四川大學

国家级大学生创新创业训练计划

创新类项目申报书

项目名称：雷震子——面向智能电网多源异构数据的桥联异常检测与溯源系统

项目负责人：吴昊

所在学院：网络空间安全学院

专业年级：2022

学　　号：2022141530050

手　　机：17751232296

电子邮箱：1173590144@qq.com

指导教师：何俊江

项目起止年月：2023 年 11 月至 2024 年 10 月

项目参与学生人数：5 人

四川大学教务处制

2023 年　11 月

填写说明

一、凡申报**四川大学“国家级大学生创新创业训练计划”**必须填写本申报书。创新类项目是本科生个人或团队，在导师指导下，自主完成创新性研究项目设计、研究条件准备和项目实施、研究报告撰写、成果（学术）交流等工作。

二、“项目所属一级学科和代码”参考《普通高等学校本科专业目录和专业介绍（2012 年）》。

三、**“项目开展支撑平台”**指支撑本项目开展的国家级和省部级重点实验室（中心、平台等）、国家双创示范基地平台、教学实验中心（实验室）、企业、事业或其他单位等，表中填写平台名称，可以多个。

四、**“项目组成员”**人数原则上不超过五人，应排序。

五、**“项目成熟度**”请参考附件《项目成熟度量表》。

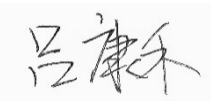
六、本书应该填写完整、内容详实、表达准确，数字一律填写阿拉伯数字。

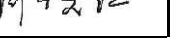
七、报送申报书的电子文档至负责人所在学院。

<table>
<tr><td>项目名称</td><td colspan="3">雷震子——面向智能电网多源异构数据的桥联异常检测与溯源系统</td></tr>
<tr><td>项目属性</td><td colspan="3">☑ 面上项目 ☐2035 特区子计划项目 ☐交叉学科子计划项目</td></tr>
<tr><td>申请类别</td><td colspan="3">☑a 科学探索与工程技术类 ☐人文艺术与社会科学类
☐软件信息与文创类 ☐智能装备与医疗器械类
☐生物医药与新材料类</td></tr>
<tr><td>申请经费</td><td>10000 元</td><td>起止时间</td><td>2023 年 11 月至
2024 年 10 月</td></tr>
<tr><td>项目所属一级学科和代码</td><td colspan="3">0809 计算机类</td></tr>
<tr><td>项目开展支撑平台</td><td colspan="3">四川大学网络靶场协同创新中心</td></tr>
<tr><td>项目来源（可多选）</td><td colspan="3">☐十大重点支持领域的项目
☑ 进课题组、进实验室、进科研团队参与的项目
☐国家级和省部级重点实验室（中心、平台等）、国家双创示范基地平台支持申报项目
☐“2035 特区子计划”命题的项目（2035 特区子计划）
☐交叉学科创新项目
☐“青年红色筑梦之旅”计划项目
☐基于前期研究实践成果、继续深入研究实践的创新项目
☐高水平课题
☐其他_____</td></tr>
<tr><td rowspan="2">2035 特区子计划项目情况（非 2035 特区子计划项目可不填）</td><td>命题名称</td><td colspan="2"></td></tr>
<tr><td>校内指导老师姓名（非交叉学科子计划项目一般仅允许一位指导老师）</td><td colspan="2"></td></tr>
<tr><td>所属重点支持领域（可不选）</td><td colspan="3">选择 1 项：E
A.不填
B.泛终端芯片及操作系统应用开发
C.重大应用关键软件
D.云计算和大数据
E.人工智能
F.无人驾驶
G.新能源与储能技术
H.生物技术与生物育种
I.绿色环保与固废资源化
J.第五代通信技术和新一代 IP 网络通信技术
K.社会事业与文化传承</td></tr>
</table>

负责人之前参与大创项目情况	无
项目成员之前参与大创项目情况	无

项目负责人基本信息			
姓名	学号	专业年级	所在学院
吴昊	2022141530050	2022	网络空间安全
性别	手机	电子邮箱	身份证号
男	17751232296	1173590144@qq.com	320583200310223816

项目组成员基本信息				
序号（含排序）	1	2	3	4
姓名/性别	吴桐曦	吕康禾	林宸	林熙金
学号	2021141230137	2022141210140	2022141530052	2022141010277
专业年级	2022	2022	2022	2023
所在学院	网络空间安全	数学学院	网络空间安全	网络空间安全
手机	18828104656	18787651471	15058294692	15902813608
电子邮箱	2460477175@qq.com	2190387852@qq.com	1073646717@qq.com	1044813659@qq.com
身份证号	360102200308282857	533222200311070033	330281200403256812	511323200311202813
签名				

指导教师 1 基本信息 （非交叉学科子计划项目一般仅允许一位指导老师）			
姓名	所在学院或单位	研究方向	职称/职务
何俊江	网络空间安全	网络智能攻防、未知攻击检测、隐私保护、免疫动力学模型	助理研究员
性别/年龄	手机	电子邮箱	签名
男	15198130461	hejunjiang@scu.edu.cn	

指导教师2　基本信息 （交叉学科子计划项目需填写第二指导老师）			
姓名	所在学院或单位	研究方向	职称/职务
性别/年龄	手机	电子邮箱	签名

项目内容概述(限 200 字以内)

随着通信技术的发展和电力物联网的建设，智能电网逐步兴起。然而智能电网环境下形成的**海量异构电力数据给智能电网异常数据检测带来了困难**，限制了智能电网发展的同时，也带来了极大的电网安全隐患。

为此，本项目基于**生物免疫桥联抗体原理**，研究设计一种针对**智能电网多源异构数据的异常检测与溯源系统**，其中包括**异构数据检测器生成，免疫检测器桥联，基于桥联检测器的异常数据检测与溯源的理论模型和基于以上理论设计的原型系统**。

项目特色创新点概述（限 100 字以内）

- **检测器生成改良：基于抗原密度实现异构检测器高效生成。**
- **桥联检测器综合应用：基于抗体偶联实现检测器桥联、数据桥联检测与异常溯源。**
- **检测溯源原型系统开发：开发面向智能电网多源异构数据的检测溯源原型系统。**

项目成熟度评估（项目成熟度请参考附件《项目成熟度量表》）

目前项目成熟度自评估为 4 级；

预期结题时项目成熟度达到 8 级。

项目组成员分工

姓名	承担工作内容
吴昊	统筹规划，技术，文本
吴桐曦	技术，文本，美工
吕康禾	技术，文本
林宸	技术，文本
林熙金	技术，文本

申报立项正文（含一、二、三、四、五）限 **4000-8000** 字

一、项目简介（研究内容、目的意义、具体目标、国内外研究现状分析及评价等）

1. 研究内容：

(1) 基于抗原密度的智能电网异构检测器生成方法研究

研究抗原（数据样本）子空间模板的表达方法，包括子空间模板的确定及其维度计算；研究基于子空间模板划分抗原集及子空间内抗原密度的计算方法；研究基于抗原空间密度计算实现在高密度低维子空间中高效检测器的生成方法，其中包括免疫耐受计算方法及免疫抑制计算方法；研究免疫检测器饱和度的计算方法。

(2) 面向多源异构智能电网的桥联检测器生成、检测以及溯源方法研究

研究异构特征检测器的选择方法，包括基于互信息最大化原理和可学习聚类矩阵的检测器聚类方法；研究基于生物免疫原理的桥联检测器生成方法；研究基于桥联检测器的免疫检测方法；研究桥联检测器判断数据为异常状态时异构异常数据的溯源方法。

(3) 基于免疫桥联的智能电网异构数据检测与溯源原型系统设计研究

研究原型系统架构设计细节，包括各模块功能的划分，核心算法模型的训练调整与优化；研究原型系统设计的相关技术细节，包括编程语言，开发所需的库，开发框架的选择；研究原型系统模块

功能的实现，包括设备拓扑结构、运行状态的获取，数据库的搭建，数据获取与预处理，检测结果可视化等方法。

2. **目的意义：**

随着电力物联网建设和通信技术的发展，电力系统中的数据种类更加丰富，信息传递更加便捷，为电力系统大数据应用带来了良好的数据基础。然而这也使得电力系统数据更加多源异构化、复杂化，为电力系统大数据的集成与分析带来了新的挑战。

