得每个人既是内容的制造者，也是内容的浏览者。能源互联网借鉴信息互联网自底向上的理念，增加分布式可再生能源的灵活接入和就地消纳。微网等能量自治单元相当于局域网，主干网相当于广域网，能源路由器负责能量的转发，最终实现能源双向按需传输和动态平衡使用。能源互联网可以打破传统电网的自然垄断，衍生出诸多商业模式，形成独特的生态系统，具有强大的生命力。图 1.2^[1]是能源互联网的架构图。

在利用互联网的技术方面，主要是岁能源互联网中的数据利用大大增加，利用数据辅助网络态势的评估、诊断和决策。虽然从传统电网到智能电网的发展过程中，信息技术得到了广泛的应用，但其中更多的是简单的监测和反馈，对数据的深度挖掘还不够。随着机器学习和云计算的发展，数据传输、存储和计算的成成本已经非常低廉。能源互联网综合利用电网状态数据、新能源发电数据、用户需

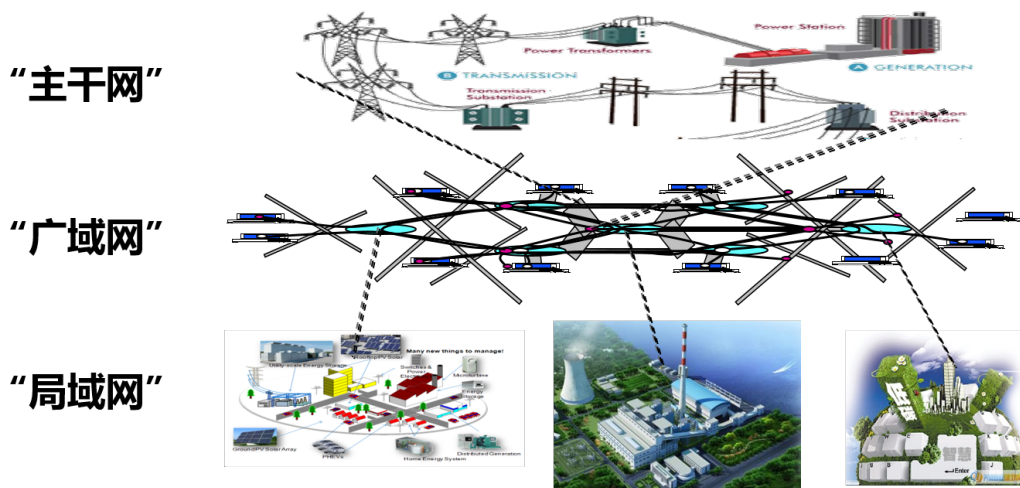


图 1.1 能源互联网架构图

求数据、电力市场数据、电力企业管理数据，实现能源需求和供给的匹配，同时保证整个网络的安全和稳定，最终打通并优化能源生产和能源消费端的运作效率。图 1.2是能源互联网的应用场景。

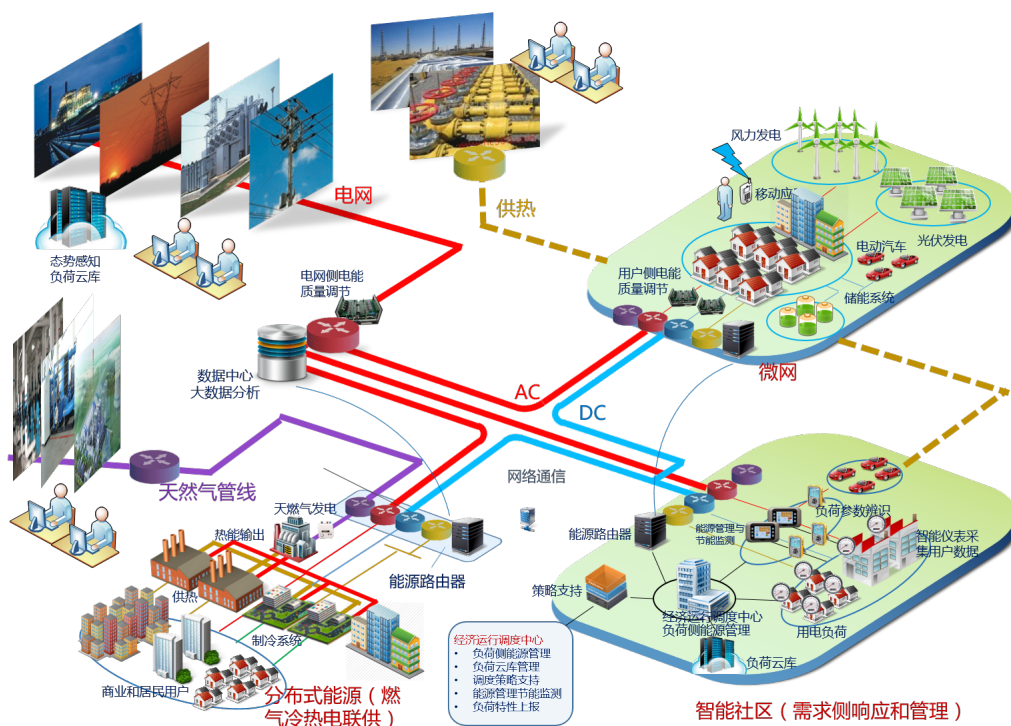


图 1.2 能源互联网应用场景

与互联网相比，能源互联网面对的难点和挑战主要有：

1) 全世界可利用的风能、太阳能等可再生能源资源具有非常强的随机性、间断性和波动性。目前将它们成功的并入电网，并将其与能源需求匹配并且高效率

用，还是非常困难的。

2) 能源无法像信息一样，低成本地进行存储，因此需要能源需求和能源供给时刻相互匹配。同时，也要尽可能地开发新的储能技术，来实现一定的缓冲和调节。

3) 电力网络是最复杂的人工非线性系统，当大规模可再生能源接入后，会变得更加地脆弱和不稳定。如何高效利用监测数据，实现更精准的感知、诊断和控制，增强其鲁棒性，最终保证能源互联网的安全、稳定。

1.1.2 研究综述

“共享经济之父”、著名畅销书作家在其著作《第三次工业革命》^[2]使得能源互联网收到大家的广泛关注。他提到未来的能源系统主要有这样几个特点：首先是以可再生新型清洁能源为主要一次能源；第二是广泛支持分布式发电系统和分布式储能系统的接入；第三是通过借鉴互联网中的信息技术实现广域的能源共享和互联互通。目前学者的研究主要集中在能源互联网的架构和设计、能源路由器和路由算法、大规模储能、能源互联网中的大数据技术、能源市场和电力交易等几个方面。

在能源互联网的架构方面，文献^[3]提出了以能源路由器作为信息与能源技术设施一体化的解决方案，并对基于能源路由器的能源互联网应用场景给出了详细描述。文献^[4]通过借鉴信息互联网中软件定义的思想，在能源信息基础设施一体化的背景下，提出软件定义的能源互联网信息通讯的具体实施方法。

能源路由器和路由算法方面，文献^[5,6]中提出的“能源交换机”的概念可以说是能源路由器的雏形。能源交换机是多种类型能源的转换、汇集、存储的中信单元，并提出了基于能源交换机的新型电网架构。文献^[7]提出了能源路由器所必须的功能，包括能量流的动态调整和能量设备间的实时通信。同时展现了能源路由器的基本架构和初步的仿真实验。文献^[8]提出采用多绕组结构的多端口能源路由器。通过改变绕组匝数实现大范围电压等级的结构。与传统母线结构的能源路由器相对比，能量在不同端口之间传输所经过的变换器相对较少，系统效率更高。

储能技术方面，文献^[9]回顾了过去的在能量存储方面的工作，系统总结了各个阶段储能材料、储能技术和应用场景的不同；文献^[10]针对全球能源互联网的特点和发电、输电、配电等各个环节的不同需求，提出了储能配置方案和配置原则；文献^[11]提出了面向调度技术和协调规划的储能设计，把储能和能量分发结合，设计能源路径和分发策略，同时考虑了能源交易机制在储能中的应用。

“数据驱动”是能源互联网最重要的标签，大数据分析将是能源互联网的支

撑性技术。在数据存储和管理、云计算、深度学习等技术飞速发展的今天，数据的获取、存储和分析的成本已经非常低廉，以数据为支撑的应用已经渗透到每一个行业。“数据驱动”的本质是利用基于数据的反馈，提供更优质的服务、更有针对性的营销和更有效率的管理，形成精细化的闭环的商业模式。能源互联网大数据分析的研究在智能电网时代就已经开始了，文献^[12-14]总结了在智能电网时代电网大数据的特点，从算法、平台、应用等多个方面分析电网大数据技术的现状和挑战。但传统电网大数据技术更多的是简单信息监测和反馈，对数据的深度挖掘还不够。文献^[1]介绍了能源互联网与电力大数据之间的相互关系以及能源互联网大数据分析的典型应用。文献^[15]把大数据服务定位于分布式电源用户，通过“虚拟化+大数据+云计算”的技术架构构建电源云服务与大数据分析平台。

能源市场和电力交易方面，文献^[16]总结了能源互联网背景下衍生的各种新的商业模式，并对其特征和应用场景进行系统分类，再次基础上进一步提出建立能量耦合、价格耦合、衍生交易“三位一体”的能源互联网市场架构新模式；文献^[17]提出去中心化的电力交易管理方法，通过区块链这种分布式的存储技术，数据保存由所有节点共同维护，有效解决了电力交易过程中的信任问题，降低运行成本，保护用户隐私；文献^[18]通过研究丹麦两类重要的能源互联网技术理念，即基于服务的灵活性交易，以及跨行业能源系统的集成设计、规划和运行，提出了未来能源互联网的交易和商业模式。

1.1.3 典型能源互联网系统

目前对能源互联网的理解主要有以下三种：

1) 强调通过数据实现对电网的优化管理，以C3 Energy、Opower等公司为代表。将信息网络架构的传统电网结构之上，通过数据采集、数据挖掘、决策支持等几个环节，利用数据指导能源互联网的调度和运行。

2) 强调可再生能源发电和分布式储能装置基础上的新型电网结构，以美国FREEDM系统为代表。通过效仿互联网中的路由器，通过电力电子技术和储能技术来实现能源路由器。通过能源路由器规划、转发和控制，实现能量和信息的双向流动。

3) 强调国与国之间的跨区域电网的互联，以国家电网为代表。特点是建设以特高压电网为主干骨架，实现清洁能源在全球范围内的共享。

下面会选取几个典型的能源互联网系统进行介绍。

1.1.3.1 C3 Energy

C3 Energy(www.c3energy.com)是一家通过大数据、社交网络、机器学习、和云计算来提高电网的安全性、可靠性、可持续性和功率传输效率的软件公司。它通过整合来自电网公司等公用事业部门和其他第三方的超过22种数据,包括公用事业公司拥有的能耗数据、仪表数据,第三方或用户的企业运营情况、建筑物特性、地理信息数据等,形成自己的数据分析引擎(C3 Energy Analytics Engine),提供电网实时监测和即时数据分析。

C3 Energy有两大类软件——C3电网分析(C3 Energy Grid Analytics)和C3用户分析(C3 Energy Customer Analytics),以及超过20种应用。

C3 电网分析(C3 Energy Grid Analytics)逐步形成了9种成熟的应用或解决方案:收入保障(C3 Revenue Protection),断电分析、预测及恢复(C3 Outage Analysis, Prediction & Restoration),智能仪表控制(C3 AMI Operations),智能电网可靠性及安全(C3 Customer Reliability & Safe),财产及系统风险(C3 Asset & System Risk),电压优化(Voltage Optimization),需求响应分析(C3 Demand Response Analytics),变电站自动优化(C3 Substation Automation)和伏安优化(Volt/VAR Optimization)。图 1.3 是C3 Energy电网分析的界面。

C3 用户分析(C3 Energy Customer Analytics)是双向的,一方面针对电网公司的公共事业公司,辅助其设计合理的需求响应方案;另一方面面向客户,帮助其解决用电,进行能耗管理。

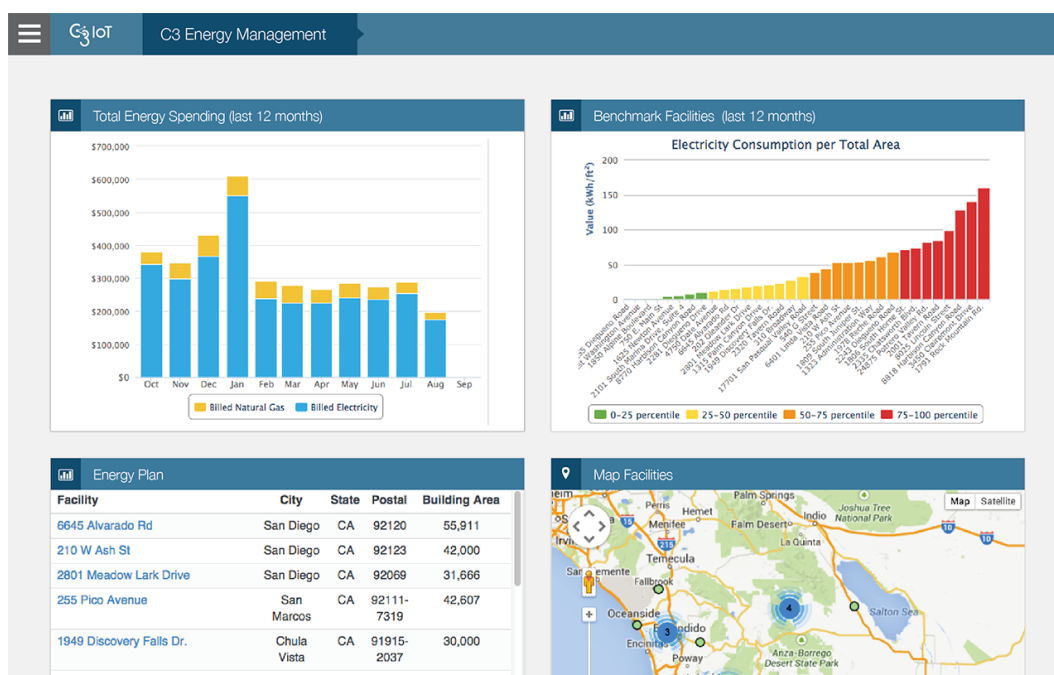


图 1.3 C3 Energy的电网分析界面

1.1.3.2 FREEDM

未来可再生电能传输及能量管理系统（Future Renewable Electric Energy Delivery and Management System, FREEDM）^[19]是由美国North Carolina State University（NCSU）提出的新型智能微型电网模型。它从2008年起有美国国家科学基金会（National Science Foundation, NSF）资助开展。FREEDM主要有几个特点：

- 非常好的可扩展性：任何电能或者电源设备都可以即插即用
- 信息网络与能源网络一体化：拥有安全和升级的通讯中心，通过信息网络管理能源网络
- 可脱离主电网独立运行：电能质量高，功率因数高，鲁棒性强

在传统的电网结构中，电能的流动是单向的：只能从电网公司流向用户；而在FREEDM系统中，能量的流动是双向的，每个用户基数能源的消费者，也可以是能源的攻击者。用户自己的光伏的新能源设备可以将无法消纳的多余能源上传电网，从而获得收益。图 1.4是FREEDM系统的架构和关键部件。

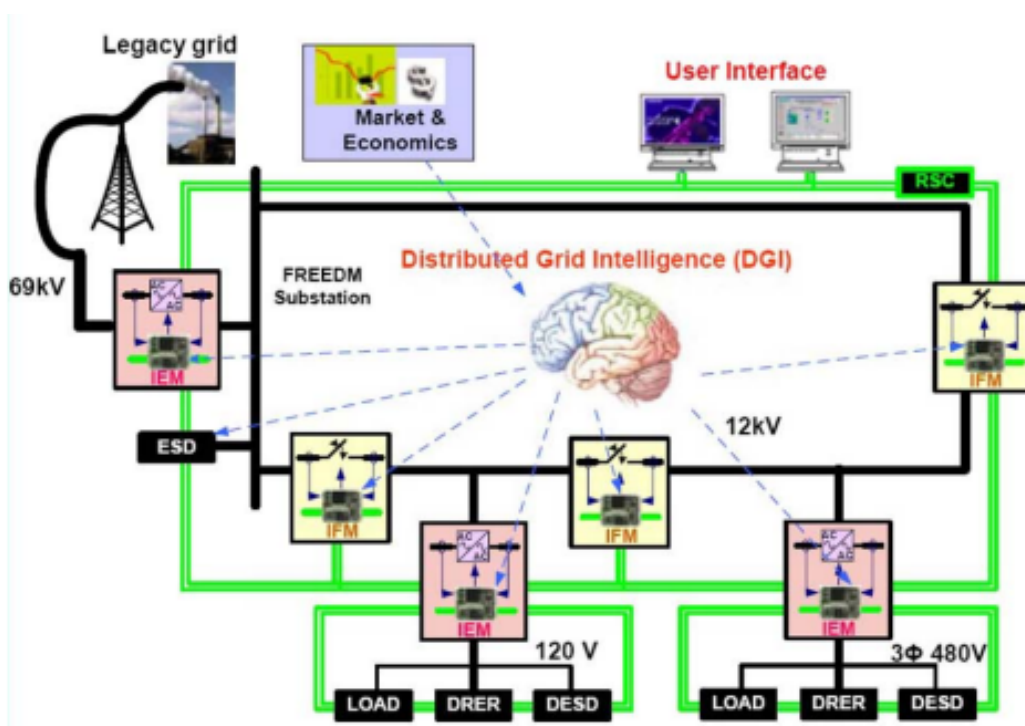


图 1.4 FREEDM系统的关键部件

1.1.3.3 能源互联网在中国的实践

能源互联网在中国引起了广泛的关注，国家电网、清华大学、国防科技大学、天津大学纷纷开始从事能源互联网的研究。国家电网公司下属的智能电网研究院

更名为全球能源互联网研究院，研究建设以特高压电网为主干骨架，实现以风能、太阳能等一系列清洁能源在全球范围内的共享。如何通过图 1.5 是国家电网提出的全球全球能源互联网架构图^[20]。

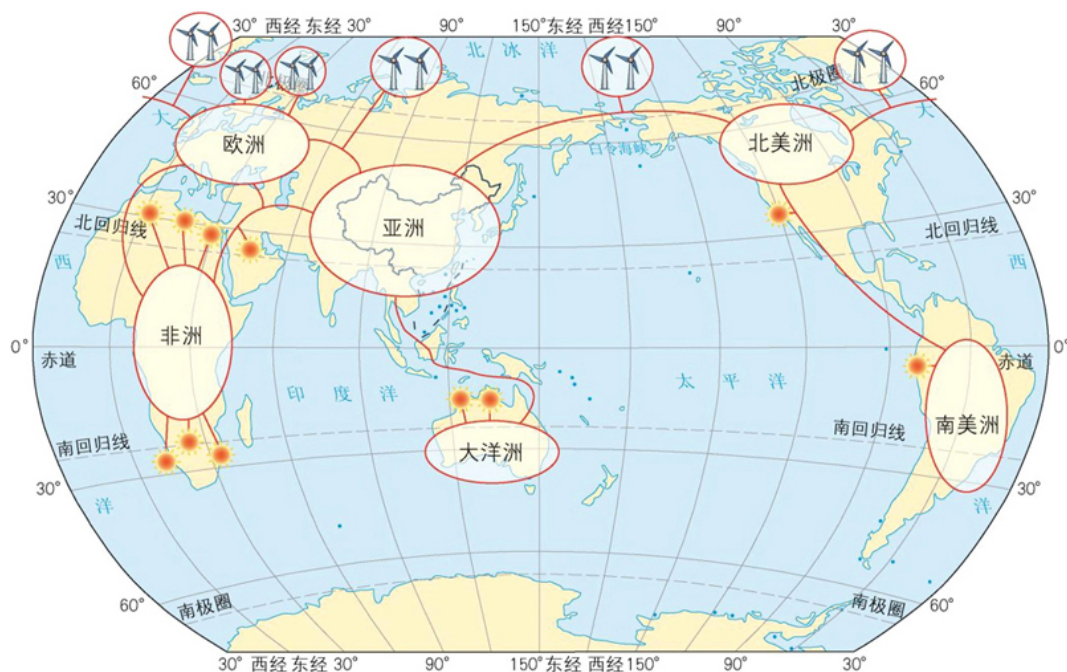


图 1.5 全球能源互联网架构图

清华大学成立了专门的能源互联网研究机构：智中能源互联网研究院。研究院下设多能流系统混合仿真中心、大数据与信息技术中心、柔性交直流技术研究所、储能与控制技术研究所等10个研究所和创新中心。开展了海淀北区、合肥、嘉兴等多个能源互联网示范项目。同时，清华还承办了首届IEEE能源互联网大会（IEEE ICEI 2016），收到投稿238篇，录用65篇，还举办了“能源互联网架构与设计”，“能源互联网中的能量转换与路由”，“能源互联网的信息和通信”，“能源互联网的能源管理系统”，“能源市场和交易”等多个workshop，促进了学术交流和发

1.1.4 态势感知

态势感知^[21-23]是一种典型的基于数据的应用。因为它的面向对象主要是大规模复杂系统，这些系统的各个部分之间物理关系十分复杂，其中有一些规律现在仍然没有发现和掌握，因此很难进行精准的建模仿真分析。基于数据的方法把复杂系统视为黑箱，直接对输入输出进行拟合。随着人工智能和深度学习的发展，计算机对高维非线性关系的理解能力越来越强，为态势感知的发展做出了铺垫。

在能源互联网领域，态势感知基于广域动态安全监测系统、多种数据库和数据挖掘、动态参数辨识、超实时仿真、可视化等技术的集成，综合解决电力系统安全分析经典难题^[24]，实现电力系统运行关键动态数据的在线测量、数据处理、分析，达到对主要动态行为的测量、辨识、显示、预警、控制。下面围绕能源互联网态势感知的四个层次：获取、理解、预测、展示，分别阐述数据在其中的关键作用。

(1) 获取

即为数据的采集和整理。能源互联网中的数据主要包括电网运行监控数据：如电流、电压、有功、无功、谐波等；电力市场数据：如交易电价、售电量、用电客户等；电力企业管理数据：如ERP、一体化平台、协同办公等。随着广域动态相量测量系统(WAMS)的大量安装和电力市场的逐步放开，能源互联网大数据将呈现出典型的“体量大 (Volume)、类型多 (Variety)、速度快 (Velocity)、价值密度低 (Value)”的大数据“4V”特征。这四个特征也对平台和算法提出了相应的要求：“体量大”要求平台有更大的吞吐量 (表1)；“类型多”要求更好的数据集成能力；“速度快”要有更高的计算速度和实时性；“价值密度低”要求进行数据清理、降维等预处理。

(2) 理解

即为网络的评估和诊断。正常运行时，通过第一步获取的数据评估了解电力系统组件状态；及时发现电力系统中隐藏的问题。

(3) 预测

已知某时刻电网的状态，预测 $T+1$, $T+2, \dots$, $T+N$ 等未来时刻电网的状态。预测是态势感知中至为重要的一步，如果知道网络未来的趋势，就可以采取相应的措施，防止进一步恶化和大规模事故。

(4) 展示

展示即网络态势的可视化。由于人眼对图形的敏感度远远大于对数据的敏感度，因此用图形展示出的信息更加直观、易于理解、更容易反映出整个网络的全局信息。

这四个层次的具体应用如表 1.1所示：

表 1.1 态势感知各层的具体应用

获取层	理解层	预测层	展示层
数据清洗	设备状态评估	网络稳定性预测	可视化
数据降维	电能质量评估	微网负荷预测	
数据集成	系统稳定性评估	新能源发电预测	

1.1.5 基于大数据的能源互联网的态势感知

由于发电侧和用电侧不确定性的提高，能源互联网相对于传统电网，面临的困难就是如何将用电需求与发电供给相互匹配以及如何保持整个网络的安全和稳定。我们提出利用大数据的方法进行能源互联网的态势感知研究，主要基于以下3点：(1)相对于过去的研究，我们目前新增的研究资源就是大量来源于电网的新安装的传感器数据，基于大数据的方法可以充分发挥这些数据的价值。(2)能源互联网是非常复杂的非线性系统，传统的建模仿真方法对进行很多假设和简化，无法反应整个网络的真实情况。基于数据的方法直接对系统的输入输出进行拟合，对非线性复杂系统有非常好的模拟能力。(3)随着云计算技术和各种开源分布式计算平台的发展，进行大规模机器学习的成本大大下降。

我们研究的能源互联网的态势感知，主要集中在三个环节：评估、诊断和决策，三个环节之间的逻辑关系如图 1.6。评估主要是预测系统故障后的暂态稳定性，是态势感知的第一个环节，对接下来故障的隔离和采取有效措施具有重要意义。第二个环节是诊断。当暂态电压的评估和和预测结果已经为不稳定时，并且通过故障诊断定位到故障位置。第三个环节是决策，当找到故障位置后，接下来就要采取紧急控制措施，包括切负荷、切设备、补偿无功功率等，供调度员选择。

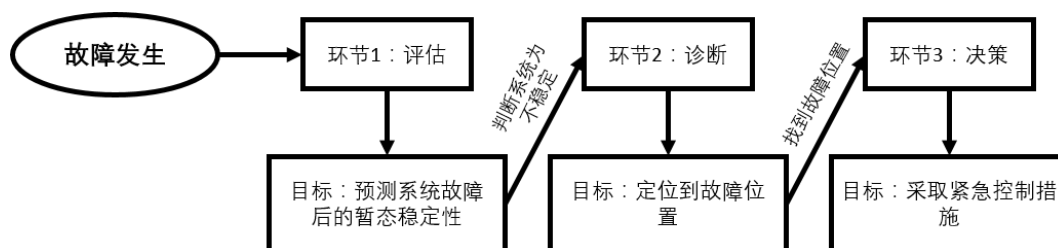


图 1.6 能源互联网态势感知的三个环节逻辑关系

1.2 论文研究内容与本人贡献

本研究基于能源互联网的数据特点，面向能源互联网中的高级应用——态势感知，提出有针对性的大数据算法和平台，最终构建一个高效、安全、稳定的能源互联网架构。第2章、第3章、和第4章分别对应能源互联网的评估、诊断和决策支持。本文的贡献主要有以下几点：

(1)提出了基于时延的电压稳定性预测模型。传统的电压稳定性分析模型主要有分类结果可解释性差、特征选取和特征表达困难、无法在线运行等缺点。因此我们的模型通过引入网络特征、延时预测和移动窗口，把特征时段和预测时刻分

开，得到预测结果后，可以有充分的时间进行无功补偿等措施。开发了在线电压稳定性分析系统（RVSAS）。RVSAS通过HDFS进行数据存储，Spark进行数据训练，Storm进行在线预测，Kafka负责参数传递和模型更新。实现了在线预测、快速迭代的电压稳定性分析平台。

(2)提出了有监督的LLE算法，并将其运用于能源互联网故障诊断。流形学习中的LLE算法是一种典型的无监督学习方法，本文把线性判别中的类内间距和类间间距引入，实现了有监督的LLE算法。利用LLE算法降维后数据在空间中的分布，进行能源互联网故障诊断，并取得了很好的效果。

