

密级： 保密期限：

# 北京邮电大学

## 硕士学位论文



题目： 基于移动协作感知平台的异构数据融合的研究与实现

学 号： 2012111491

姓 名： 周雪松

专 业： 软件工程

导 师： 王文东

学 院： 网络技术研究院

2015 年 3 月 25 日

### 独创性（或创新性）声明

本人声明所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：\_\_\_\_\_ 日期：\_\_\_\_\_

### 关于论文使用授权的说明

学位论文作者完全了解北京邮电大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属北京邮电大学。学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）

非保密论文注释：本学位论文不属于保密范围，适用本授权书。

本人签名：\_\_\_\_\_ 日期：\_\_\_\_\_

导师签名：\_\_\_\_\_ 日期：\_\_\_\_\_

# 基于移动协作感知平台的异构数据融合的研究与实现

## 摘 要

随着通信技术的快速发展和硬件水平的提升,移动智能终端设备也变得日益普及。移动智能终端设备内置了大量传感器,并且具有强大的计算能力,在这样的背景下,移动协作感知平台应运而生。

随着智能终端设备的快速发展,移动协作感知系统也变得越来越广泛,该系统可以通过用户收集大量的感知信息,并提取其中有用的信息以满足不同的应用。然而在实际应用过程中,会面临着很多的挑战。首先,移动终端节点的移动具有随机性,节点数量多且分布不均匀;其次,不同节点拥有的资源存在着差异,节点的异构性导致了提供感知服务的差异性;最后,数据的收集过程受参与者的影响,参与者不同的文化和教育背景对收集的感知数据在可信度方面有影响。最终收集的数据将会存在很大的时间和空间间隙,也会存在数据在小范围内冗余度很高,用这样的数据集去评估某一地区的环境指标都会变得很困难。为了解决以上的问题,本文将对移动协作感知系统中数据融合策略进行研究。

本文从介绍移动协作感知平台和数据融合的相关技术出发,结合加权平均的思想,利用时间和空间结合的克里金算法,最终在移动协作感知平台上实现了数据融合算法。首先介绍几种数据融合算法,通过对数据融合算法的分析和相关论文的阅读,拟采用克里金算法来解决该问题。然后对移动协作感知平台上数据融合模块进行需求分析和设计,包括数据存储设计、接口设计、网格设计,同时利用 GeoLife 数据集和 GreenOrbs 数据集来生成北京邮电大学范围内的实验数据集,该数据集模拟了移动感知系统收集的数据。最后针对克里金算法只对空间维度思考的片面性和单一性,本文提出了时间和空间结合的克里金算法,并进行实验验证,最终在该平台上实现了算法。

**关键词:** 移动参与式感知 数据融合 数据插值 克里金算法

# RESEARCH AND IMPLEMENTATION OF HETEROGENEOUS DATA FUSION ON MOBILE COLLABORATION SENSING PLATFORM

## ABSTRACT

As the fast development of communication and hardware technology, mobile smart terminals are becoming increasingly popular. These smart terminals are integrated with a large number of advanced sensors and have powerful computing ability, the mobile collaborative sensing platforms emerged in this context.

Mobile collaborative sensing system is becoming widely used by people due to the widespread use of smart phones, this system collects a lot of information provided by the smart phone users which can be used to extract useful information in order to meet the need of different applications. However, lots of challenging problems would occur in the process of deployment and operation in actual environments. Firstly, mobile nodes are random, and a large number of nodes are not uniform and provide services with dynamic uncertainty. Secondly, different nodes have different resources, the heterogeneous character of the system leads the diversity of the sensing services. Finally, participants with the background of different cultures and educations influence the precision and reliability of the collected data, therefore the collected data sets will produce gaps in time and space and will be redundant in a small range. Using these data sets to evaluate the environmental index in a particular area would become very difficult. In order to solve the above problem, this thesis will study the data fusion strategy in mobile collaborative sensing system.

In this thesis, after learning data fusion technologies and the weighted average algorithm, we implement the algorithm on the mobile collaboration sensing platform using temporal and spatial kriging

algorithm. First of all, after referring to quantities of data fusion algorithm and papers, we find that kriging method can solve this problem efficiently. Then, we introduce the design of data fusion module which includes database design, interface design and grid design. After that, by using GreenOrbs and GeoLife datasets, we generate the data sets of Beijing University of Posts and Telecommunications, and these generated data sets are used to simulate collected data on mobile sensing system. Finally, since the ordinary kriging algorithm just considers the space dimension and ignores the time factor, we propose temporal and spatial kriging algorithms, we test it using experiments and implement the algorithm on our platform.

**KEY WORDS:** participatory sensing; data fusion; data interpolation; kriging;

---

# 目录

第一章	绪论	1
1.1	课题的研究背景	1
1.2	课题的研究内容和意义	2
1.3	本文描述的主要工作	3
1.4	本文的组织结构	4
第二章	相关技术的研究	6
2.1	移动协作感知技术	6
2.1.1	移动协作感知技术概述	6
2.1.2	移动协作感知平台的整体框架	7
2.1.3	移动协作感知系统的应用	9
2.2	数据融合技术	11
2.2.1	数据融合概述	11
2.2.2	数据融合功能模型	12
2.2.3	数据融合的层次	14
2.2.4	数据融合的方法	17
2.3	数据融合技术在移动协作感知平台中的应用	23
2.4	本章小结	23
第三章	移动协作感知平台数据融合模块的需求和设计	25
3.1	移动协作感知平台的需求和设计	25
3.1.1	移动协作感知平台的需求分析	25
3.1.2	移动协作感知平台的整体设计	26
3.2	数据融合模块的需求	29
3.2.1	数据融合模块的需求分析	29
3.3	数据融合模块的设计	31
3.3.1	数据融合相关表的设计	32
3.3.2	数据融合模块接口设计	33
3.3.3	网格划分策略	34
3.4	本章小节	35
第四章	克里金算法改进和模拟数据生成	37
4.1	基于空间模型的克里金算法的分析	37

4.2	基于空间模型的克里金算法的改进.....	38
4.2.1	Product 模型介绍.....	39
4.3	模拟数据集生产.....	40
4.3.1	GeoLife 数据集描述.....	40
4.3.2	GreenOrbs 数据集描述.....	41
4.3.3	模拟数据集的生成.....	43
4.4	本章小结.....	50
第五章	基于时间和空间模型的克里金算法的验证与实现.....	52
5.1	建立数据模型.....	52
5.2	基于时间和空间模型的算法的实验评估.....	55
5.2.1	算法对比分析.....	55
5.2.2	算法准确率和性能比较.....	58
5.2.3	算法的误差分析.....	59
5.3	基于时间和空间的克里金算法的实现.....	60
5.4	本章小结.....	62
第六章	结论及展望.....	63
6.1	结论.....	63
6.2	展望.....	63
参考文献	.....	65
致 谢	.....	69
攻读学位期间发表的学术论文目录	.....	70

## 第一章 绪论

### 1.1 课题的研究背景

通信技术和计算机技术的飞速发展，也推动了移动智能终端设备的迅猛发展和普及，移动电话由原来只具有简单的通话功能发展到现在集成了各种其他功能，并且具有很强的计算能力、感知能力和通信能力。移动终端领域的快速发展和普及也促使了相关的研究者们探索出了一种新型的数据收集方式——参与式感知（Participatory Sensing）<sup>[1-4]</sup>，也叫移动协作感知（Mobile Participatory Sensing）。该方法的提出使付出低成本就可以获取大范围内的感知数据成为了可能，也使用户自觉上传收集到的感知数据的收集方式成为了现实。移动协作感知的基本思想就是将用户随身携带的移动智能手机作为数据采集设备，收集环境中的感知数据信息，并且在收集到的感知数据基础上为其他应用提供数据服务。这种新型的数据收集方式相比于传统的传感器收集方式在很大程度上方便了研究人员的工作，而且平台更加的简单灵活。

移动协作感知网络技术相比于固定部署的传感器网络技术，具有很多优点：

（1）部署成本比较低，移动协作感知技术使用移动设备终端的感知和无线通信功能进行数据的收集和上传，硬件设备只需有移动设备终端、无线网络和服务即可搭建一个简单的平台；（2）移动智能终端用户的移动性决定了移动协作感知平台收集的数据可以史无前例的覆盖时间和空间范围，用户的时间和空间随机性导致了数据的时间和空间范围；（3）移动协作感知平台的搭建和开发部署难度低，便于大规模部署，具有很好的可扩展性，用户的智能终端都可以成为该平台的一部分，加入简单，无成本。（4）用户是该平台运行的基础，主要是靠手中的智能终端设备提供数据，实现数据的感知收集，但反过来也可以使用收集到的数据进行其他的应用，所以也是最终的受益者。

尽管移动协作感知平台有着广阔的应用前景和诸多优点，但在相关平台的实际部署和应用过程中依然会面临着很大的挑战和不足。首先，节点也就是用户随身携带的移动智能终端设备，具有移动的随机性和不确定性，节点的数量比较多而且分布是不均匀的，并且数量也是不可控制的，从而导致节点提供的服务也具有动态的不确定性；其次，移动设备节点所剩余能量和计算能力也是受限制的，节点所携带的电量和用来进行数据采集所占的资源也是有限的；再次，不同的节点所拥有的资源也会存在着很大的差异，节点的异构性也将会导致提供的感知服



务具有很大的差异性；最后，收集数据的过程也会受到用户参与的影响，不同用户具有不同的文化背景和教育背景等因素，这也会对收集的感知数据在精确度、可信度、时延、均匀分布等方面产生一定影响，并且用户的参与热情也对数据质量有很直接的影响<sup>[6]</sup>。对于以上该平台存在的缺点和不足，也有不少研究者正在进行这方面的研究和探索，极力解决这方面的问题。

基于以上移动协作感知平台自身特性导致的收集到的数据不均匀、不准确、冗余性高、缺失度大等不足，着手对数据准确度和弥补数据缺失等方面入手进行深入的研究。目前也有一些公司、大学实验室以及科研机构开始研究移动协作感知平台，在这样的背景下，我们可以参考别人的研究成果和研究现状，在此基础上进行更加深入的探索和研究，移动协作感知系统未来会大大满足人们的需求，并可以削弱人们的投资成本，具有很好的实用性，这些优点将会使移动协作感知在未来会变得更加的普遍和适用。基于以上的研究背景作为出发点，本文也将继续对移动协作感知平台上存在的问题进行深入一步研究。在互联网高速发展和大数据时代来临之际，数据对于人们变得越来越重要，不仅仅是因为数据本身所代表的知识，更是因为数据内部所隐含的信息，因此收集数据的平台将会变得很重要，并且移动协作感知平台将会有很好的应用前景。

## 1.2 课题的研究内容和意义

随着近几年移动协作感知平台的提出和快速发展，越来越多的人注意到移动协作感知平台的可用性、易搭建性和扩展性等诸多优点，越来越多的研究者也将目光转向了移动协作感知平台，并在该平台上进行应用开发和基础研究，由于目前该领域发展的还不完善，许多问题需要解决，在本论文中就是对移动协作感知平台上数据融合问题进行的研究。

由于移动协作感知网络中的终端节点具有移动性、随机性、且分布不均匀和异构性等特点，这些特点都将对移动协作感知平台的设计提出了很大挑战。异构性将会导致数据的不准确、不均匀和不可靠，必须进行相应的冗余数据处理，修补缺失数据，最终得到真实可靠的数据，以保证后续工作的持续进展。本项目的研究是基于协作的移动感知平台，从提高移动协作感知系统的数据集的全面性、可靠性和真实性等方面入手，对移动协作感知平台上数据融合的问题进行理论分析、建模计算，算法验证和最后的算法实现。从基础算法进行分析研究，为了算法的可靠性和高效性，并进行算法的扩展，以便数据计算结果能够更有实用价值。数据融合是移动协作感知平台收集的数据的重要数据处理过程，经过数据融