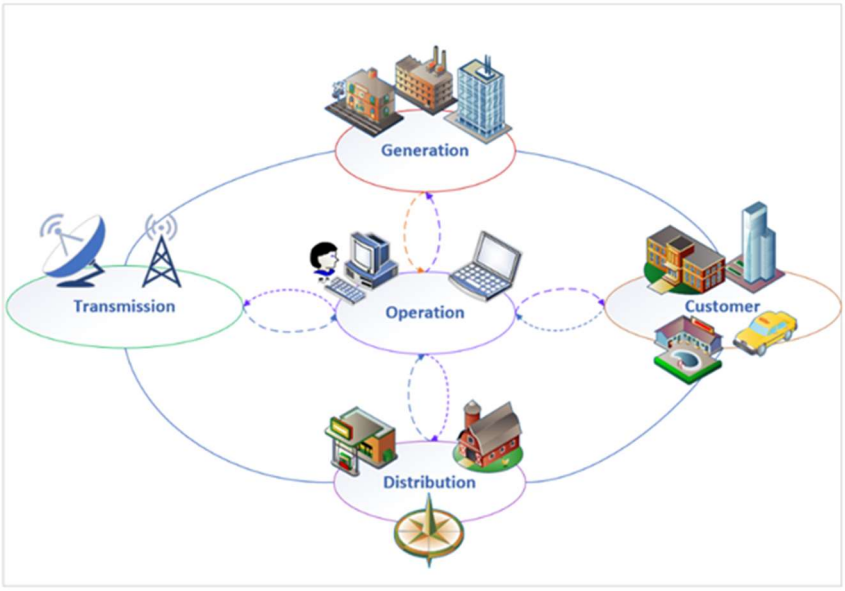


图 1：智能电网概念模型

从智能电网中操作域来看，如图 1，其会接收到来自发电域、运输域、分配域、客户域的大量数据。这些数据由于各域间数据模型、数据标准、访问接口、实现平台以及采用的数据采集技术等方面的不同，数据本身呈现多源异构的特点，即数据的来源、类型、结构、格式复杂多样，如表 1。这极大增加了操作域对数据内部异常情况检查的难度，导致智能电网网络空间安全隐患大大增加。然

而，目前智能电网的大量多源异构数据并没有高效的检测系统，导致对实时电网数据异常的检测存在较大困难，给不法分子提供了攻击机会。

表 1：数据标准间的异构

数据来源	数据内容	数据位置	数据标准
EMS 数据	运行信息、电网信息	调度主站	IEC 61970 等
变电数据	运行信息	变电站子站	IEC 61850
配电数据	分布式能源信息	配电子站	IEC 61970 等
计量	计量、付费及符合控制	用户侧	IEC 62056

以云南某牵引变电站为例，如图 2，其辅助监测系统会收到来自隔离开关、避雷器、牵引变压器、电流电压互感器等设备的数据信息，各数据间又普遍存在异构问题，大大增加了系统进行异常检测的难度。

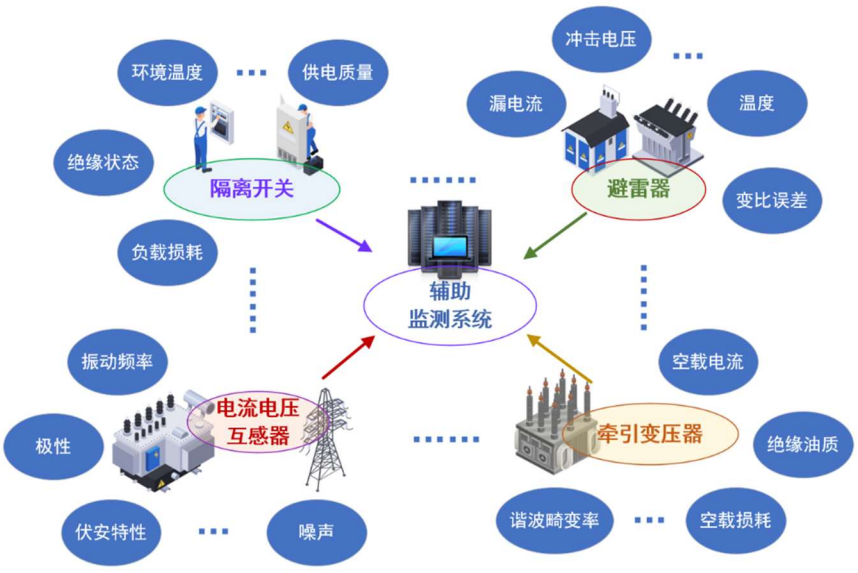


图 2：云南某变电站异构数据来源

智能电网对于多源异构数据检测目前还缺少有效的异常检测系统，这也导致了在社会生活、经济发展、国防安全等方面存在诸多问题。以 2022 年 3 月 3 日发生的“‘3·3’中国台湾电网大停电事故”

为例[1]，事故造成停电用户约549万户，合计减少约1050万千瓦的供电能力，岛内厂商损失金额达60亿元新台币（约合13.5亿人民币），总计损失达上百亿新台币。此外，2020年4月，葡萄牙跨国能源公司（天然气和电力）EDP（Energias de Portugal）遭 Ragnar Locker勒索软件攻击，赎金高达1090万美金；在2020年5月5日，委内瑞拉国家电网干线遭到攻击，造成全国大面积停电，全国11个州府均受严重影响； 2020年6月，巴西的电力公司Light S.A被黑客勒索 1400 万美元的赎金。若能及时对电网各个单元做到精确识别，并有效提高多源异构数据的处理识别能力，可大大降低各类电网事故的发生的概率。

3. 国内外研究现状分析及评价

针对智能电网存在多源异构数据导致异常检测困难的问题，国内外学者做了大量研究并取得以下成果：2016年上海交通大学电气工程学院副研究员韩蓓等人[2]针对智能电网大数据环境下导致电力系统负荷波动的诸多因素存在多源异构性的问题,利用多核函数来对其多源异构特性进行差异化处理和融合,并描述了影响因素的内在分布特性并应对其变化,提高了负荷预测精度；2017 年四川省电力公司潘可佳、王鑫等人[3]提出基于 SOA 的多源异构数据融合架构,有效提升电力系统数据融合效率与数据应用能力；2021 年澳大利亚迪肯大学信息技术学院 GE 教授等人[4]开发了一种用于多分类的具有嵌入层的前馈神经网络模型，其中嵌入层用于编码高维分类特征，并利用迁移学习对高维特征进行编码以构建二元分类器，提

高了分类精度。

此外，一些研究者通过人工免疫算法对智能电网的运行和检测进行了优化。2011 年吉林大学通信工程学院赵继印教授等人[5]针对电力变压器单一故障和多故障诊断问题,模拟生物免疫系统,提出一种两级分类器级联的诊断算法，提高了电力变压器故障诊断的准确率和速度；2018 年阿根廷国家科学技术调查委员会 Diego Lizondo 等人[6]提出了一种基于人工免疫网络算法的分布式需求侧管理系统并研发了容错分布式自控制系统，解决了因为数以千计设备同时使用导致的热带和亚热带气候中出现的电力峰值负荷问题；2021 年武汉大学电气工程学院杨军教授等人[7]提出了一种基于改进人工免疫网络的单相接地故障辨识方法，该方法在小样本训练的情况下能对故障与正常扰动进行识别,训练时间短,不需要人为设置阈值;同时,在中性点接地方式变化、信号存在噪声以及系统拓扑结构变化时,该方法适应性依然良好。

虽然有关智能电网数据安全方面的研究已经产生诸多成果，但仍存在以下问题：

1) 数据量较大时融合速度慢、高维数据运算复杂度高，导致检测效率不高。

2) 由于多源异构数据直接融合会损失一定特征信息，导致检测准确度不高。

3) 传统深度神经网络无法反映具体错误数据，导致检测没有

异常溯源能力。

此外，王臻、刘东等人[8]在《新型电力系统多源异构数据融合技术研究现状及展望》一文中指出现有智能电网多源异构数据一般处理方法为数据融合再处理，这一处理模式**继承了数据融合方法的局限**。以特征级融合为例，其涉及提取特征，特征融合等步骤，不仅增加时间消耗，降低检测效率，且融合后的数据会有信息缺失。

表 2：现行数据融合方法的局限性统计表

检测方法	局限
专家系统	1)知识获取及完备性验证难度较大;2)容错能力较差;3)不具备学习能力
模糊理论	1)系统较大时融合速度较慢:2)维护复杂;3)不具备学习能力
人工神经网络	1)样本需求大，收敛速度慢;2)重复利用性不好;3)融合结果缺乏解释性;4)无法反映具体错误数据;5) 外推时误差较大;6)不适用于大型网络
遗传算法	1)需建立数学模型;2)交叉和变异算子对结果影响较大
DS证据理论	1)数据需要独立;2)证据合成规则的合理性和有效性难以判断
加权平均法	需要对系统和数据具体含义的详细理解
卡尔曼滤波算法	数据要求同构
贝叶斯网络	数据需满足概率分布
聚类分析	高维数据运算复杂度高

