

填表说明

一、本报告中相关的技术或数据如涉及知识产权保护、军工项目保密等内容，请作脱密处理。

二、请用宋体小四字号撰写本报告，可另行附页或增加页数，A4纸双面打印。

三、表中所涉及的签名都必须用蓝、黑色墨水笔，亲笔签名或签字章，不可以打印代替。

四、同行专家业内评价意见书编号由工程师学院填写，编号规则为：年份4位+申报工程师职称专业类别(领域)4位+流水号3位，共11位。

一、个人申报

(一) 基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】

1. 对本专业基础理论知识和专业技术知识掌握情况(不少于200字)

知识掌握：

1. 基础知识：系统掌握机器学习、深度学习及工业控制（PLC）等领域的核心知识，熟悉RNN、LSTM等序列模型，CNN、U-Net、YOLO 等图像处理模型，精通 Transformer 架构及各类生成式模型，包括扩散模型、生成对抗网络（GAN）等。同时，熟练掌握 PyTorch、TensorFlow 等主流深度学习训练框架，并具备扎实的 Python、C/C++ 编程基础。

2. 大模型训练与优化：能够跟踪国内外技术前沿，熟悉行业最新标准、生产流程、政策法规及先进工艺。深入研究并掌握大模型训练与调优技术，熟练使用 LoRA、P-Tuning、KV-Cache 等微调方法，以及 DeepSeek R1 提出的 MLA（Multi-Head Latent Attention）架构。精通低精度训练、分布式训练技术及大模型强化学习（RLHF）。具备丰富的大模型检索增强生成（RAG）经验，并精通知识图谱构建与优化技术。

3. 工程实践与默会性知识：在实际工程实践中积累了丰富的隐性知识，包括项目管理、故障排查、性能优化等方面的经验。能够在复杂工程环境下灵活应用所学知识，并具备跨学科问题解决能力。

4. 交叉学科融合能力：具备多领域交叉研究能力，能够将控制系统与计算机视觉相结合，具有医学因果预后模型等技术创新。

专业技术能力：

1. 工程实践经验：在中组部指定的企业实践过程中，完整参与多个项目的全生命周期，包括项目立项、方案设计、技术实现及最终落地。作为核心技术人员，承担关键研发任务，具备扎实的工程实践经验，能够独立应对压力与挑战，在实际工作中不断提升自身适应能力和决策能力。

2. 技术工具与数据分析：精通工程建设和研究所需的专业软件及多种操作系统，熟练掌握数据爬取、数据分析、自动化处理技术。能够综合运用先进仪器设备、企业现场数据采集方法以及大数据分析工具，为工程项目提供可靠的数据支撑。

3. 项目关键问题解决能力：具备强大的问题分析与解决能力，能够利用 RAG 技术构建医学专用大模型及数据库，优化智能问答系统，提高医学知识的精准检索和智能生成能力。

4. 团队协作与沟通能力：具备良好的团队合作精神和沟通能力，能够有效指导团队成员，促进跨学科合作。例如，在医学体检报告 OCR 识别项目中，与团队紧密配合，共同推动项目顺利落地。

5. 工程创新与市场研究：在企业实践过程中，曾主导市场竞品调研，并基于调研结果制定创新技术方案。具备敏锐的工程创新意识，能够结合行业趋势进行产品优化及技术迭代，推动技术成果转化。

6. 国际化视野与跨文化交流：具有国际视野，关注全球技术发展趋势，具备跨文化沟通与合作能力。

2. 工程实践的经历(不少于200字)

2023.02 - 2023.08 | 之江实验室 | 基于时序数据的阿尔茨海默症疾病演变轨迹分析
在之江实验室期间，围绕阿尔茨海默症（AD）疾病的演变轨迹分析开展研究，重点探索基于因果发现和动态建模的方法，以提高疾病干预的科学性和可解释性。

