北京工业大学国家级大学生创新创业训练计划创新训练项目申请书

项目编号 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

项目名称 \_\_\_物联网多源数据存储优化及模式抽取 \_\_

项目负责人 \_\_ 王煜 联系电话 \_\_\_ \_\_13521715651 \_ \_\_

所在学院 \_\_ \_樊恭烋学院 \_\_

学号\_\_\_\_\_15074403\_\_\_ 专业班级 \_\_\_ 150001\_\_\_\_\_\_\_\_

指导教师 \_\_\_ \_王丹\_\_\_\_\_\_\_陈军成\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

E-mail \_\_\_\_\_\_\_\_\_\_\_\_\_sallywangyu@163.com\_\_\_\_\_\_\_\_\_\_\_\_ \_\_

申请日期 \_\_\_\_\_\_\_ \_\_\_\_2016年11月30日 \_\_\_

起止年月\_ \_\_ 2017年1月 至 2017年12月 \_ \_

北京工业大学

填 写 说 明

1.本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要。

2.申请人可以是个人，也可为创新团队，首页只填负责人。“项目编号”一栏不填。

3.本申请书为A4纸印刷，左侧装订成册。可自行复印或加页，但格式、内容、大小均须与原件一致。

4.负责人所在学院认真审核，经初评和答辩，签署意见后，将申请书报送至校团委。

1. 基本情况

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项目  名称 | **物联网多源数据存储优化及模式抽取** | | | | | | | |
| 所属  学科 | **理工类** | | | | | | | |
| 申请  金额 | 10000元 | | | 起止年月 | | 2017年 1 月  至 2017年12 月 | | |
| 负责人姓名 | 王煜 | | 性别 | 女 | 民族 | 汉 | 出生年月 | 1997年4月 |
| 学号 | 15074403 | | | 联系电话 | 手机：13521715651 | | | |
| 指导  教师 | 姓名 | 职称 | 学院 | 联系电话 | 手机：13311157499 | | | |
| 王丹 | 教授 | 信息学部 |
| 指导  教师 | 姓名 | 职称 | 学院 | 联系电话 | 手机：13811022812 | | | |
| 陈军成 | 讲师 | 信息学部 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 负责人曾经参与科研的情况 | | 作为负责人承担北京工业大学“星火项目”并成功结题  《运用脑电信号监测系统预警睡眠状态下疾病的突发》 | | | |
| 指导教师承担科研课题情况 | | 王丹老师：  完成北京市自然科学基金项目、北京市教委以及多项企事业合作项目。  作为项目骨干成员，参与国家重点基础研究发展计划（973计划）、核高基专项、北京工业大学科研基地等项目的研究。  近几年在国内重要学术期刊和国际国内学术会议上发表学术论文50余篇。  美国伊利诺伊大学香槟分校（UIUC）和美国加州大学河滨分校（UCR）访问学者。北京市中青年骨干教师，第六届国家级教育教学成果二等奖1项。  陈军成老师：  《基于错误发现能力的GUI测试用例集约简方法》  《管理决策大数据平台构建与关键技术》 | | | |
| 指导教师对本项目的支持情况 | | 王丹老师在LINUX操作系统和软件应用具有丰富的经验，可为本项目中软件开发部分的指导。  陈军成老师在数据提取、数据规范、数据库构建方向有深入研究，可为本项目组在数据库构建以及格式规范方面给予指导。 | | | |
| 项目组主要成员 | 姓名 | 学号 | 专业班级 | 所在学院 | 项目中的分工 |
| 邵嘉琳 | 15071004 | 150001 | 樊恭烋学院 | 数据模式抽取 |
| 张蓬郁 | 15020008 | 150001 | 樊恭烋学院 | 数据管理 |
| 江旻宇 | 15071001 | 150001 | 樊恭烋学院 | 软件开发 |
| 张洪滨 | 15071019 | 150001 | 樊恭烋学院 | 数据库构建 |

