**赛尔网络下一代互联网技术创新项目**

**结题工作报告**

|  |  |
| --- | --- |
| 项目编号： | NGII20170703 |
| 项目名称： | IPv6环境下支持隐私保护的医疗大数据挖掘方法研究 |
| 技术领域： | IPv6智能医疗 |
| 承担单位： | 陕西师范大学 （盖章） |
| 项目负责人： | 王寒 |
| 指导教师： | 李蜀瑜 |
| 项目期限： | 一年 |
| 填报日期： | 2019年 05 月 12 日 |
| 联系人： | 陈竑毓 |
| 联系电话： | 18072210523 |

**中国教育和科研计算机网CERNET网络中心制**

**二〇一七年五月**

**填 写 说 明**

一、请严格按照表中要求填写各项，用A4纸打印（单双均可），按照要求逐一签字盖章。

二、本报告的“项目名称”和“项目编号”等项目基本信息应与《任务合同书》一致；已做调整的按照调整批准后的信息填写。

三、本报告第一次出现外文名词时，要写清全称和缩写，再出现同一词时可以使用缩写。

四、编写人员应客观、真实地填报，尊重他人知识产权，遵守国家有关知识产权法规。

**项目基本信息**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | | | IPv6环境下支持隐私保护的医疗大数据挖掘方法研究 | | | |
| 技术领域 | | | IPv6智能医疗 | | | |
| 项目期限 | | | 一年 □两年 | | | |
| 成果类型 | | | □硬件 □软件平台 软件应用系统 □APP □网站 专利 □技术标准 开发源码 软件著作权 论文论著 □其他 | | | |
| 申请  单位  信息 | | 单位名称 | | 陕西师范大学 | | |
| 通讯地址 | | 西安市长安区西长安街620号 | 邮政编码 | 710119 |
| 所在地区 | | 陕西省西安市 | 组织机构代码或统一社会信用代码 | 43523253-x |
| 校科技主管部门 | | 陕西师范大学科学技术处 | 联系电话 | 029-85310340 |
| 电子信箱 | | kjc@snnu.edu.cn | 传真号码 | 029-85310230 |
| 项目  申请人  信息 | 学生 | 姓名 | | 王寒 | 性别 | 男 |
| 出生日期 | | 1994-01-01 | 年级 | 2016级硕士研究生 |
| 所学专业 | | 软件工程 | 所在院系 | 计算机科学学院 |
| 固定电话 | | 029-85310552 | 移动电话 | 13892900811 |
| 传真号码 | | 029-85310554 | 电子信箱 | wanghan061@163.com |
| 证件类型 | | 身份证 | 证件号码 | 610481199401010037 |
| 指导教师 | 姓名 | | 李蜀瑜 | 性别 | 男 |
| 年龄 | | 41 | 职称/学位 | 副教授**/**博士 |
| 专业领域 | | 计算机软件 | 所在院系/部门 | 计算机科学学院**/**网络信息中心 |
| 移动电话 | | 13572128758 | 固定电话 | 029-85310552 |
| 传真号码 | | 029-85310554 | 电子邮箱 | lishuyu@snnu.edu.cn |
| 证件类型 | | 身份证 | 证件号码 | 610302197801202011 |
| 资助经费（万元） | | | | | 10.00 | |

# 项目计划执行情况

1. 项目基本情况

本项目依托新兴的智慧医疗行业，设计了IPv6环境下支持隐私保护的疾病预测系统。本项目的基本情况如下：

1. 基于深度学习的隐私保护癫痫发病预测算法的设计

针对癫痫病实时发病预测的需求，基于深度学习LSTM模型设计一种预测算法，该算法根据当前时刻患者的实时ECG数据进行二分类，用于预测患者当前处于发病/间歇状态。同时考虑到当前大数据分析技术的隐私保护要求，结合差分隐私技术提供对训练数据集的隐私保证。该算法的设计已完成。

1. 基于小波聚类的隐私保护心脏病、帕金森患病预测算法的设计

针对心脏病和帕金森病的患病预测需求，结合差分隐私模型设计了一种满足隐私保护的小波聚类算法，该算法根据心脏病和帕金森病数据集，考虑组内共同特征与患者的关联性，通过多标签聚类预测患者患病状态。同时考虑到存储空间消耗低以及隐私保护要求，采用了元胞数组降低存储空间并结合差分隐私技术为提供隐私保护。该算法的设计已完成。

1. IPv6环境下支持隐私保护的疾病预测系统的架构设计与实现

考虑实际应用场景，设计一个IPv6环境下的疾病预测系统，并将上述两个隐私保护疾病预测算法应用到该系统中，系统的功能模块包括账户管理、挂号管理、文件管理、算法模型训练及其可视化、基于预测算法的疾病实时预警、医患社区等多个方面。该项工作已完成。