合处理后的数据对于环境保护领域、公共卫生领域、灾难救援、交通运输、全民的身体健康等方面都具有广阔的应用前景。

本文就是利用移动智能终端作为传感器进行数据收集，首先将采集外界环境信息并保存本地，最终通过无线通信技术将所有收集数据以一定的格式上传至服务器后端，以便进行下一步的数据融合和数据挖掘等处理工作。服务器需要的是基于数据而产生的知识，最终完成数据融合、数据挖掘工作，从而达到数据融合的目的。数据融合的工作对于移动协作感知平台来说很重要，由于移动协作感知平台自身的特性，就会导致收集的数据的不均匀、不真实、冗余性高、缺失度大等特点，这样的数据集对于其他应用来说将会削减其内在价值，如果使用这样的数据集去进行环境监测和评估，也都会对结果产生影响，因此数据融合是非常必要的数据处理过程。融合之后的数据对于环境监测和其他应用都会有帮助，并且也提高了数据的内在价值。

我国国民经济和社会发展的“十二五”规划中重要强调要加大环境的保护力度，建立健全完整的服务体系。考虑到在我国的城市环境中具有人口密集度大和移动智能设备普及性高的特点，并且移动协作感知技术的部署成本极低，在公共设施服务、环境保护、灾难救援和提升全民健康水平等方面有着广阔的应用前景。例如在公共卫生领域，可以通过收集参与者身体状况的相关数据来监控疾病的传播人群；在交通运输领域，通过参与者收集的实时车速信息、位置信息等，来监控城市道路的交通拥堵状况，为交通管理部门提供了重要的决策依据；在环境保护方面，通过参与者收集的环境感知数据或图片，计算获取城市细粒度的实时空气质量的信息，例如 PM2.5；在灾难救援中，利用参与者提供的受灾地域的照片及位置信息等相关信息，可以分析出灾害在不同区域地段所造成的损失程度，为灾难救援决策的制定提供一定的指导。因此，移动协作感知平台上的研究工作对于国家的未来建设具有非常重要的意义。

### 1.3 本文描述的主要工作

- (1) 移动协作感知平台架构的需求分析和设计，然后对该平台中数据融合模块进行需求分析和设计。这一部分是整个工作的基础，也是本文的出发点。具体包括对整个移动协作感知平台架构的硬件需求、软件需求和具体结构设计、服务器端的数据融合模块的功能分析、外部接口设计、数据存储格式设计、内部结构设计，还有数据融合过程中的网格划分策略。
- (2) 模拟数据集的生成。对于移动协作感知平台上关键算法的研究，数据是算

法研究的基础支撑，对于该系统搭建之初，还没有该平台收集的数据集，因此采用生成模拟数据集的方式进行算法的验证。根据现有微软的 GeoLife 数据集和“绿野千传”项目的 GreenOrbs 数据集，通过对两个数据集的截取、平移、插值处理和计算，最终生成模拟的数据集，数据集的位置是北京邮电大学校园范围，该数据集模拟了移动协作感知平台收集数据集的方式。

- (3) 数据融合算法的验证和实现，通过阅读论文了解到克里金算法在移动协作感知系统中可以解决数据缺失和冗余的问题。通过对移动协作感知平台收集数据的性质和克里金算法的特性进行分析，时间维度对移动协作感知平台上数据融合计算也是一个非常重要的因素，从而提出基于时间和空间模型的克里金算法，并引入 product 模型，最终用生成的数据集进行算法的验证，并在该平台上实现了该算法。

## 1.4 本文的组织结构

第一章：绪论。首先介绍了课题的研究背景和研究内容及意义；然后介绍本文的主要工作内容，最后指明本文的后续章节的内容安排。

第二章：相关技术，包括移动协作感知技术的研究和数据融合理论的探索。移动协作感知技术的概述和系统架构、数据融合的概述、数据融合的功能模型、数据融合的层次、以及数据融合的方法，对数据融合相关的基本知识有所了解。为决定移动协作感知平台所用数据融合的方法打下基础。

第三章：移动协作感知平台和数据融合模块的需求分析和设计，首先介绍移动感知平台的基本需求，并呈现了该平台的框架设计；然后介绍移动协作感知平台中数据融合模块的基本需求，并也呈现了相应的设计工作，包括接口设计和数据表设计，以及网格设计等。

第四章：移动协作感知平台数据融合所采用的算法以及改进。首先是对克里金算法进行介绍，并针对算法和平台特性进行相应的改进，提出了基于时间和空间模型的克里金算法；然后介绍了模拟数据的生成，先介绍了微软亚洲研究院项目的 Geolife 数据集和无锡清华信息科学与技术国家实验室物联网技术中心的 GreenOrbs 项目的数据集，并针对两种类型的数据集进行分析和计算处理，最终生成模拟的数据集，并且该数据集模拟移动感知系统收集数据的过程。

第五章：基于时间和空间结合的克里金算法进行实验验证和实现。利用生成的模拟数据进行数据建模，然后利用 java 语言和 MATLAB 程序进行仿真验证，并最后在该平台上实现了该算法。

第六章：论文工作的总结和未来工作的展望，首先对自己在本文中工作的总结，并也提出了一些不足的地方，然后对未来的发展方向和前景进行分析，未来移动协作感知平台的工作将会更加具有挑战性，也会更有意义。

## 第二章 相关技术的研究

### 2.1 移动协作感知技术

#### 2.1.1 移动协作感知技术概述

通信技术和计算机技术的快速发展,诞生了互联网技术。随着近几年互联网技术的快速发展,又产生了无线传感器网络,给人们的生活带来了翻天覆地的变化。随着人们追求简单化,易操作化的意愿,近几年又出现了移动协作感知的概念,移动协作感知技术主要就是通过用户随身携带移动智能终端来收集环境中的感知数据,通过无线网络进行感知数据的上传,并用服务器进行数据存储,从而实现感知数据的收集。该技术主要就是基于传统的无线传感器网络技术发展演化而来的。

新型的数据收集方式在很大程度上方便了研究人员的工作,相比于传统的传感器网络平台,移动协作感知平台具有更加简单、灵活等诸多优点。首先,和传统的传感器网络的工作方式不同,移动协作感知平台不需要事先部署特定的传感器设备,它将普通用户的手持或者车载智能设备终端所嵌入的多种传感器作为基本的感知单元,保证了数据收集的多元性和及时性,实现了来源于任务参与者的感知动作和来源于任务需求方的感知任务的交互、感知数据的有效收集和分析处理,进而形成具有明显交互特性的“移动协作感知网络”。利用用户有意识或者无意识的参与,移动协作感知平台具有用户的覆盖面广、数据的资源丰富、不需要额外的硬件设备投资、具有很好的可扩展性等优点,并能够向基于移动协作感知网络的其他应用提供全面的、便捷的、深度的数据支持。移动协作感知平台在完成大规模或者复杂的社会感知任务<sup>[5]</sup>方面具有广阔的应用前景。

传统的传感器网络的特点主要是用固定传感器来采集某一环境参数的数据信息,并最终将采集数据上传到服务器的某个节点,该网络主要是以“传感器”为中心,但是事先需要人为的参与和设置,需要进行前期传感器节点的位置部署和环境部署,其次就是要设置传感器采集数据和上传数据的频率,还要设置传感器上传数据的目标节点等信息。而“移动协作感知网络”则是以“用户”为中心,用户携带着移动智能设备终端,用户的移动性影响着传感器节点的位置和分布,数据主要由用户收集,用户可以决定数据的收集时间、收集频率和上传数据的时间,数据经过用户选择后,最后上传到服务器端。用户的参与才是移动协作感知平台运行的基础,用户的兴趣爱好和行为习惯,都影响着数据信息的采集。在该

平台中数据不仅仅是数据的贡献者,用户也会因为共享的数据而开发的相关应用而受益。

移动协作感知平台提供了一般性的系统概念和系统框架,它使用用户随身携带的移动智能终端设备作为传感器采集设备,进而形成了传感器网络。该平台提高了用户的参与度,并且鼓励用户诚实地遵守系统的规则,这样可以提高数据的可靠性,并且鼓励用户参与的积极性,这就要求系统要保证用户的隐私不能泄露。该平台主要就是实现以计算机、传感器等设备为中心到以用户为中心的转变,通过用户的兴趣爱好、行为习惯等来实现对周围感知环境数据的采集上传,为下一步进行数据处理和数据挖掘等工作提供数据支撑,也为其他类似移动协作感知平台上算法的研究提供了数据支持。在互联网高速发展和大数据时代来临之际,数据对于人们变得越来越重要,不仅仅是因为数据本身所代表的知识,更是因为数据内部所隐含的信息,因此收集数据的平台将会变得很重要,并且移动协作感知平台将会有很好的应用前景。

### 2.1.2 移动协作感知平台的整体框架

对于移动协作感知平台的结构,其包括移动智能终端设备、中央服务器、网络环境中的基础设施和各种传感器设备等装置。传感器不但可以集成于移动智能终端设备中,还可以通过短距离无线通信设备接口和移动智能终端相连接。移动智能终端设备之间、移动智能终端和附近的基础设施之间通过短距离无线接口进行通信,一定区域内的所有移动智能终端设备就可以构成一个自组织网络。移动协作感知平台搭建比较简单,并且基础设施投资也比较少,每个人携带的移动智能手机就可以作为移动终端设备,接入到移动协作感知平台也很方便。相比于固定传感器网络,成本更加低廉,搭建起来更加方便,有很好的扩展性,具有更好的使用价值和研究价值。

图 2-1 展示了移动协作感知网络的系统层次架构图,系统主要由采集信息的智能通信设备,其中里面内置了各种传感器,例如声音、光照、噪声、温度、GPS、加速度等,采集的信息通过无线网络将数据传输到服务器端进行数据的存储和处理,在通过客户端进行数据的展示和应用呈现。移动协作感知平台的框架分成四层进行描述,从下至上依次为:感知层、网络传输层、数据处理层和应用层。每一层都有不同的职责和功能。感知层是利用用户携带的各种智能终端设备,实现对数据的实时采集;网络传输层主要为数据传输提供可靠的服务,在移动感知系统中,由感知主体将移动设备采集到的感知数据安全可靠地上传到服务器;数据处理层主要是服务器端对接收到的数据进行分析加工处理的过程;应用层主要是在采集的数据信息的基础上,加工过的数据在以用户需求的角度反馈给用

户，让用户成为最终的受益者。

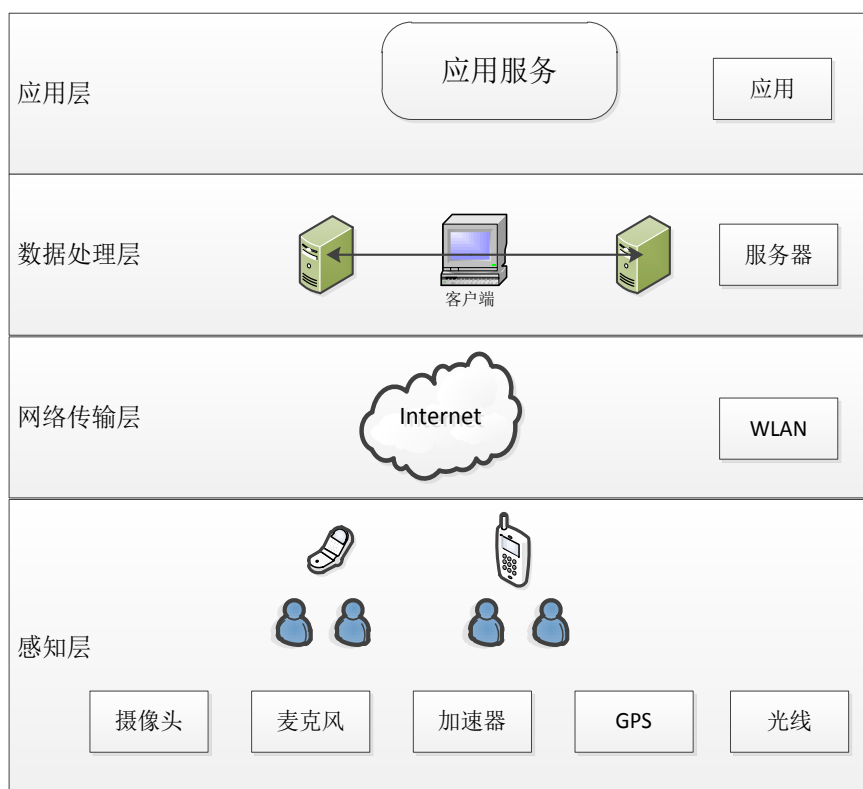


图 2-1 移动协作感知网络的层次架构图

图 2-1 所示的移动协作感知网络的层次结构图展示了四层结构，其中感知层和应用层都是在客户端，数据处理层在服务器端，网络传输层属于客户端和服务器的通信接口，各个层的具体模块划分如图 2-2 所示。