- 提出了一种结合因果发现与动态建模的疾病干预效应分析算法，用于量化不同干预措施对阿尔茨海默症发展轨迹的影响；

- 通过反事实推断模拟不同治疗方案下的病程演变，为临床医生提供辅助决策支持；实验结果表明，该方法能够有效提升干预策略的合理性，有助于个性化治疗方案的制定。

2023.08 - 2024.08 | 上海微创集团（中组部项目） | 大模型项目组

在上海微创集团期间，深入参与医学大模型与知识图谱相关项目，负责知识图谱构建、医学信息抽取、智能诊断与生成式 AI 研究，推动医学领域 AI 赋能应用的落地。

1. 医学知识图谱的构建：设计医学知识图谱

schema，整合多源医学数据，包括爬取的医学文献、医学指南及真实医疗数据；采用高效的数据清洗、实体抽取与关系构建方法，完成大规模医学知识图谱构建；知识图谱涵盖疾病、症状、检查、药物等医疗概念，支持智能问答、医学检索和辅助诊断。

2. 基于知识图谱的医学体检报告 OCR 解析与检查：利用 OCR 技术

自动提取体检报告信息，解析关键医学指标；基于构建的知识图谱，将 OCR 提取的医学数据与标准医学检查项进行匹配；结合医学规则和知识推理技术，判断健康状况，自动生成个性化检查建议和诊断意见，辅助医生决策。

3. 基于知识图谱的大模型检索增强生成系统：结合 知识图谱 + RAG+

可微调大模型，构建医学领域智能问答与知识检索增强系统，解决医学大模型回答过程中幻觉问题，提高问答的准确性、可解释性和可控性；设计前端交互界面，实现用户医疗问答和健康助手小程序，为医学领域提供智能化的知识服务。

3. 在实际工作中综合运用所学知识解决复杂工程问题的案例（不少于1000字）

基于知识图谱的大模型检索增强生成系统——解决医学问答幻觉问题的工程实践背景与挑战

在医学问答系统中，确保回答的准确性、可靠性和可解释性至关重要。然而，现有的大语言模型（LLM）普遍存在“幻觉”问题，即在缺乏相关知识或无法完全理解问题时，仍然会生成看似合理但实际错误的回答。这一问题主要源于以下几个方面：

1. 意图理解不完善：模型可能无法精准把握用户查询的真实意图，导致回答偏差。

2. 训练数据局限性：现有数据可能包含错误、主观偏见或覆盖不足，影响模型的知识广度和深度。

3. 知识边界限制：大模型的参数和训练数据决定了其知识上限，无法突破其训练时未接触的信息。

4. 知识回忆能力不足：即使模型包含某些知识点，也可能因推理能力不足而无法正确调用。

针对这些问题，本项目提出 基于知识图谱的检索增强生成（Retrieval-Augmented Generation, RAG）

方案，以优化大模型的知识获取能力，减少幻觉现象，并提升医学问答的准确性和可解释性。

核心技术方案

1. 知识图谱构建与优化

本项目通过爬虫技术采集医学相关数据，并结合结构化处理方法，构建规模达 28,000+ 实体、380,000+ 关系 的高质量医学知识图谱，提供精准的语义支持。

- 挑战：

(1)

许多医学网站采用动态加载技术，传统爬虫难以直接获取完整数据，可能出现数据缺失、重复或空值问题。

(2) 部分医学知识由人工录入，存在数据结构不一致、属性缺失等情况。

•解决方案:

(1) Selenium

浏览器模拟: 通过模拟用户操作(如点击、滚动、等待加载等), 确保网页内容完全加载后再进行爬取, 从而获得完整的医学数据。

(2) AC

自动机匹配算法: 用于精准识别文本中的疾病名称、症状属性, 并补全实体间的关联关系, 确保知识的完整性和一致性。

2. 嵌入模型与高效检索

为提升知识检索的精准度, 本项目采用嵌入模型(Embedding Model)

对用户查询与数据库内容进行编码, 并进行语义匹配。

•挑战: 传统基于关键词匹配的检索方式难以捕捉用户查询的语义信息, 导致查询结果不准确。

•解决方案:

(1) 采用高性能的语义嵌入模型(如gte-large-

zh), 将用户输入与知识库进行向量化编码, 并通过最近邻搜索(ANN)高效检索相关医学知识。

(2)

灵活可配置的接口设计: 用户可根据本地服务器的显存和计算能力, 自主选择最适合的嵌入模型, 实现计算资源的优化利用。

3. 大模型加载与多样化部署

考虑到不同用户的计算环境 and 应用需求, 系统支持三种大模型部署方式:

(1)