**二、**立项依据（可加页）

|  |
| --- |
| 1. **研究目的**   传感器是物联网中的基础组件，是采集数据的关键模块。传感器通常以一定的频率搜集数据，并将数据上传至相应的数据服务器。传感器类型众多，采集的数据格式多样，并且通常具有时空属性。  现有物联网数据系统主要存在以下两个方面的问题：  1）缺乏统一的数据规范，对时空属性支持不足；  2）海量数据查询效率低下。  本项目针对这两个问题，研究设计并实现一套物联网分布式时空数据管理系统。   1. **研究内容**   针对物联网传感器采集数据中缺乏数据规范和时空属性处理问题，本项目的研究内容包括：   1. 传感器数据规范设计。传感器通常分为两类，一类为地理位置相对静止传感器，即地理空间相对固定，传感器采集的数据随时间变化而变化，其采集的数据为时值型；另一类为移动传感器，其属性值随着时间变化和地理位置变化而变化，其采集的数据为时空值型。针对这两类传感器数据，研究设计一套相应的传感器数据规范，适应上述两类传感器数据； 2. 传感器数据模式提取。现今大多数传感器数据是连续型数据，如温度传感器、风力传感器和GPS传感器等；然而传感器采集的数据是离散的数值点，由此产生两个问题：   1）如何获取非采样点的数据；  2）如何从海量的采样点数据中高效查询数据？  本项目针对两类传感器类型，拟利用Bernstein- Bézier等方法分析获取传感器数据的数据模式，解决上述两类问题；   1. 分布式数据存储、查询优化。在传感器规范和数据模式提取的基础上，本项目将研发一套分布式时空数据系统，解决或部分解决海量物联网传感器数据的存储、查询优化等。 2. **国、内外研究现状和发展动态**   传感器是物联网的基础模块，用于检测环境数据、社会数据等。形式各异的传感器采集的数据具有数据规模庞大、数据类型各异、自动产生、实时性强、价值高等特点，是典型的大数据。  针对这类数据，许多传统的商家如水电气数据采集公司依然采用传统的关系型数据库如Oracle[1]、MySql[2]等实施存储、查询等管理。然而，随着传感设备的快速增加以及采集频率的增加，关系型数据库已不再适用这类场景，于是纷纷转向NoSql型数据库如Redis[3]、MongoDB[4]等，NoSql数据库支持BASE特性，具有良好的横向扩展能力，解决了传统关系数据库中的不足。  面对海量数据，如何快速查询数据，依然是当前的一个研究热点。当前，众多研究利用BTree、RTree、地域分割等方法，结合MapReduce模型，以求快速获取满足条件的数据。然而，MapReduce是一种蛮力型的计算方法，是一种典型的以空间和功耗换时间的方法，处理海量数据时，功耗巨大[5]。  针对上述问题，本项目底层存储采用成熟的NoSQL数据库，根据传感器数据特性，构建传感器数据的数据模式，利用数据模式计算查询值，规避从海量数据中查询所需要的庞大遍历。  参考文献：  [1] www.oracle.com/cn/  [2] <https://www.mysql.com/>  [3] <https://redis.io/>  [4] https://www.mongodb.com  [5]罗乐，刘轶，钱德沛.《内存计算技术研究综述》.软件学报，2016,27(8):2147-2167.http://www.jos.org.cn/1000-9825/5103.htm   1. **创新点与项目特色**   本项目的创新点包括以下三个方面：   1. 传感器数据类型规范的设计与研制。 2. 数据模式提取。 3. 设计并实现一套分布式时空数据系统。 4. **技术路线、拟解决的问题及预期成果**   本项目的系统架构如下图所示：    图1 项目系统架构  本系统分为四个层次，每层的功能如下：   * 传感器层：传感器层包括各类传感器，主要实现数据采集功能，实现传感器上传至云存储端的接口； * 云存储层：设计传感器数据规范，按照规范实现传感器采集数据的原始数据存储，对已有的开源数据库如Reddis、Mongodb、Titan、Neo4J等进行扩充，向上提供数据增删改查等功能； * 数据模型提取层：根据原始数据特征，利用Bernstein函数、牛顿插值算法、曲面（或曲线）逼近算法等设计并实现相应的机器学习算法，提取传感器数据的数据模式（其基本模式为attr = f(t),即传感器数据的属性attr值为时间t的函数），并向上提供数据的操作接口； * 数据接口层：根据数据模型或外部使用方式，进一步对数据接口进行规范和优化。   本项目拟解决的关键问题包括：   1. 扩展性良好的数据规范； 2. 准确率高、误差率低的数据模型（在本项目中为连续函数）。   本项目的预期成果包括：   1. 原型系统； 2. 论文1-2篇； 3. 专利或软件著作权1-2项。 4. **项目研究进度安排**   2017年1月—2017年3月：程序设计语言golang、NoSql数据库学习，数据规范制定及基本的数据方案的研发；  2017年4月—2017年9月：利用已学习的数学知识和基本的机器学习算法实现数据模式提取并研发原型系统；  2017年10月—2017年12月：完善系统，撰写相应的论文和软件著作权或专利。  **（七）已有基础**  1、与本项目有关的研究积累和已取得的成绩  本项目组成员已完成以下两个项目的研究：  《蓝牙脑电传感器收集的脑电波数据分析及应用》  《图像模式识别及其原理研究》  2、已具备的条件，尚缺少的条件及解决方法  本项目已具备条件：  本项目依托信息学部大数据研究室，研究室近50台服务器、20台网络存储服务器为本项目提供硬件基础；研究室在时空大数据方面的研究基础为本项目提供理论及方法指导。  本项目尚缺少条件：  缺乏各类传感器数据，少量数据可以从公网下载，大量的数据需要相关部门授权或者购买。  解决方法：  对于交通、环境等监测数据，可以和本校交通学院与环能学院合作，联合研发。 |

三、经费预算

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 开支科目 | 预算经费（元） | 主要用途 | 阶段下达经费计划（元） | |
| 前半阶段 | 后半阶段 |
| 传感器 | 2000 | 采集数据 | 1500 | 500 |
| 相关书籍 | 2000 | 学习语言和数据库搭建 | 1000 | 1000 |
| 网络费用 | 2000 | 数据爬取  下载查阅文献 | 750 | 1250 |
| 论文发表、专利或软件著作权 | 4000 | 版面费  软件著作权费用 | 1000 | 3000 |

四、指导教师意见

|  |
| --- |
| 导师（签章）：  年 月 日 |

五、学院意见

|  |
| --- |
| 负责人（签章）：  年 月 日 |

六、学院**评审意见**

|  |
| --- |
| 评审教师（签章）：  年 月 日 |

**七、**学校意见

|  |
| --- |
| 负责人（签章）：  年 月 日 |