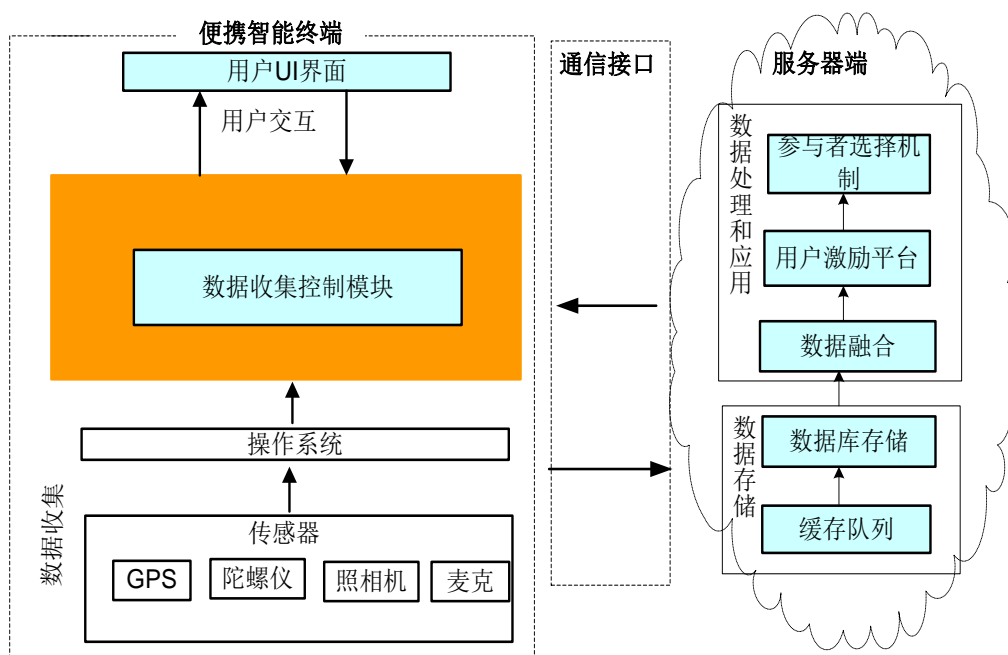


图 2-2 移动协作感知网络的整体结构图

图 2-2 展示了移动协作感知平台网络的整体结构图，从左到右依次分为三部分：客户端（智能移动终端）、通信接口、服务器端。

智能终端上从下至上的结构分别为：硬件传感器层、操作系统、数据采集模块和应用层。硬件传感器包括光照传感器、GPS 传感器、照相机和声音传感器等；操作系统包括 android、ios 或者其他智能终端操作系统；数据采集模块主要就是用来采集感知环境中的数据。

服务器端从下至上包括两层：数据存储模块和数据处理模块。数据存储模块主要是数据先上传存到缓存队列，然后再存储到数据库中；数据处理模块主要是先从数据库中读出数据进行数据融合，数据缺失时可以触发激励模块下发激励，又要启动参与者选择模块，进行下发激励的用户筛选，最后完成下发激励的任务。

通信模块是终端和服务器的通信桥梁，主要在服务器和客户端进行通信。

对于整体结构，主要用到了几个技术，参与者选择技术、激励机制、数据融合技术等。

- (1) 参与者选择机制是因为参与者的感知能力存在差异、参与者的轨迹不可控，协作感知系统需要从大量参与者中选取符合要求的用户来完成感知任务。
- (2) 激励机制主要就是选取合适的用户进行激励下发，该机制往往是用来保障用户的参与度，提高数据质量。
- (3) 数据融合模块主要就是提高数据质量，去除数据中错误数据，针对异构数据进行融合统一。

### 2.1.3 移动协作感知系统的应用

近几年，在移动互联网高速发展的大前提下，移动协作感知这个概念变得越来越火热，也有越来越多的人将视线转向了移动协作感知平台。从固定传感器网络到移动协作感知平台的转变，这将会是一个时代的大趋势，虽然现在移动协作感知平台处于刚初期阶段，但是未来带给人们利益的潜力是巨大的，现阶段已经有一些该平台上的应用。该平台上应用的方向主要可以分为两个方面：

(1) 面向个人的终端感知服务的应用，该应用不仅仅收集用户个人的位置、轨迹行为等信息，还有很多其他的用户信息。(2) 面向终端群的感知服务的应用，该应用是指利用收集的许多参与者感知数据来完成一些公共的服务。国外已经有很多研究者们开展了这方面的研究和探索工作，利用随身携带的智能手机进行城市噪音污染、空气 PM2.5、湿度、氮化物含量等事件的监测，以及感知人们的行为习惯、心情变化和参与的不同类型活动。

移动协作感知平台是目前移动物联网研究领域的热点之一。在部分国外相关



文献中, Urban sensing<sup>[6]</sup>、Community Sensing<sup>[7]</sup>、Volunteer Sensing<sup>[8]</sup>、Public Sensing<sup>[9]</sup>、Mobile Crowdsensing<sup>[10]</sup>等一些相关感知的研究与移动协作感知都相近。近年来,已经涌现了一批创新型的移动协作感知平台上的应用。按照移动协作感知平台上收集数据内容的不同,可以将移动协作感知平台归为以下三类:

(1) 环境感知相关的应用,如空气质量监测、温度监测、湿度监测、噪声监测等;(2) 公共设施服务感知相关的应用,如交通导航、交通路况、路线指引、预估公共汽车到站时间、停车空位引导等;(3) 社会化感知的应用,也就是以参与者为中心的感知相关应用,如监测并共享用户的饮食习惯、用户喜好、用户朋友圈等。在文献<sup>[11]</sup>介绍了使用车辆中的智能终端设备来收集交通路况信息,共享给其他车辆使用。文献 Common Sense<sup>[12]</sup>介绍了移动智能终端通过无线蓝牙接口收集空气质量的感知数据,还能够检测出附近的多种空气污染物含量。文献<sup>[13]</sup>采用了一种更加通用的方法来监测空气质量,该方法就是通过移动智能终端用户拍摄天空的照片,并对其色调等特征进行对比分析,可以粗略地判断空气质量的好坏。文献 NoiseSPY<sup>[14]</sup>介绍了通过移动智能终端的麦克风传感器来收集周围环境的噪声信息,进而绘制出城市的实时噪声等级图。文献<sup>[15]</sup>根据骑车人的当前位置、车速、消耗的热量等情况,可以了解骑车人的运动轨迹和身体状况,收集到的感知数据由骑车人进行数据分享,并和其他骑车人的数据进行合并,或者和其他的环境参数进行结合,例如空气质量和交通特性等,对骑车爱好者群体进行详细的跟踪描述。Micro-Blog<sup>[16]</sup>主要利用智能手机的传感器来增强多媒体博客。文献 LiveCompare<sup>[17]</sup>和 PetrolWatch<sup>[18]</sup>介绍用户能够采用拍照价签的方式,对不同商场的同类商品价格进行方便比较。文献<sup>[19]</sup>描述了通过外围传感器设备和多摄像头的合作来实现对移动物体追踪的功能。在文献<sup>[20]</sup>中提出了通过将视频传感器用无线自组织的方式组建网络,并用分布式协作的方式来追踪运动的物体。在文献<sup>[21]</sup>提出了一个名为 MoVi 的应用,该应用通过多个移动智能终端收集视频片段,通过组织处理后来识别社会事件。文献<sup>[22]</sup>搭建了由多个无人驾驶飞行器组成的一个协作感知平台,通过飞行器间的相互合作来获取地面的图像信息。

国内在这方面的研究相对比较少,主要有上海交通大学的智能车辆技术研究所在基于车辆感知技术的车队控制方面开展了部分相关研究<sup>[23]</sup>,还有清华大学的信息科学与技术国家实验室已经开始探讨群体智能感知在未来的重要性<sup>[24]</sup>。

以上对移动协作感知平台研究现状的列举可知,目前国外的研究成果相比于国内比较多。虽然移动协作感知方向在 2009 年到 2013 年间涌现出了大量的论文,其中论文方向主要是以协作感知平台提供服务的先导研究和对协作感知系统的整体架构、中间件、相关算法的研究。然而,目前还没有发现使用普通参与者收集数据,实际应用价值有望进一步提高。

## 2.2 数据融合技术

### 2.2.1 数据融合概述

近年来随着计算机技术和硬件技术的飞速发展，外加和通信技术的紧密结合，不断满足各个领域的需求，在我国重要的军事领域里，二者的结合应用在这方面起到了至关重要的作用，也不断创新出新的技术。数据融合技术<sup>[25]</sup>，一种新兴的数据处理技术，该技术在近 10 年的军事领域中得到了惊人的发展，应用广泛。

数据融合技术就是利用计算机技术和通信技术对收集到的时序数据进行有用信息的提取，并以一定的目的进行信息分析、处理、过滤、综合等操作，得到最终满足决策需求和评估任务的结果的一种信息处理技术。

数据融合技术主要就是经过数据的采集、数据传输、数据存储、数据处理、数据综合等一系列过程，最终得到满足人们需求的结果，该结果可以发挥环境的监测，战场上的态势评估，交通规划等的作用，辅佐人们进行判断、评估、做出最终决定。数据融合技术早起主要用在战场上，以便准确的获得各种有用的信息，包括战场的作战情况，包括目标位置、目标的威胁等级、目标的属性等。根据对这些信息的分析可以得出对战场的决策战略、任务分配等信息。在未来的战场将主要以技术和数据为主导，所有的决策都将会基于对数据的应用处理。信息处理技术已经成为了战争中关键所在，大量的、准确的、不间断的数据传输和准确的数据处理结束成为了战争中的首要目标。战争上情况的复杂多变，且影响因素也颇多，这就要求指挥员要在很短的时间内做出重要决策和计划出指挥战略，并对作战部队进行有效的指挥控制。这一系列的需求都需要有数据处理技术作为基本保证，否则在信息时代只靠指挥官的单一判断和实时分析会变的力单势薄，必须要结合先进的信息技术增强作战实力。数据融合技术不仅仅可以用在地面，还可以用在水下，主要用来大范围的监视和监测。例如水下作战时，对付敌人使用的隐身术，例如消声技术、低雷达截面等，主要进行大面积的监测活动。

数据融合技术本身有很多优点，例如高速性能和低成本等。这些优点不仅仅让该技术在军事领域得到广泛应用，在商业领域、自动化领域、环境领域、交通领域、家庭等方面都有很好的应用前景。可以用在自动化领域中，也可以用在商户和家庭的防盗防火安全。对于环境信息收集经处理后得到的结果可以用来进行环境监测和环境规划；对于交通信息的收集也可以来在实时导航方面提供便利。在军事领域研究的一些复杂的数据融合应用也应用在其他方面，例如城市规划、资源管理、气候监测、地质分析等，这些大量的信息可以在各个领域之间进行信