(3)提出了基于数据的能源互联网无功补偿方案。传统的电网决策往往是依据运维人员的经验。本文通过对大规模能源互联网的多组故障方案和无功补偿方案进行仿真和学习，大大提高了无功补偿的效果。同时，开发了电力仿真软件数据格式化工具包，大大方便了数据挖掘工作的开展。

1.3 章节安排

第2章是关于能源互联网的暂态电压稳定性分析，是态势感知的第一步。首先综述了以往暂态电压稳定性的研究，总结了其缺点和不足，然后提出基于时延模型的电压稳定性预测模型，并在CEPRI-36bus系统上进行样本仿真和预测。最后根据提出的模型，利用开源分布式计算工具，搭建了在线电压稳定性分析系统（RVSAS）。

第3章是能源互联网的故障诊断，是态势感知的一个重要环节。通过对流形学习中的LLE算法进行改进，提出了有监督的LLE算法。利用降维后样本在空间中的分布进行故障诊断。本章的重点是如何把已知的故障类型（标签）引入到LLE中，以及如何通过确定算法中的两个重要参数近邻数 k 和内嵌维数 d 。

第4章是能源互联网的决策支持，是态势感知的最终目的。通过对大规模能源互联网的添加故障后和进行无功补偿后的仿真，利用机器学习方法找出最优的无功补偿方案。同时根据目前主流的电力仿真软件，开发了将其数据进行格式化的工具包，方便后人利用电网仿真软件的数据进行机器学习。

第5章是对全文工作的总结以及对未来工作的展望。

第2章 能源互联网暂态电压稳定性预测

2.1 本章引言

自从上世纪70年代，全世界各地的一系列重大的停电事故带来了巨大的经济损失。导致这些事故的主要原因是调度人员无法准确评估和预测故障后的系统暂态稳定性，使得没有采取有效的措施进行故障隔离，避免事故扩大。因此，电压稳定性评估和预测成为电力系统运行中的关键问题。随着能源互联网的到来，网络复杂程度越来越高和负荷接近极限，电压稳定性的评估和预测变得更加困难。

传统暂态稳定性评估都是基于时域仿真^[25]或者能量函数方法^[26]。时域仿真模型通过建立和求解一系列高阶微分方程，来还原系统的真实运行情况。时域仿真方法主要有三个问题：首先是当系统规模和复杂程度增大时，计算复杂度和求解时间会迅速上升，在真实系统中难以应用。虽然可以通过分布式计算等方法加快仿真速度，一个大规模电网的仿真时间仍要超过10s以上。其次，模型中的关键参数对最终结果的影响非常大，但是在实际建模和仿真过程中，我们经常会对模型中的关键参数做假设和简化，因此时域仿真中的模型无法准确还原真实系统；第三，时域仿真都是通过设置一些典型故障完成的，这些故障很难覆盖真实运行过程中的实际情况，使得代表性和准确性受到影响。

另外、根据应用场景，电压稳定性评估主要可以分为离线评估、在线评估和实时评估三个阶段，如表 2.1。离线评估是在日前或者更长的时间尺度下，通过典型故障的离线仿真来检验电力系统在扰动下能否维持稳定。在线评估指在SCADA和状态估计基础上得到的“在线潮流”，目标是搜索安全隐患、为预防和控制提供早期依据。但在线评估全部计算仍然基于离线模型、参数进行，因此大误差问题仍未解决。为了系统的绝对可靠性，只能靠提高保守度，加大稳定裕度，造成浪费。另外，对于SCADA系统所使用的RTU设备，其数据采集缺乏统一时标；断面数据同步性差（2-3秒）；估计周期较长（为分钟级）；迭代计算时可能不收敛或收敛到局部极值点；难以高效处理多重错误同时出现的情况。这在本质上导致了传统安全稳定分析的低效性。

根据上面的分析，本章主要提出的是基于全PMU的状态测量数据的电压稳定性评估和预测。PMU数据的优点主要有带有时标，同步性好；数据更新周期短（子站可到10ms）；可获得电压电流，相量数据等。但仍然有一定的不足：如存在误差和不良数据；数据精度不够，难以直接应用等，需要进一步改进。另外，实时评估目的是达到网络状态的实时感知和预警，真正做到防患于未然。

表 2.1 电压稳定性评估的三个阶段

	离线评估	在线评估	实时评估
评估时刻	日前或更早	超短期	故障后短暂时段
评估目的	运行方式校核	发现安全稳定隐患、为预防控制提供参考	判断故障发生后系统稳定性，为系统紧急控制提供参考
评估方法	典型故障扫描仿真	典型故障扫描仿真	根据系统故障后的测量数据

本章通过机器学习的方法，充分利用基于PMU的WAMS系统产生的电网状态数据，来进行能源互联网暂态电压稳定性的分析和预测。本章的具体安排如下：2.2节主要是目前的基于大数据方法的电压稳定性预测的综述，同时对这些综述进行总结，提出其中存在的问题；2.3节提出了我们的电压稳定性预测模型，主要有引入网络特征、延时预测、移动窗口三个特点；2.4节是具体的算法实施流程，包括样本生成、类别标记、特征压缩等几个部分；2.5节是基于上面的方法，结合使用多种开源大数据分析框架，开发了在线电压稳定性分析系统。

2.2 暂态电压稳定性预测

2.2.1 传统暂态稳定预测方法

传统的暂态稳定性评估方法主要有四种，分别是基于电力系统暂态仿真、直接法（能量函数法）、仿真与直接法的混合方法、通过判断系统的故障响应。

电力系统暂态仿真的核心思想是求解大规模微分方程组，计算速度慢、时间长。因此后续的研究主要集中在通过串行和并行^[27]两个角度改进计算速度。但是计算速度还远远达不到实时评估和预测的要求。

直接法（能量函数法）通过系统的能量函数得到能量值，然后将能量值和系统的临界能量值进行对比，得到系统的稳定裕度。相比于暂态仿真方法，能量函数法的计算速度大大提高，基本可以满足实时暂态评估的预测的要求。但其缺点是在复杂系统中能量函数非常难以构造，因此在现实情况中应用是非常困难的。

暂态仿真和直接法的混合方法最早由文献^[28]提出。该方法首先通过时域仿真得到故障后系统的早期轨迹，然后再应用能量函数来计算系统的稳定裕度。该方法的可以避开计算时的不稳定平衡点，增加收敛速度。因此可以在提供稳定裕度的同时，也有不错的计算速度。

基于系统故障响应的暂态稳定评估主要是利用系统故障后的轨迹，同时结合

一系列判据来进行暂态稳定评估。该方法的数据主要来自广域测量系统（WAMS）中的向量测量装置（PMU），得到大量状态量的轨迹。然后通过发电机动能-功角曲线、功率-功角特征曲线、角速度-功角曲线等一系列判据来进行判断。该方法的主要优点是通过较少的测量信息给出准确的结果，因此速度快。但是仅仅在小系统的测试上有非常好的效果。当系统非常复杂时，轨迹中的特征变得不明确，判据也变得非常复杂。

2.2.2 基于数据的暂态稳定预测和评估

2010年以前，已经有基于数据的电压稳定性分析的研究^[29-32]，其中，选择的方法主要包括SVM， Decision Tree等。2010年后，随着机器学习理论的发展，更多新的方法被应用于电压稳定性的分析，包括前向人工神经网络^[33]，多层感知器神经网络^[34]，多层回归^[35]，增强学习^[36]等。表 2.2.2从应用背景、机器学习算法、测试系统、特点等五个方面对文献进行系统的梳理和总结。

参考文献	应用问题	机器学习方法	测试系统	特点
[29]	暂态稳定性	SVM, MLP	New England 39-bus system	在线决策
[30]	估计长期电压稳定裕度	ANN	New England 39-bus	用于决策最好的PMU 安装地点
[31]	电压安全评估	Decision tree	Eastern AEP system	定期自动更新决策树
[32]	故障后的暂态稳定性预测	SVM	New England 39-bus system and Venezuelan power network	提出利用电压等级，发电机速率，转子角作为特征
[33]	在线电压稳定性评估	Feed-forward ANN	IEEE 30-bus system	前向反馈神经网络；利用发电机和负荷节点的电压、有功、无功作为特征
[34]	在线电压稳定性裕度估计	Multi-linear Regression Models	IEEE 30-bus system	两阶段，第一阶段利用无功功率和电压稳定性裕度的关系建立多个回归模型，第二阶段训练利用哪个回归模型
接下页				

接上页				
参考文献	应用问题	机器学习方法	测试系统	特点
[35]	在线电压稳定性检测	Multilayer Perceptron (MLP) neural network	IEEE 30-bus system	利用电压和无功生成新特征
[36]	稳定性估计	主动学习, ANN	IEEE 3-machine 9-bus system and IEEE 10-machine 39-bus system	通过概率分布建模, 减少训练所需要的样本数量
[37]	电力系统调度和电压稳定性	Function Optimisation by Reinforcement Learning	IEEE 30-bus system and IEEE 57-bus system	解决电力系统中数据维数高的问题
[38]	电压稳定性裕度评估	Extreme Learning Machine	IEEE 118-bus system	相对于传统ANN, 速度更快, 不需要参数调整
[39]	电压稳定性增强	Multilayer ANN Model with Error Back Propagation Learning	A 52 buses system	针对分布式电网

2.2.3 存在的问题

通过上面的分析, 可以发现在过去的研究中, 主要有两类样本不均衡、分类结果解释性差、忽略时间特征和空间特征、无法分布式运行等4类问题, 下面分别具体阐述这几类问题。

1) 正负样本不均衡

在两类分类问题中, 机器学习方法要求有正样本和负样本两类训练样本。然而, 在电网中, 绝大部分时刻都是稳定的, 只有在极少数时刻会出现波动甚至崩溃, 因此两类样本的比例很不平衡。同时, 异常样本的获取成本很高, 使得没有

充足的正样本来支持分类边界。有偏的样本比例使得训练出的分类器更加偏向所占比例更大的样本类。举个例子，比如样本中正样本占99%，负样本占比1%。那么如果分类器即使不做任何训练和学习，直接预测为正样本，也可以达到99%的正确率，真正训练一个分类器反而达不到这样好的效果。由于电网稳定和不稳定两类样本的不平衡，使得传统暂态稳定性预测的训练样本都来自于仿真软件。如果想使用电网真实样本，需要设计出基于单类样本的分类器^[40-42]。

2) 分类结果不可解释

大部分机器学习的缺点是模型的可解释性降低了。很多特征在经过复杂的变换和映射之后，已经不具有物理意义。如果模型的可解释性较强，我们就可以知道电网中的关键节点、薄弱节点、波动是如何在网络中传播的，最后去指导电网的规划和建设。另外，我们选择基于数据的方法后，就放弃了对问题的现有经验和理解，转向直接对输入输出的关系进行拟合。因此，我们希望在特征选取步骤中可以重新引入已有知识和经验，“设计”出更能反映问题本质的好特征。

3) 忽略时间特征和空间特征

传统的机器学习方法都是针对于向量数据集，每个样本没有关联，完全独立。而在能源互联网的数据集中，不同节点在在时间和空间上具有强关联性，因此算法设计关键是引入时间维度和空间维度特征。时间维度即时间序列分析，要考虑网络以前的状态对现在和未来的影响。空间维度即网络信息，这样才能发现一些故障是如何在网络中传播的和网络中的关键节点，从而进一步指导电网未来的规划。要反映更长时间段多变量的绝对值和变化趋势（一次微分是相邻两点之间的趋势），需分析的数据由一个时间断面变成多个时间序列的组合。因此如何科学的选择特征和表达特征非常关键。

4) 无法在线运行

随着数据的规模越来越大，算法复杂度越来越高，计算的效率也是衡量算法的重要准则。可分布式在线运行的算法更容易充分利用大数据平台的优势，实现高效的分析和处理。

2.3 电压稳定性预测模型

通过上一小节中的文献综述和总结可以看出，过去对暂态稳定性的预测主要是基于发电机的功角，使用的数据类型和特征选取比较单一。另外，每一个节点的学习样本只来自于这个节点的历史数据，忽略了空间特征。在本章中，使用PMU数据，同时，训练每一个节点分类器的时候，将特征选取的时段和预测时间节点拉开，形成一种延时的预测方法。模型设计思想主要有三个特点：

1) 引入网络特征

在过去预测电压稳定性的工作中，特征仅仅是来自于时域。也就是说，针对一个节点，训练数据仅仅来自与这个节点过去的行为。然而，几乎所有的不稳定事件都是由于电网中某一处的故障或者负荷的波动引起的。这些故障或者负荷首先引起局部的电压跌落，之后传播到整个网络。因此，其他节点的历史数据含有大量目标节点未来行为预测的有效信息。在我们的模型中，针对每一个特定节点，整个网络的历史数据都作为特征输入到训练器中进行学习（图 2.1）。目标是捕捉到电压不稳定的一些早期特征，同时辅助理解电压波动在电网中的传播规律，为未来电网中的规划准备。

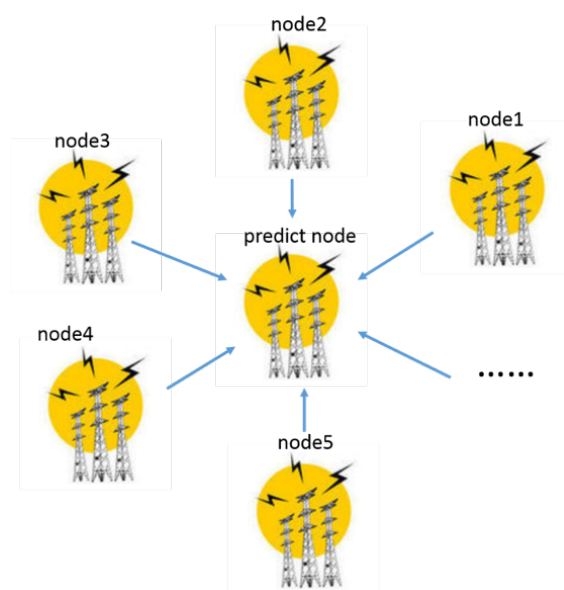


图 2.1 引入网络特征示意图

2) 延时预测

电压暂态稳定性分析的最终目标是实现稳定性预测。希望充分分析和利用已有数据来预测整个系统未来几秒钟或者几分钟是否会维持稳定。然而，在过去的工作中，没有把特征时段和预测时刻分开。由于临近预测时刻的数据包含了非常明确的预示稳定的信息，因此这种方法的逻辑性不够有说服力。

如图 2.2所示，我们的主要贡献是在预测时刻和特征时段之间设置了一段间隔。预测时刻要滞后于特征时段几秒钟。这样，分类器可以学习出一些早起的不稳定迹象。我们可以充分利用这几秒钟的时间间隔，提早采取措施（比如加入动态无功补偿装置SVG等）来稳定整个网络。

3) 移动窗口

在线系统中实现是暂态电压稳定性预测的最终目标，所有的特征收集是通过

一个“移动窗口”来实现的。当整个程序启动后，数据连续流入特征矩阵中。当特征矩阵被填满后，在线预测模块开始输出预测结果。同时，特征时段向前移动，最初的特征被丢弃，同时新的特征加入到特征矩阵的末尾。整个特征时段伴随着预测时刻不停地向前移动如图 2.2所示。

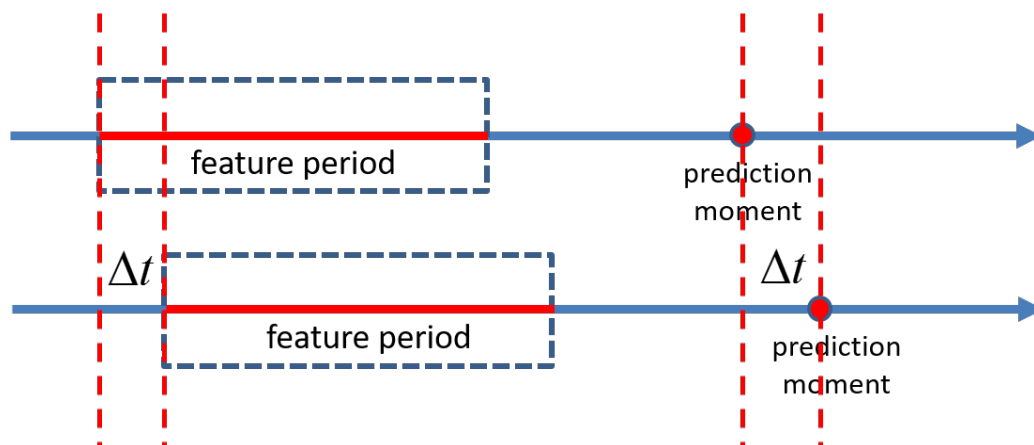
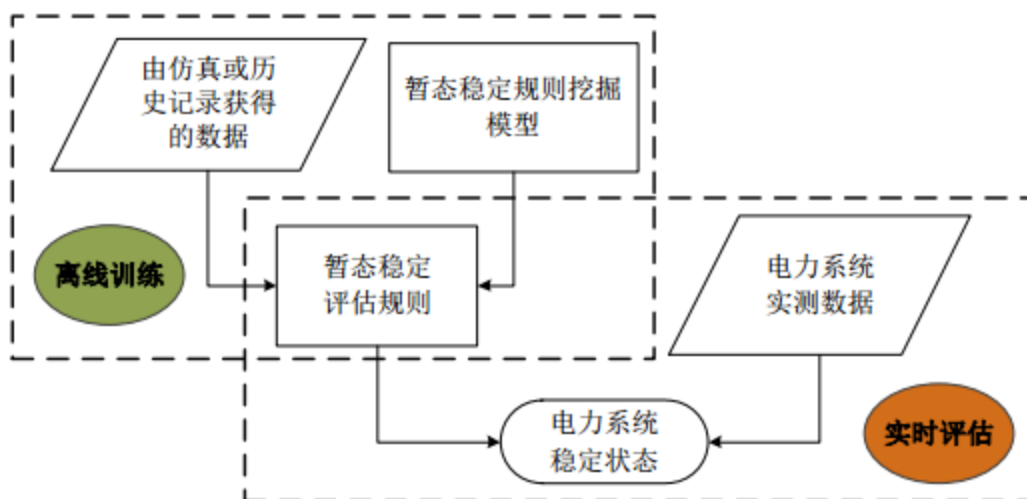


图 2.2 预测模型示意图

4) 离线训练和在线预测 基于数据的暂态电压稳定性评估和预测主要由两部分组成：离线训练和在线预测。如图 2.3所示。暂态稳定的评估规则是通过训练已有



数据和标签得到的，在线预测通过这组规则给出实时的电压稳定性预测结果。同时，新的数据会加入到训练库中，定期重新执行一次训练，更新模型和参数。

2.4 预测结果与分析

2.4.1 样本生成

在2.1节中，我们已经提到过，在真实的电网中，不稳定事件是非常罕见的。因此，稳定和 unstable 两类样本非常不平衡，导致有偏的分类器和预测结果。为了避免这个问题，我们的训练数据集是通过电力系统仿真软件PSASP（Power System Analysis）产生。PSASP^[43]全名是电力系统分析综合程序，是一款由中国电力科学研究院自主研发的电力仿真软件。

我们选择其中的示例系统CEPRI 36-bus系统（图 2.4），其中共有36根母线。选择其中的9条母线作为我们的研究目标，分别是bus16，bus18，bus19，bus20，bus21，bus22，bus23，bus29，bus9。

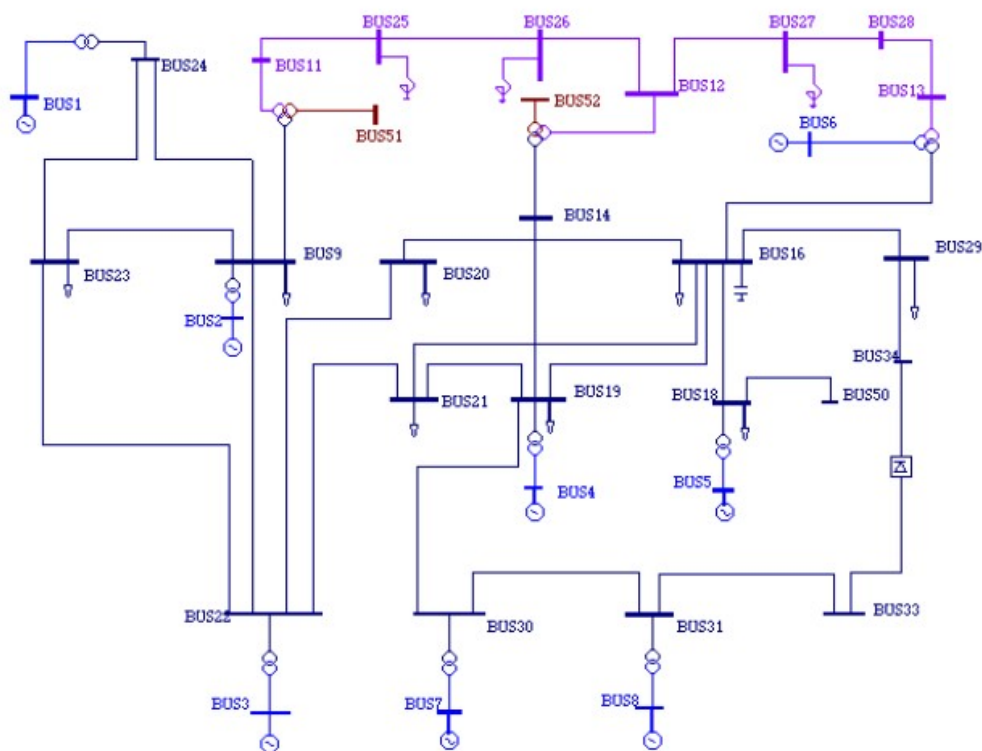


图 2.4 CEPRI 36-bus 系统

想要通过机器学习系统来对一个真实系统的输入输出进行学习，需要足够数量的样本，要能覆盖现实中所有可能发生的情况。因此，首先根据真实的运行记录或者经验，列出它所有的可能的运行条件或者故障。这些运行条件或者故障包括负荷比例、故障类型、故障地点、故障切除时间，或者其他运行条件。通过这些运行条件的自由组合，并输入到仿真软件中进行仿真，可以获得大量的样本。

在实际仿真中，我们选择的参数包括：总负荷的比例（80%,100%,120%）；感应电动机动态符合所占比例（40%，60%，80%）；故障地点在线路全长比例（0%，20%，40%，60%，80%）；故障切除时间（近端0.05s，远端0.1s和0.35s，0.4s）。

样本数量是否足够取决于特定系统的规模和复杂程度。系统的复杂程度越高，输入输出之间的非线性关系就越复杂。如果样本数量不够，使得样本在样本空间非常稀疏，很难准确的找到有效的分类边界。同时，如果在常规运行条件下非常难找到运行不稳定的例子，可以考虑增加更加极限情况，比如连锁故障或者增加动态负荷的比例。

在每次仿真过后，仿真软件可以输出不同母线的电压、电流、有功和无功。图 2.5和图 2.6 分别是典型稳定和不稳定样本的电压曲线

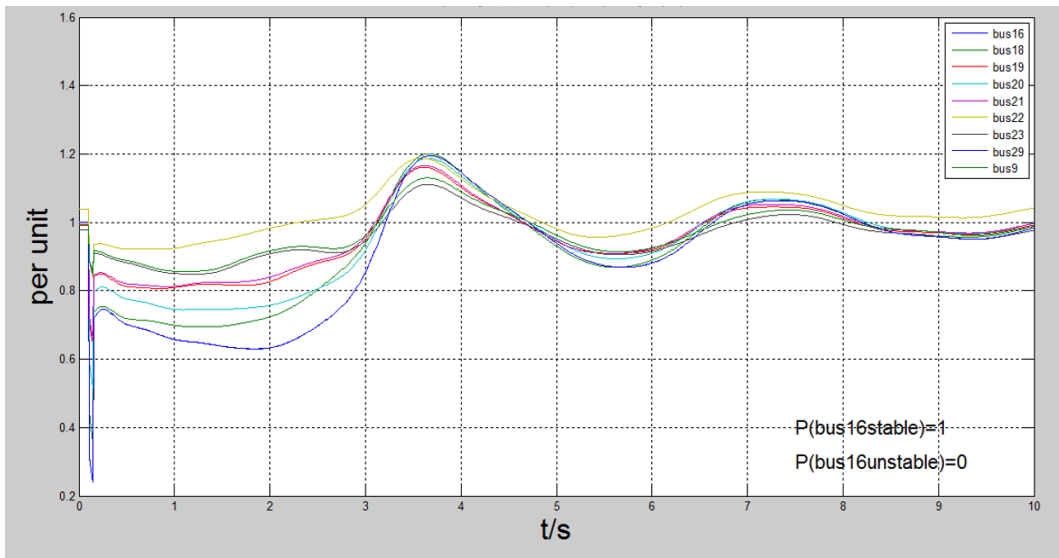


图 2.5 典型稳定样本电压曲线

2.4.2 算法流程

算法具体步骤包括：利用采样节点的测量值，建立样本矩阵；根据所述样本矩阵中各样本中的电压值标记各样本的稳定性标记值；压缩所述样本矩阵的维度，并利用压缩维度后的样本矩阵训练分类器；利用所述分类器预测下一时刻的电网状态稳定概率。下面具体阐述每一个步骤。

步骤1：建立样本矩阵

获得的数据为仿真软件的输出数据（或者部署在关键测点的同步相角测量单元PMU采集电网实时数据），这些数据包含电网中关键节点每个节点的电压 U 、有功 P 、无功 Q 、电流 I ；选取一个特征提取时段 T ，把这个时间段内的电压、有功、无功、电流的值排成一

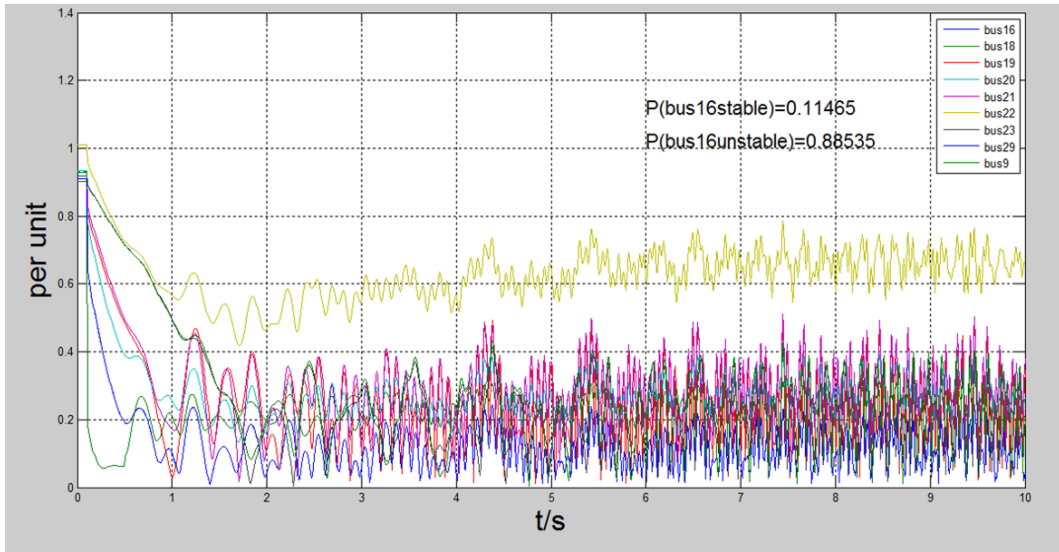


图 2.6 典型不稳定样本电压曲线

排 $U_{1,1}, P_{1,1}, Q_{1,1}, I_{1,1}, U_{1,2}, P_{1,2}, Q_{1,2}, I_{1,2}, \dots, U_{1,M}, P_{1,M}, Q_{1,M}, I_{1,M}$ 。最终组成一个样本矩阵

$$\begin{bmatrix} U_{1,1} & P_{1,1} & Q_{1,1} & I_{1,1} & \cdots & U_{1,M} & P_{1,M} & Q_{1,M} & I_{1,M} \\ U_{2,1} & P_{2,1} & Q_{2,1} & I_{2,1} & \cdots & U_{2,M} & P_{2,M} & Q_{2,M} & I_{2,M} \\ U_{3,1} & P_{3,1} & Q_{3,1} & I_{3,1} & \cdots & U_{3,M} & P_{3,M} & Q_{3,M} & I_{3,M} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ U_{N-1,1} & P_{N-1,1} & Q_{N-1,1} & I_{N-1,1} & \cdots & U_{N-1,M} & P_{N-1,M} & Q_{N-1,M} & I_{N-1,M} \\ U_{N,1} & P_{N,1} & Q_{N,1} & I_{N,1} & \cdots & U_{N,M} & P_{N,M} & Q_{N,M} & I_{N,M} \end{bmatrix} \quad (2-1)$$