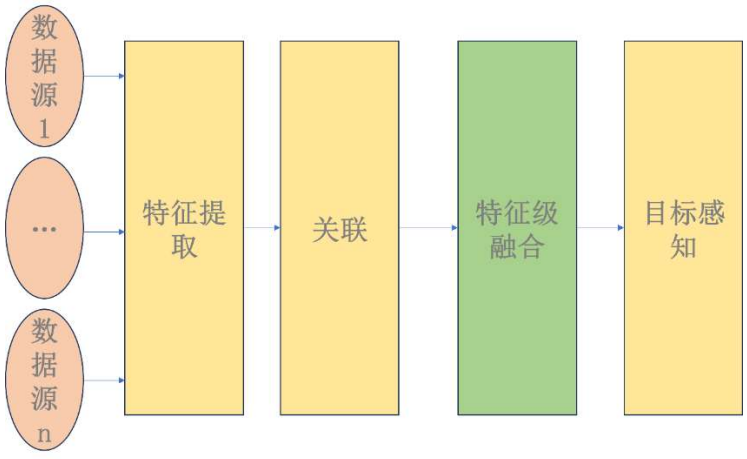


图 3：特征级融合示意

因此，本项目以牵引变电所数据检测过程为背景，研究设计一

种**针对智能电网多源异构数据的异常检测系统**，用免疫算法生成检测器，选取特征检测器后基于生物免疫偶联原理实现桥联检测器生成，利用桥联检测器对数据进行不同维度的检测，**有效避免单一数据融合过程，使得检测高效准确**。通过对已收集的真实数据集训练异常检测模型，可实现异常数据溯源。并能基于此设计**一个自适应、自完善，高宽容，强动态的异常检测系统**。

具体目标：

- **社会效益**：提高电力系统的稳定性，从而提高社会稳定性。能有效防止大规模停电等问题，保护关键基础设施，从而保障国家安全。
- **经济效益**：通过减少因电力系统故障引起的经济损失来显著的提升全社会经济效益。降低运营成本，提高电力行业的经济效益。
- **具体指标**：在**时间效率方面**，检测多源异构数据时，相比现行检测方法平均检测效率提升 10%。在**检测准确度方面**，检测多源异构数据时，相比现行检测方法平均准确度提升 20%。在**检测器自适应和宽容度方面**，检测多源异构数据时，相比现行检测方法平均检测器自适应和学习能力提升 30%。

参考文献：

[1] 雷傲宇,周剑,梅勇等.“3·3”中国台湾电网大停电事故分析及启示[J].南方电网技术,2022,16(09):90-97.DOI:10.13648/j.cnki.issn1674-0629.2022.09.011.

[2] 吴倩红,高军,侯广松等.实现影响因素多源异构融合的短期负荷预测支持向量机算法[J].电力系统自动化,2016,40(15):67-72+92.

[3] 潘可佳,王鑫,杨帆等.面向电力大数据的多源异构数据融合技术研究[J].机械与电子,2017,35(09):7-11.

[4] GE Mengmeng, SYED N F, FU Xiping, et al. Towards a Deep Learnign-Driven Intrusion Detection Approach for Internet of Things[EB/OL]. (2021-02-26)[2022-08 -30]. https://doi.org/10.1016/j. comnet.2020.107784.

[5] 郑蕊蕊,赵继印,赵婷婷等.基于遗传支持向量机和灰色人工免疫算法的电力变压器故障诊断[J].中国电机工程学报,2011,31(07):56-63.DOI:10.13334/j.0258 -8013.pcsee.2011.07.011.

[6] Diego Lizondo, Sebastian Rodriguez, Adrián Will, Victor Jimenez, Jorge Gotay, An Artificial Immune Network for Distributed Demand-Side Management in Smart Grids, Information Sciences, Volume 438,2018,ISSN 0020-0255,https://doi.org/10.1016/j.ins.2018.01.039.

[7] 刘雯静,杨军,陈振宁等.基于改进人工免疫网络的配电网单相接地故障辨识方法[J].科学技术与工程,2021,21(21):8909-8915.

[8] 王臻,刘东,徐重酉等.新型电力系统多源异构数据融合技术研究现状及展望[J].中国电力,2023,56(04):1-15.

二、研究技术路线及可行性分析

1.技术路线

基于以上研究内容，我们可拟实现一种面向智能电网异构数据实时监测与异常溯源系统，技术路线如图 4 所示。其中，主要通过三种理论研究实现了三个理论模型。包括面向智能电网数据的异常检测器生成模型，面向多源异构数据的桥联检测器生成、检测与溯源模型，面向牵引变电所数据监测原型系统设计模型。详细设计如下：

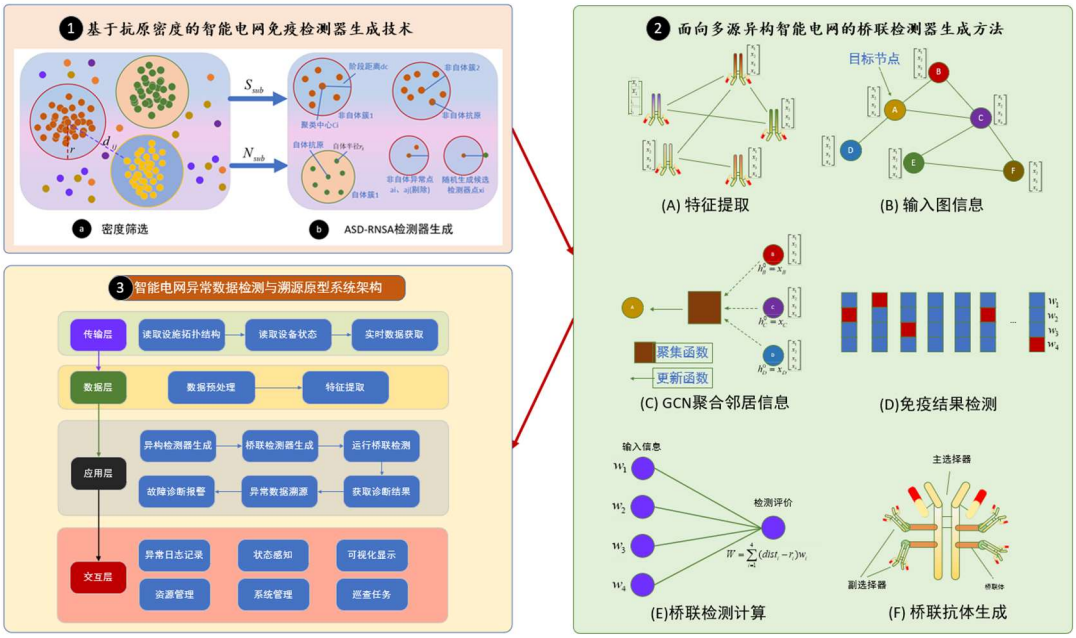


图 4：技术路线流程图

1.1 基于抗原密度的智能电网异构检测器生成技术

本项目研究的数据实例来源于智能电网中的多源异构数据，数据样本经预处理实现向量化后，其在高维空间中的分布稀疏且不均匀，大部分数据处于低维子空间（S_{sub}）内。为此，我们采取基于

抗原密度的实值负选择算法来生成子检测器，如图 5 所示，技术流程可分为密度筛选和异构检测器生成两个部分。

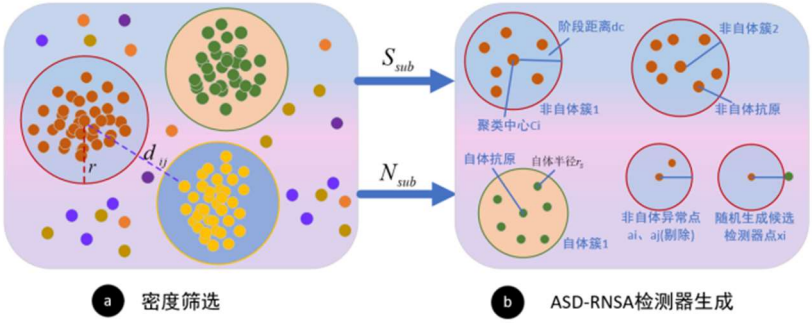


图 5：异构检测器生成技术图

密度筛选：首先基于子空间表示模型，对抗原数据集进行特征提取，获得子空间模板，利用方差解释率方法计算得出子空间的维度。然后基于子空间模板，利用欧式距离计算每个抗原与子空间模板的相似度，实现抗原集的划分。对于子空间采用离散化方法，计算每个离散空间内数据点的个数，以实现抗原密度的计算。以上步骤可对低维子空间内抗原实现有效捕捉，解决传统 NSAs 不考虑抗原分布，无法有效区分抗原的缺点。

异构检测器生成：首先利用抗原密度，基于免疫耐受计算，实现在高密度低维子空间中生成高效检测器，如图 6 所示。为消除冗余检测器，选择抗体抑制率作为终止条件，通过设定饱和阈值，计算免疫检测器饱和度，采用惩罚机制，对子空间内免疫检测器进行动态调整，获得在各个子空间下合格的检测器集合。以上步骤解决了传统 NSAs 算法覆盖时存在的采样点数量不足、多样性不足的缺点，避免生成过程在高维空间过早收敛的弊端。