息共享。

### 2.2.2 数据融合功能模型

数据融合的功能模型是从数据融合过程出发描述整个融合过程中所用到的数据库，所实现的功能，以及融合时各个部分的组成部分。

#### (1) 数据融合的 JDL 模型

数据融合兴起和快速发展是其应用到了很多的领域，即使在同一相同的领域，不同的应用对数据融合里面的相同的功能概念也有不同的定义，为了让开发者和军事应用者之间方便交流，美国国防部在 1986 年成立了一个数据融合工作组，该组织致力于统一数据融合的专业术语。JDL 创建了统一的数据融合专业术语表，并提出了 JDL 功能模型<sup>[26-28]</sup>。

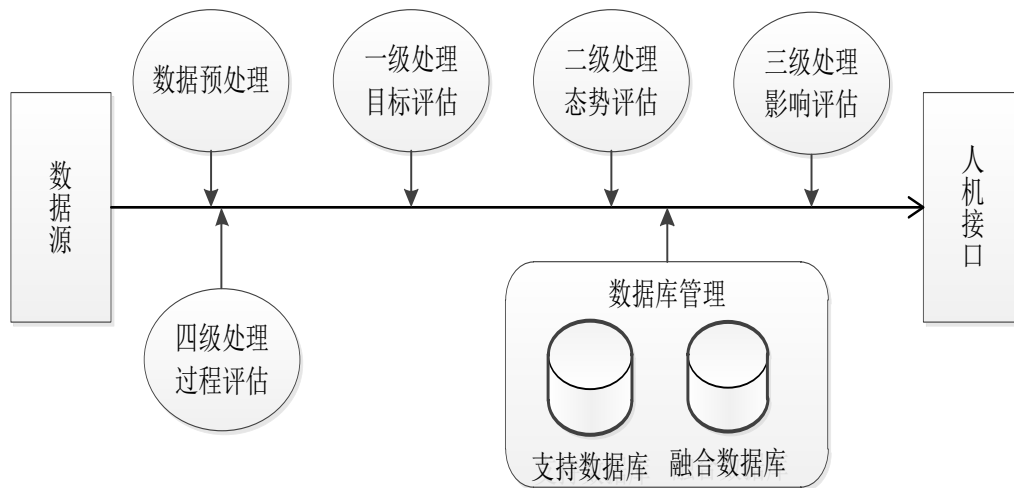


图 2-3 数据融合功能模型

该功能模型将多源传感器信息融合的过程分为几个部分：数据源、数据的预处理部分、目标评估、态势评估、影响评估、过程评估、数据库存储以及人机接口。

数据源：也就是数据来源，包括与数据融合模块进行连接的传感器集群，主要是提供原始的、未经过任何处理数据集。

目标评估：通过对目标物体的运动位置、速度和其他参数估计出目标进行更加准确的描述。该过程主要执行了四个重要的功能：将采集到的数据都统一到一个系统坐标下和一个测量单位下；精确的计算和估计目标的位置和其他各参数；应用统计理论去处理数据；再精确计算和目标有关的各个参数的特征值。

态势评估：态势评估就是对态势的抽象和评定。态势抽象就是根据不完整的数据集去构造一个模拟的态势表示；态势评定就是对态势抽象的理解和分析。

影响评估：基于对目前情况的分析预测未来，基于假设的参与者行为的影响

而进行的评估。例如在作战进行中，根据敌人的政治信仰和所处环境以及其他情况对敌人的意图和下一步进行估计。

**过程评估：**这是一个更高级的处理阶段。通过建立一定的优化指标，对整个融合过程进行实时监控和评价，而实现多传感器自适应信息获取和处理，以及资源的合理分配，最终提高系统的性能。

**人机接口：**允许人们输入诸如命令、请求、评估、操作员的报告等，HCI的功能在于通过警报!显示以及在地形图上动态刷新目标位置和信息来报告数据融合的结果。

**数据库管理：**是数据库最基本的功能，主要就是对数据库中数据进行修复、存储、保管、压缩、查询和数据加密等。

JDL 功能模型是一个统一的数据融合模型，但是在实际的应用中，可以根据实际需要对该模型进行修改和完善，以满足应用的需求。

## (2) 数据融合的 white 模型

1988 年 white 针对军事系统又提出了另一个数据融合处理模型<sup>[29]</sup>，在这个模型中，数据融合被分为三个等级：一级主要就是融合位置和标志估计；二级是敌军和我军的态势评估；三级是敌军和我军的兵力威胁估计。如图 2-4 所示。

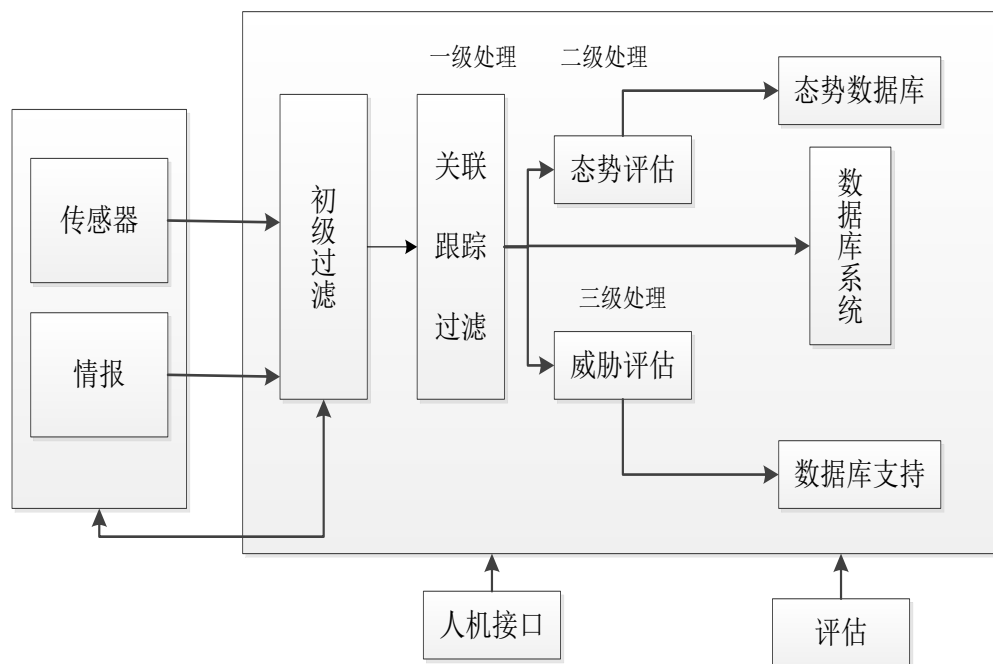


图 2-4 数据融合 white 模型

- (1) 一级处理：主要就是融合位置和标志估计。经过初级过滤之后，获取到对融合目标有意义的数据项，进行数据关联、数据跟踪和过滤。
- (2) 二级处理：也就是态势评估。根据不完整的数据集对态势环境进行模拟还原。

(3) 三级处理：威胁评估。根据目前采集的信息情况和敌方的所处环境和各方面情况去评估敌军的威胁。

### 2.2.3 数据融合的层次

数据融合按照融合对象的不同，可以分为三个层次，从低到高依次为：数据层、特征层、决策层。数据层的融合对象是原始数据，特征层的融合对象为提取的特征信息，决策层的融合对象为提取的识别信息。

#### (1) 数据层融合

数据层融合<sup>[30]</sup>是处于层次模型中最底层的融合技术，该技术主要就是针对原始数据进行分析处理的。由于该技术一般用在图像处理领域，因此也被称作为像素层融合技术。在图像处理领域，通常在成像传感器系统中通过多个传感器所采集的模糊图像进行综合处理得到更加清晰的图像，或者辨认出图像中所想要的目标属性。图 2-5 介绍了数据层融合的流程。

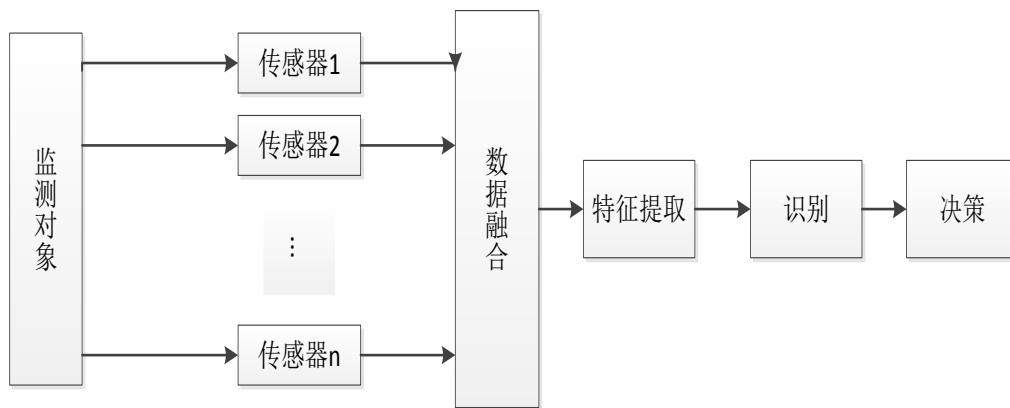


图 2-5 数据层融合流程

从图 2-5 所示的数据层融合流程可知，数据层融合是对多传感器采集的原始信息上进行融合处理，处理之后再继续进行后续的特征提取、识别和决策过程。

数据层融合方法具有很多的优点，主要有（1）能够保持更多的原始数据，并且信息的损失度小，提供的信息比其他两个层面要更详尽一些；（2）由于传感器本身在处理数据时容易受外界环境噪声影响，并且传感器收到的数据也具有很多的缺点，例如数据量大时耗时就是变长，并且实时性也会变差，从而代价就会变高，因此这就需要传感器在处理数据时具有很好的容错能力。（3）为了得到更好的融合结果，前提就是原始数据就需要满足一定的校准精确度，这也就需要满足每一个传感器信息都是来自同质的传感器。（4）数据级的融合应用比较广泛，且主要都应用在图形图像处理领域，例如可以用于多个图像的合成、从图像中追踪物体的运动状态、多传感器遥感信息融合等方面。

#### (2) 特征层融合

特征层融合<sup>[30]</sup>是层次模型中处于中间层次的融合技术,从字面意义上,我们可以知道一定和数据的本身特征有关系。该技术主要就是在原始的收集数据中提取有用的特征信息,提取的特征信息可以目标的速度、方向、温度、图像的灰度,图像中物体的边缘等信息,根据提取到的特征信息再进行下一步的原始数据分类聚合等处理。由于该融合方法也是以目标为导向的,这就要求提取的特征信息和最终决策结果有关系,因此数据融合的目标就是能够尽量准确地做出决策和评估态势,这就要求提取的特征信息满足数据融合所需要的信息。图 2-6 展示了特征层融合的流程。