其中第一个下标代表第几个样本，第二个下标代表第几个时间采集点。

步骤2：类别标记

对于有监督的机器学习方法，需要有数据标签。对于每一组样本，用0标记稳定，用1标记不稳定。区分稳定和不稳定的标准是特定时刻电压U是否恢复到标准值的0.8倍。因此，若样本中的电压值小于标准电压值的0.8倍，则该样本的稳定性标记值为0；若样本中的电压值大于等于标准电压值的0.8倍，则该样本的稳定性标记值为1。

步骤3：特征压缩

用主成分分析方法（Principal Component Analysis, PCA）^[44]进行数据降维与压缩。其核心思想是通过正交变换，将一组可能存在线性相关性的变量转换为一组线性不相关的变量，即将其投影到方差最大的方向上，从而选出最重要的变量。最终达到剔除数据中的冗余和特征降维与压缩的作用。

所有数据组成一个 $N \times n$ 的矩阵 X , N 代表样本的个数, n 代表每一个样本的维数, 矩阵的列向量为 x_1, x_2, \dots, x_n , 目标是把维数从 n 降到 m ;

首先进行样本数据归一化:

$$x_i(t)' = \frac{x_i(t) - \mu(x_i)}{\sqrt{\delta^2(x_i)}} \quad (2-2)$$

其中, $x_i(t)'$ 为样本矩阵 X 中的第 i 列向量 t 时刻测量值得归一化后的数值, $x_i(t)$ 为样本矩阵 X 中的第 i 列向量 t 时刻测量值, $\mu(x_i)$ 是 x_i 的均值, $\delta(x_i)$ 是 x_i 的标准差; t 为采样时刻。

计算归一化后数据的协方差矩阵 XX^T 的特征值和特征向量, 将所有的特征值按大小排序, 选择最大的 m 个特征值对应的特征向量 $\alpha^1, \alpha^2, \dots, \alpha^m$, 则投影降维后的数据为 $y(t) = [\alpha^1, \alpha^2, \dots, \alpha^m]^T X$;

降维后的矩阵为 P , $P = [\alpha^1, \alpha^2, \dots, \alpha^m]^T X$, 其中, P 是一个 $N \times m$ 的矩阵, N 为样本个数, m 为维数。

图 2.7是把数据降到2维后, 九条母线在空间中的分布图(其中星号代表不稳定, 圆圈代表稳定)。可以看出, 稳定和不稳定的节点具有很好的可分性。

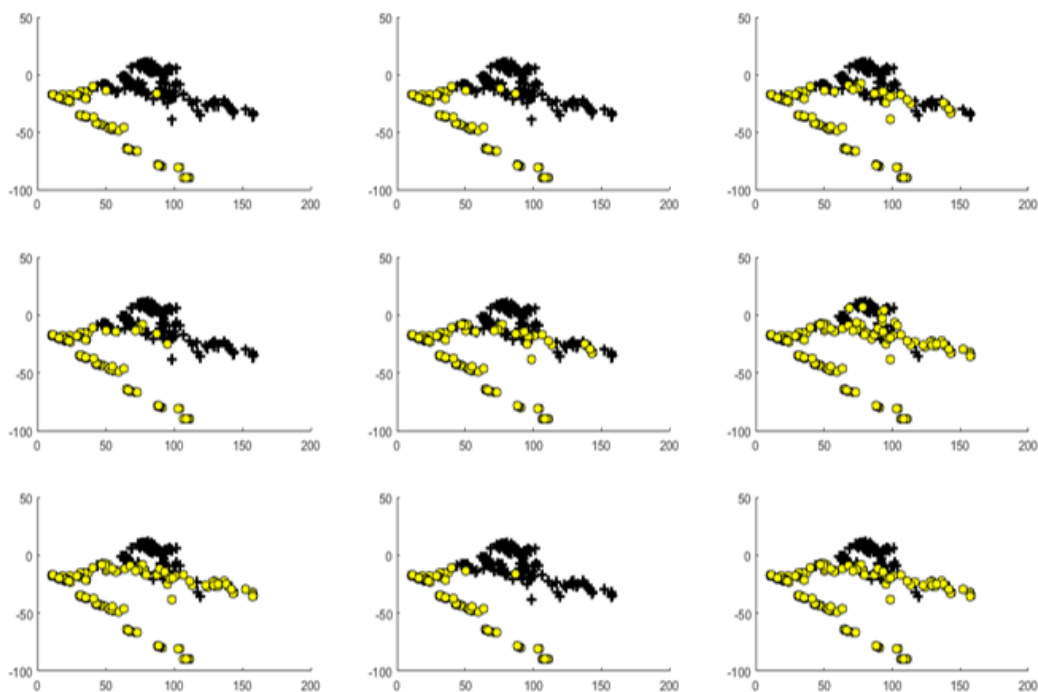


图 2.7 pca 降维后九条母线的样本分布

步骤4: 分类器训练

采用分类器是Logistic Regression^[45]，Logistic Regression是一种广义的回归方法，主要通过构造一个Cost函数（损失函数）来表示预测的输出与训练数据类别之间的偏差，然后通过梯度下降方法，找出最优参数，也就是分类器的训练过程。之所以选择Logistic回归作为电压暂态稳定性预测的分类器，主要基于以下两点原因：

1) Logistic回归得到的参数是具备物理意义的，代表输出对每一个维度的敏感性。这样可以看出重要节点和重要变量，方便知道未来的电网规划。

2) Logistic回归输出的结果是soft的，也就是说代表了稳定和不稳定的概率，而不是强硬的判断当前是稳定或者不稳定，这对电网的态势感知是非常重要的。

步骤4.1 降维后的样本 P 是一个 $N \times m$ 的矩阵， N 为样本个数， m 为维数。每一个样本分别为 $p^{(1)}, p^{(2)}, \dots, p^{(N)}$ ，共 N 个。每一个样本对应的电网状态（稳定或者不稳定）分别为 $y^{(1)}, y^{(2)}, \dots, y^{(N)}$ ，取值为0或者1, 0代表不稳定，1代表稳定。

建立logistic目标函数

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N \text{Cost}(h_{\theta}(p^{(i)}), y^{(i)}) \quad (2-3)$$

其中，当 $y = 1$ 时， $\text{Cost}(h_{\theta}(p), y) = -\log(h_{\theta}(p))$ ；当 $y = 0$ 时， $\text{Cost}(h_{\theta}(p), y) = -\log(1 - h_{\theta}(p))$ 。由于 y 的取值只能取0和1， $J(\theta)$ 可以进一步写成

$$\begin{aligned} J(\theta) &= \frac{1}{N} \sum_{i=1}^N \text{Cost}(h_{\theta}(p^{(i)}), y^{(i)}) \\ &= -\frac{1}{N} \left[\sum_{i=1}^N y^{(i)} \log h_{\theta}(p^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(p^{(i)})) \right] \end{aligned}$$

步骤4.2 其中， θ 一个 m 维的向量，代表特征中每一个维度的权重。然后通过梯度下降，寻找 $\hat{\theta}$ ，使得 $J(\theta)$ 最小。重复步骤

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad (2-4)$$

其中

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{N} \sum_{i=1}^N (h_{\theta}(p^{(i)}) - y^{(i)}) p_j^{(i)} \quad (2-5)$$

因此为了得到 $\min_{\theta} J(\theta)$ ，重复步骤

$$\theta_j = \theta_j - \alpha \frac{1}{N} \sum_1^N (h_{\theta}(p^{(i)}) - y^{(i)}) p_j^{(i)} \quad (2-6)$$

步骤5：预测部分

表 2.2 logistic回归寻找参数详细流程

算法1：logistic回归寻找参数详细流程

- 1: input: 训练集 $X = x_1, x_2, x_3, \dots, x_m$ ，训练集标签 $Y = y_1, y_2, y_3, \dots, y_m$
- 2: output: 最小化 $J(\theta)$ 的 $\hat{\theta}$
- 3: 建立目标函数

$$J(\theta) = -\frac{1}{N} \left[\sum_{i=1}^N y^{(i)} \log h_{\theta}(p^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(p^{(i)})) \right] \quad (2-7)$$

- 4: repeat:
- 5: for j from 1 to m

$$\theta_j = \theta_j - \alpha \frac{1}{N} \sum_{i=1}^N (h_{\theta}(p^{(i)}) - y^{(i)}) p_j^{(i)} \quad (2-8)$$

- 7: until: $\sum_{i=1}^N (h_{\theta}(p^{(i)}) - y^{(i)}) p_j^{(i)} < \varepsilon$
 - 8: $\hat{\theta}_j = \theta_j$
-

当预测启动后，首先要经历数据启动阶段，此阶段是在填充特征矩阵，没有输出；特征矩阵填充完成后，所有的特征形成队列 \hat{x} 。

利用步骤4中得到的 $\hat{\theta}$ ，预测网络稳定概率为

$$\frac{1}{1 + e^{-\hat{\theta}^T \hat{x}}} \quad (2-9)$$

不稳定概率为

$$1 - \frac{1}{1 + e^{-\hat{\theta}^T \hat{x}}} \quad (2-10)$$

特征时段向前滑动，最初的特征被抛弃，新特征补充在队尾，如图 2.2所示。每隔一定时间间隔 Δt ，要把新收集来的数据与以前的数据一起，重新返回到步骤4训练分类器，更新参数。整个算法的流程如图 2.8所示。

表 2.3 电压稳定性预测算法流程

算法2: 基于时延的能源互联网在线电压稳定性预测
1: input: 训练集 $X = x_1, x_2, x_3, \dots, x_m$, 训练集标签 $Y = y_1, y_2, y_3, \dots, y_m$, 实时在线数据 \hat{x}
2: output: 暂态电压稳定概率 p_{stable}
3: 对训练集利用PCA进行降维, 降维后的数据集为 $P = p_1, p_2, p_3, \dots, p_m$
4: 利用降维后的数据集 $P = p_1, p_2, p_3, \dots, p_m$ 和训练集标签 $Y = y_1, y_2, y_3, \dots, y_m$, 建立 $J(\theta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \left(\frac{1}{1+e^{-\theta^T p_i - y_i}} \right)^2$
5: 通过梯度下降, 找 $\hat{\theta}$ 使得 $J(\theta)$ 最小
6: 填充特征矩阵, 形成特征队列 \hat{x}
7: repeat:
8: 稳定概率 $p_{stable} = \frac{1}{1+e^{-\hat{\theta}^T \hat{x}}}$
9: 特征时段向前滑动, 最初的特征被抛弃, 新特征补充在队尾

2.4.3 算法结果与分析

理论上机器学习的算法验证应该被分为训练集、验证集和测试集。训练集主要用于训练模型，通过样本数据集建立一个分类器。验证集是对学习出来的模型，调整分类器的参数，如在神经网络中选择隐藏单元数、网络结构等。测试集主要是测试训练好的模型的分辨能力，如分辨率等。由于获取样本的成本较高，总样本数量较少，所有的样本只被分为训练集和测试集。同时，通过交叉验证^[46]的方法来避免出现过拟合。

在本方法中，有两个重要的参数，分别是特征时段的长度和特征时段距离预测时刻的间隔。特征时段的长度代表了所选取特征数量的多少，与计算成本和计算复杂度正相关。特征时段距离预测时刻的间隔代表了提前多长时间就可以发现早期不稳定的特征。如果发现特征时段距离预测时刻的间隔较长时，就可以得到比较高的正确率，说明了通过对早期特征的挖掘，就可以从中发觉不稳定的迹象，从而达到态势感知和提前预警的目的。这段宝贵的间隔也可以为尽早加入电网治理措施（如无功补偿等）留出时间。

因此在训练和预测过程中，设置特征时段长度分别为3.0s、3.5s、4.0s、4.5s、5.0s、5.5s、6.0s、6.5s、7.0s、7.5s、8.0s、9.5s，设置特征时段分别为0.5s、1.0s、

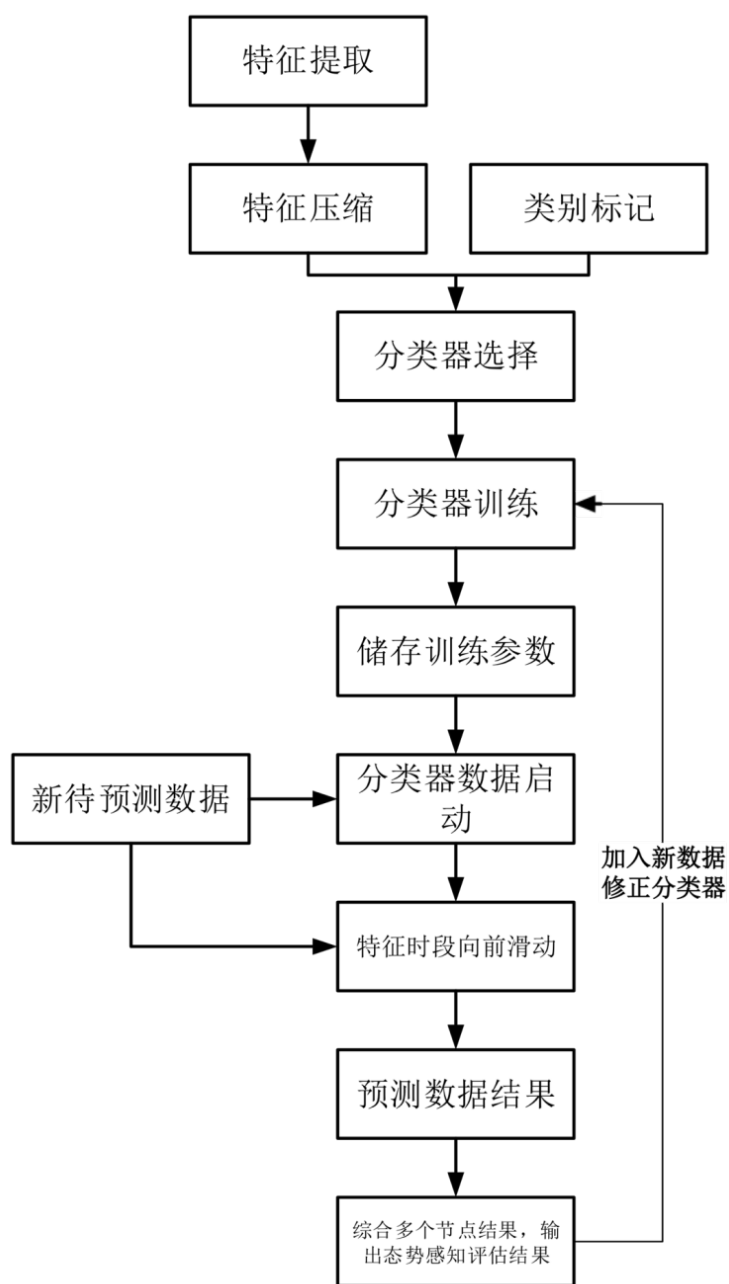


图 2.8 电压稳定性预测算法流程图

1.5s、2.0s、2.5s、3.0s、3.5s、4.0s、4.5s、5.0s。与如果通过交叉验证方法得到的在9条母线上的测试正确率的如图 2.9所示，在bus16上的正确率如表 2.4所示（表格中的空数据是由于仿真时段输出的总长度为10s）。

从图中可以看出，随着特征时段与预测时刻时间间隔的增大，9条母线上的正确率呈现降低的趋势。反映了越接近预测时刻的特征，与预测时刻是否稳定的相关性越大。但是在0.5s-2.5s这个区间内，准确率下降的速度较慢，准确率下降的绝对值并不明显。从bus9，bus16，bus21，bus23的曲线可以看出，在0.5s-2.5s区间，

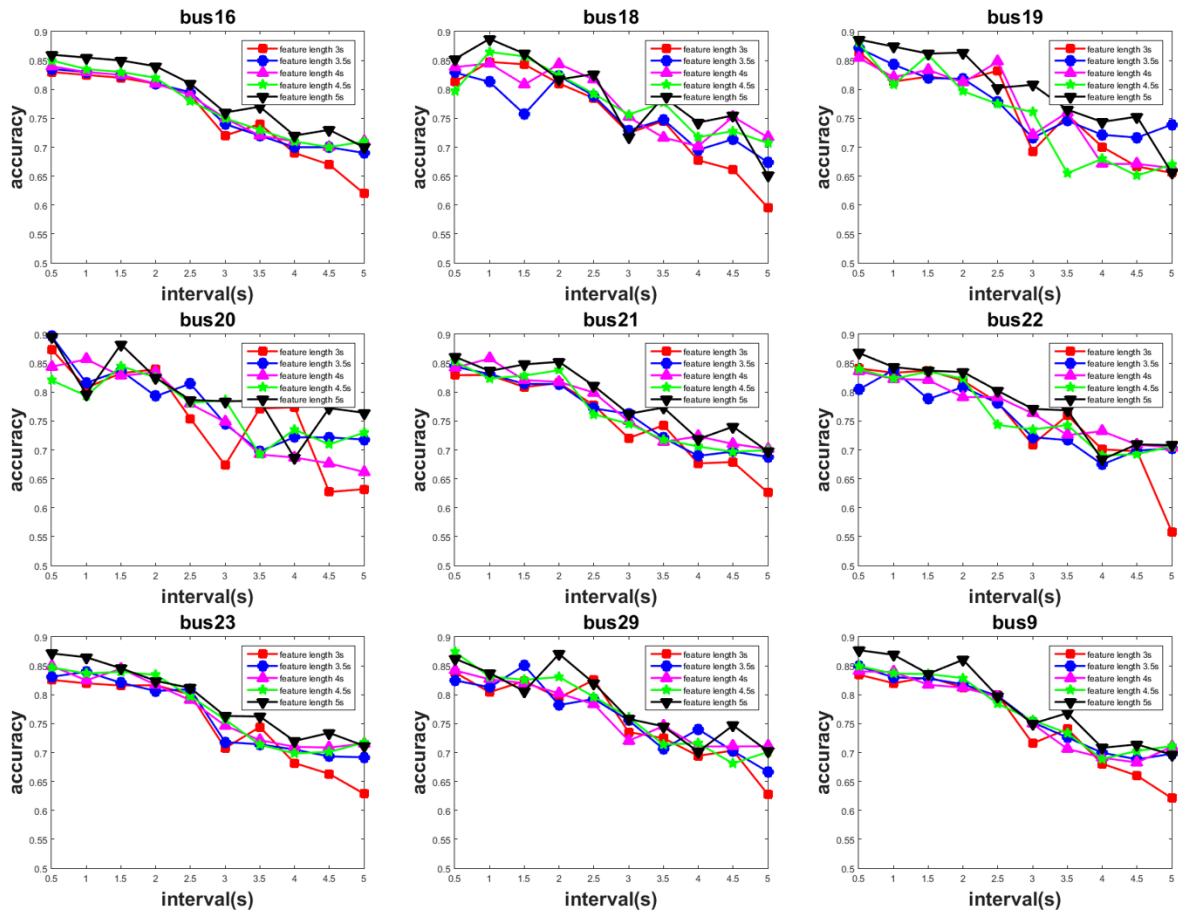


图 2.9 9 条母线上正确率

准确率大约普遍从0.85下降到0.8。同时，在2.5s-5s这个区间内，准确率下降的速率明显加快。因此，综合以上两条特征，面对较短的特征时段与预测时刻的间隔范围内，间隔的增大带来的信息损失和预测准确度的下降并不多，在一个特定的时间点，系统是否稳定的特征就已经非常明确。随着间隔的减小，并不会带来绝对信息量的增加。

另外，可以从图中看出，随着特征时段长度的增加，准确率有轻微的下落，但是远不如特征时段与预测时刻间隔的变化明显。综合特征时段长度的变化和特征时段与预测时刻间隔的变化，得到9条母线上正确率的变化得到曲面图 2.10。可以看出，特征时段越长，特征时段与预测时刻间隔越短，正确率越高（黄色的部分）；反之，特征时段越短，特征时段与预测时刻间隔越长，正确率越低（蓝色的部分）。

2.5 在线电压稳定性分析系统

在这章中，我们开发了具有时延的暂态电压稳定性分析模型。在模型设计中，

表 2.4 bus16母线正确率

interval feature length	0.5s	1.0s	1.5s	2.0s	2.5s	3.0s	3.5s	4s	4.5s	5s
3.0s	0.83	0.83	0.82	0.81	0.79	0.72	0.74	0.69	0.67	0.62
3.5s	0.84	0.83	0.83	0.81	0.80	0.74	0.72	0.70	0.70	0.69
4.0s	0.84	0.83	0.83	0.81	0.79	0.75	0.72	0.71	0.70	0.71
4.5s	0.85	0.84	0.83	0.82	0.78	0.75	0.73	0.71	0.70	0.71
5.0s	0.86	0.86	0.85	0.84	0.81	0.76	0.77	0.72	0.73	0.70
5.5s	0.84	0.83	0.85	0.83	0.81	0.76	0.74	0.72	0.74	-
6.0s	0.85	0.82	0.83	0.81	0.79	0.74	0.70	0.73	-	-
6.5s	0.86	0.85	0.86	0.82	0.79	0.77	0.77	-	-	-
7.0s	0.85	0.88	0.87	0.84	0.80	0.81	-	-	-	-
7.5s	0.90	0.85	0.86	0.82	0.83	-	-	-	-	-
8.0s	0.92	0.93	0.91	0.88	-	-	-	-	-	-
8.5s	0.89	0.92	0.86	-	-	-	-	-	-	-
9.0s	0.96	0.96	-	-	-	-	-	-	-	-
9.5s	0.98	-	-	-	-	-	-	-	-	-

一个重要特征是把离线训练和在线预测相结合。在本节，我们将利用上章中的模型开发一套应用系统——在线电压稳定性分析系统。总体的设计思想是充分利用已有的开源分布式计算框架，根据能源互联网数据分析的需求和数据特征，来组合各个模块。由于目前的开源分布式计算框架对于低等的存储、消息处理等提供了很好的保证，同时提供了诸多功能结构供上层调用，使得我们可以专注在面向能源互联网的功能设计上。

2.5.1 系统架构

在线电压稳定性分析系统（Real-time Voltage Stability Analysis System, RVSAS）是通过上面所描述的机器学习方法来进行能源互联网电压稳定性分析和预测的平台。实时性是对RVSAS处理电网数据的基本要求。传统的电网大数据平台^[47-50]只考虑了大数据的批处理，而没有考虑实时性和在线处理的要求。考虑到数据的特点和整个算法的流程，对RVSAS平台的基本要求如下：

- 对网络状态流数据的实时分析：由于电网的状态数据是实时不间断产生的，因此RVSAS应该可以连续处理数据流并且预测电网故障发生之前预警网络接下来的行为。
- 预测参数的快速训练和学习：传统的有监督学习的训练过程是离线的，并且批处理的；在新的应用背景下，要求训练系统要有高性能、要可用性、高容

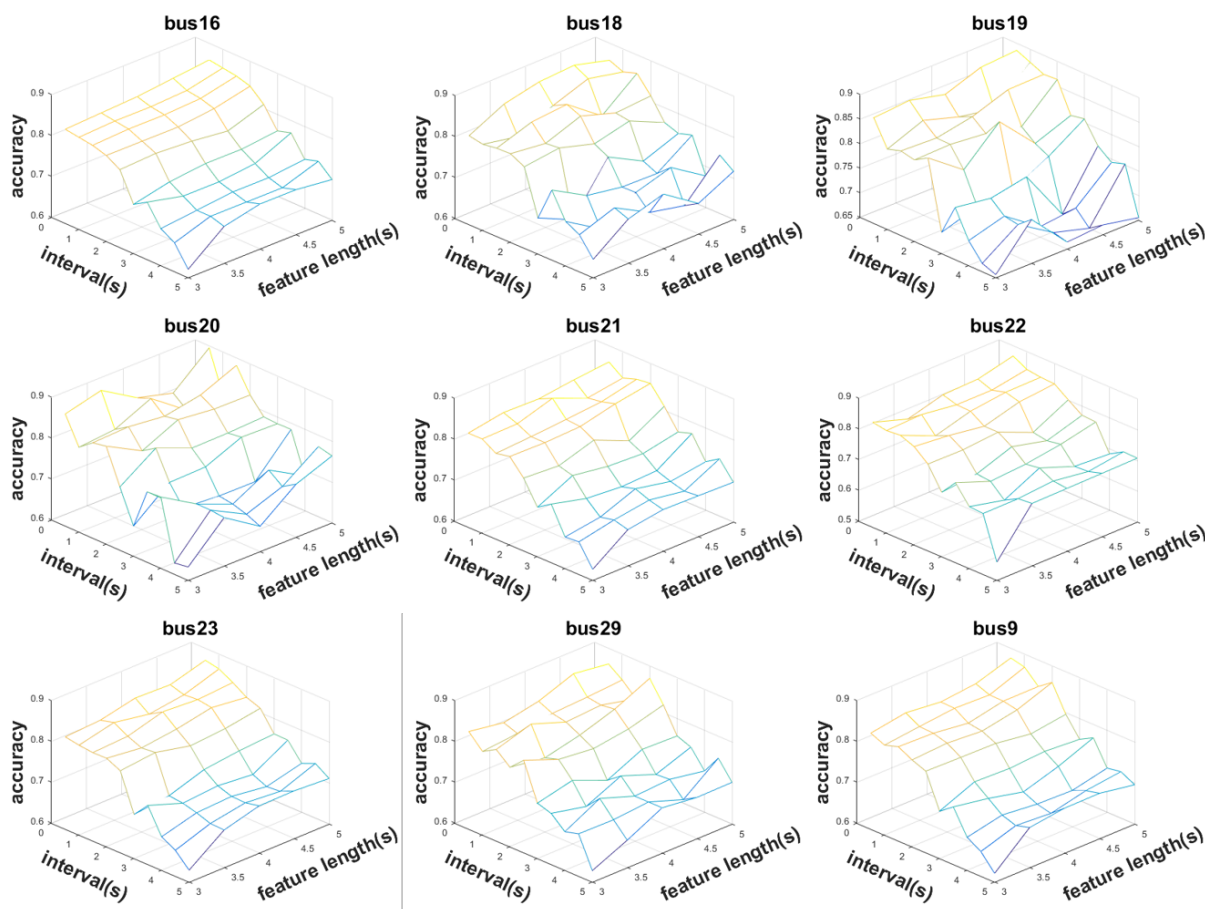


图 2.10 9 条母线正确率随着特征时段长度和预测间隔的变化

错性，训练的频率更加频繁。

- 实时的模型更新：在RVSAS中，网络结构的变化被考虑进来。训练模型定期进行训练，并将训练结果传递给预测模型，更新在线预测的参数。这样，刚刚产生的数据也可以马上加入到样本中，反应网络的最新情况，提高整个预测的准确度。