本地加载: 支持用户在本地运行大模型, 实现数据完全私有化, 适用于高隐私需求的医疗场景。

(2) Ollama 部署: 基于 Ollama 进行轻量级推理, 降低本地计算资源消耗。

(3) API 调用: 集成外部大模型 API, 如 OpenAI GPT-4、Claude、Qwen或 Deepseek, 实现高性能云端推理。

该设计方案确保系统高度自定义与可扩展性, 满足不同硬件条件和应用场景的需求。

4. 大模型微调框架

大模型虽然具备强大的自然语言处理能力, 但其在生成 Cypher

查询语言和医学实体提取方面仍存在优化空间。因此, 我们设计了高效的微调框架, 以提升大模型在医学领域的适应性。

•挑战:

(1) 现有大模型多为通用预训练模型, 在医学问答和数据库查询方面可能存在不足。

(2) 直接微调完整模型需要大量计算资源, 难以大规模部署。

•解决方案: 提供多种轻量级微调方案, 用户可根据需求灵活选择:

(1) BitFit: 仅调节偏置参数, 实现低计算成本的优化。

(2) LoRA (Low-Rank

Adaptation): 通过低秩适配技术, 在显存占用较低的情况下实现模型微调。

(3) Prompt-Tuning: 优化模型的提示词设计, 使其更好地理解医学查询任务。

5. 多策略检索增强生成(RAG)架构

设计了三种检索增强生成策略, 结合知识图谱、自然语言处理(NLP)和深度学习, 提升医学问答系统的准确性和可解释性:

(1) 基于 Cypher 结构化查询:

•采用预设 Cypher 查询模板, 结合大模型填充或向量化索引, 返回结构化的知识三元组。

•优势: 准确率高, 适用于医学诊断、病理分析等高精度任务。

(2) 基于提示词工程 (Prompt Engineering) :

- 先调用大模型获取知识图谱结构, 再让模型基于该结构生成 Cypher 查询, 检索后返回增强知识。

- 优势: 查询效率高, 适用于快速响应的医疗辅助系统, 但对模型性能要求较高。

(3) 基于传统 NLP 方法:

- BERT

进行意图识别, 结合CRF (条件随机场) 命名实体识别, 自动填充对话策略, 实现智能问答。

- 优势: 适用于对话式医疗咨询, 提升系统的语义理解能力。

6. 前后端设计与用户交互

- 后端: 采用 Flask 和 FastAPI 设计高效 API, 支持高并发访问。

- 前端: 采用 Streamlit

构建交互界面, 支持用户自由切换大模型、检索模式、参数配置, 实时优化系统体验; 允许用户根据实际需求调整模型加载方式、查询优化策略, 提升系统灵活性。

应用场景

本项目的设计和实施, 是一个典型的跨学科工程问题求解案例, 在实际工作中, 利用机器学习、知识图谱、NLP 等多种技术手段, 综合解决复杂医学问答系统的挑战, 目标场景包括:

- 医学智能问答: 辅助医生快速查询医学知识, 提供精准的病症分析和诊断建议。

- 临床决策支持: 结合病历数据, 提供基于知识图谱的医学推荐, 提高医疗决策效率。

本项目通过知识图谱 + RAG +

可微调大模型的融合方法, 为医学问答系统提供了高准确性、低幻觉风险、可自定义配置的智能增强方案。无论是本地化部署、API

调用, 还是自定义微调, 都能满足不同场景的需求。

(二) 取得的业绩(代表作)【限填3项, 须提交证明原件(包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等)供核实, 并提供复印件一份】

1. 公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】

成果名称	成果类别 [含论文、授权专利(含发明专利申请)、软件著作权、标准、工法、著作、获奖、学位论文等]	发表时间/授权或申请时间等	刊物名称/专利授权或申请号等	本人排名/总人数	备注
基于时序数据的因果发现与疾病发展轨迹预测系统及方法	发明专利申请	2023年12月21日	申请号: CN 2023117639 00.7	2/3	

2. 其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】

成果名称: 基于知识图谱的大语言模型检索增强生成系统

成果类别: 样机/产品

本人排名/总人数: 1/3

(三) 在校期间课程、专业实践训练及学位论文相关情况	
课程成绩情况	按课程学分核算的平均成绩： 86 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间： 1.5 年(要求1年及以上) 考核成绩： 82 分
本人承诺	
<p>个人声明：本人上述所填资料均为真实有效，如有虚假，愿承担一切责任，特此声明！</p> <p style="text-align: right;">申报人签名：张文卓</p>	