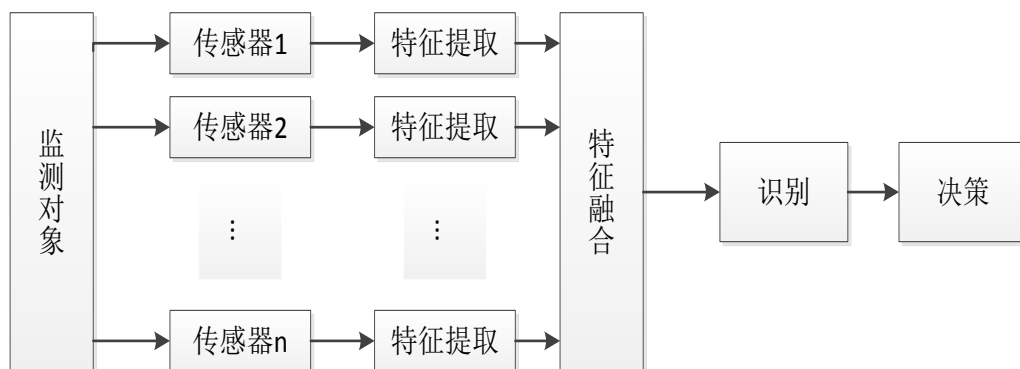


图 2-6 特征层融合流程

从图 2-6 所示的特征层融合流程可知,特征层融合首先要对多传感器采集的原始数据进行特征提取,再进行特征融合,融合之后再继续进行后续的认识和决策操作。

特征层数据融合的优点很多,容易实现实时数据处理,并且具有较高的精确度。有许多监测系统都用到了用到了这个层次的融合,例如情报系统、指挥系统、通信系统和控制系统。

### (3) 决策层融合

决策层融合<sup>[30]</sup>是层次模型中最高层次的融合技术。该技术主要就是根据一定的处理策略和决策的可信性,最终做出最优的决策规划,以便系统能够具有更好的实时性和容错性,在决策层处理之前,传感器内置的信号处理装置已经完成了决策任务和分类任务。这种方法能够在一个或者多个传感器失效的时候正常工作,不受传感器崩溃的影响,也增加了系统的健壮性。和特征层融合相似,决策层融合也是以结果为导向的,要将得到准确的决策结果最为问题的立脚点和出发点,因此就需要在提取原始数据的信息时,必须满足决策融合所需要的信息,进而达到最终的目的。图 2-7 展示了决策层融合的流程。

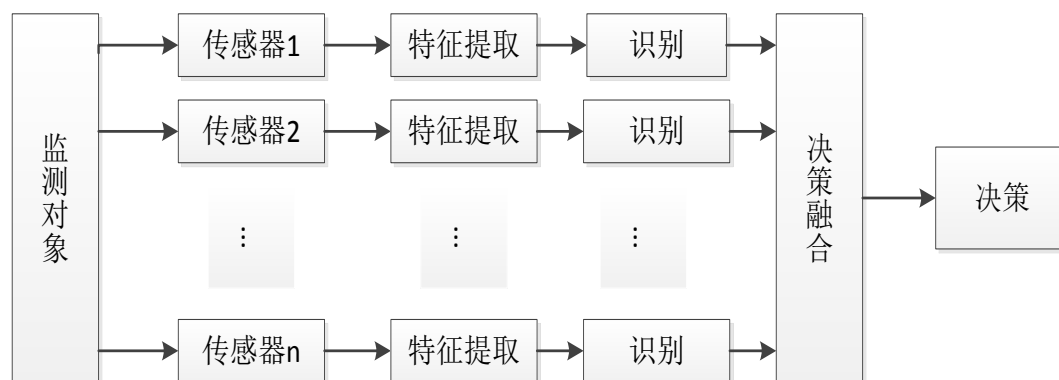


图 2-7 决策层数据融合流程

从图 2-7 所示的决策层融合流程图可知，决策层融合首先对多传感器采集的数据进行特征提取，再进行识别，然后再进行决策融合，最后做出决策。

决策层数据融合的有诸多优点。具体如下：（1）系统对传输数据信息的带宽要求比较低；（2）具有很高的灵活性和准确性；（3）能有效地反映环境不同角度的不同信息，对传感器的分离程度大；（4）当有传感器出现问题影响到正常工作时，可以通过适当的融合策略，可以让系统获得正常的的数据结果，就有很强的纠错能力。

#### （4）三个融合层次优缺点的对比

通过上文对数据融合三个层次的了解，为了更好的明确三个融合层次之间的差异，我们在几个方面进行了对比分析，具体的分析情况如下表 2-1 所示。

表 2-1 三个融合层次优缺点的对比分析

	数据级融合	特征级融合	决策级融合
处理信息量	最大	中等	最小
信息量损失	最小	中等	最大
抗干扰性能	最差	中等	最好
容错性能	最差	中等	最好
算法性能	最难	中等	最易
融合前处理	最小	中等	最大
融合性能	最好	中等	最差
对传感器的依赖程度	最大	中等	最小

从表 2-1 可知，对于数据层融合，处理信息量大，信息量损失小，抗干扰能力和容错能力差，融合前对数据处理的任务量最小，融合性能也是最好的，但是对传感器具有很强的依赖性；特征级融合在各个方面都是处于中间水平的；而决策级和数据层融合在各个方面都是完全相反的。

### 2.2.4 数据融合的方法

数据融合方法有很多种，可以按照数据融合的水平对算法进行划分，也可以按照经典算法和现代算法来进行划分。图 2-8 展示了数据融合算法按照融合水平进行的分类，图 2-9 展示了数据融合算法按照经典算法和现代算法的分类。

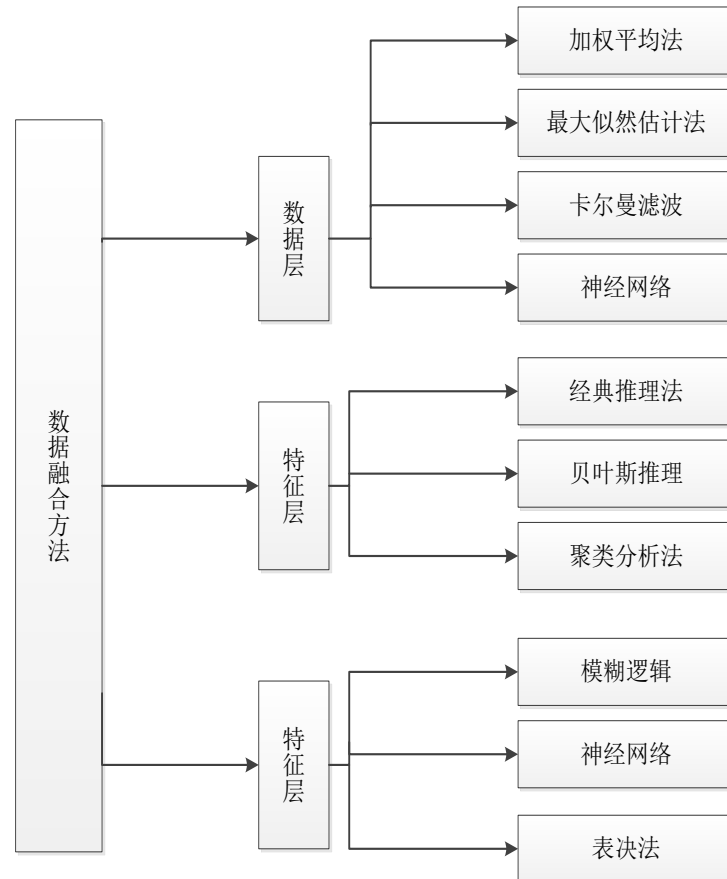


图 2-8 数据融合算法的分类

图 2-8 展示了将数据融合的算法按照数据融合的水平进行分类，数据层融合主要有加权平均法、最大似然估计法、卡尔曼滤波法、神经网络。特征层融合主要有经典推理法、贝叶斯推理法、聚类分析法。决策层融合主要有模糊逻辑法、神经网络法、表决法。



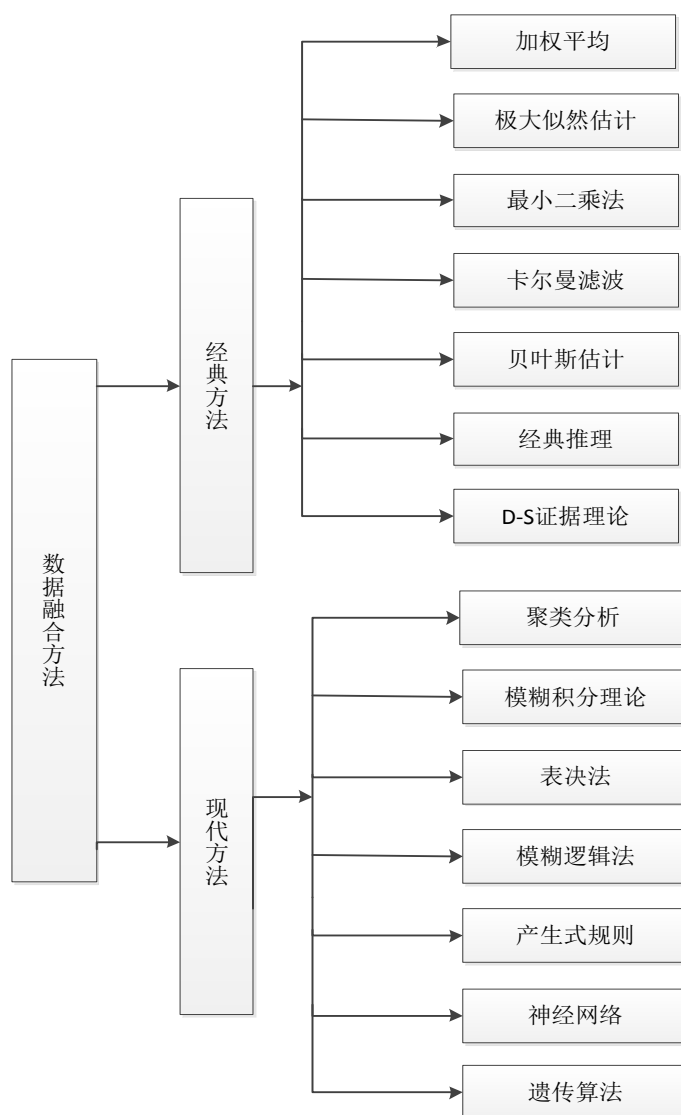


图 2-9 数据融合的方法划分

图 2-9 展示了将数据融合算法按照经典方法和现代方法进行分类。在实际应用过程中加权平均的思想比较常见，卡尔曼滤波算法也在许多应用中使用。

数据融合的算法比较多并且都比较复杂，为了更好的了解数据融合技术，我们将对比较常用的算法进行简单的描述，以便在以后的学习过程中能够进行更深入的研究，下面就对几种常用的算法进行简单的介绍。

#### (1) 加权平均法

加权平均法<sup>[31]</sup>是数据层融合的方法，该方法是最简单的融合方法，其主要思想就是利用周围的测量点来加权计算得出估测点的数据值。该方法实现简单，且应用方便，但是主要的难点就是在确定每个传感器所得的数据值在估测点中所占的权重，具体的也就是分析传感器后得到的传感器所占的权重信息。算法的具体计算过程如图 2-10 所示。

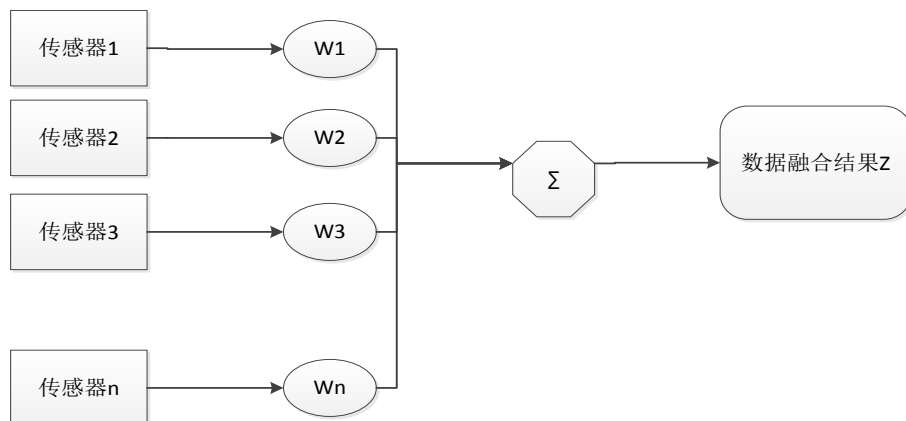


图 2-10 加权平均算法计算过程图

该算法的应用场景描述如下：假设有  $N$  个传感器节点对环境中的某一参数进行监测，监测结果的测量值为  $x_i$ ， $i=1,2,3\cdots N$ ，假设事先已经对传感器进行分析得出的权重为  $w_i$ ，则对环境估测值进行加权平均计算，得到数据融合的结果，具体的计算公式为：

$$X = \sum_{i=1}^N w_i x_i$$

## (2) 卡尔曼滤波法

1960 年，卡尔曼 (kalman) 发表的论文将卡尔曼滤波算法<sup>[32]</sup>带入到人们的视野，这篇论文主要描述的是用递归的方法解决离散数据的线性问题，其研究的基础和出发点是因为维纳滤波，维纳滤波理论起源于 40 年代，从理论和实际场景的应用上来讲，维纳滤波在解决问题的时候都要用到无限过去的的数据，这样的处理方式对于实时性很好的方法来说很有挑战性。基于这样的背景，卡尔曼滤波理论才应运而生，卡尔曼滤波把系统的状态空间模型引入到了滤波理论当中，并且用一套递推公式来描述。该算法的主要思想就是：通过分析系统之后，得到系统的状态模型，然后利用前一时刻系统的估测最优值和最有协方差，还有此刻系统的测量值来更新系统的现在时刻的估测最优值和最优协方差。通过不断的更新最优测量值和最优协方差，就实现了系统的不断递归更新。卡尔曼滤波器的功能强大，应用广泛，即使在不知道系统的确切模型性质时，依然可以估计出系统的当前状态和过去状态，还能够预测系统的未来状态。