基于上述的几点要求，RVSAS采用开源分布式计算工具来架构整个分析平台，他们的分工如下：

- Hadoop HDFS：电网训练数据的存储
- Spark：分析模型的快速训练和机器学习
- Storm：电网流式数据的实时处理
- Kafka：训练模块和分析模块之间的数据管道，勇于传递参数
- D3.js：web前端进行数据可视化

整个RVSAS的架构如图 2.11所示。

最初的离线训练数据存储存储在HDFS中。模型的训练部分由Spark完成，训练得到的参数被传递到流式计算的Storm在线分析模块中。在线分析模块主要分为两

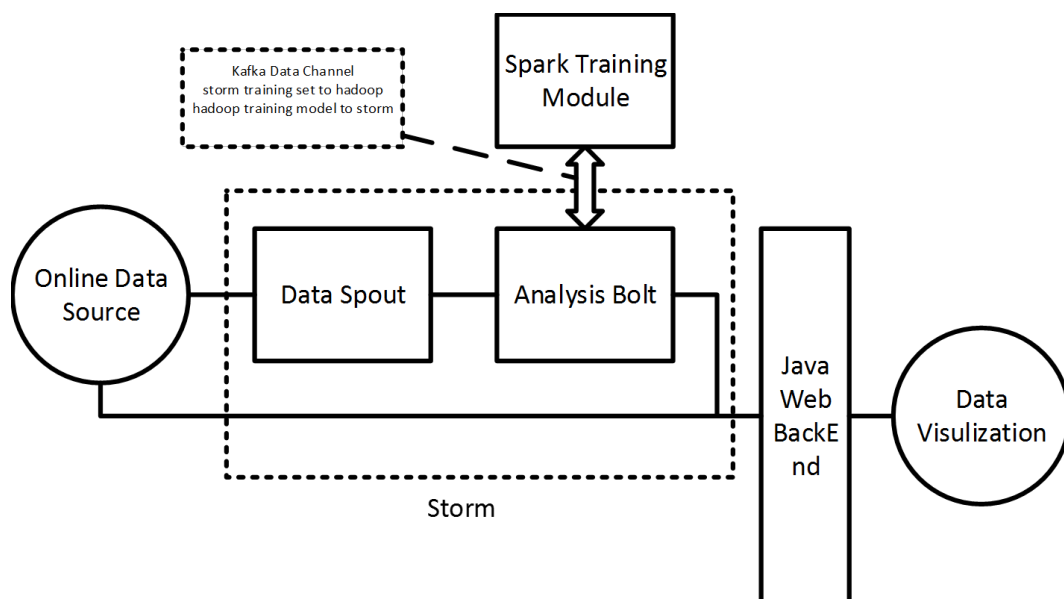


图 2.11 RVSAS 系统架构图

个部分：Data Spout和Analysis Bolt。Data Spout负责收集在线电网数据同时传递到数据存储模块。Analysis Bolt通过训练模块得到的参数，结合最新的电网状态进行电网稳定性的预测。在这个过程中，在线的数据每隔一段时间通过Kafka传递给Spark训练模块进行模型的重新训练，得到更新后的参数。更新后的参数同样由Kafka传递回Analysis Bolt。

为了直观生动的反映电力网络系统的整体状态，可以对相关参数和向量进行三维显示。三维显示图中，不同颜色表示不同的状态，可以直观的观察系统处于稳定状态和不稳定状态时的电压、有功、无功的大小分布情况，从而对于稳态判断和态势感知有一定的辅助作用，最终辅助支持决策。图 2.12和图 2.13是RVSAS可视化模块的效果展示。

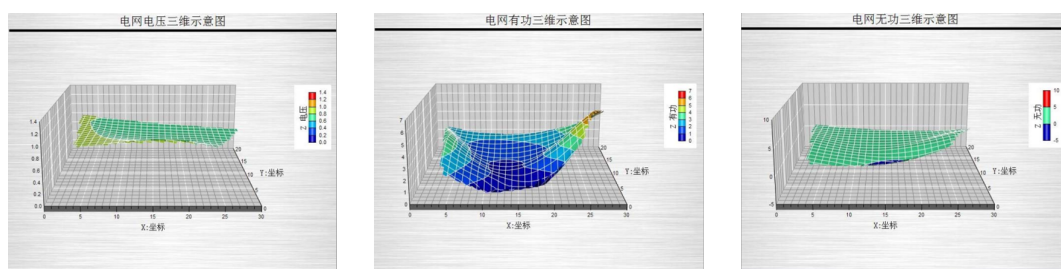


图 2.12 稳定情况下的RVSAS 的可视化模块

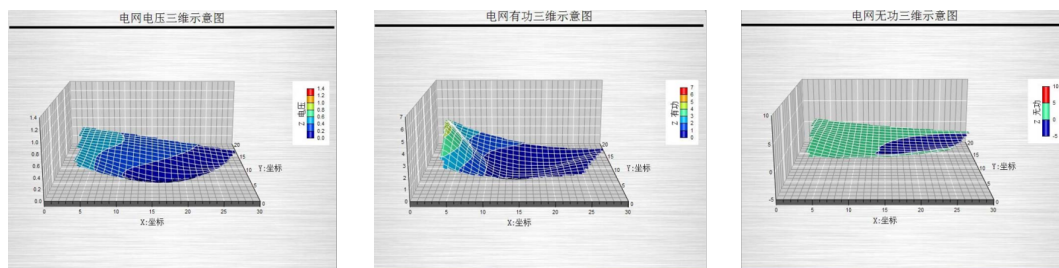


图 2.13 不稳定情况下的RVSAS 的可视化模块

2.5.2 离线学习与参数更新：Spark

Spark是UC Berkeley AMP lab所开源的类Hadoop MapReduce的新型并行计算平台和框架。与Hadoop一样，Spark也是基于Map reduce算法实现的分布式计算，因此具有Hadoop所具有的的优点。但它本身并不需要读取分布式存储系统，因此计算速度更快。Spark的特点如下：

- 支持比Map 和Reduce 更多的函数
- 优化任意操作算子图（operator graphs）
- 可以帮助优化整体数据处理流程的大数据查询的延迟计算
- 提供简明、一致的Scala, Java和Python API
- 提供交互式Scala 和Python Shell

2.5.3 在线稳定性预测：Storm

Storm是一个流式分布式实时大数据计算平台。与Hadoop批处理不同，Storm是一个流式计算框架，因此它的任务是永不停止的。相比于Hadoop注重吞吐量，Storm更加注重计算延迟和处理速度。Storm的主要特点如下：

- 支持多种主流编程语言：包括Clojure、Java、Ruby 和Python等，若要增加对其他语言的支持，只需要增加一个简单的Storm 通讯协议即可完成
- 横向扩展能力强：由于计算任务是在多个计算之间并行完成的，因此具有非常好的水平扩展能力
- 确保消息被处理：Storm消息处理的可靠性非常强，使得保证了每一个消息都可以确保被处理，处理失败的消息会负责任地从消息源重试
- 处理速度快：Storm的流式计算框架使得其处理时延远远小于批处理平台。
- 具有本地模式：本地模式在处理消息过程中完全模拟集群模式，方便任务的开发和调试

整个Storm的架构和topology架构分别如图2.14所示，主要分为Spout, Bolt, Tuple等几个模块，下面分别介绍其功能与作用：

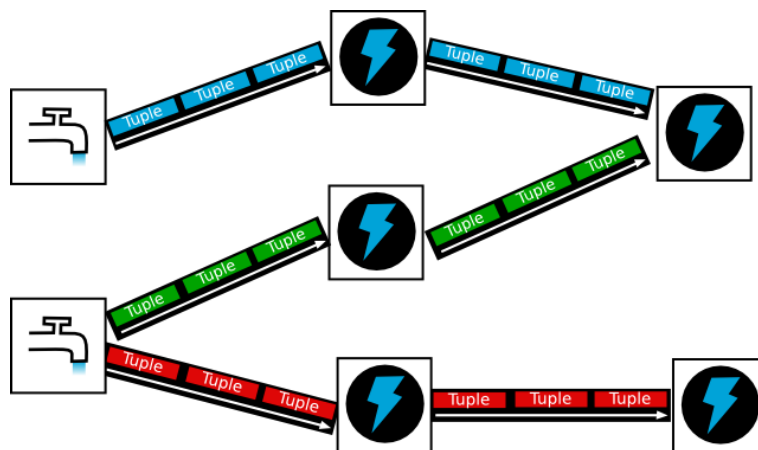


图 2.14 Storm 系统架构图

- Spout: Spout是一个topology中产生数据流的组件，通过不断调用nextTuple()函数，来从外部数据源读取数据，并成为内部数据源
- Bolt: Bolt是一个topology中执行数据处理的组件，通过不断调用execute()函数，Bolt可以执行任何函数、合并、读写数据库等一系列操作
- Tuple: 是Storm中消息传递的基本单位
- Stream: 源源不断传递的tuple 就组成了stream。一个stream 是一个没有边界的tuple 序列

2.5.4 参数传递: Kafka

Kafka是分布式发布-订阅消息系统，最初由LinkedIn公司开发，之后成为Apache项目的一部分。Kafka是一种快速、可扩展的、设计内在就是分布式的、分区的和可复制的提交日志服务。Kafka与传统消息系统相比，有以下不同：

- 扩展性强: Kafka是一个分布式的消息系统，可以随时通过增加机组来增加处理数据的规模
- 支持多订阅者: 并且当消息发送和传播失败时，可以自动平衡消费者
- 吞吐量高: Kafka可以同时为发布和订阅提供高吞吐量
- Kafka: 训练模块和分析模块之间的数据管道，勇于传递参数
- 将消息持久化到磁盘，因此可用于批量消费，例如ETL，以及实时应用程序

Kafka的架构如图 2.15所示。主要由话题、生产者和消费者组成。在RVSAS系统中，Kafka主要承担参数传递和模型更新的任务，如图 2.16。电网新增数据通过Kafka定期传递给Spark进行模型训练，训练完成后更新的参数通过Kafka传递给Storm进行实时预测。

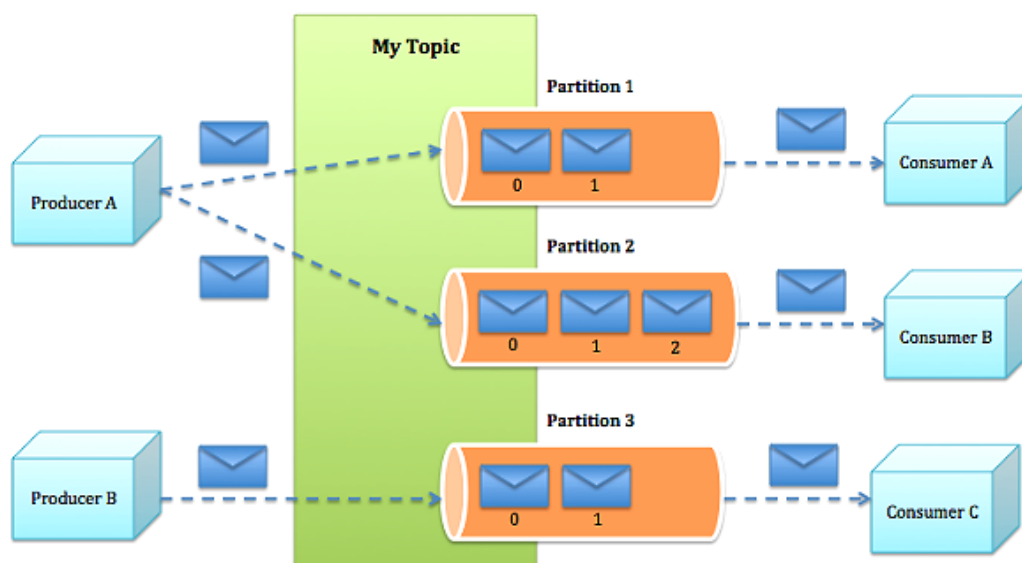


图 2.15 Kafka 系统架构图

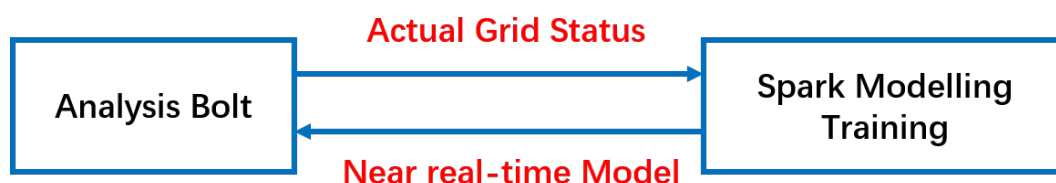


图 2.16 Kafka 进行参数传递和模型更新

2.6 本章小结

电压稳定性分析是能源互联网态势感知的第一步，也为接下来的故障诊断和决策选择做出了铺垫。本章的贡献和创新点主要有以下三点：

建立网络的延时电压稳定性预测模型。通过在特征时段和预测时刻中间设置间隔，使得分类器充分学习网络中电压不稳定的早期特征和迹象，最终在预测效果上实现提前预警，这个间隔也为各种保护措施的投切提供了可能。同时引入空间特征。传统的电压稳定性分析的训练数据只来自于目标节点的历史数据。通过引入空间特征，整个网络的数据都作为一个特定节点的训练数据。使得我们可以提早发现外源性的电压波动，同时理解各种故障在网络中的传播规律。

通过对CEPRI-36母线系统仿真出的样本进行训练和预测，延时电压稳定性分析模型得到了非常好的效果。我们发现随着特征时段和预测时刻间隔的增大，预测准确率逐渐下降，证明越靠近预测时刻的特征，越能反应暂态电压不稳定的迹象。但同时0.5s-2.5s这个范围内，正确率下降的速度很慢，反映了增量信息不多。通过权衡正确率和间隔大小，可以实现高效的提前预警。

第3章 能源互联网故障诊断

3.1 本章引言

大规模的区域互联使得局部电网的故障对整个系统的影响越来越大。如果对电网故障处理不当，会造成非常严重的社会经济损失。电网故障诊断是通过对各级各类保护装置、断路器的保护动作信息，以及来自电网状态监测装置的状态信息进行分析和挖掘，根据保护器件的逻辑关系和相关调度人员的经验来推断故障的类型和位置^[51]。当电网发生故障时，快速、准确的故障诊断对于整个电网的恢复具有非常重要的意义。

根据所使用的数据类型不同，电网故障诊断的发展大致可以分为四个阶段。在第一阶段，由于测量手段有限，可以获取的数据类型和数量非常少，这个阶段的故障诊断主要依靠人工实现。基于调度员经验的故障诊断的可靠性和效率都很低，故障定位要占到整个故障处理时间的三分之一以上。第二阶段主要依靠数据采集与监视控制系统^[52]（SCADA, Supervisory Control And Data Acquisition），采集的数据类型主要是保护装置的动作信息和断路器的跳闸信息。当电网中某一设备发生故障时，会导致与其相关的一系列保护装置动作和断路器跳闸，将故障设备从系统中切除。在第二阶段中，基本已经实现了有效的故障诊断，但是由于不采集暂态波形数据，无法完成对故障信息的直接分析。在第三阶段中，由于故障录波器的使用，克服了这个问题。通过采集故障时暂态录波信息，加强了对故障信息的直接分析。结合保护装置信息和断路器信息，取得了更好的诊断效果。在第四阶段，广域信息系统^[53]（WAMS, Wide Area Measurement System）兼具了SCADA系统和故障录波系统的功能。其前置PMU单元以高频率的采集电网电流、电压信息，同时计算出功角、有功、无功等信息。WAMS最大的特点是通过全球定位系统（GPS）校对，可以保证各个监测点数据的同步性。由于WAMS的使用大大丰富了可供使用的数据类型和数量，传统基于逻辑推理方法的实施难度很大，因此一批基于机器学习的智能方法得到广泛应用。目前，可供电网故障诊断使用的数据类型对比如表 3.1。

基于数据电网故障诊断使用的方法包括专家系统，ANN，模糊算法，贝叶斯网络等等。下面对他们一一进行阐述，并对比各种方法的优缺点进行对比。

专家系统^[54-56]（Expert System）是一种基于知识和规则，以推理为主要手段的诊断方法。它是把保护装置、断路器的动作信息和相关调度人员的经验用规则表达出来，形成故障诊断的专家知识库。当有故障发生后，把监测到的保护装置、

表 3.1 电网故障诊断的四个发展阶段

数据类型	人工数据	SCADA数据	故障录播数据	PMU数据
采集方法	人工收集	实时上传	故障后启动	定时采样
采集间隔	-	<0.5s	1000Hz/s	50-100Hz/s
采集地点	全网	全网	110kV以上	220kV 以上
上传方法	主动	主动	被动	被动

断路器的动作信息输入到知识库中，结合一系列规则进行正反推理，得出故障类型。专家系统方法的优点是电网中保护动作和故障之间的关系可以用直观的、模块化的规则表达出来，同时解释能力比较强。缺点是当电网规模比较大时，构建知识库难度很大；每当网络结构发生变化时，就要更新知识库，主动学习能力差；对于保护装置误动作的鲁棒性和容错性差。能源互联网网络结构的复杂性远远超过传统电网，在这种背景下，专家系统这种基于规则的方法已经不再适用。

人工神经网络^[57,58]（ANN, Artificial Neural Network）是一种模拟人脑处理信息的数学模型。在电网故障诊断中，输入层是所有保护装置和断路器的状态，输出层是电网中各个原件的状态，包括母线、线路和变压器等。数字0表示原件状态正常，1表示存在故障。人工神经网络的优点是可以表达非常复杂的非线性映射；有一定的泛化能力，可以自组织、自学习；适合于并行处理。人工神经网络算法的缺点是需要大量带有标签的样本，在实际电网中，获取此类样本是非常昂贵的；把系统当做黑箱处理，解释能力差；收敛速度慢，不适合大型网络；当电网拓扑结构发生变化时，需要重新学习。

贝叶斯网络^[59,60]是一种基于概率的不确定性推理网络，对于处理复杂电网中由于不确定因素引起的故障和数据不完备时具有很好的效果。之所以有不确定性因素存在，是因为保护装置和断路器经常会误动作，因此这些动作数据和故障事件并不是一一对应的。贝叶斯网络将保护装置和断路器的动作数据与故障之间描述为一个概率分布，提高了诊断的准确性和容错性。缺点是需要大量的样本才能计算出相应的概率分布。

基于优化算法^[61]的故障诊断是通关简历保护装置动作和故障之间的约束关系，使得问题变为0-1整数规划（0-1代表原件的两个状态，0为正常，1为故障）问题，然后通过优化算法求解该问题的最优解。

从上面的总结可以看出，过去电网故障诊断所使用的的数据类型主要还是保护装置和断路器的动作数据，使用的方法主要是基于逻辑，并没有充分利用网络中电气量的测量数据。如今PMU安装数量越来越多，其产生数据的价值没有被充分发挥。本章进行故障诊断主要使用电气量的测量数据，方法由原来的基于逻辑

变为基于数据，是典型的模式识别问题。虽然数据量的增大带来有效信息的增加，同时也使得数据中的冗余大大增加，给数据挖掘和故障诊断增加难度。因此，降维和特征提取非常重要。本章提出了基于流形学习的能源互联网故障诊断方法。通过充分利用已知的标签信息，实现有监督的LLE算法。

本章的具体安排如下：3.2节主要介绍降维和流形学习；3.3节主要阐述LLE算法是如何应用于电网故障诊断的，主要是阐述如何把标签信息应用到LLE之中，使其变为一种有监督的降维算法；第3.4节主要关于LLE算法中两个重要参数的选择：近邻数 k 和内嵌维度 d ；第3.5节描述了样本产生的过程及实验结果的分析；第3.5节是本章小结。

3.2 流形学习

3.2.1 降维

降维就是通过某种映射关系，将高维空间中的数据映射到低维空间中。之所以要进行降维是因为在很多应用场景下，可获取样本的数量很少，当样本的维度很高时会引发“维度灾难”，导致样本在高维空间中不可分或者过拟合。降维可以降低计算和存储的成本，同时减少噪声和数据中的冗余，提高分类和故障诊断的准确率。另外，人类直观只能理解三维世界，因此降维有利于我们进行数据可视化和加强对数据的理解。

降维方法主要可以分为线性和非线性，区别是所对应的映射是线性还是非线性的。线性降维方法主要包括主成分分析（PCA, principal component analysis），线性判别分析（LDA, linear discriminate analysis），独立成分分析（ICA, independent component analysis）等。然而，电力系统是非常典型的复杂非线性系统，因此在进行复杂电力系统故障诊断时需要使用非线性降维方法。非线性降维方法主要包括LLE（Locally linear embedding），Isomap，Kernal PCA等。表 3.2是主要降维算法的一个对比。

故障诊断问题是一个典型的模式识别中的分类问题，因此就需要采集大量的样本和特征进行数据训练。在能源互联网的应用背景下，原始特征空间的维数可能高达几千维甚至上万维。如果直接在原始空间上进行分类器训练，可能回直接带来两个问题：

- 1) 样本在高维空间中是不可分的，导致相当多在低维空间中具备良好性能的分类器在高维空间中无法取得很好的效果。

- 2) 如果想让样本在高维空间中可分，分类器会学习出非常复杂的非线性关系。极限状态就是直接记忆每一个样本点的分类结果。这种情况下，盲目追求准确率

表 3.2 主要降维算法对比

Algorithm	Linear	G/L	Supervised
PCA	Linear	Global	Unsupervised
LDA	Linear	Global	Supervised
ICA	Linear	Global	Unsupervised
LLE	nonlinear	local	unsupervised
Isomap	nonlinear	global	unsupervised
Laplacian	nonlinear	local	unsupervised

会使得在训练集和验证集上正确率非常高，但是分类器的泛化能力和推广能力很差，出现所谓的“过拟合”和“过学习”的现象。在能源互联网的应用背景下，2.2节中已经提到，样本的获取成本非常高，目前大多通过在仿真系统中设置典型故障获得。另外当网络规模增大时，特征数量呈指数形式增加。因此，获得的样本数量远远低于样本维数，增加了过拟合发生的可能性。

为了解决“维数灾难”问题，尽量利用较少的维度获得原始高纬度空间中较多的信息。通过降维方法从原始的特征空间中寻找与分类目标最相关的特征，提高分类器的训练效果和训练速度。

3.2.2 流形学习

流形 (Manifold)，是“局部具有欧氏空间性质的空间”，具体包括各种维数的曲线曲面。所谓流形学习，就是通过映射把一组高维数据在低维空间中重新表示，因此本质是一种降维算法。流形学习之所以有效，是因为很多高维数据是低维流形在高维空间中扭曲的结果，如图 3.1。因此在高维空间看似杂乱无章没有规律的数据，通过流形学习降维后，可以更直观地表示器流形结构。

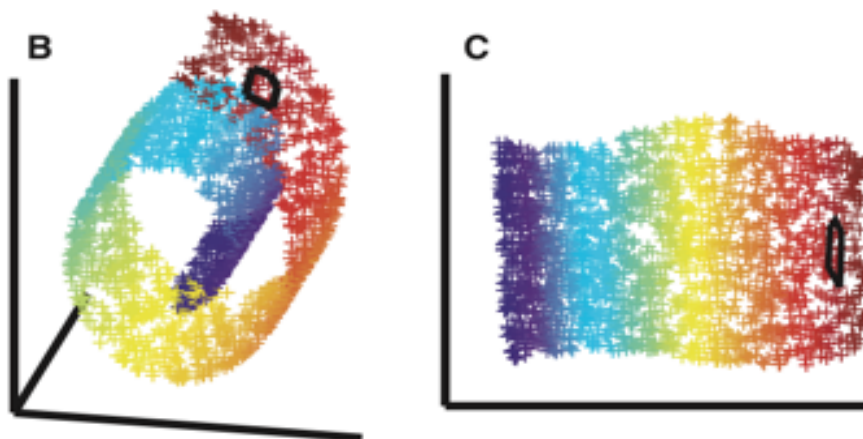


图 3.1 高维流形在低维空间中嵌入

如图 3.2是石膏像和手写数字利用ISOMAP算法降维后的结果。在原始的特征空间中，每一个石膏像都是一张 64×64 灰度图，因此如果把位图按照行或者列排列起来，就得到一个4096维的向量。在这个4096维的欧式空间中，每一个点对应一张石膏人脸的图片。但是通过观察我们可以发现，由于拍摄对象是同一个石膏，不同图片之间只有两个自由度：人脸的姿势和光线的角度。从图中可以看到，映射到二维空间后的数据已经按照这两个自由度进行排列，也就是说这是一个嵌入到4096维空间中的二维流形。在手写数字降维结果中，效果是类似的，最终由“底部圆圈”和“顶部弧度”就可以把这些图像合理组织，按照规律进行排列。

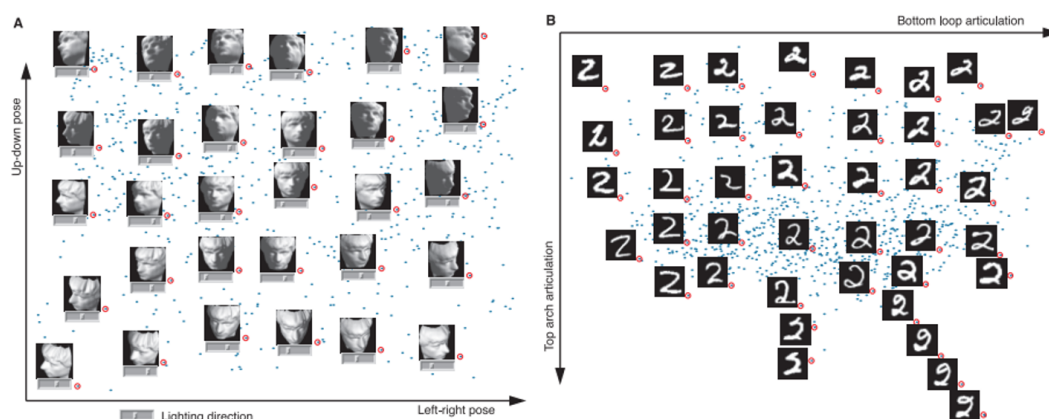


图 3.2 石膏像和手写数字利用isomap降维

流形学习算法的用途主要有两种，第一种将其直接视为一种降维算法，利用降维去除其在原数据空间中的冗余，在低维空间中进行分类聚类；第二种将在原来欧式空间中的算法加以改造，间接的对其流形结构加以利用。典型的流形学习方法包括LLE、Isomap、Laplacian Eigenmaps等，下面分别介绍。

3.2.3 LLE

LLE^[62]是一种无监督的降维方法，其核心思想是将高维流形上的紧邻点映射到低维空间中的近邻点，并保持其局部几何特征，从而达到降维的目的。不同于其他流形学习方法，LLE算法的前提假设是每一个点都是局部线性的，即每一个样本点都可以用它的近邻点线性组合来表达。