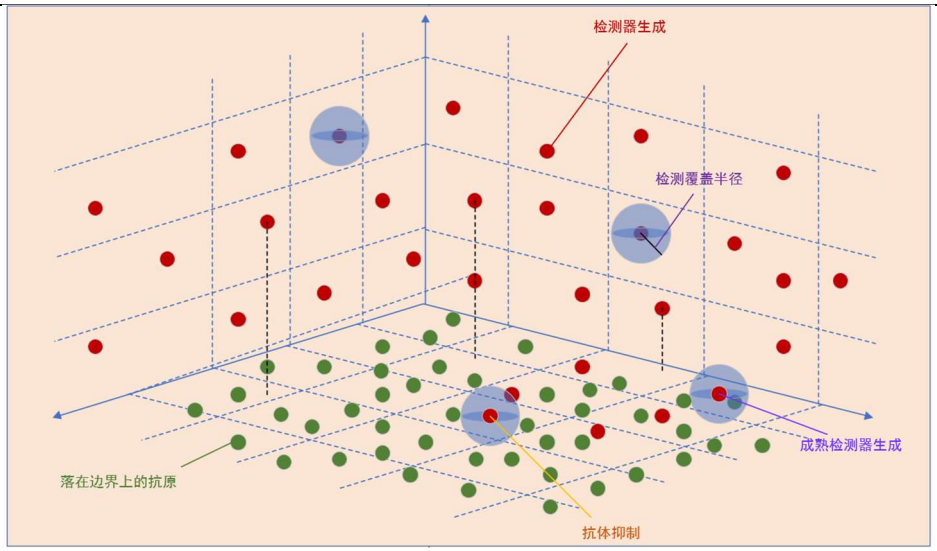


图 6: 检测器生成示意图

上述操作算法流程如表 3 所示：

表 3：异常检测器生成算法流程

输入：抗原训练集，抗体抑制率 sp，成熟检测集 D
输出：成熟检测器 D
S_{sub}：子空间集
N_{sub}：子空间的数量
$S_{sub}(N_{sub})$：S_{sub} 中的第 N_{sub} 个子空间，$S_{sub}(N_{sub}) = < x_1, x_2, ..., x_n >$，其中 n 为空间维度，$x_i$ 为 0 或 1
ρ：抗原空间密度集，$\rho = < \rho_1, \rho_2, ..., \rho_{N\text{sub}} >$，$\rho_{\text{Nsub}}$ 是 $S_{sub}(N_{sub})$ 的抗原空间密度
$D(k)$：D 在子空间 $S_{sub}(N_{sub})$ 下的子集
Cnt：抗体抑制计数

Step 1 $D = \emptyset, S_{sub} = \emptyset, k = 0, \rho = 0$；
Step 2 对于抗原训练集中每个自体抗原 ag
Step 2.1 由公式（1）通过 $ag < x_1, x_2, ..., x_n >$ 计算 $temp < t_1, t_2, ..., t_n >$
Step 2.2 如果 S_{sub} 中无 $temp$，$k = k + 1, S_{sub}(k) = temp$；
Step 2.2 反之若 $S_{sub}(k) = temp, \rho_k = \rho_k + 1$；
Step 3 对于 S_{sub} 中每个子空间 $S_{sub}(k)$；
Step 3.1 $D(k) = \emptyset, Cnt = 0$；
Step 3.2 随机生成候选检测器 $d_{new} < d_1, d_2, ..., d_n >$；
Step 3.3 由公式（2），由 $S_{sub}(k) = < x_1, x_2, ..., x_n >$ 对 $d_{new} < d_1, d_2, ..., d_n >$ 进行编辑，$x_i = 0$ 时 $d_i = 0$；
Step 3.4 如果 d_{new} 与抗原训练集中任何自身抗原匹配，转到 **Step 3.2;**
Step 3.5 如果 d_{new} 与 $D(k)$ 中的任何检测器匹配，$Cnt = Cnt + 1$；
Step 3.5.1 如果 $\frac{Cnt}{||D(i)||} \geq sp$，(其中 $||D(i)||$ 表示 i 号子空间中成熟的检测

器数),$D=[D;D(k)]$，然后退出算法；

Step 3.5.2 否则进入 **Step 3.2**；

Step 3.6 $D(k)=[D(k);d_{new}]$,转到 **Step 3.2**;

其中 Step 2.1、Step 3.3 公式如下：

$$temp=<t_1,t_2...t_n>,\text{其中}t_i=\begin{cases}0,\ if\ \ x_i=0\\1,\ if\ \ x_i\neq 0\end{cases} \tag{1}$$

$$d_{new}=d_{new}*S_{sub}(i) \tag{2}$$

1.2 面向多源异构智能电网的桥联检测器生成、检测、溯源技术

基于上述步骤生成的各子空间中的检测器集合，我们可以对检测器集合进行聚类处理，获得特征检测器，再基于生物学双特异性抗体和偶联机理，将特征检测器映射拼接，生成桥联检测器，技术路线如图 7 所示。具体设计细节如下：

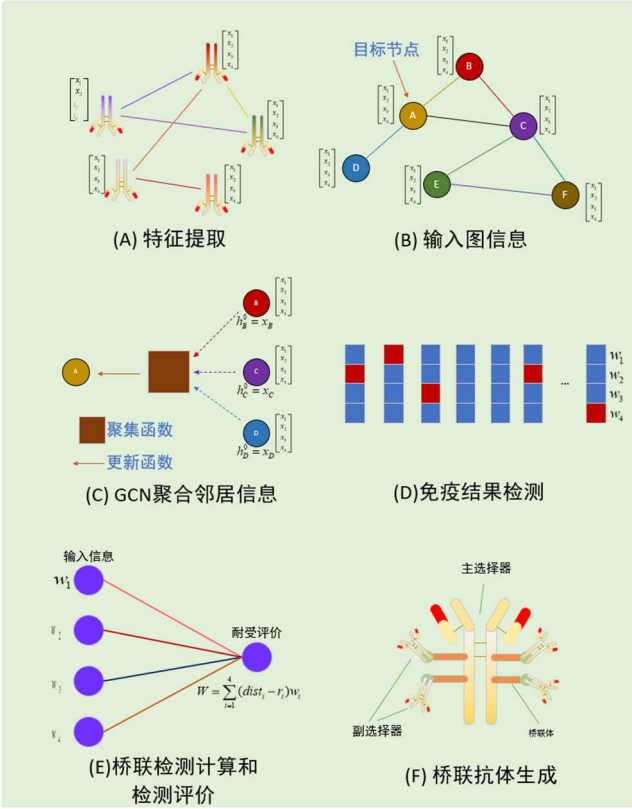


图 7：桥联检测器生成、检测、溯源技术图

为避免检测器数量过多时，直接拿检测器集合对异构数据检测的时效不够理想的弊端，我们使用图卷积网络编码网络结构和节点特征到低维向量空间，并通过重建网络结构的方式学习节点表，基于互信息最大化原理进行节点与全局对比进行具有聚类一致性优化的节点聚类，聚类流程如图 8。

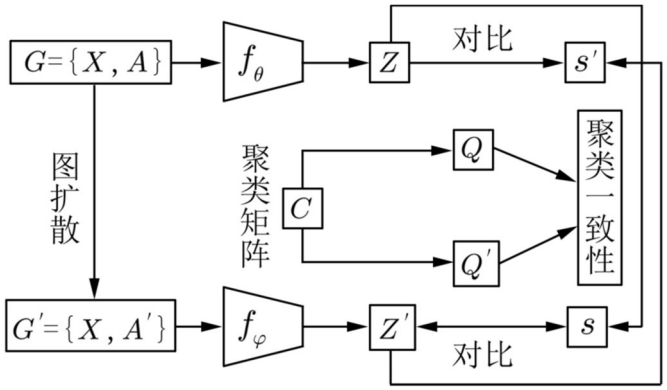


图 8：检测器聚类流程图

首先，输入特定子空间中生成的检测器作为原始图节点，对原始图$G = \{X, A\}$的网络结构A使用图扩散方法得到扩散图$G' = \{X, A'\}$。然后，将两个图被置于不共享权重的图卷积网络和提取节点表示$\mathbf{Z}$和$\mathbf{Z}'$，并经过平均池化获得全局图表示$\mathbf{s}$和$\mathbf{s}'$。接着，基于互信息最大化原理，通过最大化节点表示$\mathbf{Z}$和全局图表示$\mathbf{s}'$，节点表示$\mathbf{Z}'$和全局图$\mathbf{s}$表示间的一致性最大化两个图间的互信息，使节点表示$\mathbf{Z}$和$\mathbf{Z}'$同时学习局部和全局邻居信息。同时，在检测器空间中预先构造一个可学习的聚类矩阵$\boldsymbol{C} \in \mathbb{R}^{d' \times k}$，它由$k$个聚类质心向量$\boldsymbol{c}_i \in \mathbb{R}^{d' \times 1}$组成，节点表示$\mathbf{Z}$和$\mathbf{Z}'$经过$\boldsymbol{C}$得到聚类分配表示$\boldsymbol{Q}$和$\boldsymbol{Q}'$，最大化聚类分配表示间的一致性可以挖掘节点表示间潜在的语义信息，

将语义相似的节点聚类到同一个簇。通过在各维度下使用聚类，可以生成不同维度的特征检测器。

但针对智能电网运作环境下操作域同时处理大量异构数据的情况，我们需要对不同异构数据对应的检测器进行映射拼接。对此，我们基于生物学抗体偶联原理（图 9 左）对已有的特征检测器进行桥联（图 9 右）。