1. 项目执行情况（含研究目标、研究内容调整和变动情况）

**研究目标**

分析IPv6、隐私保护技术与医疗大数据挖掘的融合，研究疾病预测系统的架构设计和层次划分，首先，针对隐私保护的需求，研究医疗大数据挖掘方法与差分隐私技术的有机结合；其次针对疾病实时发病预测的需求，以癫痫病为例，引入基于深度学习LSTM模型的发病预测算法，学习具有一般性的癫痫发病规律，为疾病的实时发作状态判断提供支持；接着针对疾病患病预测的需求，以心脏病和帕金森为例，引入基于小波聚类的患病预测算法，为疾病患病预测研究提供依据；最后将理论研究与实际应用相结合，设计实现具有实际效用的应用系统。

**研究内容调整和变动情况**

无

# 项目研究取得的创新点与突破

1. 基于深度学习的LSTM模型，设计了一种支持隐私保护的医疗大数据分类算法。算法通过改进模型的损失函数，以向梯度值中加入正则项的方式提高了模型泛化能力，并向模型更新过程中的计算的平均梯度添加噪声实现对训练数据集的隐私保护。结合癫痫实时发病的预测需求，设计实现支持隐私保护的癫痫发病预测算法，算法根据患者当前实时的ECG数据，对比模型学习到的癫痫病发病规律，预测患者当前的癫痫发病状态，弥补了传统的人力判断不及时的不足，为发病后医生或患者在第一时间采取预防保护措施提供依据。
2. 基于差分隐私模型和小波聚类算法，设计了一种支持隐私保护的医疗大数据聚类算法。算法使用元胞数组来改进数据存储方式以降低存储空间，并结合Laplace机制分别在量化步骤和网络识别步骤添加噪声扰动以达到数据隐私保护，最后对加噪后数据进行小波变换并利用广度优先搜索实现患病标签查找。结合心脏病和帕金森病患病统计预测需求，在有效降低存储空间的基础上，考虑组内共同特征和患者的关联性，预测患者患病状态，便于医生掌握当前患者状态从而进行有效诊断。
3. 基于IPv6网络环境，设计并实现了一个支持隐私保护的疾病预测系统，利用上述算法设计实现疾病预测模块并应用于系统中，通过对该系统的部署和实际运行来验证研究成果的有效性。

# 项目研究取得的成果及成果转化与应用情况

1. 项目研究取得的成果（成果请参照任务合同书中“预期成果和考核指标”逐条说明）

|  |  |  |
| --- | --- | --- |
| **成果类型** | **预期成果** | **实际成果** |
| IPv6支持程度 | 本项目成果在纯IPv6网络环境下运行 | 已按要求完成 |
| 软件应用 | 验收技术指标：设计并实现了一个支持隐私保护的医疗大数据挖掘系统，同时最多允许100个并发用户，分类和聚类准确率达到85%以上。  是否开源：是 | 已按要求完成 |
| 专利 | 申请专利1项 | 申请专利2项，已受理 |
| 软件著作权 | 申请软件著作权2项 | 申请软件著作权5项 |
| 论文论著 | 在国内学术期刊或国际会议发表论文3篇 | 在国内核心学术期刊发表论文6篇 |
| 开放源码 | 拟开放全部源码并上传到Github开源社区 | 已开放全部源码并上上传到Github开源社区 |

1. 成果转化与应用情况

本项目的软件系统IPv6环境下支持隐私保护的疾病预测系统针对当下IPv4网络存在地址获取不足、安全可信度差、服务质量低等制约智慧医疗发展的问题提出了解决方案。其中数据平台、预警平台等功能可以很好的实现疾病实时预测以及患病状态预测，解决了患者隐私泄露等安全问题，为医生诊断提供了理论保障和有效依据；管理平台、医患社区等功能可以很好的解决当前面临的医疗资源短缺的问题，为更多的患者提供更优质的医疗资源。该系统很好的利用了医疗大数据挖掘技术，大大提高了患病诊断准确率，为更多患者提供了医疗保障。

# 人才培养情况

项目团队共培养硕士生6人，其中已毕业2人，在校4人，毕业生王寒同学已签约浦发银行总行科技部开发服务中心。截止目前，团队分别围绕隐私保护分类方法和隐私保护聚类方法两个研究方向，发表论文6篇，申请软件著作权5项，专利2项；并参与2次下一代互联网大赛创新大赛，分别获得2016年“第二届下一代互联网技术创新大赛”地区初赛二等奖，2017年“第三届下一代互联网创新大赛”全国决赛网络技术应用甲组三等奖。

# 预算完成情况

# 希望继续获得资助，提出项目下一步工作计划

1. 算法在提供隐私保护的同时，会降低其效用，在下一步的工作中，计划改进差分隐私结合方式，实现在同等隐私保护程度下，提升数据效用。
2. 系统当前支持的用于癫痫发病预测的深度学习算法的模型结构及优化方法等已经固定，在下一步的工作中，拟提供多种学习模型、优化方法、激活函数等供用户自主选择，通过构建不同模型结构能实现对更多疾病的支持，以及提高预测准确度。

# 相关人员与部门意见

项目负责人（签字）：

指导教师（签字）：

主管部门（负责人签字或部门签章）：

年 月 日

# 承担单位意见

单位盖章：

年 月 日