LLE算法的步骤如图 3.3所示。归纳起来就是“找近邻、分解、重建”：

- 寻找每个样本点的 k 个近邻点
- 通过 k 个样本点，重建该样本点的局部权值矩阵
- 通过局部权值矩阵，将该样本点映射到低维空间，得到输出值

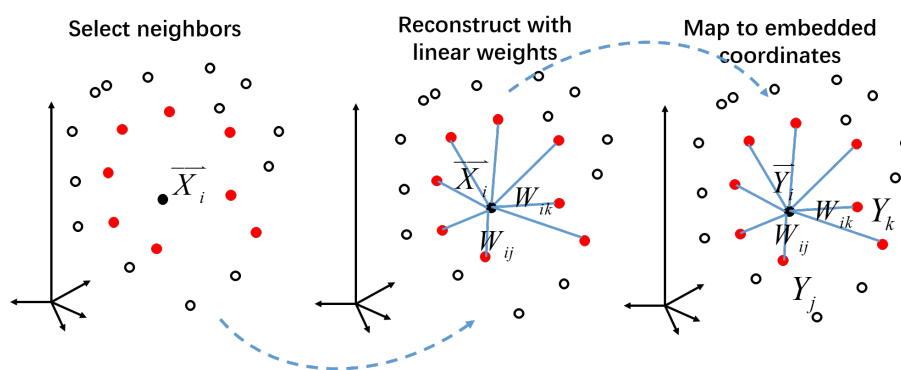


图 3.3 LLE算法具体步骤

3.2.4 ISOMAP

ISOMAP算法^[63]通过构造数据点之间的距离矩阵 A ，然后通过MDS进行降维。核心步骤如下：

- 找出每个样本点的近邻点，并将其彼此之间的距离填入 A ，非近邻之间的距离初始化为正无穷
- 根据最短路径算法（Dijkstra算法），求出任意两点之间的距离
- 将矩阵 A 作为MDS的输入，求出降维后的输出。

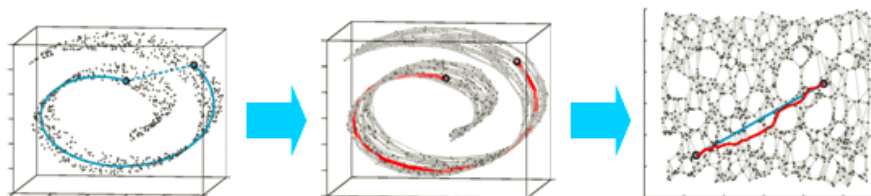


图 3.4 ISOMAP算法具体步骤

3.3 基于有监督LLE算法的能源互联网故障诊断

在参考文献中，已经有文献把流型学习方法应用于故障诊断。文献^[64]提出了基于扩展lle方法的非线性系统故障诊断方法，它的特点是将故障状态与高维空间分布结合起来，通过确定数据点在空间超球内的分布完成故障的检测，在这个过程中将超球的确定与lle算法中基于核函数的样本外数据扩展相结合，减少了计算量，提高了算法的实时性。文献^[65]针对改进的局部线性映射算法解决了非线性数据的特征映射问题。首先，通过线性拟合改进了基于分型锥估计的内在维数问题。然后，将故障状态与空间分布结合起来，通过确定数据点在空间超球内的分布完成故障检测。文献^[66]通过考虑故障类内和类间的离散度，提出一种基于最佳分类

效果的 k 和 d 综合参数选择方法。目前在电网和能源互联网领域，还没有应用流形学习算法进行故障诊断的额研究。

前面已经提到，LLE算法是一种典型的无监督方法，因此如何把已知的故障标签加入到训练过程中，使得算法最终的输出是诊断的故障类型而不是降维后的数据。目的是使用已知的标签或类别信息引导降维，使得降维后的数据最能体现和目标问题最相关的特征。我们的思路是把线性判别中的Fisher准则引入其中，使得特征提取和分类是融合在一起的，希望在选取和创造特征的过程中，有效利用类别信息。本章通过调整LLE中的两个重要参数：近邻数 k 和内嵌维度 d 来使得在降维后最小化类内间距，最大化类间间距。最终通过样本的低维空间中的分布来判断故障类型。

下面介绍具体的算法流程，如图 3.5。其中“步骤3：确定近邻数 k 和内嵌维度 d ”是核心步骤，我们在3.4节会将其展开，具体阐述。

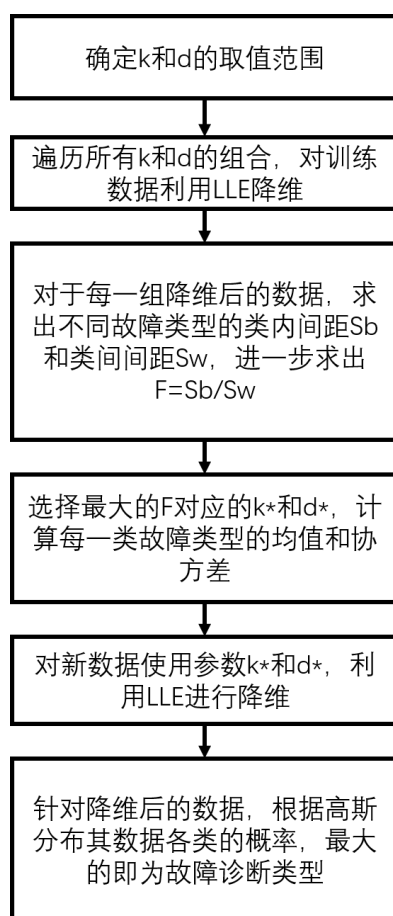


图 3.5 基于有监督LLE算法的能源互联网故障诊断流程图

步骤1：收集训练样本数据

收集样本后，对特征进行筛选和处理，特征可以是电压、电流、有功、无功

及其衍生量。要求每个训练样本标记有故障类型。故障类型可以限定为有限的几种，比如断路故障包括单项断路、两相断路和三相断路等。

步骤2：利用LLE算法对所有训练进行降维

步骤2.1 找出每个样本的 k 个近邻点。把每个样本点欧式距离最近的 k 个样本点找出， k 即所谓近邻数。

步骤2.2

计算出样本点的局部重建权值矩阵 W ，首先定义重构误差：

$$\varepsilon(W) = \sum_i \|\vec{X}_i - \sum_j W_{ij} \vec{X}_j\|_2^2 \quad (3-1)$$

于是，目标函数

$$\min \varepsilon(W) = \sum_i \|\vec{X}_i - \sum_j W_{ij} \vec{X}_j\|_2^2 \quad (3-2)$$

其中

$$\sum_j W_{ij} = W_i^T \mathbf{1}_k = 1 \quad (3-3)$$

得到

$$\begin{aligned} \varepsilon(W) &= \sum_i \|x_i - \sum_j W_{ij} x_j\|_2^2 \\ &= \sum_i \left\| \sum_j W_{ij} x_i - \sum_j W_{ij} x_j \right\|_2^2 \\ &= \sum_i \left\| \sum_j W_{ij} (x_i - x_j) \right\|_2^2 \\ &= \sum_i \|(x_i - x_j) W_i\|_2^2 \\ &= \sum_i W_i^T (x_i - x_j)^T (x_i - x_j) W_i \end{aligned}$$

其中 $W_i = (w_{i1}, w_{i2}, \dots, w_{ik})$

令矩阵 $C_i = (x_i - x_j)^T (x_i - x_j)$ ，则上式进一步转化为 $\varepsilon(W) = \sum_i W_i^T C_i W_i$

利用拉格朗日法将优化目标转化为

$$\varepsilon(W) = \sum_i W_i^T C_i W_i + \lambda(W_i^T \mathbf{1}_k - 1) \quad (3-4)$$

对 W 求导并令其值为0, 得到

$$2C_i W_i + \lambda \mathbf{1}_k = 0 \quad (3-5)$$

即 $W_i = \lambda' C_i^{-1} \mathbf{1}_k$, 其中 $\lambda' = -\frac{1}{2}$ 为一个常数。利用已知约束 $W_i^T \mathbf{1}_k = 1$, 对 W_i 归一化, 最终的权重系数为

$$W_i = \frac{C_i^{-1} \mathbf{1}_k}{\mathbf{1}_k^T C_i^{-1} \mathbf{1}_k} \quad (3-6)$$

步骤2.3 将所有样本点映射到低维空间中, 映射条件满足:

$$\min \phi(Y) = \sum_i \|y_i - \sum_j w_{ij} y_j\|_2^2 \quad (3-7)$$

约束条件为

$$\sum_i y_i = 0; \frac{1}{m} \sum_i y_i y_i^T = I \quad (3-8)$$

将目标损失函数矩阵化:

$$\begin{aligned} \phi(Y) &= \sum_i \|y_i - \sum_j w_{ij} y_j\|_2^2 \\ &= \sum_i \|Y I_i - Y W_i\|_2^2 \\ &= \text{tr}(Y^T (I - W)^T (I - W) Y) \end{aligned}$$

其中 tr 为迹函数。令 $M = (I - W)^T (I - W)$, 则目标优化函数变为

$$\phi(Y) = \text{tr}(Y^T M Y) \quad (3-9)$$

通过拉格朗日函数来加入优化条件 $Y^T Y = mI$ ，得到

$$\phi(Y) = \text{tr}(Y^T M Y) + \lambda(Y^T Y - mI) \quad (3-10)$$

对 Y 求导并令其为0，得到 $MY = \lambda Y$ ，是标准的特征分解问题，则降维后的特征矩阵 Y 为 M 的最小的 m 个非零特征值相对应的特征向量。在处理过程中，由于 M 的最小特征值为0不能反映数据特征，此时对应的特征向量为全1。因此通常取第2到 $m + 1$ 间的特征值所对应的特征向量组成列向量 $M = (y_2, y_3, \dots, y_{d+1})$ ，作为输出结果，即一个 $N * m$ 的数据表达矩阵 Y 。其中 N 为样本个数。

表 3.3 LLE算法进行数据降维

算法3: 利用LLE算法进行数据降维
1: input: 样本集 $X = x_1, x_2, x_3, \dots, x_m$ ，近邻数 $\forall k$ ，内嵌维度 d （降维后的维数）
2: output: 降维后的样本矩阵 Y
3: for i from 1 to m ，按照欧氏距离作为度量，计算并找出 x_i 最近的 k 个近邻 $(x_{i1}, x_{i2}, \dots, x_{ik})$
4: for i from 1 to m ，求出局部协方差矩阵 $C_i = (x_i - x_j)^T (x_i - x_j)$ ，求出对应的权重系数向量
$W_i = \frac{C_i^{-1} \mathbf{1}_k}{\mathbf{1}_k^T C_i^{-1} \mathbf{1}_k}$
5: 由权重系数向量 W_i 组成权重系数矩阵 W ，计算矩阵 $M = (I - W)^T (I - W)$
6: 计算矩阵 M 的前 $d + 1$ 个特征值，并计算这 $d + 1$ 个特征值对应的特征向量 y_1, y_2, \dots, y_{d+1}
7: 由第2个特征向量到第 $d + 1$ 个特征向量所组成的矩阵即为输出低维样本集矩阵 $Y = (y_2, y_3, \dots, y_{d+1})$

步骤3: 确定近邻数 k 和内嵌维度 d （详见3.4节）

步骤4: 在降维后的空间统计各类样本的均值和协方差矩阵 利用步骤3中确定的最优的 k 和 d ，利用LLE方法进行降维。假设降维后的数据为

$$Y = y_1, y_2, \dots, y_N \quad (3-11)$$

其中 Y 是一个 $N \times m$ 的矩阵， N 是样本个数， m 是内嵌维度。每一类故障类型的均值

为

$$c_i = \frac{1}{N_i} \sum_{y_j \in \phi_i} y_j, i = 1, 2, \dots, s \quad (3-12)$$

协方差矩阵为

$$\Sigma_i = E[(y_j - c_i)(y_j - c_i)^T | (y_j \in \phi_i)] \quad (3-13)$$

因此，每一个故障类型，都有其对应的均值向量和协方差矩阵

步骤5：对新样本（未知故障类型的样本）进行故障诊断 当有新样本首先利用步骤3中确定的最优的 k 和 d ，利用LLE方法进行降维，降维后的数据为 x ， x 为一个 d 维的向量。根据步骤4中得出的均值和协方差矩阵，假设样本服从高斯分布，求出属于各个类别的概率值，概率值最大的类别即为诊断的故障类别。

$$\max N(x|c_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x - c_i)^T \Sigma^{-1}(x - c_i)\right] \quad (3-14)$$

3.4 参数选择：近邻数 k 和内嵌维度 d

在上小节已经提到，LLE是一种典型的无监督降维算法，因此如何将已知的故障类型加入其中，使其变成一种有监督的方法是本章的重点。我们的思路是通过LLE中两个重要参数：近邻数 k 和内嵌维度 d 的选择，来最大化不同故障类型数据的类间间距、最小化不同故障类型数据间的类内间距。使得通过LLE降维后的在低维空间中的数据具有非常好的可分性。降维后的数据我们不会再训练分类器，而是通过不同故障类型数据的样本分布来直接确定新数据的故障类型。

近邻数 k 是LLE中的第一个重要参数。LLE算法前提假设是每一个点都是局部线性的，即每一个样本点都可以用它的近邻点线性组合来表达，在高维向低维映射的过程中，保持了样本数据之间的近邻关系。 k 的取值过大使得局部线性的范围过大，会影响整个流形的平滑性，无法很好的体现其局部特征。而当 k 的取值过小，LLE就很难保证样本的低维空间中的拓扑结构，把连续的流形变为脱节的子流形。

另外根据3.3节中LLE算法的具体步骤可知，算法的复杂度为 $O(kN)$ ，其中 k 为近邻数， N 为样本个数，因此整个算法的时间复杂度和 k 是呈正相关的，因此 k 的

表 3.4 对新样本进行故障诊断

算法4: 对新样本进行故障诊断

- 1: input: 最优近邻数 k^* 和内嵌维度 d^* , 训练集数据 X , 预测集数据
- 2: output: 预测集故障类型
- 3: 训练集数据利用LLE算法进行降维得到 $Y = y_1, y_2, \dots, y_N$
- 4: for i from 1 to n
- 5:

$$\Sigma_i = E[(y_j - c_i)(y_j - c_i)^T | (y_j \in \phi_i), c_i = \frac{1}{N_i} \sum_{y_j \in \phi_i} y_j, i = 1, 2, \dots, s \quad (3-15)$$

- 6: 预测数据利用LLE进行降维, 得到 x
- 7: for i from 1 to n
- 8: 利用高斯分布求 x 属于各个类别的概率

$$N(x|c_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x - c_i)^T \Sigma^{-1}(x - c_i)\right] \quad (3-16)$$

- 10: 概率最大类别即对应故障诊断类别

趋势如果过大, 不但破坏真实的局部流形结构特征, 还会额外增大计算量, 浪费是时间和效率。

输出数据的维度 d (即内嵌维度) 是第二个重要的参数, 它决定了利用了LLE算法降维嵌入后, 低维空间中的流形局部结构能否充分描述高维空间中的局部集合特征。即在保持原始高维空间中的数据信息损失最小的情况下, 寻找原始高维空间中数据的内在规律与本质特征, 进而寻求一个降维后输出的最低维 d 对原始高维空间中的数据进行合理的、有效的低维可视表示。最低维数 d 的取值过大将会使降维后的数据中含有过多的冗余, 反之如果 d 的取值过小, 使得在高维空间中彼此分开的点在低维空间中交叠。

算法的具体步骤如下 (步骤3.1 步骤3.5), 核心思想是通过遍历不同的 k 和 d 的组合, 最大化 $F = \frac{S_b}{S_w}$ 。

步骤3.1

根据经验, 选择 k 和 d 的取值范围

步骤3.2

在 k 和 d 的参数范围中, 分别选择一个, 组成参数组合, 带入到步骤2中利

用LLE算法进行降维，得到降维后的数据集 Y

步骤3.3

并且利用降维后的数据，利用附录中的公式计算评价指标Fisher准则
假设降维后的数据为

$$Y = y_1, y_2, \dots, y_N \quad (3-17)$$

其中 Y 是一个 $N \times m$ 的矩阵， N 是样本个数， m 是内嵌维度

有 s 个不同类别的故障，分别为 $\phi_1, \phi_2, \dots, \phi_s$ ，在每一个类别中，均值向量为

$$c_i = \frac{1}{N_i} \sum_{y_j \in \phi_i} y_j, i = 1, 2, \dots, s \quad (3-18)$$

定义所有类别的类内离散度矩阵

$$S_i = \sum_{y_j \in \phi_i} (y_j - c_i)(y_j - c_i)^T, i = 1, 2, \dots, s \quad (3-19)$$

因此，混合类内离散度矩阵即为所有类内离散度矩阵的求和

$$S_w = S_1 + S_2 + \dots + S_s \quad (3-20)$$

类间离散度矩阵为

$$S_b = \sum_{i=1}^s (c_i - c)(c_i - c)^T \quad (3-21)$$

改进的Fisher准则为

$$F = \frac{S_b}{S_w} \quad (3-22)$$

步骤3.4

选择另一组参数组合，重复步骤3.2和步骤3.3中，得到所有参数组合的 F 。

步骤3.5

选择最大的 F 所对应的参数组合 k 和 d ，在这组参数下，在新的降维空间中，类内间距最小，类间间距最大。

最优近邻数 k 和内嵌维度 d 的选择算法具体流程见表 3.5

表 3.5 最优近邻数k和内嵌维度d的选择

算法5: 最优近邻数k和内嵌维度d的选择

```

1: input: k取值范围 $k_1$   $k_m$ , d的趋势访问 $d_1$   $d_n$ 
2: output: 最优近邻数 $k^*$ 和内嵌维度 $d^*$ 
3: initialize:  $i = 1, j = 1, F = 0$ 
4: repeat:
5:     repeat:
6:         利用 $k_i$  和 $d_j$ 对数据进行降维, 降维后数据为 $Y = y_1, y_2, \dots, y_N$ 
7:         计算类内离散度矩阵 $S_w$  和类间离散度矩阵 $S_b$ 
8:          $F_{temp} = \frac{S_b}{S_w}$ 
9:         if  $F_{temp} > F$ 
10:             $F = F_{temp}$ 
11:             $k^* = k_i, d^* = d_j$ 
12:             $j = j + 1$ 
13:         until  $j > n$ 
14:          $i = i + 1$ 
15: until  $i > m$ 

```

3.5 仿真样本

本章的测试系统的IEEE 39-bus 系统。通过固定系统故障前的运行方式, 进行暂态仿真, 获得暂态稳定数据如表 3.8和3.9。使用的仿真工具是Matlab中的psat工具包, 它的优点是可以通过循环的方式改变故障的方式, 提高样本生成的速度。

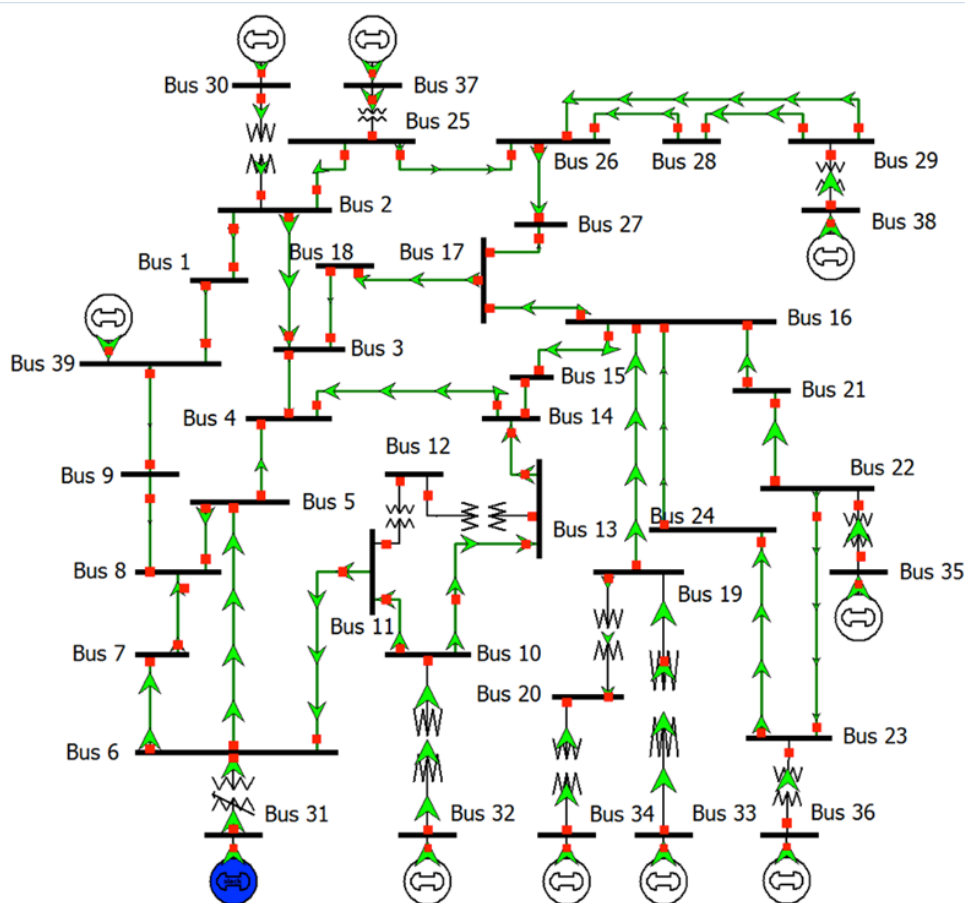


图 3.6 IEEE-39 bus系统

(1)故障类型的设置

故障诊断是本章的目标，因此必须在样本仿真过程中，必须通过设置不同的故障类型，从而获得不同故障类型下的样本。本章主要通过设置线路和不同位置的三相短路故障，同时组合不同故障切除时间。故障位置距离起点线路的全场分别为0%，20%，40%，60%，80%，共5种；故障切除时间分别为近端0.05s，远端0.1s和近端0.35s，远端0.4s，共2种。最终组合出10中故障类型如表 3.6。

(2)运行条件的设置

针对每一个故障类型都需要有足够的训练样本。本章是通过设置总负荷的比例和其中感应电动机动态负荷的占比实现的，因为这两个变量是可以连续变化的。最终每一个故障类型下，都仿真出500组样本，10个故障类型，共计5000组样本。

(3)备选特征

仿真软件输出的数据包括故障后10s的电压、电流、有功、无功等数据。IEEE 39-bus系统共有39条母线、46条线路，10个发电机。因此我们选取母线电压幅值（39个），母线电压相角（39个），线路无功功率（46个），线路有功功率（46个），

表 3.6 故障类型设置

故障标号	故障位置	故障切除时间
1	起点距离线路全长0%	近端0.05s, 远端0.1s
2	起点距离线路全长20%	近端0.05s, 远端0.1s
3	起点距离线路全长40%	近端0.05s, 远端0.1s
4	起点距离线路全长60%	近端0.05s, 远端0.1s
5	起点距离线路全长80%	近端0.05s, 远端0.1s
6	起点距离线路全长0%	近端0.35s, 远端0.4s
7	起点距离线路全长20%	近端0.35s, 远端0.4s
8	起点距离线路全长40%	近端0.35s, 远端0.4s
9	起点距离线路全长60%	近端0.35s, 远端0.4s
10	起点距离线路全长80%	近端0.35s, 远端0.4s

发电机功角（10个），发电机转速（10个），发电机有功输出（10个），发电机无功输出（10个），共计210个特征作为LLE算法降维的备选特征。

3.6 故障诊断结果与分析

本章的目的是利用有监督的降维方法实现能源互联网的故障诊断。通过对上一小节中产生的样本进行训练，证实了有监督的LLE算法对于可以将高维数据中的关键特征提取出来，使得整个样本集在低维空间中有非常好的可分性。下面就对算法中的几个重要参数分别讨论实验结果。

(1)故障类型个数

本章仿真样本的故障类型共有10类，是通过组合5中故障位置和2种故障切除时间获得的。分别其中的2、3、5、8类故障，固定近邻数 $k=10$ 、 $d=2$ ，降维后不同故障类型在空间中的分布如图 3.7。

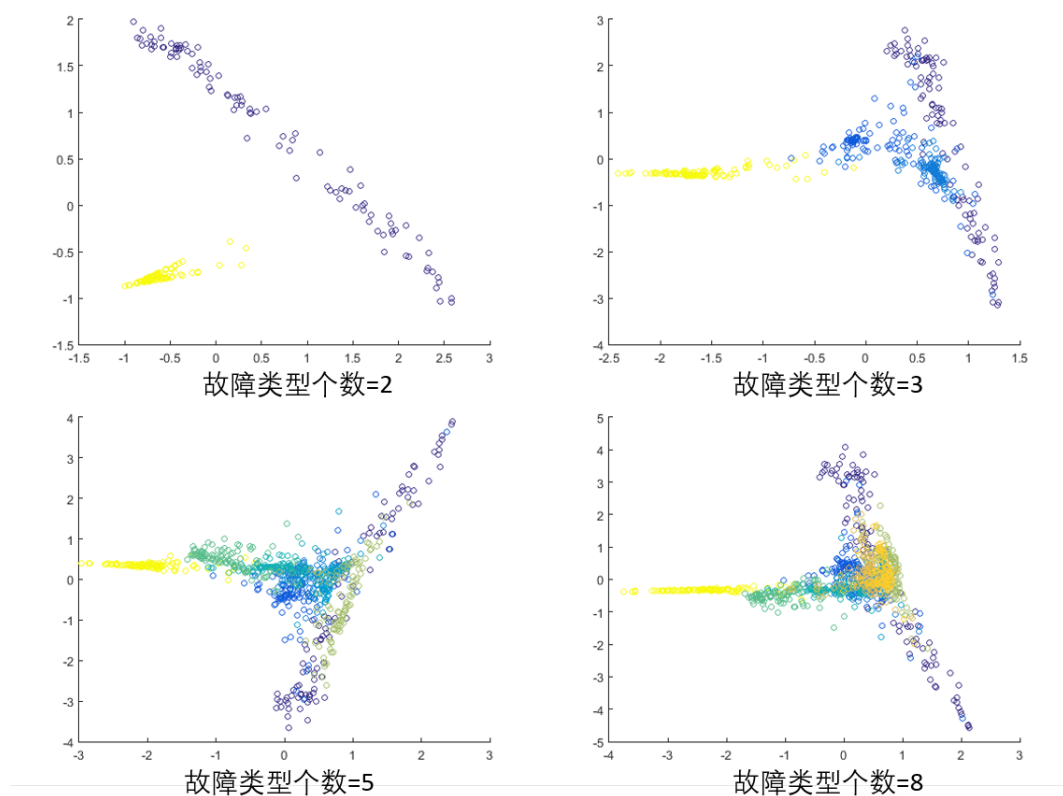


图 3.7 故障类型个数变化时降维效果

从图中可以发现，当故障类型只有2类时，两类故障显示出了非常好的可分性（黄色和紫色）；当故障类型是3类时，蓝色和紫色代表的两类故障已经有所重叠，但基本可分；当故障类型是5类和8类时，位于中间的几种故障类型完全重叠在一起，基本不可分。