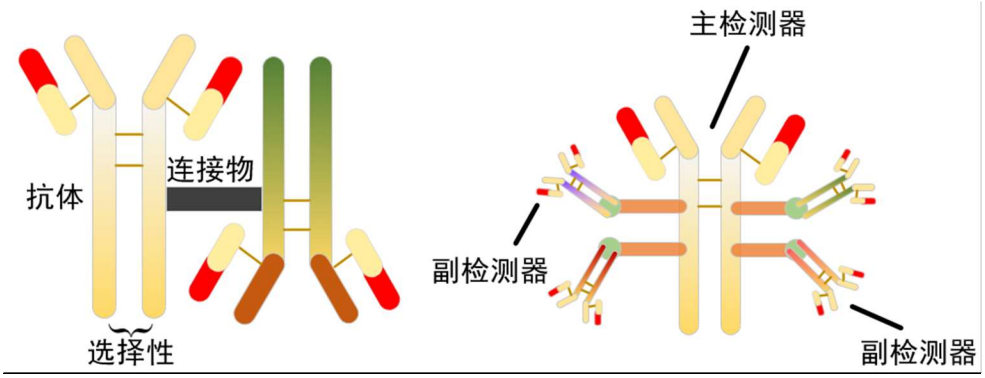


图 9：检测器的桥联技术图

结合抗体偶联原理和计算机免疫学中对于异常数据检测原理，我们实现了基于免疫偶联原理的检测器桥联方法，名词对应关系如表 4 所示：

表 4：抗体偶联与检测器桥联原理名词对应关系

抗体偶联	检测器桥联
抗原	待检测的异构数据
抗体	主检测器
连接器	桥联器
毒素	副检测器
靶细胞	异常的异构数据
抗原抗体特异性结合	桥联检测器工作
毒素注入靶细胞	副检测器检测到异常数据

其中主检测器可视为生成的桥联检测器，用于对操作域中同时涌入的大量异构数据进行检测，副检测器则为上文中生成的对于不同维度的异构数据的特征检测器集合。抗体偶联中连接器的设计，可在异常检测中类比成桥联器的生成。桥联器不仅可以将不同维度的特征检测器进行向量映射，生成一个桥联检测器，同时处理来自多个维度的异常数据。基于桥联抗体的检测机制如图 10 所示。

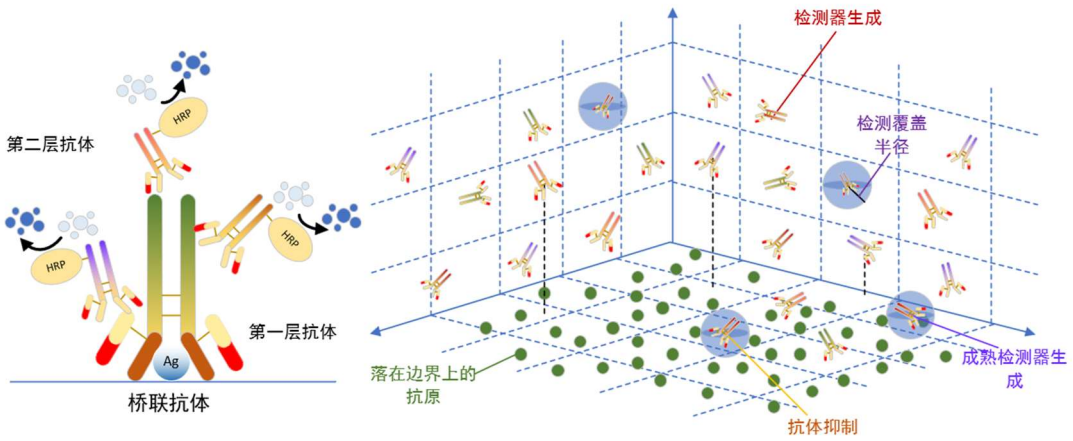


图 10：基于桥联抗体的检测机制图

桥联检测器的检测不同于单维度下的特征检测器的工作，我们通过神经网络训练，在每一个特征检测器的检测计算结果上赋予权重，具体检测步骤如下：

首先，利用已有的多维异构数据自体非自体数据集，对于不同维度下特征检测器$d_1, d_2, ..., d_n$分别进行欧式距离计算，并对于每个计算结果随机赋予一个初始化权重$w_1, w_2, ..., w_n$。当输入新的多维异构数据集时，根据公式（3）和（4），使用桥联检测器进行总检测计算，结合步骤一赋予的权重进行加权运算，我们会得到一个检

测评价W。

$$dist_i = \sqrt{\sum_{i=1}^{n}(Detect_i - Anti_i)^2} \tag{3}$$

$$W = sigmoid(\sum_{i=1}^{n}(dist_i - r_i)\, w_i) \tag{4}$$

然后，我们将计算得出的检测评价W使用 sigmoid 激活函数（公式 5 所示）输出类别标签，并与该数据的真实标签进行比对，由公式（6）计算其二值交叉熵作为损失函数。

$$\sigma = \frac{1}{1+e^{-x}} \tag{5}$$

$$L(y, f(x)) = H(p, W) = -\frac{1}{N}\sum_{i=1}^{N} p(y_i \mid x_i)\log[q(\hat{y}_i \mid x_i)] \tag{6}$$

由损失函数计算得到的损失值，我们计算参数矩阵$W^{(l)}$的梯度$\frac{\partial L(y,\hat{y})}{\partial W^{(l)}}$，通过梯度寻找模型给出的检测评价$W^{(l)}$与数据异常判断标签$p$的误差，最后将误差$\delta$反馈回每一个特征检测器$d_i$在检测评价计算中对应的权重参数$w_i$。

经过误差反复迭代后训练出桥联检测器，此时我们再输入一条多维异构数据进行检测计算得到检测评价W。若$W \geq 0.5$，检测为自体抗原，即检测数据正常，反之$W < 0.5$，抗原则为非自体，检测数据异常。并且对于每一个维度i下的特征检测器d_i，我们可以判断其是否满足$(dist_i - r_i)w_i < 0$，若满足则可认为当前数据异常的维

度为i，基于此可以实现桥联器的设计，实现异常数据的溯源。

1.3 基于免疫桥联的智能电网异常数据检测与溯源原型系统设计

基于以上步骤可以生成面向智能电网中异构数据的桥联检测器。利用桥联检测器检测与溯源原理，我们可以设计一种针对异常数据检测与溯源的原型系统，系统架构如图 11。架构分为数据传输层、数据处理层、核心算法应用层以及人机交互层，各层功能如图所示。