卡尔曼滤波算法在数据融合层次模型中处于数据层融合，是一种低层次融合的方法，因为只对前一时刻的信息比较关注，所以可以用在实时系统中，对个多传感器的测量值进行数据融合处理。如果系统是不断变化的，并且系统中传感器的误差具体高斯白噪声模型的特性，那么卡尔曼滤波算法就是最好的数据融合处理工具。卡尔曼滤波器的原理如下：

## (1) 卡尔曼滤波器原理介绍

卡尔曼滤波算法就是用来估算离散时间过程的系统状态，也就是从时间  $k$  时刻的系统状态估算时间  $k+1$  时刻的系统状态，具体的离散随机方程如下所示。

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B} \mathbf{u}_k + \mathbf{w}_k \quad \text{其中 } \mathbf{p}(\mathbf{w}) \sim \mathbf{N}(0, \mathbf{Q})$$

假设系统的观测值为  $\mathbf{z} \in \mathbf{R}^m$ ，根据观测值和系统的当前状态得到系统的测量方程为：

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \quad \text{其中 } \mathbf{p}(\mathbf{v}) \sim \mathbf{N}(0, \mathbf{R})$$

其中系统的状态变化过程中的噪声用  $\mathbf{w}_k$  表示，系统的观测噪声用  $\mathbf{v}_k$  表示，并且假设都符合正态分布的白噪声。

## (2) 卡尔曼滤波器算法

卡尔曼滤波器就是用递归的方法来估测系统的状态。在系统初始化过程中先随机假设初始化系统的状态，通过观测值的不断递归校正，最终系统方程会趋于平衡。因此可以将卡尔曼滤波分为两个部分：时间更新方程和测量更新方程。时间更新方程主要就是根据系统前一时刻得到的最优状态和最优协方差得到当前时刻的系统估计值和协方差，我们称之为先验估计。测量更新方程主要就是利用测量值去校正系统的最优值和最优协方差，我们称之为后验估计。时间更新方程也可以看做是预估方程，测量方程可以看做是校正方程。具体的预估方程和校正方程如表 2-1 所描述。

表 2-2 卡尔曼滤波更新方程

离散卡尔曼滤波器时间更新方程
$\hat{\mathbf{x}}_{k+1}^- = \mathbf{A}_k \hat{\mathbf{x}}_k + \mathbf{B} \mathbf{u}_k$ $\hat{\mathbf{P}}_{k+1}^- = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^T + \mathbf{Q}_k$
离散卡尔曼滤波器状态更新方程
$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$ $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-)$ $\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-$

从表 2-2 中可知，时间更新方程描述了如何根据前一时刻的最优协方差和最优估计值得到当前时刻的协方差和估计值，实现从  $k-1$  时刻到  $k$  时刻的推进。测量更新方程描述了如何利用当前时刻的测量值去校正当前时刻的协方差和估计值得到当前时刻的最优协方差和最优估计值。具体的流程描述如图 2-11。

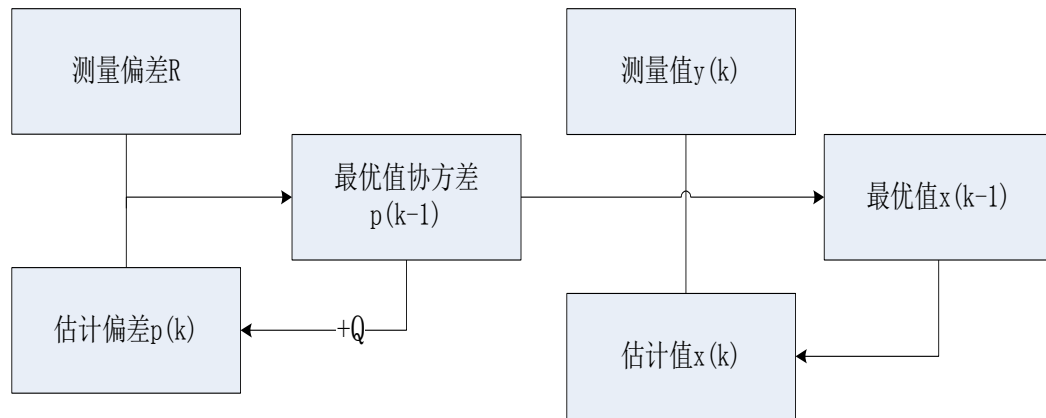


图 2-11 卡尔曼滤波融合过程

### (3) 贝叶斯估计

贝叶斯统计<sup>[33]</sup>是由英国数学家贝叶斯在 19 世纪提出的，该方法也是数理统计学中的一个重要分支。在 20 世纪 50 年代，H. Robbins 为代表的研究者们将经验贝叶斯方法和经典方法结合起来，并应用在计量经济学模型估计中，这种结合方法的使用得到了广泛的关注和应用。贝叶斯估计法对经典计量经济学模型估计方法的主要贡献源于它不仅仅利用样本的信息，还利用非样本的信息。

### (4) 证据推理方法

1967 年 Dempster 首先提出了证据理论这一概念，该方法也被称作为 D-S 证据推理方法<sup>[34]</sup>。证据推理方法是贝叶斯方法的补充和扩展，其主要包括三块内容：基本概率赋值函数、信任函数和似然函数等三个函数。证据推理方法的推理结构按照自上而下可以分为三级：第一级为目标合成，其主要作用就是把来自于传感器的各个观测值合成一个总的结果输出。第二级为推断，其主要作用就是获得传感器的观测结果并进行推断，并将传感器的观测值扩展成为目标报告。这种推荐的理论基础为：传感器的报告会以一定的可信度产生可信的目标报告。第三极为更新，对于每个传感器来说都存在着一定的随机误差，所以在连续的时间序列上，来自同一个传感器的连续报告比任何一个单一的数据报告都可靠，因此在数据推理和数据合成之前，先组合传感器的观测数据。

### (5) 神经网络

人工神经网络是模拟人脑神经结构而提出的信息处理模型。也被成为神经网络<sup>[35]</sup>。神经网络是一种运算模型，由大量的节点，也被成为神经元而构成。每个节点代表一种特定的输入函数，并且任何两个节点之间的连线都是一个权重值，这就是相当于神经网络的记忆。网络的输出就取决于网络的连接。对于网络自身来说就是自然界某一个算法或者是函数的近似逼近，也有可能是一种逻辑策略的表达。

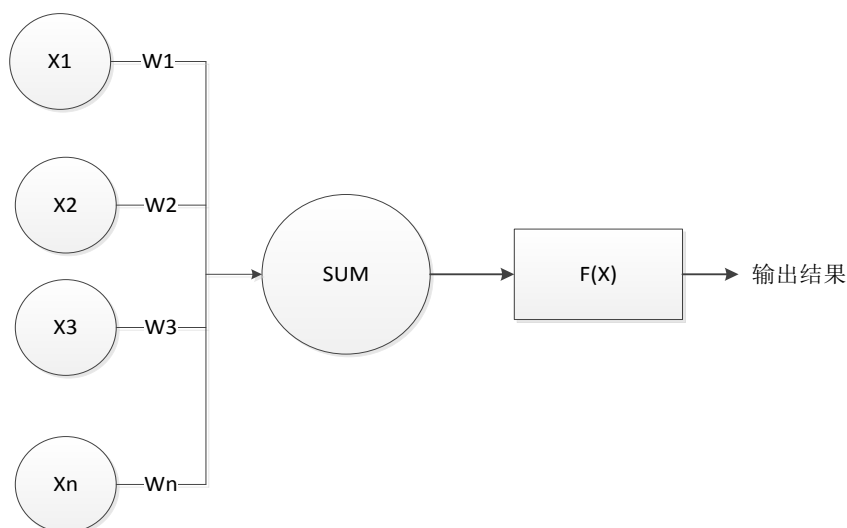


图 2-12 神经网络模型

图 2-12 展示了神经网络模型，其中  $X_1 \sim X_n$  为输入向量的各个分量， $W_1 \sim W_n$  为神经元的权重值， $F(x)$  为传递函数。由此可见，一个神经元的功能是求得输入向量与权重的乘积后，经过一个非线性传递函数得到一个标量结果。

#### (6) 数据融合算法的对比分析

数据融合的方法比较多，以上只对几种常用的数据融合方法进行了介绍，对于应用者来说，为了方便的选择适合自己的融合算法，我们将对这些算法进行对比分析，明确各个算法的适用范围，适用的信息类型，运行环境等。整理的对比分析情况如表 2-3 所示。

表 2-3 融合算法的对比分析

融合方法	运行环境	信息类型	信息表示	不确定性	融合计算	适用范围
加权平均	动态	冗余	原始读数 值		加权平均	低层次
卡尔曼滤波	动态	冗余	概率分布	高斯噪声	系统模型 滤波	低层次
贝叶斯估计	静态	冗余	概率分布	高斯噪声	贝叶斯估计	高层次
统计决策理论	静态	冗余	概率分布	高斯噪声	极值决策	高层次
证据推理	静态	冗余互补	命题		逻辑推理	高层次
模糊推理	静态	冗余互补	命题	隶属度	逻辑推理	高层次
神经网络	动/静态	冗余互补	神经元输入	学习误差	神经网络	低 / 高层次

从表 2-3 分析可知, 加权平均和卡尔曼滤波算法是底层次的融合算法, 神经网络算法既可以用在低层次, 也可以用在高层次, 其他几种算法都是高层次的。并且每种方法的融合计算方式也是不一样的。

## 2.3 数据融合技术在移动协作感知平台中的应用

随着智能终端的普及和无线传感器网络的快速发展, 数据融合技术在无线传感器网络中的应用也日趋变得成熟, 从应用通用的数据融合算法到算法的改进应用。近年来移动协作感知网络的诞生和快速发展, 数据融合技术再次受到人们的关注, 并也有一些研究和应用。

基于群智感知的语言识别系统<sup>[36]</sup>主要解决了群智感知应用场景下的背景噪声、特定人口音、信道带来的发音变形、以及多样化话语主题带来的大词汇存储问题等, 在这个系统实现过程中采用了线性插值算法对语言模型进行融合。基于 Android 平台的老人监护系统<sup>[37]</sup>主要针对收集上来的信息进行数据融合, 在低层融合中运用模糊逻辑和动态贝叶斯网络将传感器的原始数据融合成带有一定语义的信息片, 在高层融合中使用基于规则的推理方法将低级融合后的不确定性信息进行进一步的融合, 消除其不确定性, 得到精确的推断信息。基于参与式感知的气象服务系统<sup>[38]</sup>采用两级融合, 一级融合是采用平均法, 将采集数据与自动气象站数据利用平均法融合在一起; 二级融合采用倒数距离乘方法进行数据融合, 得到较准确的数据。参与式感知系统个体的移动特征识别<sup>[39]</sup>采用了贝叶斯定理进行初步的静止和移动行为的辨别。基于参与式感知的空气质量监测系统<sup>[40]</sup>主要针对利用智能终端设备采集环境信息, 由于分布的不均匀等特性, 克里金算法和马尔科夫算法进行对比分析, 然后用克里金算法实现了数据融合的功能。