(2) 近邻数 k

在4.4节中已经讨论过，近邻数 k 是LLE中的一个重要参数。LLE前提假设是每一个点都是局部线性的，即每一个样本点都可以用它的近邻点线性组合来表达，在高维向低维映射的过程中，保持了样本数据之间的近邻关系。 k 的取值过大使得局部线性的范围过大，会影响整个流形的平滑性，无法很好的体现其局部特征。而当 k 的取值过小，会把连续的流形变为脱节的子流形。

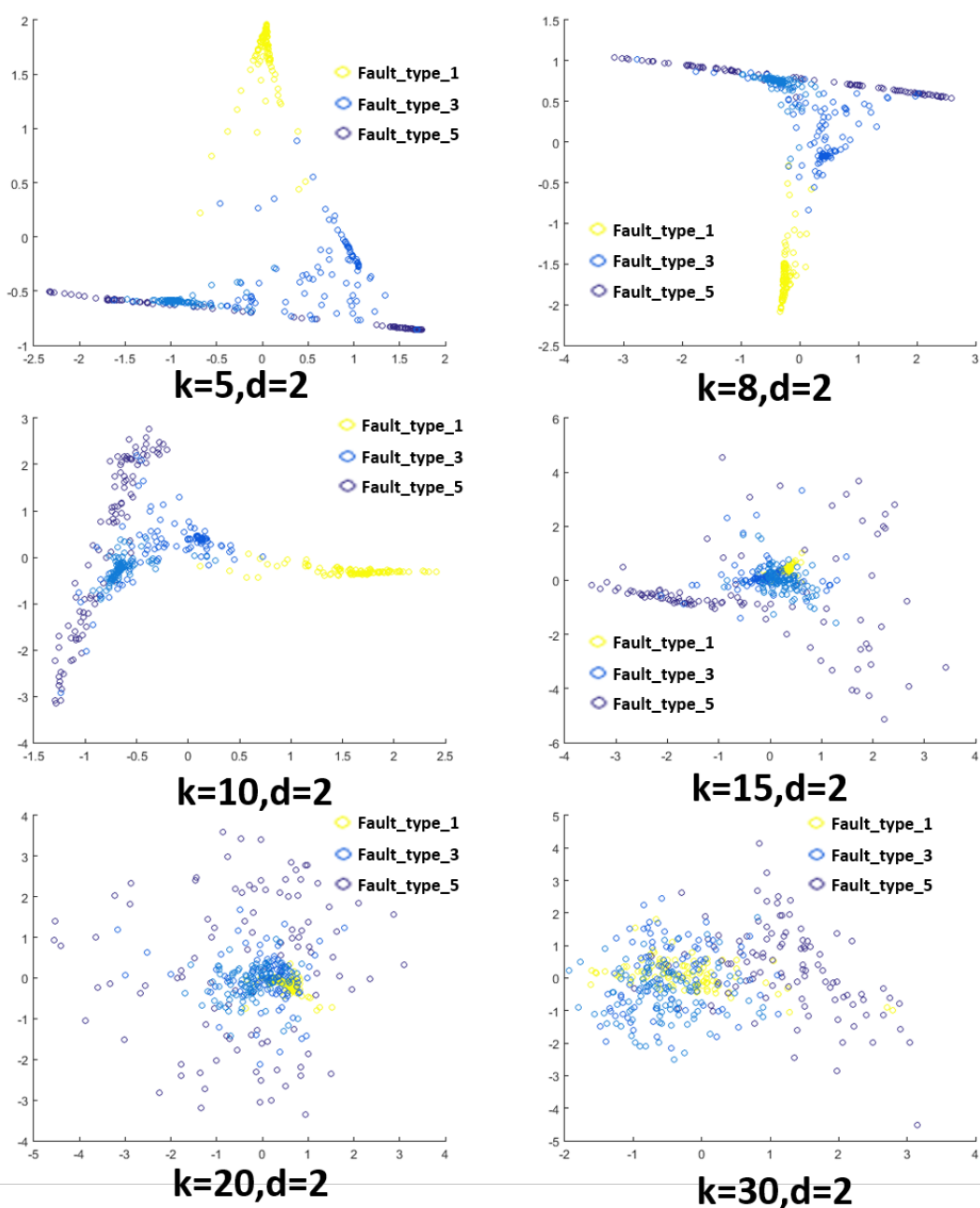
图 3.8 近邻数 k 变化时故障类型降维效果

图 3.8是当选取故障类型为3类时（分别对应表 3.6中的类型1、类型3和类型5），固定内嵌维度 $d = 2$ ，改变不同近邻数 k 时的降维效果。从图中可以看出， $k = 5$ 时，类型3和类型5有一段交叠比较严重； $k = 8$ 和 $k = 10$ 时，三种故障类型分布分散，可分性较好； $k = 20$ 时，类型5在空间中基本是均匀分布，类型1和类型3也基本交叠在一起； $k = 30$ 时，类型3和类型5基本都是均匀分布。因此，合理的近邻数 k 不宜过小，也不宜过大。

(3)内嵌维度 d

内嵌维度 d 决定了利用了LLE算法降维嵌入后，低维空间中的流形局部结构能否充分描述高维空间中的局部集合特征。即在保持原始高维空间中的数据信息损失最小的情况下，寻找原始高维空间中数据的内在规律与本质特征，进而寻求一个降维后输出的最低维 d 对原始高维空间中的数据进行合理的、有效的低维可视化表示。最低维数 d 的取值过大将会使降维结果中含有过多的噪声， d 的取值过小，致使本来不同的点在低维空间可能会彼此交叠。

由于当 $d > 3$ 时，降维后的图形无法可视化，因此前面的实验结果中都选择的是 $d = 2$ 。因此在这里我们无法通过图形化的方式展示不同的 d 的取值对应的降维后的效果图。我们通过表 3.7来表示不同 d 的取值对应的类间间距 S_b ，类内间距 S_w 和对应的F值。从表中可以看出，随着内嵌维度 d 的上升，F值先上升后下降，在 $d = 20$ 时达到最大值。也就是说当 $d = 20$ ，类间间距相对于类内间距达到最大，F值最大。因此，当我们这个样本集下，用20个左右的特征来重构样本空间是最能体现于故障类型相关特性的。此时，可以去掉样本集中大量的和故障诊断分类不相关的特征，同时样本在低维空间中的可分性最强。

表 3.7 不同内嵌维度 d 对应的F值

内嵌维度 d 取值	2	5	10	15	20	30	40	50	100
F值	6.6	10.1	12.2	16.4	22.3	18.8	9.3	7.5	6.3

3.7 本章小结

本章通过流形学习对能源互联网数据进行降维，同时实现故障诊断。LLE算法是一种典型的无监督学习方法，我们通过遍历近邻数 k 和内嵌维度 d ，使得最大化不同故障类别数据的类间间距，最小化类内间距。这样通过已知的类别和标签，引导降维后的特征选取，实现了有监督的LLE。与传统的基于数据电网故障诊断方法相比，基于流形学习方法存储的参数类型更少，预测速率更快，适合在线预测。同时，降维后的数据，可以直接进行可视化（2维或者3维），更加直观理解网络态势。

表 3.8 IEEE39节点系统算例潮流数据——母线数据

母线号	电压幅值(p.u.)	电压相角(rad)	有功输入(p.u.)	无功输入(p.u.)	有功输出(p.u.)	无功输出(p.u.)
BUS-1	1.0475	-0.16704	0	0	0	0
BUS-10	1.017	-0.11138	0	0	0	0
BUS-11	1.0125	-0.12558	0	0	0	0
BUS-12	1	-0.12585	0	0	0.085	0.88
BUS-13	1.0142	-0.12385	0	0	0	0
BUS-14	1.0117	-0.15297	0	0	0	0
BUS-15	1.0158	-0.16023	0	0	3.2	1.53
BUS-16	1.0323	-0.13573	0	0	3.29	0.323
BUS-17	1.034	-0.15315	0	0	0	0
BUS-18	1.0313	-0.16782	0	0	1.58	0.3
BUS-19	1.05	-0.05502	0	0	0	0
BUS-2	1.049	-0.12237	0	0	0	0
BUS-20	0.99096	-0.07966	0	0	6.8	1.03
BUS-21	1.0321	-0.09374	0	0	2.74	1.15
BUS-22	1.05	-0.01613	0	0	0	0
BUS-23	1.045	-0.01958	0	0	2.475	0.846
BUS-24	1.0378	-0.13364	0	0	3.086	-0.922
BUS-25	1.0575	-0.0986	0	0	2.24	0.472
BUS-26	1.0522	-0.12053	0	0	1.39	0.17
BUS-27	1.038	-0.15564	0	0	2.81	0.755
BUS-28	1.0502	-0.05924	0	0	2.06	0.276
BUS-29	1.05	-0.01108	0	0	2.835	0.269
BUS-3	1.0304	-0.17206	0	0	3.22	0.024
BUS-30	1.0475	-0.08014	2.5	1.4492	0	0
BUS-31	0.982	0	5.7284	2.0704	0.092	0.046
BUS-32	0.9831	0.0282	6.5	2.0573	0	0
BUS-33	0.9972	0.03604	6.32	1.0894	0	0
BUS-34	1.0123	0.01093	5.08	1.6699	0	0
BUS-35	1.0493	0.07046	6.5	2.1111	0	0
BUS-36	1.0635	0.11745	5.6	1.0044	0	0
BUS-37	1.0278	0.01982	5.4	0.00647	0	0
BUS-38	1.0265	0.11219	8.3	0.22659	0	0
BUS-39	1.03	-0.19388	10	0.87883	11.04	2.5
BUS-4	1.0038	-0.18596	0	0	5	1.84
BUS-5	1.005	-0.16526	0	0	0	0
BUS-6	1.0074	-0.15301	0	0	0	0
BUS-7	0.9967	-0.19147	0	0	2.338	0.84

表 3.9 IEEE39节点系统算例潮流数据——线路数据

线路始端	线路终端	线路编号	线路有功(p.u.)	线路无功(p.u.)	有功损耗(p.u.)	无功损耗(p.u.)
BUS-1	BUS-2	1	-1.1862	-0.29167	0.00451	-0.71474
BUS-1	BUS-39	2	1.1862	0.29167	0.00173	-0.76602
BUS-2	BUS-3	3	3.6468	0.92042	0.01705	-0.08006
BUS-2	BUS-25	4	-2.3375	0.81409	0.03985	-0.113
BUS-3	BUS-4	5	0.75112	1.1288	0.00259	-0.1866
BUS-3	BUS-18	6	-0.34138	-0.15228	0.00012	-0.22572
BUS-4	BUS-5	7	-1.6297	-0.04387	0.00211	-0.10164
BUS-4	BUS-14	8	-2.6218	-0.48077	0.00559	-0.0502
BUS-5	BUS-8	9	3.1773	0.58727	0.00834	-0.03092
BUS-6	BUS-5	10	4.8137	0.54575	0.00463	0.01626
BUS-6	BUS-7	11	4.2592	0.91363	0.01128	0.05955
BUS-6	BUS-11	12	-3.4366	-0.36485	0.00821	-0.04554
BUS-7	BUS-8	13	1.91	0.01408	0.00147	-0.06051
BUS-8	BUS-9	14	-0.14258	-1.0672	0.00184	-0.3606
BUS-9	BUS-39	15	0.14441	-0.70663	2.00E-05	-1.2701
BUS-10	BUS-11	16	3.4717	0.72642	0.00489	-0.02254
BUS-10	BUS-13	17	3.0283	0.369	0.00361	-0.03638
BUS-12	BUS-11	18	-0.02174	-0.42179	0.00029	0.00785
BUS-12	BUS-13	19	-0.06326	-0.45821	0.00035	0.00942
BUS-13	BUS-14	20	2.961	-0.06224	0.00767	-0.0907
BUS-14	BUS-15	21	0.32601	-0.40211	0.00027	-0.37291
BUS-15	BUS-16	22	-2.8743	-1.5592	0.00909	-0.08435
BUS-16	BUS-17	23	2.0624	-0.41274	0.00287	-0.10673
BUS-16	BUS-21	24	-3.296	0.14101	0.00821	-0.13285
BUS-16	BUS-24	25	-0.42679	-0.97549	0.0003	-0.06695
BUS-17	BUS-18	26	1.9239	0.11457	0.00245	-0.11199
BUS-17	BUS-27	27	0.13556	-0.42058	0.0001	-0.34384
BUS-21	BUS-22	28	-6.0442	-0.87614	0.02785	0.20932
BUS-22	BUS-23	29	0.42795	0.41904	0.00025	-0.19859
BUS-23	BUS-24	30	3.5384	-0.00243	0.0253	0.01103
BUS-25	BUS-26	31	0.76611	-0.17884	0.00171	-0.55352
BUS-26	BUS-27	32	2.6845	0.67439	0.00994	-0.15735
BUS-26	BUS-28	33	-1.4082	-0.21612	0.00788	-0.77517
BUS-26	BUS-29	34	-1.9018	-0.25359	0.01914	-0.92693
BUS-28	BUS-29	35	-3.4761	0.28305	0.01556	-0.10669
BUS-29	BUS-38	36	-8.2477	0.79408	0.05234	1.0207
BUS-6	BUS-31	37	-5.6364	-1.0945	0	0.92983

第4章 能源互联网无功补偿

4.1 本章引言

在前面几章中，我们分别实现了态势感知中的评估和诊断环节。当暂态电压的评估和预测结果已经为不稳定时，并且通过故障诊断定位到了故障位置，接下来就要采取紧急控制措施，包括切负荷、切设备、补偿无功功率等。目前紧急控制的标准做法是，离线对典型故障进行仿真，同时搜索响应的控制策略，供调度员选择。由于仿真速度的限制，很难实时针对当前系统的紧急情况给出有针对性的紧急控制措施。

本章在第2章和第3章的基础上，进一步实现基于数据的能源互联网无功补偿方案优化设计。通过分别训练评估网络和策略网络，分别实现无功补偿后网络状态的评估和最优无功补偿方案的搜索。策略网络首先通过对系统设置大量典型故障，然后对每一种故障组合不同的无功补偿方案进行仿真，对补偿后的网络状态进行评估。然后把典型故障结合无功补偿方案作为输入，补偿后电网状态作为输出，训练分类器。当有新的故障时，组合各种无功补偿方案输入到分类器中，最终选择评分最高的方案。

本章的具体安排如下：4.2节主要介绍无功功率、无功补偿的意义和传统的无功工厂策略；4.3节介绍基于大数据的无功补偿优化设计模型；4.4节是澳门电网的整体情况和样本的生成过程；4.5节是几个典型的算例分析，验证算法的效果；4.6节开发了面向BPA的数据处理工具包，提高样本生成和数据处理的速度；4.7节是本章小结。

4.2 无功补偿

真实电网中输出的功率主要包括两部分：有功功率和无功功率。有功功率通过直接消耗电能将其转化为其他形式的能量，诸如机械能、热能、化学能等。无功功率^[67-70]消耗电能将其转换为另一种形式的能量，这种能量的在网络中与电网进行周期性的转换，并且是电气设备作用的前提条件。

一般来讲，无功功率的损失对电网的影响主要体现在以下几个方面：

1)增加电压的损耗，引起电网中电压的波动和闪变。有很多工业设备对电压的波动非常敏感，电子设备会发生误动作，同时也使得用户获得的实际功率大大减少

2)增加输电的损耗。目前电力系统中，无功功率的损耗大约占到总发电容量的20%-30%，大大增加了发电机组和输配电线路的容量，造成严重的浪费。

通过并联无功补偿设备是目前解决无功消耗的最优方案。然而，目前非线性规划和线性规划法是求解无功规划优化的最常用方法，这两种方法对电网系统进行建模，建立目标函数和约束函数。这两种方法或多或少都具有计算量大、收敛性差、稳定性不好的特点。因此，在现实应用中，主要依靠调度员的经验进行补偿，如集中补偿（整个系统的统一补偿）和分散补偿（用电负荷旁就地补偿），缺点是比较粗略，无法获得最优效果。

静止同步补偿器（Static Var Generator, SVG）是当今无功补偿的最常用的设备，它的基本原理为：将基于可关断电力电子器件（如IGBT，GTO等）的电压源型逆变器(Voltage Sourced Converter, 简称VSC)并联在电网上，通过调节逆变器交流侧输出电压的幅值和相位（如图4.1），迅速吸收或者发出所需要的无功功率，即可实现快速动态调节无功的目的。因此，SVG相当于一台静止的同步调相机。

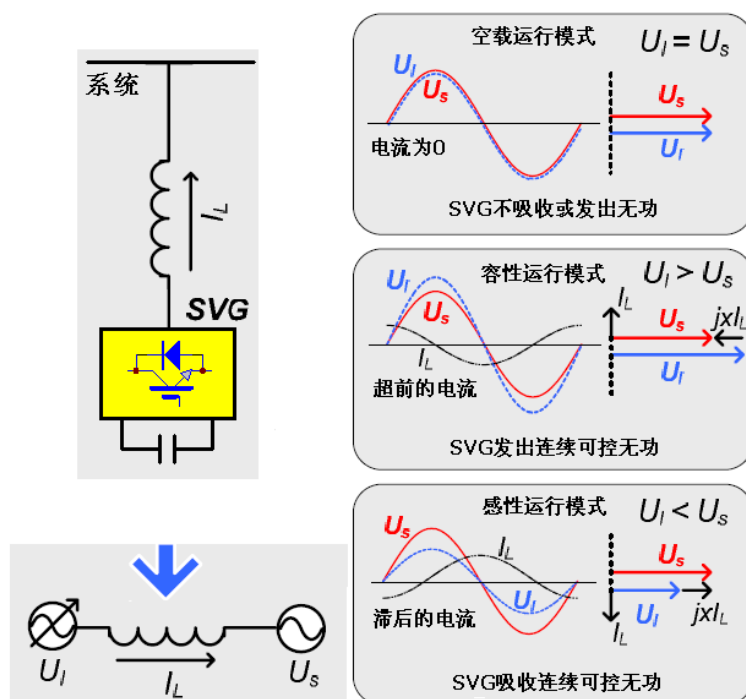


图 4.1 SVG基本原理

SVG相对于传统的固定电容器补偿、机械开关投切电容器、晶闸管投切电容器补偿(TCR)方式，具有以下优势：

1) 动态连续平滑补偿，更高的响应速度。SVG可跟随负载变化，动态连续补偿功率因子，可以发无功，也可吸收无功，避免了传统无源补偿时的无功倒送情况。

2) 电流源特性, 输出无功电流不受母线电压影响, 而传统SVC具有阻抗型特性, 输出电流随母线电压线性降低。SVG用于电压控制时具备很大的优势, 系统电压越低, 越需要动态无功支撑电压, SVG输出无功电流与系统电压没有关系, 可改善系统瞬时稳定性。而SVC输出无功电流的能力随系统电压的降低而下降。

3) 采用PWM控制, 不仅不产生谐波, 而且能在补偿无功功率的同时动态补偿谐波。

4) 具备抗谐波功能, 更保障系统安全

SVG为有源型补偿装置, 相当于一个可控电流源, 只补偿基波无功电流, 系统谐波电流不会造成补偿设备损坏; 同时避免电容器组可能造成的谐波放大, 防止系统其他设备及补偿设备因谐波过电压而损坏。

4.3 基于大数据无功补偿优化设计

4.3.1 从AlphaGo获得的启示

AlphaGo是由Google开发的一款围棋人工智能程序, 2016年3月, 该程序与围棋世界冠军、职业九段选手李世石进行人机大战, 并以4:1的总比分获胜, 由此获得大家的关注。下围棋和能源互联网故障后的决策支持有一定的相似之处, 都是需要评估当前的局势, 可供选择的决策方案是有限的。两者的最终的目标就是在评估当前的局势下, 寻找决策空间中的最优方案。另外, 每一步决策做出后, 系统的状态都会发生改变, 在做出一下一次决策前, 需要对当前整个系统的局势重新做出评估。因此, 每一次决策做出后, 都要重新评估系统的局势和状态并且对决策方案进行调整, 是一个连续的决策过程。

围棋是一类完全信息的博弈游戏。然而, 其庞大的搜索空间, 以及局面棋势的复杂度, 使得传统的搜索算法在围棋面前都望而却步。整个AlphaGo最核心的两部分分别是策略网络 (Policy Network) 和估值网络 (Value Network), 如图4.2。

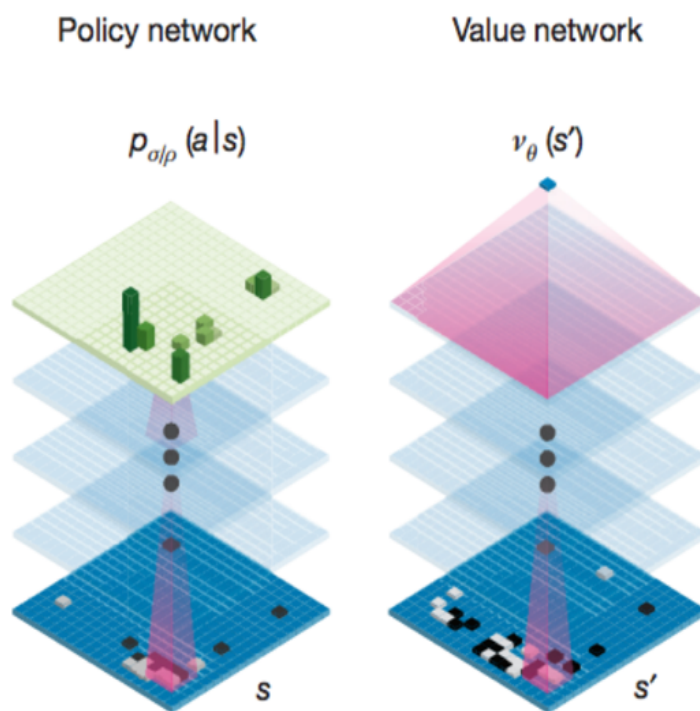


图 4.2 AlphaGo的策略网络和评估网络

Value Network主要是用来评估当前的局势。Value Network最终求得一个棋面的价值函数 $v^p(s)$ ，即在给定的一组对弈策略 p 的情况下，从状态 s 出发，最终的赢棋概率。

Policy Network用来模拟人类专家的走子。其通过网上收集大量人类棋手互相对弈的数据训练一个卷积神经网络。其输出层是一个Softmax分类器，用来计算在给定的棋面状态 s 下每一个位置的落子概率 $p(a|s)$ 。

4.3.2 模型设计

根据上一小节中AlphaGo的启示，因此本章基于数据的能源互联网无功补偿也是训练了两个神经网络。第一个网络是评价网络，主要作用是评估每一种补偿方案的优劣。它的输入是整个能源互联网的暂态检测数据，输出是系统趋于不稳定的概率。第二个神经网络是策略网络，它的作用是利用评价网络训练得到的结果，搜索遍历可能的无功补偿方案，得到所谓的最优策略。它的输入是系统故障后的暂态检测数据，输出是最优的无功补偿方案。两个网络的功能和输入输出如图4.3。

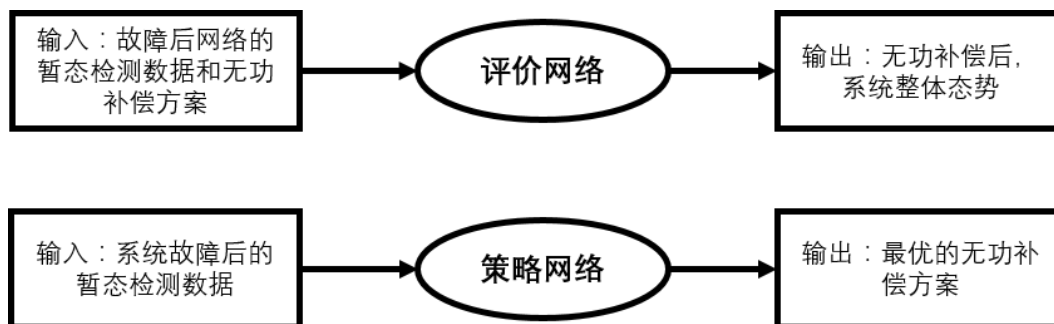


图 4.3 能源互联网无功补偿系统的策略网络和评价网络

在实际应用过程中流程如图4.4。当故障发生后，策略网络给出无功补偿方案，评估网络对每一个补偿方案进行评估从而选出最佳补偿方案。添加无功补偿后一段时间，重新把系统的暂态数据再次输入策略网络和评估网络，重新修订补偿方案。在系统达到稳定前，不停地循环下去。

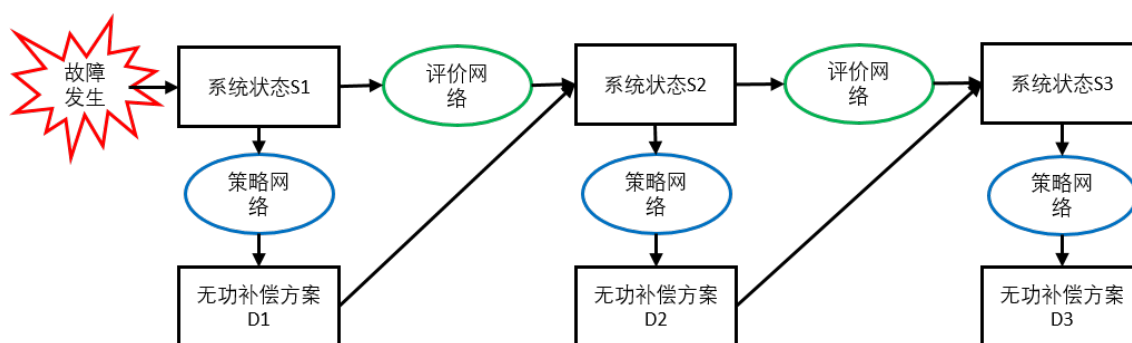


图 4.4

4.4 样本生成和训练

(1) 澳门电网概况

澳门现状高中压输配电网共有四个电压等级分别为220/110/66kV/11kV。至2013年，澳门电网共有220kV变电站2座，即鸭涌河站（CP，变电容量 $4 \times 180\text{MVA}$ ）和莲花站（CT，变电容量为 $3 \times 180\text{MVA}$ ）。110kV变电站5座，即澳北变电站（MN，变电容量 $6 \times 125\text{MVA}$ ）、路凼变电站（CI，变电容量 $3 \times 125\text{MVA}$ ）、新凼仔变电站（NT，变电容量 $3 \times 125\text{MVA}$ ）、路环B变电站（CB，变电容量 $3 \times 60\text{MVA}$ ）和澳门大学变电站（UM，变电容量 $2 \times 50\text{MVA}$ ）。66kV变电站共有18座。整个澳门电网66kV以上拓扑结构如图4.5。