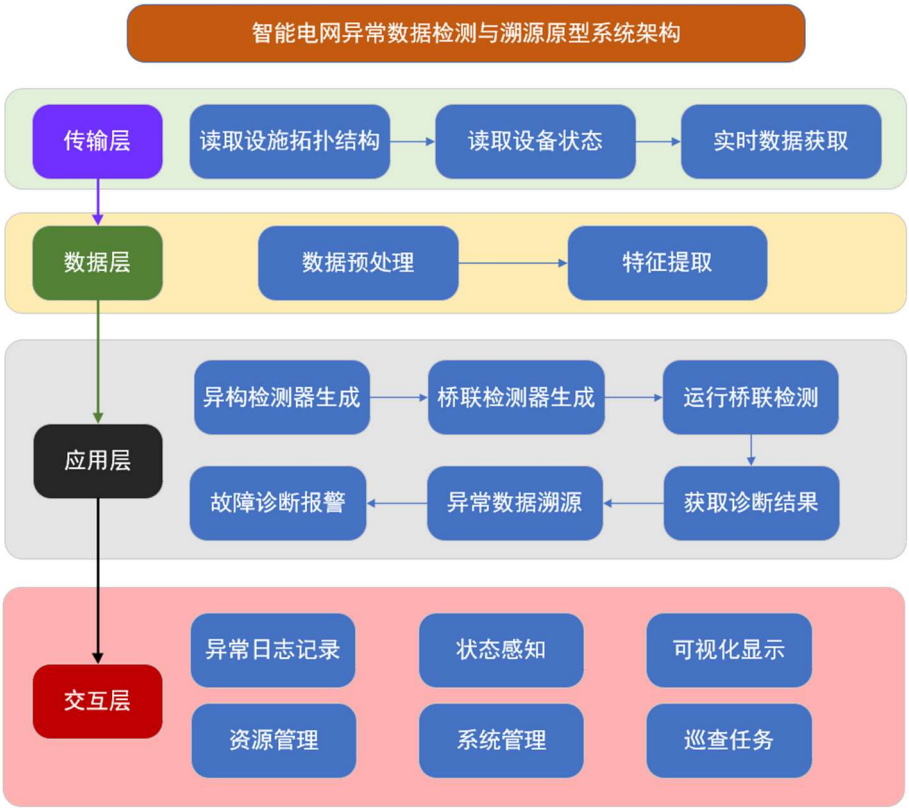


图 11：智能电网异常数据检测与溯源原型系统架构

该系统基于 SCADA 设计结构，选择 Python 作为系统开发语言，pytorch 作为选择算法训练框架，调用 Numpy，Pandas 等辅助

开发库。划分系统设计架构为以下模块：

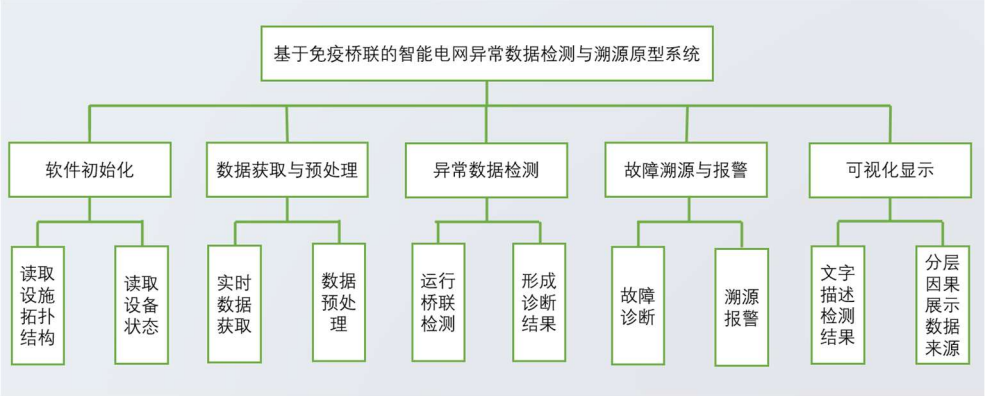


图 12：系统模块树图

①软件初始化模块：该模块调用 SCADA 接口函数读取 SCADA 数据库中的电网拓扑信息和实时设备状态，提炼后写入本地诊断用数据库中，并在人机交互界面进行集中显示。

②数据获取与预处理模块：该模块通过周期性扫描保护信息数据库，持续检查处理域中所接触的智能电网多源异构数据，将处理数据统一量纲，传入检测模块。

③异常数据检测模块：基于已生成的桥联检测器对传入的多源异构数据进行检测，并初步判断数据异常与否，形成诊断结果，若数据异常，进一步传入故障溯源与报警模块。

④故障溯源与报警模块：基于桥联检测器溯源原理对异常数据进一步故障诊断，分析异常数据来源，并进行日志记录与异常报警。

⑤可视化显示模块：用文字描述检测结果，通过分层因果图展示异常数据及来源，形成可视化结果。

本项目基于 MatLab、Python、C++集成开发环境，QT 开发框架进行设计开发，可实现高效人机交互、数据可视化采集监视等功能。效果图如下：

图 13：智能变电所异常数据检测平台效果图

2.基本原理和基本技术介绍

2.1 双特异性抗体偶联机制

双特异性抗体（即桥联抗体）是指能同时特异性结合两个抗原或抗原表位的人工抗体。双特异性抗体在自然条件下并不存在，可以通过偶联机制制备实现的。

抗体偶联是一种将高特异性抗体与高活性毒素细胞的相结合的技术，兼具药物杀伤力和强靶向性。偶联由以下三部分组成：

1.抗体：具有强特异性，能够精确对应靶点，实现抗原抗体结合。

2.毒素：能够起到杀伤细胞的作用

3.连接物：连接抗体与毒素，连接具有强稳定性。

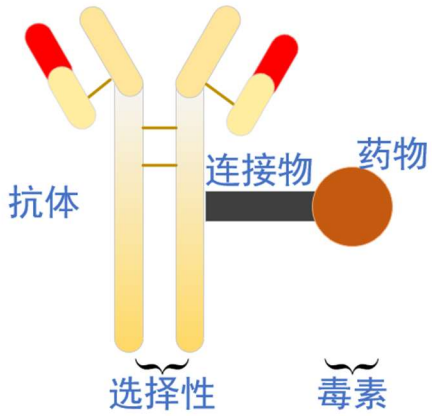


图 14：偶联组成结构

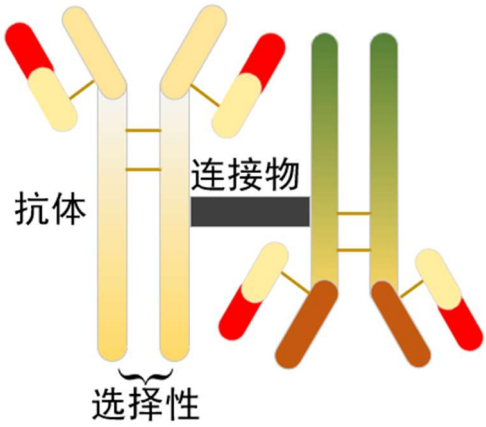


图 15：桥联抗体结构

将毒素药物换为其他具有特异性的抗体，可实现桥联抗体生成，如图 15 所示。桥联抗体可以靶向多个抗原或抗原表位，发挥的协同效应比单克隆抗体具有更多优势，同时还可介导多种特定生物学效应的发生。比如：①桥联免疫细胞与肿瘤细胞，通过招募和激活免疫细胞杀伤肿瘤细胞；②抑制或激发多个信号通路，发挥协调效应；③借助抗体双价结构，介导蛋白复合物形成，发挥生物学效应等。如图 16 所示：

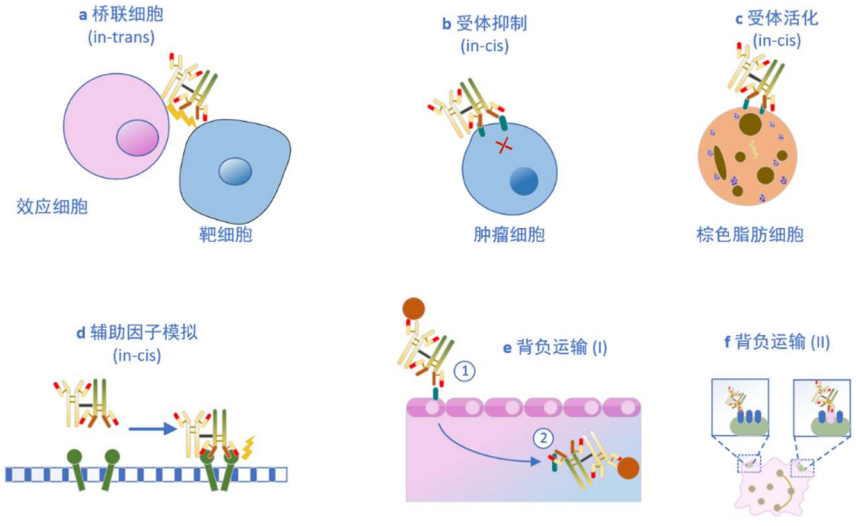


图 16：桥联抗体作用的一些实例

2.2 免疫算法（IA)

免疫算法是指以在人工免疫系统的理论为基础，实现了类似于生物免疫系统的抗原识别、细胞分化、记忆和自我调节的功能的一类算法，是一种具有生成+检测的迭代过程的群智能搜索算法。

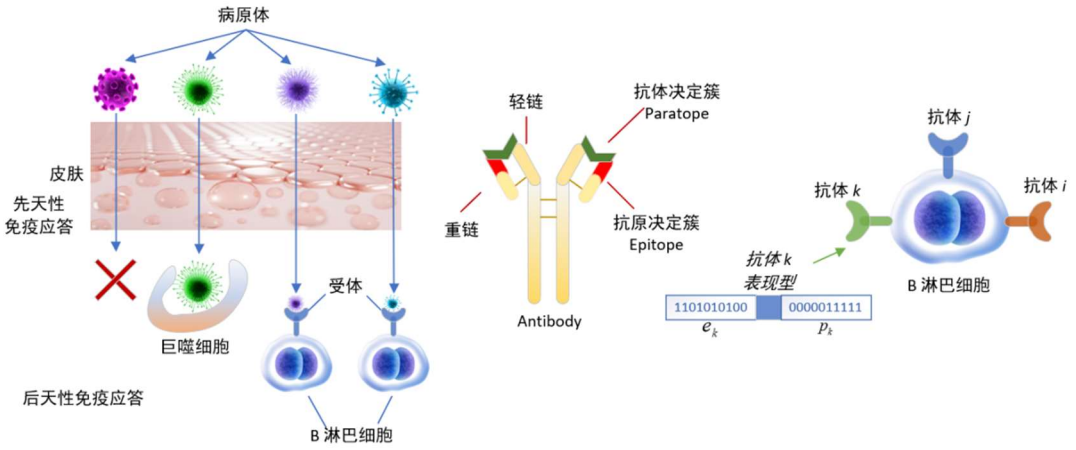


图 17：免疫算法的生物模型（左）和二进制模型（右）

在生物免疫系统（BIS）中，抗体细胞区分自身和非自身抗原以排除外部入侵。受这一机制启发，在人工免疫系统（AIS）中，数据样本被视为抗原，正常和异常样本分别被视为自身抗原和非自身抗原，检测器被视为抗体。免疫算法中的基本术语定义如下：