浙江大学研究生院
攻读硕士学位研究生成绩单

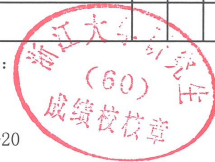
学号: 22260317	姓名: 张文卓	性别: 男	学院: 工程师学院	专业: 电子信息	学制: 2.5年						
毕业时最低应获: 24.0学分		已获得: 27.0学分		入学年月: 2022-09	毕业年月:						
学位证书号:			毕业证书号:		授予学位:						
学习时间	课程名称	备注	学分	成绩	课程性质	学习时间	课程名称	备注	学分	成绩	课程性质
2022-2023学年秋季学期	工程技术创新前沿		1.5	87	专业学位课	2022-2023学年冬季学期	智能化仪器软硬件系统设计与应用		1.0	86	专业学位课
2022-2023学年秋季学期	研究生英语能力提升		1.0	免修	跨专业课	2022-2023学年冬季学期	新时代中国特色社会主义思想理论与实践		2.0	91	专业学位课
2022-2023学年秋季学期	研究生英语基础技能		1.0	免修	公共学位课	2022-2023学年冬季学期	产业技术发展前沿		1.5	82	专业学位课
2022-2023学年秋季学期	研究生英语		2.0	免修	公共学位课	2022-2023学年春季学期	自然辩证法概论		1.0	93	专业学位课
2022-2023学年秋季学期	现代信号处理技术与应用		1.0	87	专业学位课	2022-2023学年春季学期	智能医疗技术应用案例分析		1.0	88	专业学位课
2022-2023学年冬季学期	生物医学工程方法学		1.0	92	专业学位课	2022-2023学年春季学期	工程伦理		2.0	92	专业学位课
2022-2023学年秋季学期	研究生论文写作指导		1.0	86	专业选修课	2022-2023学年春季学期	高阶工程认知实践		3.0	90	专业学位课
2022-2023学年秋季学期	数据分析的概率统计基础		3.0	74	专业选修课		硕士生读书报告		2.0	通过	
2022-2023学年冬季学期	嵌入式系统设计		2.0	83	专业选修课						

说明: 1. 研究生课程按三种方法计分: 百分制, 两级制 (通过、不通过), 五级制 (优、良、中、及格、不及格)。
2. 备注中“*”表示重修课程。

学院成绩校核章:

成绩校核人: 张梦依

打印日期: 2025-03-20



产品成果证明

证明方：上海微创卜算子医疗科技有限公司

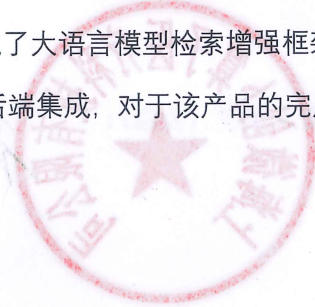
被证明方：张文卓同学

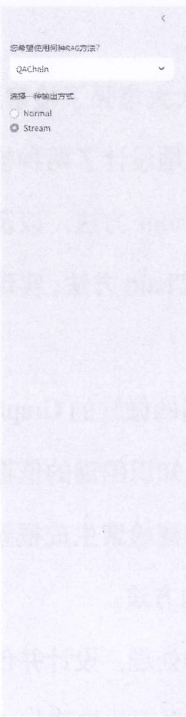
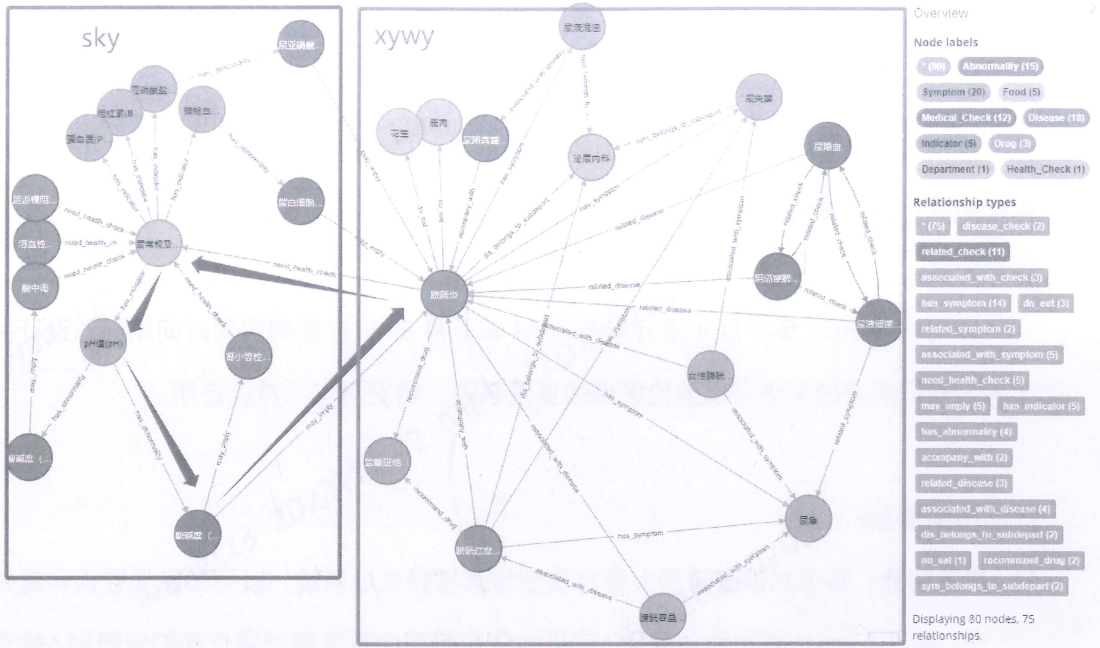
证明内容：