## 2.4 本章小结

本章节主要介绍相关的技术, 移动协作感知技术和数据融合技术。首先介绍了移动协作感知技术的概念, 系统的层次结构, 系统的整体架构; 然后介绍了数据融合的概念、数据融合的功能模型以及数据融合的三个不同融合层次, 对每个层次进行了深入的介绍, 并对三个层级进行了对比分析。然后介绍了几种数据融合的方法, 从经典方法和现代方法将数据融合算法进行分类, 并对其中几种比较基础和常用的方法: 加权平均法、卡尔曼滤波法、贝叶斯估计法进行了简单的论述, 并对几种方法进行了简单的对比分析。通过对移动协作感知平台和数据融合理论的整体了解, 对数据融合的算法有个整体的认识, 对于在移动协作感知平台

中选取合适的方法具有指导性意义。

然后对移动协作感知平台上数据融合算法的应用进行简单的介绍,了解一些目前现有基于移动协作感知平台上的应用的具体数据融合的具体数据融合的算法。

## 第三章 移动协作感知平台数据融合模块的需求和设计

### 3.1 移动协作感知平台的需求和设计

#### 3.1.1 移动协作感知平台的需求分析

随着移动协作感知网络概念的提出，许多研究者开始将目光转移到这个方向，由于该平台具有很好的易搭建性、简单性和扩展性等特点，这些特点让该平台变成一种新兴的趋势。本文是通过在北京邮电大学校园内进行实际的部署，组织学生作为参与者参与到实验中来收集数据，进行日常外界环境数据的采集，包括声音、光照强度、温度、湿度、GPS、加速度等，利用采集的数据集验证移动协作感知平台上关键算法的有效性和可靠性，收集的实验数据也可以为进一步研究提供仿真数据支撑，突破目前协作感知方向的制约瓶颈。

移动协作感知系统的理想模型具体描述如下：首先打开客户端应用，对环境中的某一点进行拍照，并将照片存储下来，此时也将收集的感知环境的信息保存到本地，这时可以对照片进行本地PM2.5数据分析，可以选择将收集到的数据上传到服务器端，服务器端先将数据存储在缓存队列中，并最终读到分布式数据库系统中。收集上来的数据分布不均匀，时间空间无序，通过数据融合处理后，提供时间和空间有序、且均匀分布的数据，可以供 web 服务器端展示。在数据融合处理过程中，如果发现数据在某一区域内存在大量缺失，就可以启动激励模块，来激励用户上传缺失数据位置的数据，这时激励模块又要调用轨迹预测模块，轨迹预测模块会告诉激励模块哪些用户会去缺失数据的地方。最后就是激励模块给这些用户下发激励，鼓励用户去收集这个地方的数据并上传，这就是整个系统的理想模型。

根据系统的理想模型可以总结出系统的用例图 3-1，对于客户端用户主要包括拍照功能、浏览照片、响应激励、查询采集信息、PM2.5 分析，拍照同时还会采集环境的其他信息，包括声音、光照、经度、纬度、时间、加速度等参数。对于 web 端的用户主要包括查看融合之后的光照信息、噪声信息、以后空气中的 PM2.5 情况，还有浏览照片，来对环境 PM2.5 进行简单的感受。



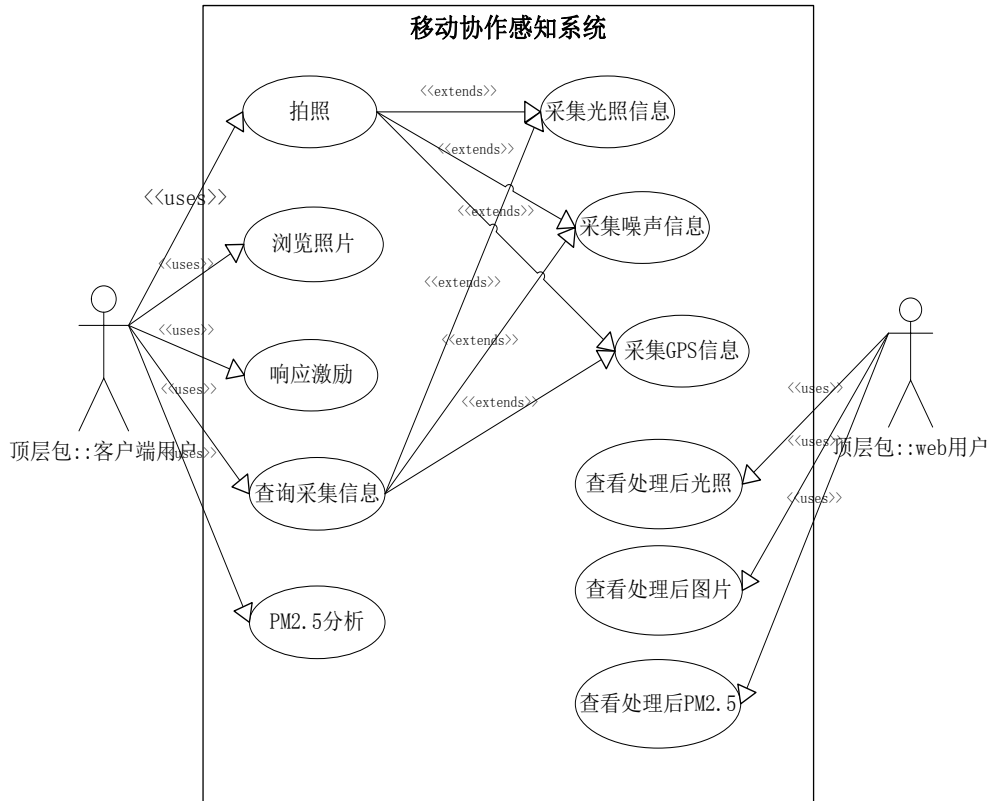


图 3-1 移动协作感知系统的用例图

移动协作感知系统的功能需求分为客户端和服务端。客户端功能分为：数据采集、数据存储、数据上传、PM2.5分析处理、UI展示；服务器端的功能分为：数据缓存、数据的分布式存储、数据融合、轨迹预测分析、下发激励、web展示等功能，我的主要工作量在数据融合模块。

移动协作感知系统的硬件设备的需求：10 台移动智能终端设备、采集数据的传感器节点、服务器、显示器设备、音频采集设备、无线路由器等网络设备。

移动协作感知系统的软件开发环境需求：客户端采用的是：Android 开发平台、JAVA 开发语言、eclipse 开发环境；服务器端采用的是 Hadoop 开发平台，JAVA 开发语言，HBase 数据库。

在整个移动协作感知系统的开发过程中，采用的是软件工程的设计流程，从需求分析到总体设计、详细设计、编码实现和测试等。

### 3.1.2 移动协作感知平台的整体设计

通过对移动协作感知平台的整体需求分析后，进行平台的设计和搭建。在北京邮电大学搭建立了一个由 10 个移动终端设备(主要是 10 台三星 S3 智能手机)、服务器、显示器设备、无线路由器等设备组成的移动协作感知与计算实验环境，利用 Android, Hadoop 等软件开发平台，为本项目的进一步研究与验证工作准备了必要条件。

图 3-2 展示了移动协作感知的系统结构图，参考了移动协作感知的基本架构进行设计的，主要有四部分组成：（1）感知数据层，主要就是移动智能终端进行数据采集和临时存储；（2）数据传输层，也就是无线网络，进行数据的传输功能；（3）数据存储处理层，也就是数据的存储操作和数据处理；（4）应用层，也就是基于感知数据提供的应用，用显示器进行展示。

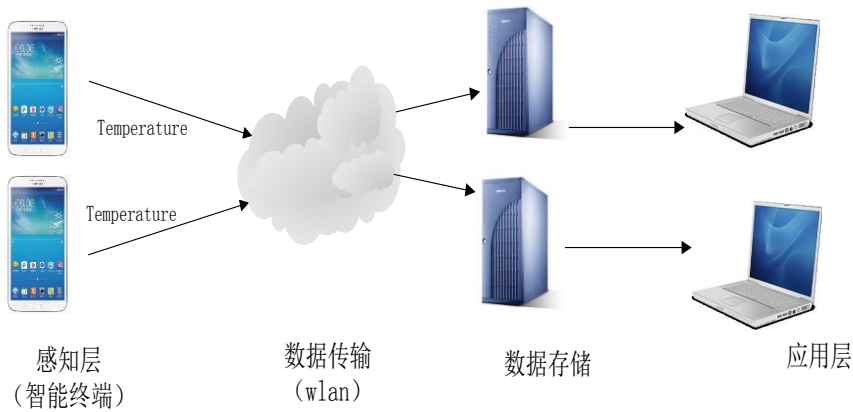


图 3-2 移动协作感知系统结构

参考了图 2-2 所示的移动协作感知平台的结构图，移动协作感知系统的物理结构图如图 3-3 所示，与图 2-2 的区别在服务器端，该图的数据处理和应用模块增加了 web 展示模块。

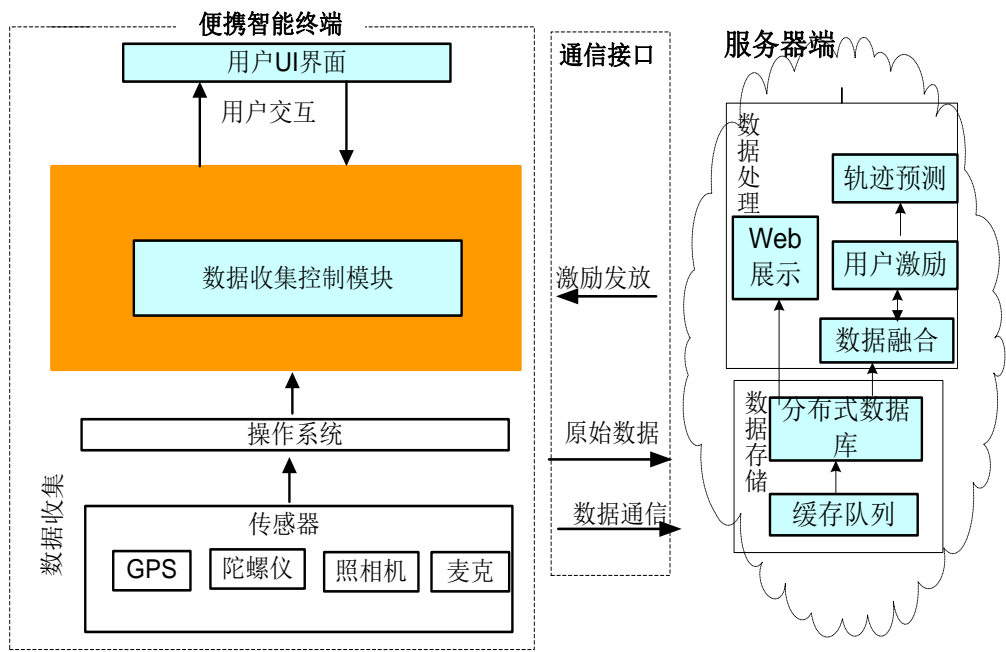


图 3-3 移动协作感知平台的物理结构图

移动协作感知系统模块按详细模块划分，主要分为两大部分：客户端和服务

器端。客户端主要包括主要分为用户 UI 模块、采集模块、PM2.5 分析模块、通信模块、客户端存储模块，服务器端主要包括 web 展示模块、分布式数据存储模块、数据融合模块、轨迹预测模块、激励模块、PM2.5 数据抓取模块。我主要负责是服务器端的数据融合模块的设计和实现。图3-4展示了该平台的详细模块。