抗原集：从特征空间中抽取的所有字符串构成抗原集，其中 n 为数据维数，i 为第 i 个归一化属性值。

$$\mathrm{AG}=\{\mathrm{AG}|\mathrm{AG}=(x_1,x_2,\cdots,x_n),\mathrm{x}_i\in[0,1]\}$$

自体/非自体集合:自体集合(self)表示所有的自体抗原，即正常数据样本，$r_s\in R^+$是自身样本的可变性阈值。非自体集合（nonself）表示所有的非自体抗原，即异常样本。

检测器:检测器(抗体)$d=<y,r_d>$，其中y为检测器 d 的位置，表示为$y=(y_1,y_2，…，y_n)$。式中，y_i表示第i个归一化属性值，其中$y_i\in[0,1]$，且$1\leq i\leq n$，$r_d\in R^+$为d的半径(检测阈值)。检测器集表示为$D=\{d\mid d=<y,r_d>\}$

亲和力:抗原和抗体之间的亲和力是由它们之间的距离决定的。在实值特征空间中，亲和度可由公式（7）计算欧几里得距离可得。

$$affinity(ag,d)=dist(ag,d.y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \qquad (7)$$

自反应性:若有一种自身抗原位于检测器 d 的检测区域即满足$Affinity(ag_{self},d)<r_s$，则 d 为自反应性检测器。

基于以上定义,可以通过负选择算法实现检测器训练(图 18 左)和异常数据检测（图 18 右）

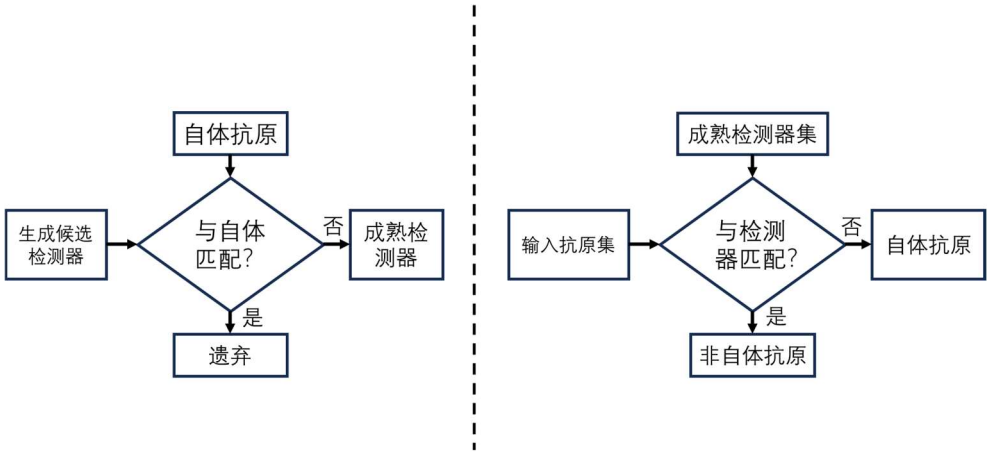


图 18：负选择算法流程图

2.3 互信息最大化聚类方法

互信息最大化聚类方法关键在于使用 JS 估计器估计互信息及可学习聚类矩阵$\boldsymbol{C}$的构造。为使用 JS 估计器估计互信息，定义优化

损失为：

$$L_{\text{node-global}} = -[I(\mathbf{Z}, s') + I(\mathbf{Z}', s)] = -\frac{1}{N}\sum_{i=1}^{N}[\log D(z_i, s') + \log(1 - D(\tilde{z}_i, s')) + \log D(z_i', s) + \log(1 - D(\tilde{z}_i', s))] \quad (8)$$

式中$D(\cdot,\cdot)$用于评估节点表示和全局图表示间的一致性，基于公式（9）使用双线性评分函数估计：

$$D(z_i, s') = \text{sigmoid}(s'\boldsymbol{W}\mathrm{z}_i^T) \quad (9)$$

在进行对比学习实现聚类时，需要构造负样本.具体方法为，按行打乱$\boldsymbol{G}$和$\boldsymbol{G}'$中的特征矩阵$\mathbf{X}$，这样可以保持图的拓扑结构不变，同时改变节点的位置。把两个扰动图$\widetilde{\boldsymbol{G}}$和$\tilde{G}'$输入各自编码器得到负样本表示$\widetilde{\boldsymbol{Z}}$和$\widetilde{\boldsymbol{Z}}'$，带入到式中计算$D(z_i, s')$和$D(z_i, s)$

可学习的聚类矩阵由 k 个聚类质心向量组成。对于原始图，根据公式（10），计算任一节点表示和 k 个质心向量间的相似。其中，$p_i(\mathbf{y} \mid z_i) \in \mathbb{R}^{k\times 1}$为聚类分配向量.

$$\boldsymbol{p}_i(\boldsymbol{y} \mid \boldsymbol{z}_i) = \text{softmax}(\boldsymbol{C}^{\mathrm{T}} \cdot \boldsymbol{z}_i^{\mathrm{T}}) \quad (10)$$

同理可以计算扩散图的$p_i'(y|z_i')$. 让z_i能预测z_i'的聚类分配向量，如下：

$$l(p_i, p_i') = -\sum_{y=1}^{k} p_i'(y|z_i')\log p_i(y|z_i) \quad (11)$$

要让z_i和z_i'彼此预测对方的聚类分配向量，定义聚类一致性损失：

$$L_{\text{consistency}} = -\frac{1}{N}\sum_{i=1}^{N}[l(p_i, p_i') + l(p_i', p_i)] \quad (12)$$

然而直接优化式(11)可能导致平凡解，使所有的样本划分到同

一个簇.

为解决这个问题，目标是让 N 个样本节点可以均匀地划分到 k 个聚类质心. 假设有 $Q=[q_1,\cdots,q_N]\in\mathbb{R}_+^{k\times N}$ 是聚类分配矩阵，为实现前面提的目标可以按式（11）优化 Q.

$$\begin{aligned}&\max_{\boldsymbol{Q}\in\mathbf{T}}\mathrm{Tr}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Z}^{\mathrm{T}})+\varepsilon H(\boldsymbol{Q}),H(\boldsymbol{Q})=-\sum_{ij}\boldsymbol{Q}_{ij}\log\boldsymbol{Q}_{ij}\\&s.t.\boldsymbol{T}=\left\{\boldsymbol{Q}\in\mathbb{R}_+^{k\times N}|\boldsymbol{Q}\mathbf{1}_N=\frac{1}{k}\mathbf{1}_k,\boldsymbol{Q}^T\mathbf{1}_k=\frac{1}{N}\mathbf{1}_N\right\}\end{aligned}\tag{13}$$

式中：Tr()为矩阵的迹，表示矩阵的主对角线之和. H 为熵函数。$\mathbf{1}_N$ 是 N 维全一向量，$\mathbf{1}_k$ 是 k 维全一向量。

式(13) 可以看作一个最优传输问题，它的解 Q 可以写成归一化指数矩阵。

$$Q=\mathrm{Diag}(\boldsymbol{u})\exp\left(\frac{C^{\mathrm{T}}Z^{\mathrm{T}}}{\varepsilon}\right)\mathrm{Diag}(\boldsymbol{v})\tag{14}$$

式中：$\boldsymbol{u}\in\mathbb{R}^{k\times1}$ 和 $\boldsymbol{v}\in\mathbb{R}^{N\times1}$ 为重归一化向量，可以使用 Sinkhorn–Knopp 算法计算 $\boldsymbol{u}$ 和 $\boldsymbol{v}$。

同理可以按式(13)计算扩散图的 $\boldsymbol{Q}'$。最后按列归一化 $\boldsymbol{Q}$ 和 $\boldsymbol{Q}'$，式(12) 可重写为：

$$\min_{p,q}L_{\text{consistency}}=-\frac{1}{N}\sum_{i=1}^{N}[l(p_i,q_i')+l(p_i',q_i)]\tag{15}$$

此外，为确保聚类质心向量尽量彼此远离，引入分离损失：

$$L_{\text{separate}}=\|CC^{\mathrm{T}}-\mathrm{Diag}(\mathbf{1}_N)\|^2\tag{16}$$

最终模型的总体优化目标如下所示：

$$L=L_{\text{local-global}}+\lambda_1L_{\text{consistency}}+\lambda_2L_{\text{separate}}\tag{17}$$

式中 λ_1 和 λ_2 是权衡系数。模型优化后，将学习到的节点表示和相加

用于节点聚类任务。

3.可行性分析

3.1 社会可行性：

随着智能电网进入大数据时代，愈发严重的数据多源异构问题暴露出传统数据采集检测系统（SCADA）部分结构功能的弊端。未来电力系统需要支持更多种通信协议，数据标准化和格式化转换将变得更加频繁且复杂。传统检测系统选用复杂的人工采集方式，对于专业技术支持要求高，这无疑增加系统的总体维护成本，不利于小型电力机构的发展。同时传统异常检测系统无法解决不同设备、传感器导致的数据多样性问题，融合后往往会出现数据跳变、数据特征缺失等问题，影响了数据检测的准确性与可靠性。不仅如此，传统的检测系统无法准确获取异常来源，不利于后期复检、维护工作的开展。为保障智能电网的稳定发展，我们急需一个数据检测系统在实时获取电力设施拓扑结构及状态数据的同时，可以对多源异构数据进行精确检测，找出异常数据来源，保证该电力机构安全稳定地运转。