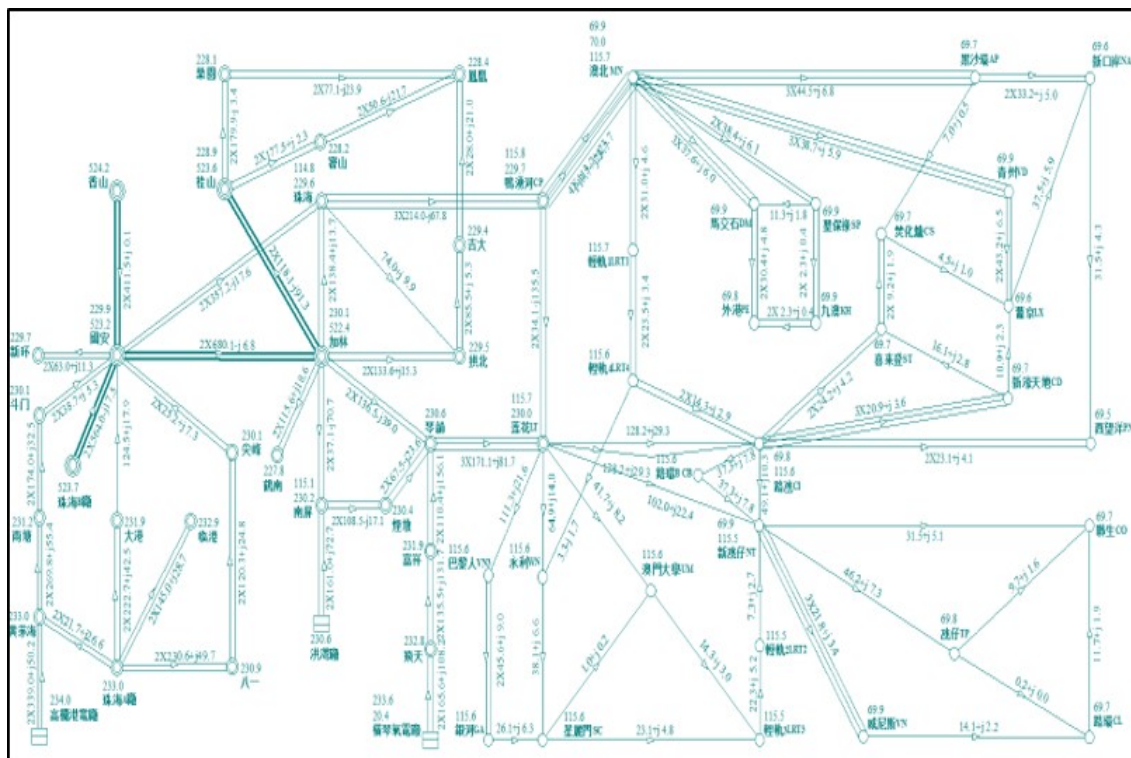


图 4.5 澳门电网各电压等级结构图

目前澳门电网容性补偿配置现状是：220kV变电站没有安装低压电容器；110 kV 变电站低压电容器配置比例为2.6%；66kV变电站低压电容器配置比例为15.39%，整体电容器安装容量偏低。抗性补偿配置现状是：220kV 变电站电抗器整体配置率为8.33%，110kV及66kV变电站没有配置电抗器补偿。整体电抗器配置率较低。因此，研究澳门电网无功补偿配置策略是非常有意义的。

(2)样本生成
 为了确保机器学习算法的可推广性和泛化性，需要确保可供学习样本的数量，可以通过在目标电网系统模型上设置不同类型的负荷模型以获得大量数据样本，具体来说可以通过通过变换电网不同的负荷来增加样本数量。如下表 4.1所示，为针对南方电网，通过组合地点一和地点二负荷，来增加样本数量。具体的故障类型和负荷的比例可以来自于真实电网运行的数据，使得最终仿真样本更加具有代表性。在具体实施例中，所述在目标电网系统模型在BPA仿真软件中设置不同类型的线路故障包括：通过在中国版本电力系统分析软件中稳定数据文件中填写基本故障卡或者简化故障模型卡。在BPA中，主要是在稳定数据文件（swi文件）中填写基本故障卡（LS卡）或者简化故障模型卡（FLT）卡完成的。

(3)样本训练

策略网络训练的输入输出如图 4.7:

表 4.1 负荷类型组合

珠海/澳门	40%电动机+60%恒阻抗	50%电动机+50%恒阻抗	60%电动机+40%恒阻抗	70%电动机+30%恒阻抗
3-4-3静态负荷模型	负荷模型1	负荷模型2	负荷模型3	负荷模型4
40%电动机+60%恒阻抗	负荷模型5	负荷模型6	负荷模型7	负荷模型8
50%电动机+50%恒阻抗	负荷模型9	负荷模型10	负荷模型11	负荷模型12
60%电动机+40%恒阻抗	负荷模型13	负荷模型14	负荷模型15	负荷模型16



图 4.6 策略网络训练的输入输出

步骤1，在目标电网系统模型上设置不同类型的线路故障；基于所述各项线路故障，设置各项电路故障相对应无功补偿方案；

步骤2，基于所述各项线路故障和无功补偿方案，利用暂态稳定仿真程序，分别获得无功补偿前后的暂态数据；

步骤3，将所述无功补偿前后的暂态数据分别作为输入输出项，利用神经网络进行训练，获得训练后的无功补偿分类器；

步骤4，将新的线路故障信息与各项无功补偿方案进行组合，利用所述训练后的无功补偿分类器进行评分，取得分最高的方案为最优方案。

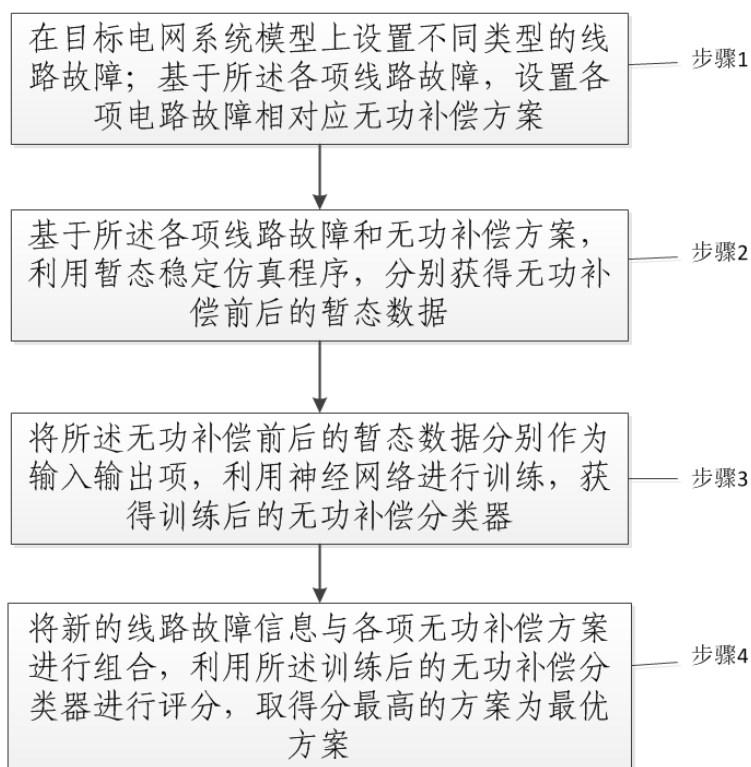


图 4.7 基于大数据的能源互联网无功补偿

4.5 算例分析

(1)算例1：负荷模型4

在第一个算例中，运行条件我们选择表4.1中的负荷模型4，即珠海是3-4-3静态负荷模型，澳门是70%电动机+30%恒阻抗。设置的故障类型为220kV线路单回线三相短路故障，故障切除时间均设置为0.1秒。图4.8为故障后的四个级别母线电压曲线。从图中可以看出，只有220kV母线电压在故障切除后2s内恢复到0.80（标么值）以上，而110kV、66kV与10kV母线没能在故障切除后10s内恢复到0.80（标么值）以上。因此，只有220kV母线暂态电压恢复到稳定要求，110kV及以下电压等级母线电压均已失稳。

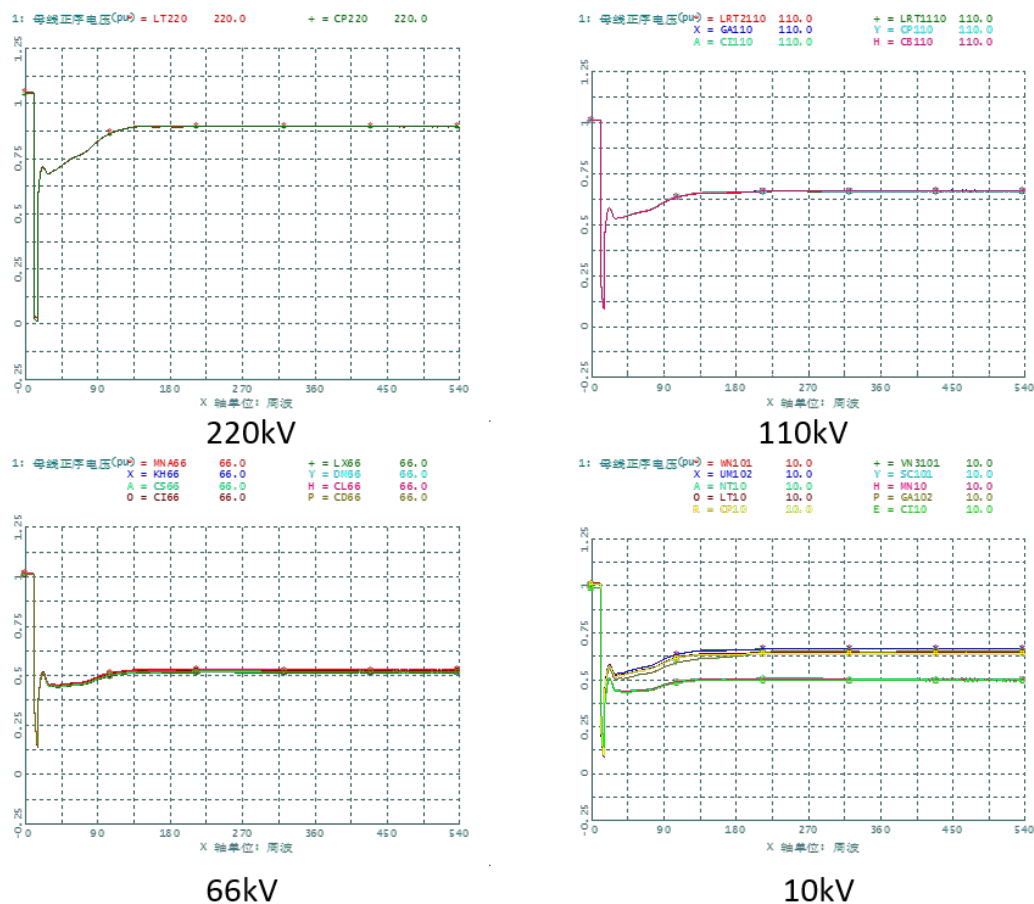


图 4.8 算例1：故障后各级别母线暂态电压

我们首先通过策略网络得出的补偿方案一为：在CP站和LT站两个母线电压为220kV的变电站分别补偿容量为490Mvar，总计补偿容量为980Mvar。图4.9为方案一补偿后各级别母线电压。从图中可以看出，220kV，110kV，66kV和10kV四个级别的母线均在7s（350周波）左右恢复到标么值0.8倍以上，基本稳定。

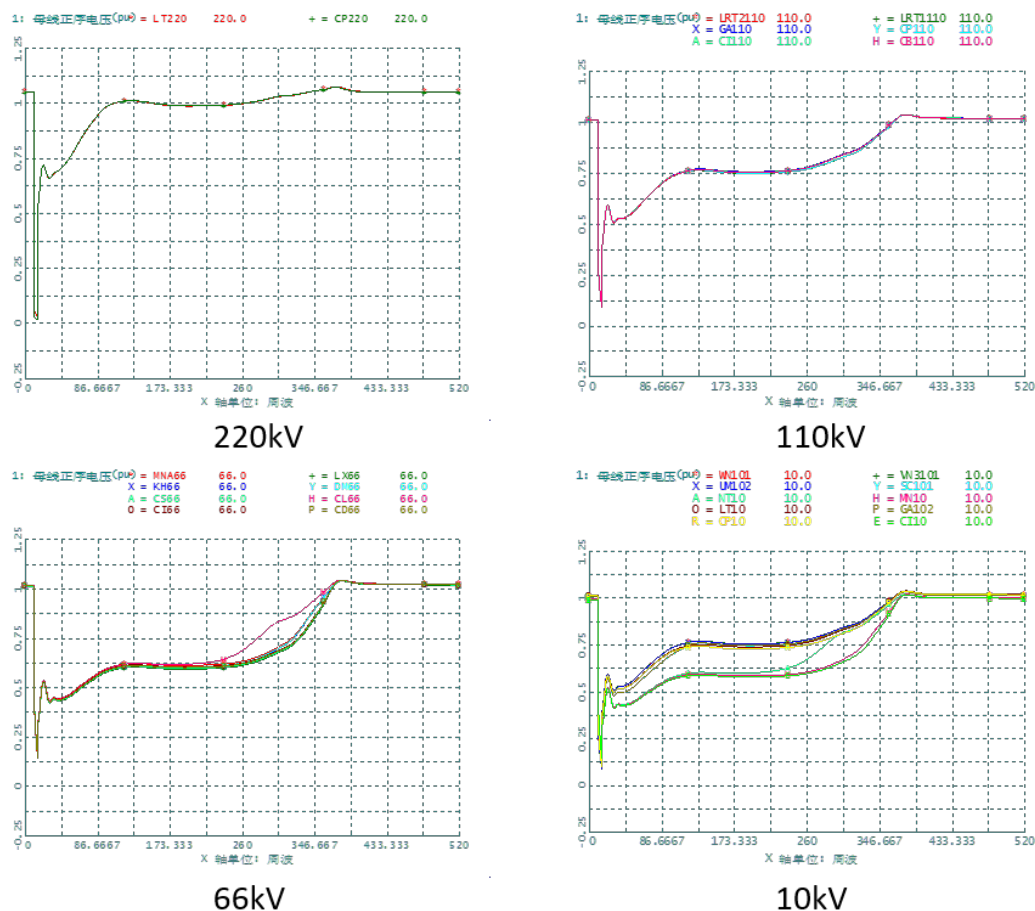


图 4.9 算例1: 方案1补偿后各级别母线电压

那我们在补偿方案一实施2s（100周波）后，通过策略网络根据目前网络的暂态数据再次得出方案二：在MN、AP两个66kV变电站补偿20Mvar，在DM、LX、NA、PE、PN、SP、ST、VN等8个66kV变电站分别补偿无功30Mvar，共计补偿280kV。补偿后效果如图 4.10。从图中可以看出，四个级别的母线普遍在3.5s左右恢复稳定，比只是用方案一提前3s恢复稳定。并且在2s变为方案2后，可以看出图中电压恢复的速率明显上升，证实了我们基于大数据的无功补偿方案的有效性。

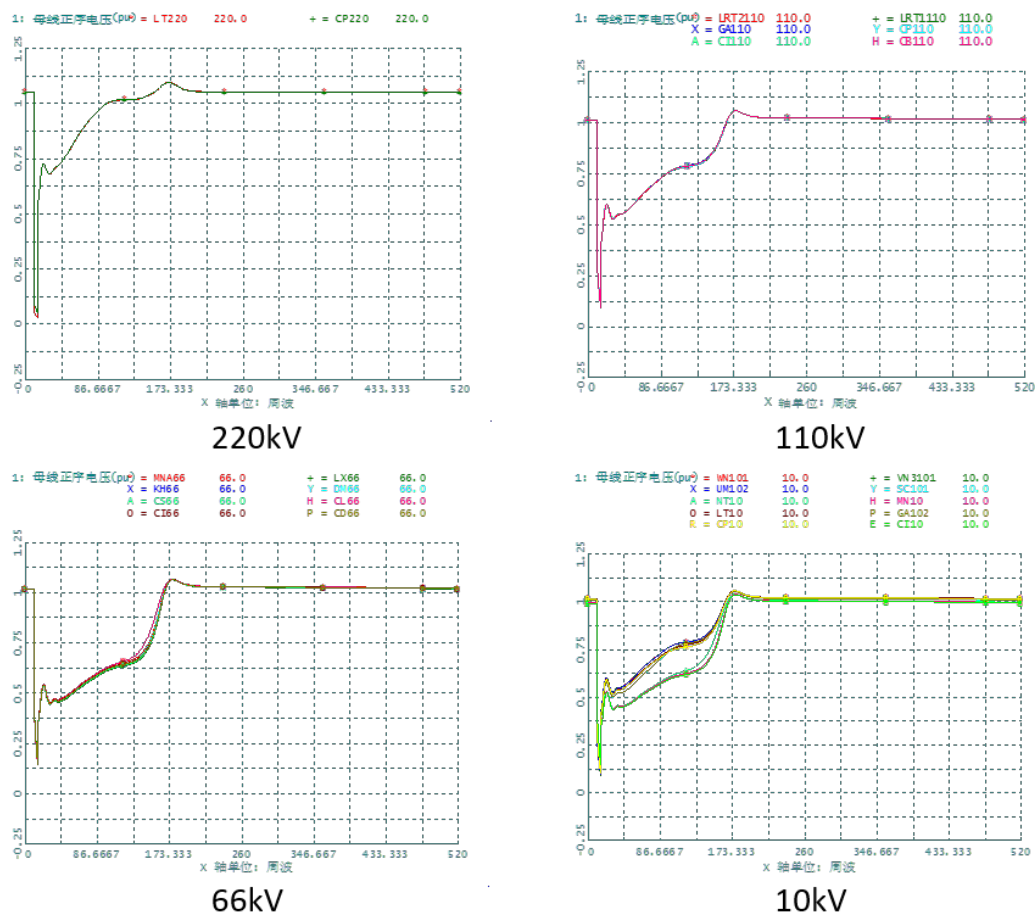


图 4.10 算例1：方案1实施2s后变为方案2各级别母线电压

(2)算例2：负荷模型8

在第二个算例中，运行条件我们选择表4.1中的负荷模型8，即珠海是40%电动机+60%恒阻抗负荷模型，澳门是70%电动机+30%恒阻抗。设置的故障类型仍然为220kV线路单回线三相短路故障，故障切除时间均设置为0.1秒。图4.11为故障后的四个级别母线电压曲线。从图中可以看出，从各个电压等级母线电压恢复曲线来看，母线电压在扰动过后经过短暂的恢复继续跌落直至电压过低计算终止，系统暂态电压失稳。

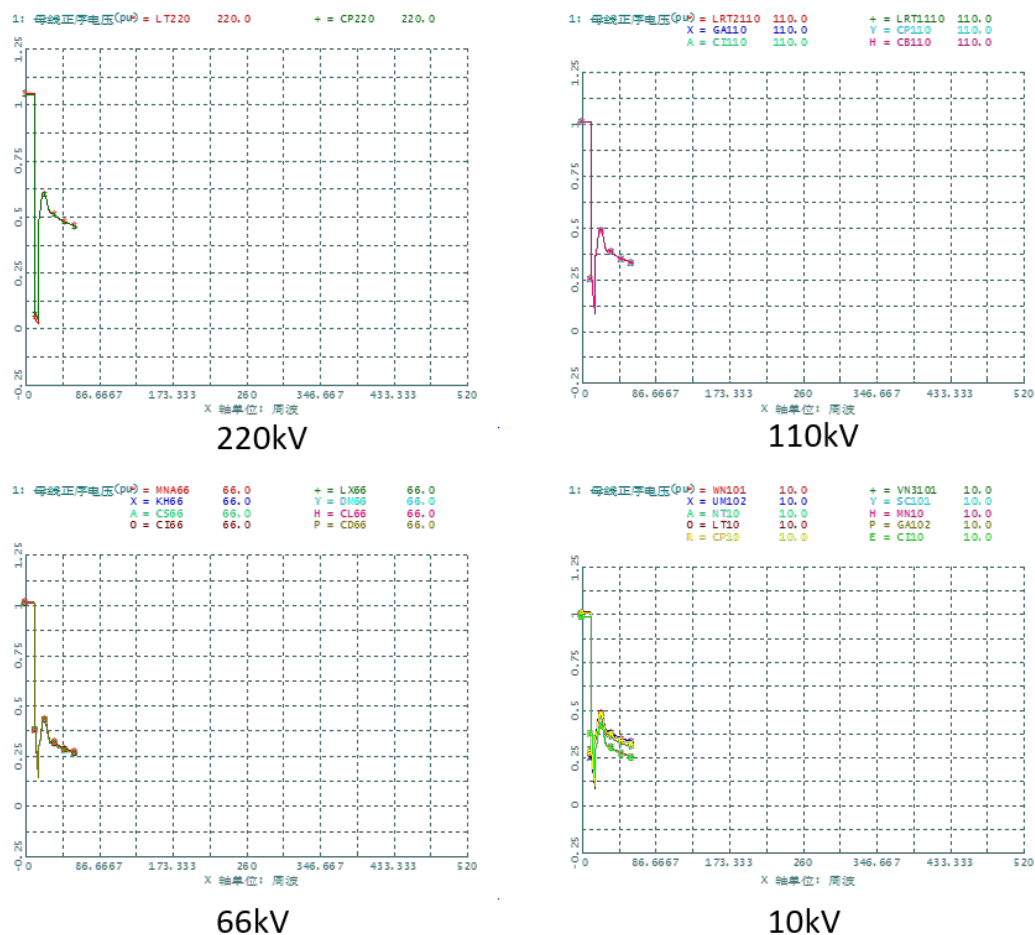


图 4.11 算例2：故障后各级别母线暂态电压

我们首先通过策略网络得出的补偿方案一为：在CP站和LT站两个母线电压为220kV的变电站分别补偿容量为490Mvar，总计补偿容量为980Mvar。图4.12为方案一补偿后各级别母线电压。从图中可以看出，220kV母线在3s(150周波)左右恢复稳定，而110kV，66kV和10kV三个级别的母线也均在7s（350周波）左右恢复到标幺值0.8倍以上。因此，证实了我们基于数据选出的无功补偿方案的有效性。

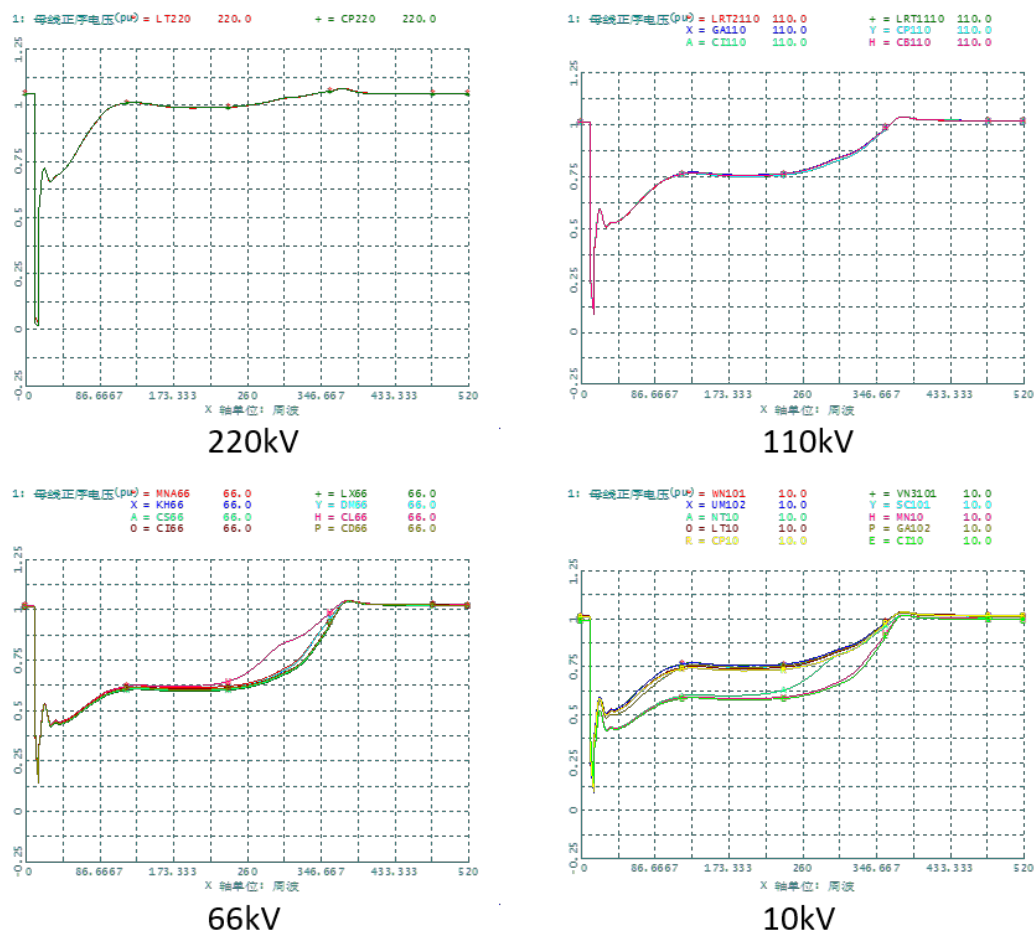


图 4.12 算例2: 方案1补偿后各级别母线电压

那我们在补偿方案一实施2s（100周波）后，通过策略网络根据目前网络的暂态数据再次得出方案二：在NA、PN、SP、ST等四个66kV变电站补偿200Mvar，在MN、AP、DM、LX、PE、VN等6个66kV变电站分别补偿无功150Mvar，共计补偿1700kV。补偿后效果如图4.13。从图中可以看出，四个级别的母线普遍在3.5s左右恢复稳定，除了220kV母线外，110kV、66kV和10kV均比只用方案1提前3s恢复稳定。

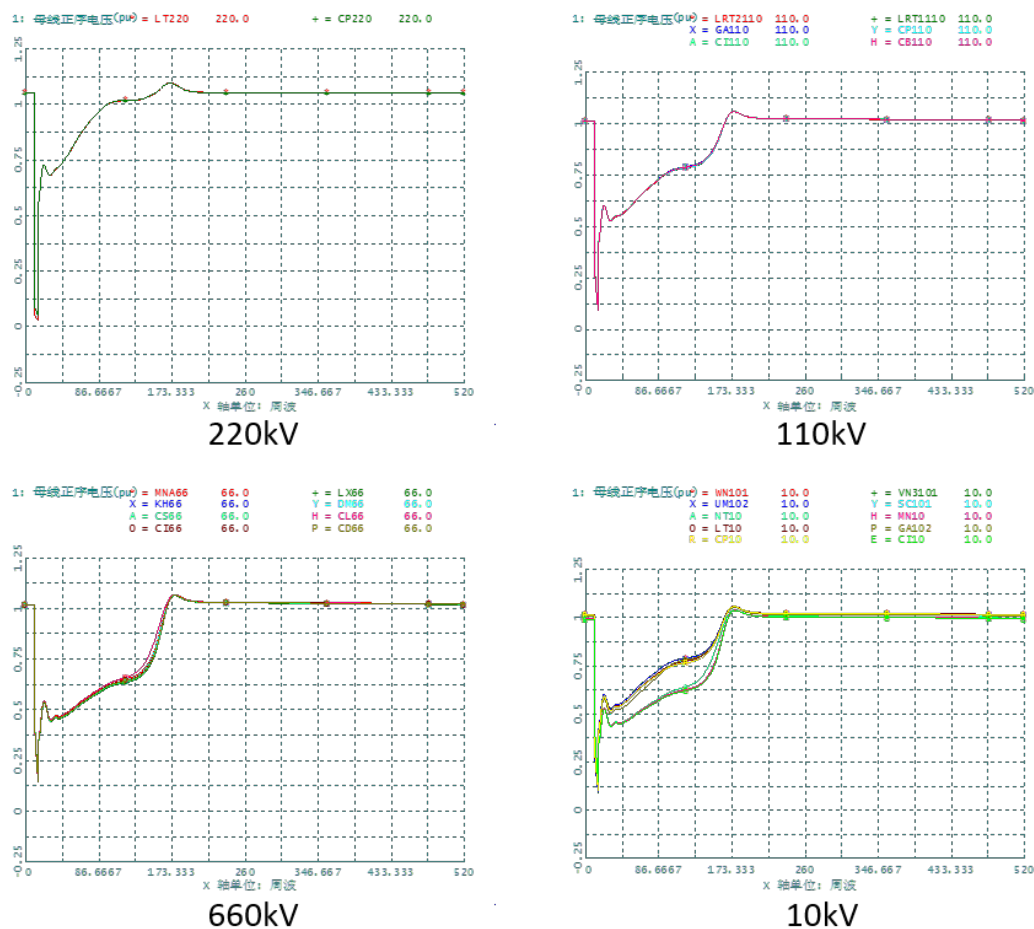


图 4.13 算例2：方案1实施2s后变为方案2各级别母线电压

4.6 面向BPA的数据处理工具包

前面已经提到，基于数据的方法本质是一个机器学习问题，核心是要有训练数据来训练分类器。由于电网获取数据的成本十分昂贵，同时由于正负样本不均衡，获取标签困难和网络安全等一系列因素，使得实际数据很难直接引用于数据挖掘。因此，基于数据的态势感知研究的第一步大多是利用仿真软件，通过组合多种故障和运行条件，仿真出大量训练样本。但是这其中有一个问题是，由于电力仿真软件往往不是针对数据挖掘设计的，都是通过界面操作的方式，很难通过自动修改运行条件，快速生成大量训练样本。同时，数据导出的格式也是非结构化的，并不便于进行数据直接处理。