该同学于 2023 年 9 月 4 日开始至 2024 年 6 月 30 日在我司实习时间期间，设计并开发了基于知识图谱的大语言模型检索增强生成系统，情况属实，特此证明。

产品相关信息：

- 1) 产品功能：**基于知识图谱的大语言模型检索增强生成系统，以可靠医学数据构建知识图谱，通过 Cypher 查询和向量化，实现大语言模型的垂直领域强化和幻觉限制，构建可进行严肃医学咨询、问答的对话智能体。
- 2) 创新性介绍：**创新性地结合知识图谱与大模型，使用向量化方式大大增强了检索速度，并且可以利用知识图谱限制大模型的幻觉和垂直领域强化。创造性地设计了两种检索增强生成方式：采取可靠 Cypher Prompt 保证查询结果的插件型 Custom 方法，以及采用 Schema 和生成 Prompt 结合来完全依靠大模型查询的智能体型 QChain 方法。实现了包含检索增强生成方法和流式输出选择功能的前端交互界面。
- 3) 社会经济效益：**知识图谱与大语言模型结合属于前瞻领域，与新推出的微软的 GraphRAG 和 Neo4j 的 llm-graph-builder 项目不同，该产品不注重从文本提取知识图谱的低稳定性方法，而是采用人工构建数据，构建高度结构化知识图谱，着重构建检索生成框架，从而使大语言模型的垂直强化水平大大提高，是以产品引用为导向的方法。
- 4) 个人贡献说明：**张文卓同学在本次实践中，完成了医学数据爬取和处理，设计并创建知识图谱，并在公司算法工程师的指导下完成了大语言模型检索增强框架的构建与 API 的创建，后续完成了前端交互页面设计并与后端集成，对于该产品的完成与完善做出了重要贡献，贡献排名：1/3。
- 5) 相关照片：**





健康助手

A Q4Chain chatbot powered by R1 or R1.5-Lite

您好，我是健康助手人模型，很高兴为您服务

介绍一下自己

我是由清华大学开发的人工智能助手，我擅长进行医学问答、报告解读和解读等。基于大量知识库训练的包括疾病和知识以帮助用户提供准确有用的信息。我也会不断学习优化更好地为您服务 (DONE)

您有什么问题呢?

Generated Cypher

```

MATCH (disease: Disease)-[:has_symptom]-(symptom: Symptom)
WHERE d.name = "感冒"
RETURN symptom

```

疾病回答

感冒的症状因个体素质和病毒不同而异,但常见的症状包括:

- 鼻塞会导致鼻腔充血从而引起。
- 喉咙痛等炎症或呼吸道感染, 咳嗽、不适等症状
- 流感会引起喉咙有时伴随发烧
- 头痛头部酸痛神经头痛等
- 肌肉酸痛肌肉组织乏力
- 免疫功能下降会产生免疫系统紊乱
- 声音沙哑

如果出现上述建议及时就医并遵医嘱进行治疗 (DONE)

实践单位负责人签字 (公章):



2024年8月13日

校内导师签字: 丁

2024年8月11日



(12) 发明专利申请

(10) 申请公布号 CN 117936106 A

(43) 申请公布日 2024. 04. 26

(21) 申请号 202311763900.7

G06N 3/0455 (2023.01)

(22) 申请日 2023.12.21

G06N 3/084 (2023.01)

(71) 申请人 浙江大学

地址 310058 浙江省杭州市西湖区余杭塘路866号

(72) 发明人 丁翥 张文卓 孙周健

(74) 专利代理机构 杭州求是专利事务有限公司 33200

专利代理师 林超

(51) Int. Cl.