3.2 技术可行性：

1）免疫算法：免疫算法采用分布式计算，可以实现高效地并行处理，处理电网数据这种大规模数据集时能保持较高的性能，对于计算机计算资源和存储资源损耗低。

2）神经网络：神经网络具有准确的表达能力、良好的泛化能力、高效的学习能力、分布的处理能力、强大的适应能力，可以帮助项目实现特征聚类和桥联检测。

3）桥联抗体：桥联抗体具有优秀的可介导时序或空间效应，生物免疫与计算机免疫机理上又具有高度相似，基于桥联抗体原理可以实现对异常数据的高效检测与准确溯源，是多学科交叉的新兴领域。

4）互信息最大化聚类方法：互信息最大化聚类方法直观反映变量依赖关系，可以有效衡量样本的相似性，得到可解释的聚类结果，为后续检测器桥联分析与应用提供支持。

5）实验平台成熟可靠：本项目依托于四川大学网络空间安全靶场，拥有各种先进的仪器设备，具有一流的实验环境与雄厚的科研实力。同时，项目组指导老师何俊江老师在计算机免疫学领域也深耕多年，经验丰富。在出色的硬件设施以及专业的指导下，本项目将能够顺利地开展推进。

三、研究基础（对项目的参与动机、已有知识储备、相关研究和训练基础

1.项目参与动机

1.1 学习意义：

参与本项目组实验的同学通过本次项目的学习与实践,可以拓展自己对于计算机领域内的相关知识，极大地提升个人综合能力：

1）在本次项目中，项目组成员自主学习计算机安全领域相关知识,对于计算机免疫学、异构数据领域的前沿知识都能有深刻的认识,可以学习到神经网络、人工智能等本科阶段不易接触到的知识，拓宽眼界、增长兴趣的同时，也为成员未来选择研究深造方向起到极大的引导作用。

2）在项目准备阶段，由于涉及核心技术与算法对于本科生而言晦涩难懂，所有成员均利用课余时间积极调研背景与国内外研究现状，研究能力得到十足提升，专业素养得到极大培养，为以后的科研工作奠定扎实基础。

3）通过本次项目，项目组成员可以培养团结协作的集体精神，通过对项目的合理化分工，提高了整体的完成效率，这对于组员未来的实习与工作打下基础。

1.2 社会意义

目前，我国正在积极构建新型电力系统，智能电网安全问题极大关联着国家安全,而多源异构数据所带来的安全隐患迟迟得不到好的

解决方法。本项目旨在通过改进检测器的生成，特征检测器聚类，以及检测器检测与异常溯源的方法,基于生物免疫通过桥联提高检测的各项性能，以期望降低异常异构数据给电网安全带来的威胁。在当下电网大数据时代，异构问题层出不穷，但目前却没有一个兼顾效率与质量的解决方法，该项目将能够填补相关方面的空白。

1.3 科学意义

随着人工智能的持续发展，深度学习、神经网络等技术不断被投入实际应用，在许多领域大放异彩。许多交叉领域也成为科学家的科研方向,本项目将生物医学中的双特异性抗体与计算机异常检测相关知识融合，实现知识转移，进一步丰富了计算机免疫学的内容。在项目中学习理解相关技术,也有助于我们在实践中对它们进行猜想与验证，我们也有机会能竭尽所能在相关前沿领域完善这些技术。

2.已有知识储备

项目成员三位来自 2022 级网络空间安全学院，一位成员来自 2022 级数学学院。所有成员均有一定编程基础，懂得 C，Python，Java 等程序设计语言的编写，部分成员有 web 应用开发（html、CSS，JavaScript），Matlab，C++等方面的知识储备。所有成员都具有一定项目实践能力，能够确保项目的顺利开展。

吴昊：

学习刻苦踏实，工作认真负责，大一学年综测年级排名第六，有学科竞赛、开发项目竞赛获奖经历,具有良好的团队领导与协调能力，

能够胜任团队中许多工作，正在深入学习免疫算法、神经网络等计算机免疫学相关知识。

吕康禾：

学习刻苦，请于思考，逻辑严密，工作认真负责，具有较好的数学基础，正在深入学习负选择算法，人工免疫方面的知识。参加过数学建模国赛、基础数学研究项目、四川大学数学学院“小火花”科研等项目。

林宸：

擅长编程，数学（曾获校级竞赛一等奖），能够熟练解决编程问题，熟悉计算机操作系统，计算机网络，计算机组成原理等相关知识，现正在深入学习数据结构与算法，数论与密码学等。曾开发过网站，并尝试租用云服务器分享搭建的网站。

吴桐曦：

学习认真刻苦，工作积极认真，有学科竞赛获奖经历。大一学年综测年级排名第五，具有成熟的沟通协作能力以及奉献精神。如今正在学习 web 安全和深度学习的内容。

林熙金：

学习踏实刻苦，勤奋好学，担任班长，具有团队精神。高中数学竞赛二等奖，乐于研究钻研。正在深入学习免疫算法、神经网络等计算机免疫学相关知识，参加过团队游戏项目。

四、研究计划和进度（就文献查询、社会调查、方案设计、实验研究、数据处理、研制开发、撰写论文或研究报告、结题和答辩、成果推广、论文发表、专利申请等工作逐项计划时间，时间节点精确到月份）

第一阶段：2023 年 8 月-2023 年 11 月，通过查询文献资料，了解国内外目前解决智能电网中异构数据相关问题的研究现状;学习双特异性抗体结构及基本原理，学习计算机免疫学、机器学习、神经网络、聚类算法等相关基础知识：包括所使用的相关算法、原理、模型以及 Python 的实现方法，实现知识迁移；初步设计原型系统框架及 UI 界面；在老师指导下撰写申报书。

第二阶段：2023 年 12 月-2024 年 2 月：搜集牵引变电所异构数据集，完成异常检测器模型的设计与优化，并基于数据集训练模型。

第三阶段：2024 年 3 月-2024 年 6 月：初步实现特征检测器的选取与桥联，以此生成检测、溯源集成系统。对系统进行大量测试，进一步提高性能以及完善体验。

第四阶段：2024 年 7 月-2024 年 8 月：总结整理项目研究成果。由项目老师指导，审核结果，准备进行结题考核。

五、项目研究支撑条件（项目所依托的重点实验室（中心、平台）、双创平台、课题组等各类单位能提供的直接支持项目开展的软硬件设施和其他校内外资源）

1.校内环境支撑：

本项目依托于四川大学网络靶场协同创新中心,靶场内置有高性能计算机 10 台，高性能计算服务器 8 台，提供了可定制、可扩展的工业系统仿真环境，具有一流的实验环境与雄厚的科研实力。

2.校外电力单位支持：

本项目前期背景调研在云南某牵引变电所指导下完成，经沟通，该牵引变电所也会在项目后续进展中提供科学指导和必要支持。

3.课题组有力的指导：

指导老师在多模态数据融合、异构数据处理、异常检测等研究领域内深耕多年，相关经验丰富。该项目与指导老师的研究方向契合，能够保证在研究过程中出现的问题都能够得到老师以及课题组内学长的有力指导。

六、预期成果形式（可多选）

（1）☑SCI 论文 1 篇

（2）□核心期刊论文__篇

（3）□会议论文__篇

（4）□内部编印期刊论文__篇

（5）□授权发明专利__项

（6）☑ 申请发明专利 1 项

（7）□创新类竞赛获奖

（8）□创业类竞赛获奖

（9）□其他　　名称：_____

七、经费预算（按申报项目目标任务需要及学校预计划拨经费进行预算，经费执行情况将与结题考核成绩挂钩）

经费预算（单位：元）

（1）仪器设备费_____

（2）耗材费 1000

（3）测试加工费 3000

（4）国内会务及差旅费_____

（5）国外会务及差旅费_____

（6）文献/知识产权事务费 3000

（7）办公费（含文印、办公用品等） 1000

（8）其他费用 2000

合计 **10000**

八、评审情况

指导教师意见： 该项目基于生物免疫抗体桥联机制，设计了一种针对智能电网多源异构数据的异常检测与溯源系统。通过模拟“单个免疫抗体可以识别不同种类病毒”机制，提出一种解 x 决多源异构数据异常检测问题，项目创新性强、可行性高。申报书撰写层次分明，结构合理，叙述清楚。团队成员组成合理，成员工作热情高，工作认真负责。 **指导教师（签名）：** 何俊江 **2023 年 12 月 3 日**
学院推荐意见： 主管院长签名：　　年　月　日
学校专家评审意见： 组长签名：　　年　月　日
学校认定意见及批准经费：

学校负责人签名：　　年　月　日