因此，本小节开发了一套面向BPA软件的数据处理工具包。主要包括自动填卡、数据归一化、数据降维、批量运行等多个函数。为未来电力大数据的研究提供了便利。

PSD-BPA^[71]是一款运用于大型电网系统的分析工具软件。在上世纪六十年代

由美国的邦纳维尔电力局（Bonneville Power Administration, BPA）开发组开发的，七十年代末成功的运用在小型机上。中国电力科学研究院在1984年把它引入我国，在此基础上经过改进，八十年代开发出windows操作系统版。进入本世纪之后，逐渐形成了中国版的BPA（PSD-BPA）。美国BPA已于1996年终止了BPA潮流和暂态稳定程序的开发和维护，如今只有中国电力科学研究院电力系统研究所在维护升级PSD-BPA。

PSD-BPA现在主要由集成平台、单线图、地理接线图、稳定曲线画图工具等组成，主要用于潮流计算和暂态稳定计算。目前，它具备了电力系统稳态、电磁暂态、机电暂态以及中长期动态、短路电流计算、电压稳定计算和频域计算等交直流电力系统全过程仿真能力。

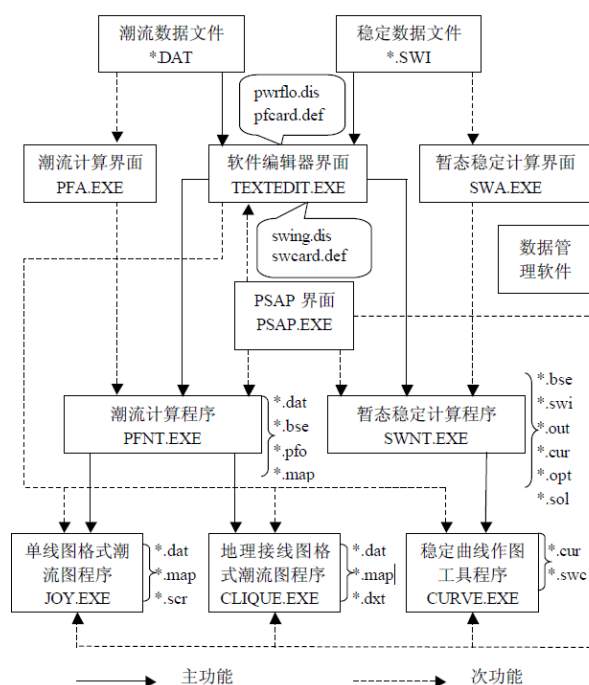


图 4.14 BPA潮流计算文件关系图

首先应由控制语句和卡片编写网络数据文件，其中控制语句主要是为了控制网络数据进行计算的开始、计算方法、显示结果和结束，卡片是记录电力系统的节点信息，传输线信息以及发电机、变压器等信息。然后根据网络数据文件进行潮流计算，生成.bse、.pfo和.map(网络拓扑结构)文件，这些文件包含了潮流运算结果和电力系统网络拓扑的二进制文件。

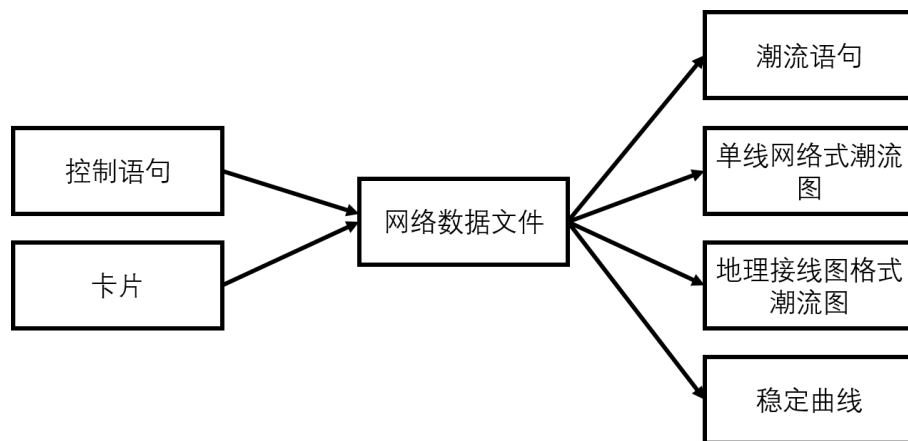


图 4.15 BPA功能

潮流计算文件.dat是由控制语句、网络数据、控制语句和（END）组成（如图4.16）。通过潮流文件进行潮流计算；然后生成.map文件，通过.map文件即可生成单线图格式潮流文件和地理接线图格式潮流文件。

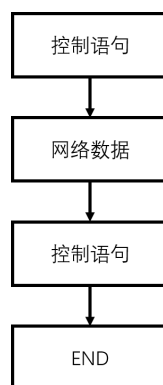


图 4.16 数据卡结构

我们的工具包主要功能和输入输出如表4.2。主要可以分为三个方面：参数修改、数据导出和数据处理。参数修改主要是修改B卡、L卡和T卡，分别对应节点参数修改、线路参数修改和变压器参数修改。参数修改的主要作用是通过设置不同的故障类型，改变网络运行条件。BPA的输出是.OUT文件，属于非结构化数据。数据导出就是从.OUT文件中分别导出电压、电流、有功和无功数据。数据处理主要包括数据降维和批量运行两个功能。

4.7 本章小结

本章通过在目标电网系统模型上设置不同类型的线路故障；基于所述各项线路故障，设置各项电路故障相对应无功补偿方案；基于所述各项线路故障和无功

表 4.2 面向BPA数据处理工具包主要功能

函数名	功能	输入	输出
B_MODIFY	节点参数修改	修改前B卡	修改后B卡
L_MODIFY	线路参数修改	修改前L卡	修改后L卡
T_MODIFY	变压器参数修改	修改前T卡	修改后T卡
VOLTAGE_EXPORT	电压数据导出	.OUT文件	VOITAGE.dat文件
CURRENT_EXPORT	电流数据导出	.OUT文件	CURRENT.dat文件
ACTIVE_EXPORT	有功导出	.OUT文件	ACTIVE.dat文件
INACTIVE_EXPORT	无功导出	.OUT文件	INACTIVE.dat文件
DIMEN_REduc	数据降维	降维前.dat文件	降维后.dat文件
BATCH_EXE	批量运行	批量数据卡	批量.OUT文件

补偿方案，利用暂态稳定仿真程序，分别获得无功补偿前后的暂态数据；将所述无功补偿前后的暂态数据分别作为输入输出项，利用神经网络进行训练，获得训练后的无功补偿分类器；将新的线路故障信息与各项无功补偿方案进行组合，利用所述训练后的无功补偿分类器进行评分，取得分最高的方案为最优方案；实现了精准、量化的决定无功补偿的地点和总量，减少人为干预，大大提高电网的稳定性和安全性的有益效果。

同时，我们还开发了面向BPA软件的数据处理工具包，大大方便了电力大数据样本生成和数据处理，为后续研究工作的开展奠定了基础。

第5章 总结与展望

5.1 论文工作总结

本文主要研究基于大数据的能源互联网的态势感知。由于能源互联网中的能量供给和需求都有非常强的随机性和不确定性，加上网络结构是一种去中心化的分布式结构，使得对整个网络的稳定性提出了非常大的挑战。由于实施的评估、诊断和控制决策对计算速度和时延要求很高，传统的基于仿真的方法难以完成，因此本文提出了基于大数据的态势感知。面向态势感知的三个环节：评估、诊断和决策，分别提出了基于机器学习模型的解决方案。

论文的主要研究成果归纳如下：

1)提出了基于时延的电压稳定性预测模型。由于传统的暂态电压评估和预测方法有特征选取和特征表达困难、分类结果可解释性差、无法在线运行等缺点。我们通过引入网络特征，找出网络中的关键节点、充分理解电压波动在电网中的传播规律，为未来电网中的规划准备。同时把特征时段和预测时刻分开，得到预测结果后，可以有充分的时间进行无功补偿等措施。同时，开发了基于时延模型的在线电压稳定性分析平台（RVSAS）。在RVSAS中，通过组合HDFS、Spark、Storm、Kafka等一系列开源大数据分析平台，实现对网络状态流数据的实时分析、预测参数的快速训练和学习、对模型实时更新和在线可视化。

2)提出了基于有监督的LLE算法，并将其应用于数据降维和故障诊断。能源互联网中数据获取和标签的成本非常高，同时样本数据的维度随着网络规模增呈指数增加，使得样本数量远远低于样本数据维度，因此必须通过降维从高维空间中提取关键、有效的描述。本文把已知的故障标签加入到LLE的训练过程中，通过调整LLE中的两个重要参数：近邻数 k 和内嵌维度 d 来使得在降维后最小化类内间距，最大化类间间距。最终通过新样本在空间中的分布确定故障类型。

3)提出了基于数据的能源互联网无功补偿方案。传统的电网决策往往是依据运维人员的经验。本文通过对大规模能源互联网的多组故障方案和无功补偿方案进行仿真和学习，大大提高了无功补偿的效果。同时，开发了电力仿真软件数据格式化工具包。目前的电网大数据分析的训练数据主要来自电力仿真软件。本文针对主流的电力仿真软件BPA，开发了数据格式化工具包，大大方便了数据挖掘工作的开展。

5.2 未来研究工作展望

能源互联网的大数据研究方兴未艾，后续的研究工作主要可以集中在以下几个方面展开：

1) 单类样本分类器。与计算机视觉等领域的数据不同，电网中的大数据主要有正负样本非常不均衡，导致我们需要利用仿真软件中的仿真样本进行数据挖掘。但是仿真软件中的系统模型并不能反映电网运行的真实情况。因此，未来如果想真正应用网络中的真实数据，可以基于电网数据的特征，开发基于单类样本的分类器。

2) 基于数据和规则的融合方法。机器学习方法虽然速度快、对复杂系统的拟合能力强，但是缺点是可解释性差，易发生过拟合。单单基于数据的方法意味着放弃我们对电网已有规律的理解。因此，下一步可以提出一种基于数据和规则的融合方法，不但可以增加模型的解释能力，最终的评估规则也可以指导未来电网的规划和运行。

3) 构建公共测试数据库。基于机器学习的方法想要发展，一个重要的前提就是不同算法之间要可以互相比较。可以互相对比的前提是，要有公认、有代表性的数据集。目前电网中的大数据研究都是独立仿真产生的，可比性较差，因此构建一个类似计算机视觉中Imagenet的数据集非常必要。

参考文献

- [1] Cao J, Yuan Z, Ming Y, et al. Survey of big data analysis technology for energy internet. *Southern Power System Technology*, 2015, 11:1–12.
- [2] Rifkin J. *The third industrial revolution: how lateral power is transforming energy, the economy, and the world*. Macmillan, 2011.
- [3] Junwei C, Mingbo Y, Dehua Z, et al. Energy internet: an infrastructure for cyber-energy integration. *Southern power system technology*, 2014, 4:1–10.
- [4] Junwei C, Jiye W, Yangyang M, et al. Software-defined information and communication technology for energy internet. *Proceedings of the CSEE*, 2015, 35(14):3649–3655.
- [5] Geidl M, Koeppel G, Favre-Perrod P, et al. The energy hub—a powerful concept for future energy systems. *Third annual Carnegie Mellon Conference on the Electricity Industry*, Pittsburgh, 2007. 13–14.
- [6] Geidl M, Koeppel G, Favre-Perrod P, et al. Energy hubs for the future. *IEEE power and energy magazine*, 2007, 5(1):24–30.
- [7] Xu Y, Zhang J, Wang W, et al. Energy router: Architectures and functionalities toward energy internet. *Smart Grid Communications (SmartGridComm)*, 2011 IEEE International Conference on. IEEE, 2011. 31–36.
- [8] Liu X, Zheng Z, Wang K, et al. An energy router based on multi-winding high-frequency transformer. *Applied Power Electronics Conference and Exposition (APEC)*, 2016 IEEE. IEEE, 2016. 3317–3321.
- [9] Farid M M, Khudhair A M, Razack S A K, et al. A review on phase change energy storage: materials and applications. *Energy conversion and management*, 2004, 45(9):1597–1615.
- [10] Zhou H, Yao J, Liu X, et al. Enabling energy storage in internet datacenters: Requirements, standards, and opportunities. *IEEE Internet Computing*, 2016, 20(5):66–70.
- [11] Feroldi D, Rullo P, Zumoffen D. Energy management strategy based on receding horizon for a power hybrid system. *Renewable Energy*, 2015, 75:550–559.
- [12] Dongxia Z, et al. Research on development strategy for smart grid big data. *中国电机工程学报*, 2015, (2015 年 01):255–255.
- [13] Peng X, Deng D, Cheng S, et al. Key technologies of electric power big data and its application prospects in smart grid. *Proc. CSEE*, 2015, 35(3):503–511.
- [14] 彭小圣, 邓迪元, 程时杰, et al. 面向智能电网应用的电力大数据关键技术. *中国电机工程学报*, 2015, 35(3):503–511.
- [15] Sheikhi A, Rayati M, Bahrami S, et al. A cloud computing framework on demand side management game in smart energy hubs. *International Journal of Electrical Power & Energy Systems*, 2015, 64:1007–1016.
- [16] Qixin C, Dunnan L, Jin L, et al. Business models and market mechanisms of energy internet (1). *Power system technology*, 2015, 39(11):3050–3056.

- [17] 邵雪, 孙宏斌, 郭庆来. 能源互联网中基于区块链的电力交易和阻塞管理方法. 电网技术, 2016, 40(12):3630–3638.
- [18] You S, Jin L, Zong Y, et al. The danish perspective of energy internet: from service-oriented flexibility trading to integrated design, planning and operation of multiple cross-sectoral energy systems. *Zhongguo Dianji Gongcheng Xuebao*, 2015, 35(14):3470–3481.
- [19] Huang A Q, Crow M L, Heydt G T, et al. The future renewable electric energy delivery and management (freedm) system: the energy internet. *Proceedings of the IEEE*, 2011, 99(1):133–148.
- [20] 刘振亚. 《全球能源互联网》. 当代电力文化, 2015, (3):102.
- [21] Endsley M R. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 1995, 37(1):32–64.
- [22] Endsley M R. Design and evaluation for situation awareness enhancement. *Proceedings of the human factors and ergonomics society annual meeting, volume 32*. SAGE Publications, 1988. 97–101.
- [23] Endsley M R, Connors E S. Situation awareness: State of the art. *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, 2008 IEEE. IEEE, 2008. 1–4.
- [24] Panteli M, Crossley P A, Kirschen D S, et al. Assessing the impact of insufficient situation awareness on power system operation. *IEEE Transactions on power systems*, 2013, 28(3):2967–2977.
- [25] Tang C, Graham C, El-Kady M, et al. Transient stability index from conventional time domain simulation. *IEEE Transactions on Power Systems*, 1994, 9(3):1524–1530.
- [26] Pai A. *Energy function analysis for power system stability*. Springer Science & Business Media, 2012.
- [27] 江涵. 大规模电力系统暂态稳定并行计算研究[D]. 浙江大学, 2012.
- [28] Maria G, Tang C, Kim J. Hybrid transient stability analysis (power systems). *IEEE Transactions on Power Systems*, 1990, 5(2):384–393.
- [29] Moulin L S, Da Silva A A, El-Sharkawi M, et al. Support vector machines for transient stability analysis of large-scale power systems. *IEEE Transactions on Power Systems*, 2004, 19(2):818–825.
- [30] Zhou D Q, Annakkage U D, Rajapakse A D. Online monitoring of voltage stability margin using an artificial neural network. *IEEE Transactions on Power Systems*, 2010, 25(3):1566–1574.
- [31] Diao R, Sun K, Vittal V, et al. Decision tree-based online voltage security assessment using pmu measurements. *IEEE Transactions on Power Systems*, 2009, 24(2):832–839.
- [32] Gomez F R, Rajapakse A D, Annakkage U D, et al. Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. *IEEE Transactions on Power Systems*, 2011, 26(3):1474–1483.
- [33] Kamalasan S, Thukaram D, Srivastava A. A new intelligent algorithm for online voltage stability assessment and monitoring. *International Journal of Electrical Power & Energy Systems*, 2009, 31(2):100–110.

- [34] Leonardi B, Ajarapu V, Djukanovic M, et al. Application of multi-linear regression models and machine learning techniques for online voltage stability margin estimation. *Bulk Power System Dynamics and Control (IREP)-VIII (IREP)*, 2010 IREP Symposium. IEEE, 2010. 1–10.
- [35] Bulac C, Triștiu I, Mandiș A, et al. On-line power systems voltage stability monitoring using artificial neural networks. *Advanced Topics in Electrical Engineering (ATEE)*, 2015 9th International Symposium on. IEEE, 2015. 622–625.
- [36] Malbasa V, Zheng C, Kezunovic M. Power system online stability assessment using active learning and synchrophasor data. *PowerTech (POWERTECH)*, 2013 IEEE Grenoble. IEEE, 2013. 1–6.
- [37] Liao H, Wu Q, Jiang L. Multi-objective optimization by reinforcement learning for power system dispatch and voltage stability. *Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, 2010 IEEE PES. IEEE, 2010. 1–8.
- [38] Zhang R, Xu Y, Dong Z Y, et al. Voltage stability margin prediction by ensemble based extreme learning machine. *Power and Energy Society General Meeting (PES)*, 2013 IEEE. IEEE, 2013. 1–5.
- [39] Kayal P, Chanda S, Chanda C. An ann based network reconfiguration approach for voltage stability improvement of distribution network. *Power and Energy Systems (ICPS)*, 2011 International Conference on. IEEE, 2011. 1–7.
- [40] Manevitz L M, Yousef M. One-class svms for document classification. *Journal of Machine Learning Research*, 2001, 2(Dec):139–154.
- [41] Ratsch G, Mika S, Scholkopf B, et al. Constructing boosting algorithms from svms: an application to one-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(9):1184–1199.
- [42] Khan S S, Madden M G. A survey of recent trends in one class classification. *Irish conference on Artificial Intelligence and Cognitive Science*. Springer, 2009. 188–197.
- [43] Zhongxi W, Xiaoxin Z. Power system analysis software package (psasp)-an integrated power system analysis tool. *Power System Technology*, 1998. Proceedings. POWERCON'98. 1998 International Conference on, volume 1. IEEE, 1998. 7–11.
- [44] Jolliffe I. *Principal component analysis*. Wiley Online Library, 2002.
- [45] Hosmer Jr D W, Lemeshow S, Sturdivant R X. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.
- [46] Golub G H, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 1979, 21(2):215–223.
- [47] Simmhan Y, Aman S, Kumbhare A, et al. Cloud-based software platform for big data analytics in smart grids. *Computing in Science & Engineering*, 2013, 15(4):38–47.
- [48] Wang D, Song Y, Zhu Y. Information platform of smart grid based on cloud computing. *Dianli Xitong Zidonghua(Automation of Electric Power Systems)*, 2010, 34(22):7–12.
- [49] Markovic D S, Zivkovic D, Branovic I, et al. Smart power grid and cloud computing. *Renewable and Sustainable Energy Reviews*, 2013, 24:566–577.

- [50] Mohsenian-Rad A H, Leon-Garcia A. Coordination of cloud computing and smart power grids. Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on. IEEE, 2010. 368–372.
- [51] Zhang X, Wei J, Zhao D, et al. Research course and prospects of power grid fault diagnosis. Power System Technology, 2013, 37(10):2745–2753.
- [52] Boyer S A. SCADA: supervisory control and data acquisition. International Society of Automation, 2009.
- [53] Xu S k, Xie X r, XIN Y z. Present application situation and development tendency of synchronous phasor measurement technology based wide area measurement system [j]. Power System Technology, 2005, 2:009.
- [54] Liao S H. Expert system methodologies and applications—a decade review from 1995 to 2004. Expert systems with applications, 2005, 28(1):93–103.
- [55] Lin C E, Ling J M, Huang C L. An expert system for transformer fault diagnosis using dissolved gas analysis. IEEE Transactions on Power Delivery, 1993, 8(1):231–238.
- [56] Park Y M, Kim G W, Sohn J M. A logic based expert system (lbes) for fault diagnosis of power system. IEEE Transactions on Power systems, 1997, 12(1):363–369.
- [57] Zhang Y, Ding X, Liu Y, et al. An artificial neural network approach to transformer fault diagnosis. IEEE Transactions on Power Delivery, 1996, 11(4):1836–1841.
- [58] Li B, Chow M Y, Tipsuwan Y, et al. Neural-network-based motor rolling bearing fault diagnosis. IEEE transactions on industrial electronics, 2000, 47(5):1060–1069.
- [59] Yongli Z, Limin H, Jinling L. Bayesian networks-based approach for power systems fault diagnosis. IEEE Transactions on Power Delivery, 2006, 21(2):634–639.
- [60] WU X, GUO C x. Power system fault diagnosis approach based on bayesian network [j]. Proceedings of Electric Power System and Automation, 2005, 4.
- [61] Lin X, Ke S, Li Z, et al. A fault diagnosis method of power systems based on improved objective function and genetic algorithm-tabu search. IEEE Transactions on Power Delivery, 2010, 25(3):1268–1274.
- [62] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. science, 2000, 290(5500):2323–2326.
- [63] Tenenbaum J B, De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction. science, 2000, 290(5500):2319–2323.
- [64] 张伟, 周维佳, 刘晓源, et al. 基于扩展 lle 方法的非线性系统故障诊断研究. 电子学报, 2015, 43(9):1810–1815.
- [65] 张伟, 周维佳, 李斌. 基于分维 lle 和 fisher 判别的故障诊断方法. 仪器仪表学报, 2010, (2):325–333.
- [66] 王江萍, 崔锦. 基于改进 lle 算法的机械故障特征压缩与诊断. 科学技术与工程, 2016, 16(13):86–91.
- [67] Chompoo-Inwai C, Yingvivanapong C, Methaprayoon K, et al. Reactive compensation techniques to improve the ride-through capability of wind turbine during disturbance. IEEE Transactions on Industry Applications, 2005, 41(3):666–672.

- [68] Corasaniti V F, Barbieri M B, Arnera P L, et al. Hybrid active filter for reactive and harmonics compensation in a distribution network. *IEEE Transactions on Industrial Electronics*, 2009, 56(3):670–677.
- [69] Steeper D, Stratford R. Reactive compensation and harmonic suppression for industrial power systems using thyristor converters. *IEEE transactions on industry applications*, 1976, (3):232–254.
- [70] Byerly R, Poznaniak D, Taylor E. Static reactive compensation for power transmission systems. *IEEE transactions on power Apparatus and systems*, 1982, (10):3997–4005.
- [71] LI G k, LI G y. The summary of power system simulation software [j]. *Journal of Electrical & Electronic Engineering Education*, 2005, 3:015.

致 谢

衷心感谢导师曹军威老师对本人的精心指导。曹老师为学勤勉、非常注重科研和产业的结合。曹老师对学生要求非常严格，同时给我创造了很多学术交流的机会，开拓了我的视野。

感谢实验室万宇鑫、陈硕师兄，杨刚、张婉璐、郝传统几位师弟师妹，和阳子婧、杨明博等几位博士后。你们为我提供了一个很好的科研氛围，帮我解决了科研道路上的困难。特别感谢电机系的刘显壮同学，由于我刚开始对电力系统的了解比较少，刘同学帮我在样本仿真和特征选取方面提供了诸多帮助。

感谢我的好朋友洪洋、孟录欧、李东珂、刘显茁、迟彤、席玉、肖永博、李益等。因为有你们，三年的硕士生活并没有感觉到枯燥乏味，你们是我人生路上最宝贵的财富。感谢老爸老妈和姐姐，你们的期许是我前进最大的动力。

我于百年校庆时入学，是清华前100年的最后一届学生。7年间，我见证了学校的各种变化：硬件更加完善；制度更加合理；更加人文化、国际化。这个时代非常浮躁，是你给了我不卑不亢的资本。

最后感谢 THUTHESIS，它的存在让我可以专注在科学研究本身，而不用在论文格式上浪费太多时间。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

个人简历、在学期间发表的学术论文与研究成果

个人简历

1992年2月9日出生于吉林省长春市。

2010年9月考入清华大学自动化系自动化专业，2014年7月本科毕业并获得工学学士学位。

2014年9月免试进入清华大学自动化系控制科学与工程专业，攻读工学硕士学位至今。

发表的学术论文

- [1] Zhao B, Cao J, Zhu Z, et al. A new transient voltage stability prediction model using big data analysis. Innovative Smart Grid Technologies-Asia (ISGT-Asia), 2016 IEEE. IEEE, 1065-1069, 2016. (EI收录, 检索号:20170403280701)
- [2] B. Zhao, Z. Yuan, J. Cao, H. Zhang, Z. Zhu, and S. Yao. Voltage Sag Diagnosis Using Big Data Analysis. Proc. Int. Conf. on Information Technologies in Education and Learning, conj. 5th Int. Conf. on Networking and Distributed Computing, Hangzhou, China, 113-117, 2015.
- [3] Power Quality Data Compression Using Principal Component Analysis. H. Zhang, Z. Zhu, S. Yao, B. Zhao, and J. Cao. Proc. 23rd Int. Conf. on Electricity Distribution, Lyon, France, Paper 0458, 2015.

研究成果

- [1] 全球能源互联网研究院, 清华大学, 国家电网公司. 一种基于机器学习的电网全局延时态势感知方法: 中国, 已受理, 申请号: 201611160493.0
- [2] 全球能源互联网研究院, 清华大学, 国家电网公司. 一种基于单类样本的电压稳定性预测判断方法, 中国, 已受理, 申请号: 201611160511.5
- [3] 清华大学, 国家电网公司, 南京南瑞继保电气有限公司, 国网辽宁省电力有限公司. 一种电网无功补偿方法: 中国, 已受理, 申请号: 201611039347.2