G16H 50/50 (2018.01)

G16H 50/70 (2018.01)

G16H 50/20 (2018.01)

G06N 3/0499 (2023.01)

G06N 3/0442 (2023.01)

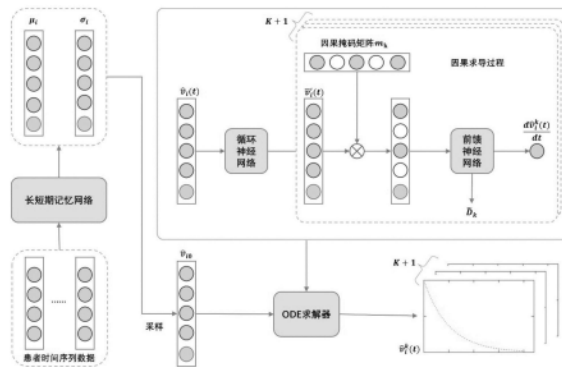
权利要求书2页 说明书11页 附图2页

(54) 发明名称

基于时序数据的因果发现与疾病发展轨迹预测系统及方法

(57) 摘要

本发明公开了一种基于时序数据的因果发现与疾病发展轨迹预测系统及方法。系统包括用于预处理患者时间序列疾病数据的数据预处理模块,用于预测患者的疾病特征导数量和疾病特征间预测关系的因果衍生模块,以及预测患者的疾病特征状态轨迹的轨迹预测模块;方法包括:对因果发现与疾病发展轨迹预测系统进行优化,完成后进行预测获得患者的疾病预测轨迹和疾病特征间预测关系矩阵,实现疾病的因果发现与疾病发展轨迹预测。本发明可基于患者慢性疾病的纵向数据挖掘出因果关系,给出因果可解释的疾病发展预测轨迹,解决深度学习进行轨迹预测不可解释和具有因果关系的父子特征互相影响的问题,从而辅助医生进行临床决策。





310013

浙江省杭州市西湖区古墩路 701 号紫金广场 C 座 1506 室 杭州求是
专利事务所有限公司
林超(0571-87911726-817)

发文日:

2023 年 12 月 21 日



申请号: 202311763900.7

发文序号: 2023122100397640

专利申请受理通知书

根据专利法第 28 条及其实施细则第 38 条、第 39 条的规定, 申请人提出的专利申请已由国家知识产权局受理。现将确定的申请号、申请日等信息通知如下:

申请号: 2023117639007

申请日: 2023 年 12 月 21 日

申请人: 浙江大学

发明人: 丁磊, 张文卓, 孙周健

发明创造名称: 基于时序数据的因果发现与疾病发展轨迹预测系统及方法
经核实, 国家知识产权局确认收到文件如下:

权利要求书 1 份 3 页, 权利要求项数: 10 项

说明书 1 份 13 页

说明书附图 1 份 2 页

说明书摘要 1 份 1 页

专利代理委托书 1 份 2 页

发明专利请求书 1 份 4 页

实质审查请求书 文件份数: 1 份

申请方案卷号: 超-231-492-徐

提示:

1. 申请人收到专利申请受理通知书之后, 认为其记载的内容与申请人所提交的相应内容不一致时, 可以向国家知识产权局请求更正。
2. 申请人收到专利申请受理通知书之后, 再向国家知识产权局办理各种手续时, 均应当准确、清晰地写明申请号。

审查员: 自动受理
联系电话: 010-62356655

审查部门: 初审及流程管理部



200101
2022.10

纸件申请, 回函请寄: 100088 北京市海淀区蓟门桥西土城路 6 号 国家知识产权局专利局受理处收
电子申请, 应当通过专利业务办理系统以电子文件形式提交相关文件。除另有规定外, 以纸件等其他形式提交的文件视为未提交。