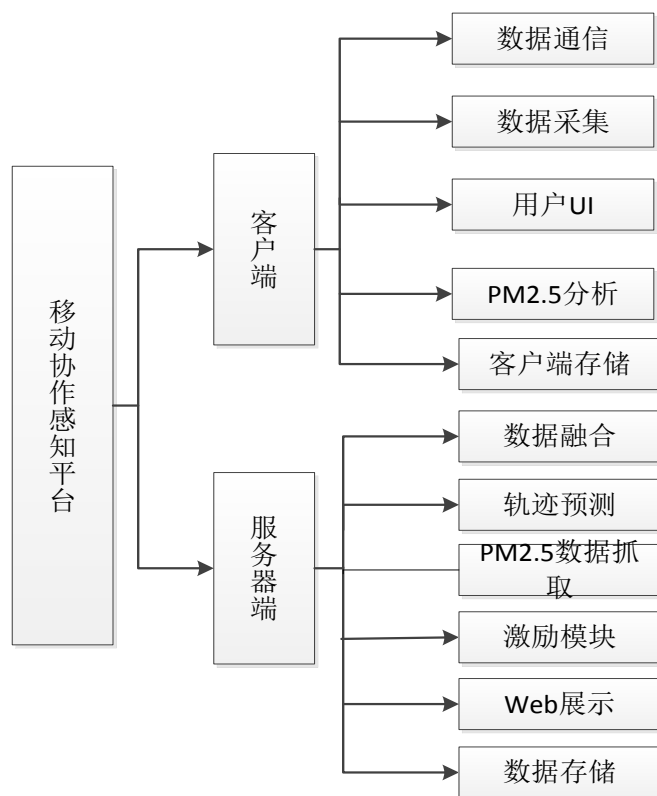


图 3-4 移动协作感知系统功能结构

图 3-4 展示了该系统所具有的所有功能模块，对于客户端各个模块之间具体的通信关系如图 3-5 所示的客户端逻辑结构图，该图具体的描述了该平台客户端各个模块之间的通信关系。

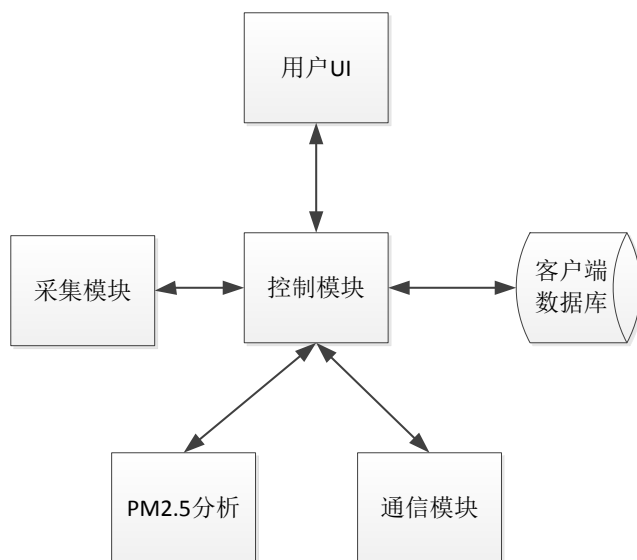


图 3-5 移动协作感知客户端逻辑结构图

服务器端的逻辑结构图如图 3-6 所示，该图描述了服务器端的各个模块之间的数据通信，其中通信模块是客户端和服务端通信的接口模块。我的主要工作在服务器端的数据融合模块，主要和分布式数据库、激励模块进行通信。

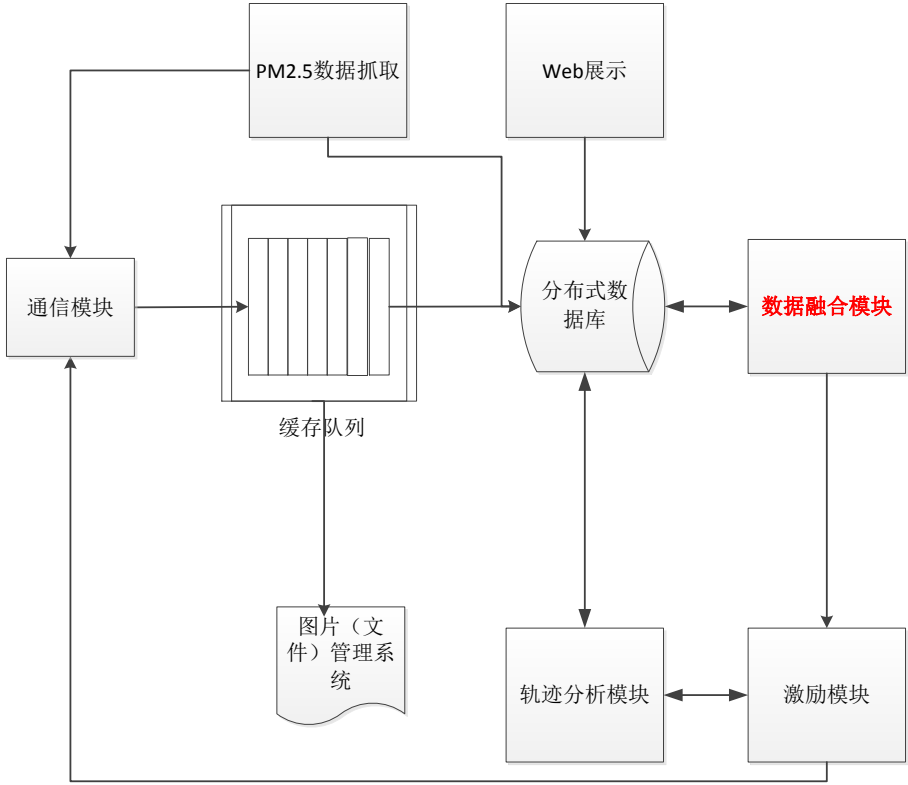


图 3-6 移动协作感知服务器端逻辑结构图

客户端和服务端端的逻辑结构图就组建了整个移动协作感知平台的系统逻辑结构图，其中通信模块是端的桥梁，负责数据通信。

### 3. 2数据融合模块的需求

在移动协作感知系统中，用户携带移动智能终端来收集数据，但是由于用户的移动性和随机性，因此收集到的感知数据分布不均匀，存在着数据冗余和缺失的问题，这样的数据集对于环境的监测和评估都会有一定的影响，因此数据融合就变得尤为重要。在本文中对数据融合模块采用的是软件工程的设计流程，首先进行模块的需求分析，然后再进行整体架构设计和模块的详细设计。

#### 3. 2. 1数据融合模块的需求分析

本课题就是对移动协作感知平台数据融合进行研究，在北京邮电大学的校园

内采集数据，由于智能终端携带者的移动性，导致了收集的数据分布不均匀；由于移动智能终端型号和操作系统等方面的差异，导致了收集数据的异构性，这都是在实际数据收集过程中存在的问题。假设收集到的数据的分布情况如数据分布比较不均匀，杂乱无章，实际结果可能比这还要糟糕，如果想要监测学校某一固定位置的外界温度情况，那该位置（不是确切位置点，粗略计算）在一段时间内可能有多条数据信息，在这种情况下，周围任何一数据点对于估测位置点的数据信息都是有幫助的，因此我们采用数据融合技术来解决该问题。在此我们通过采用网格策略，对每一个小区域进行单独融合，得到的融合结果作为这个网格的评估值，每个区域有一条融合结果来代表该区域的估测值。

数据融合模块是由时间来监控触发的，是时间进行驱动的数据融合，该模块的具体说明：（1）在一天的时间范围内，可以事先设定一个（或者几个）触发数据融合模块的时间点，当时间点到达时，数据融合模块将自动启动，进行数据处理；（2）对数据采集区域事先进行网格划分，使每一个数据采集点都在一个网格中，在从数据库中读取数据时，都是取出一个网格内一定时间范围内的数据。

数据融合模块的用例图如图 3-7 所示，包括的内容如下：（1）取数据：从分布式数据库的原始数据表中读取数据，读取的数据主要是一个网格内的一定时间范围内的数据；（2）数据计算：根据一定的数据融合算法对该数据进行计算处理，得到最终的数据融合结果；（3）启动激励模块：在数据计算过程中，如果发现数据的缺失度大于一定的阈值，则触发激励模块进行激励机制下发；（4）存入数据：将计算得到的融合结果存入到分布式数据库的融合数据表中。

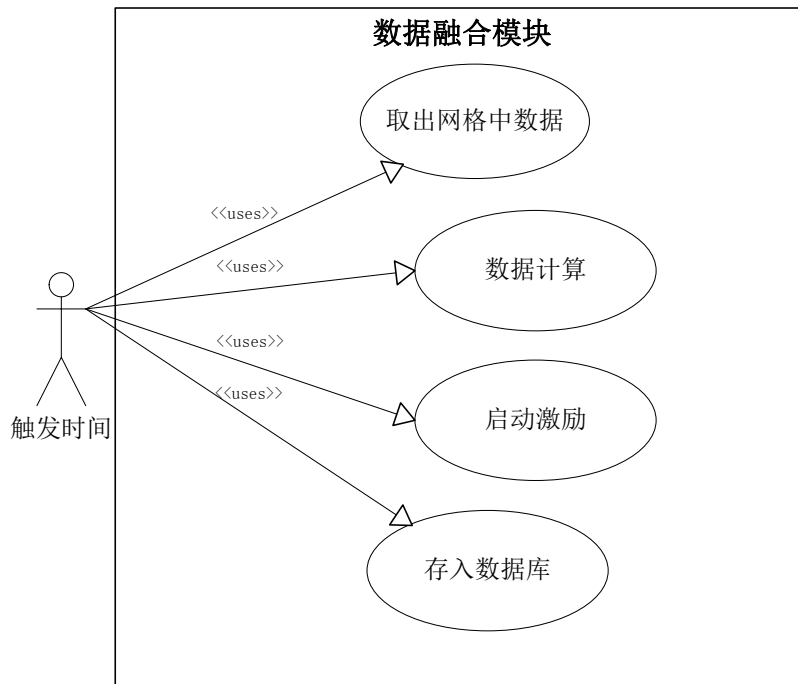


图 3-7 数据融合模块的用例图

数据融合过程中具体的数据流图如图 3-8 所示，该图描述了数据的来源和去处。

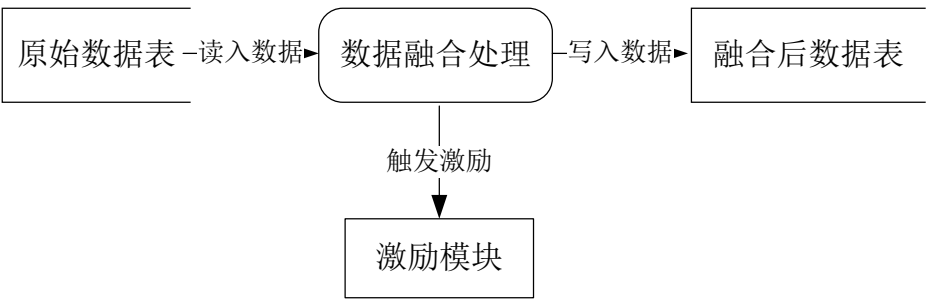


图 3-8 数据流图

数据流图展示了数据在数据融合模块的数据流向，从原始数据表中读出数据经数据融合模块处理后，最后写入融合后的数据表，在处理过程中，发现数据缺失，还会触发激励机制。

3.3数据融合模块的设计

在移动协作感知系统中，数据经过用户收集、上传，最终会存储到服务器端的分布式数据库中，对于数据融合模块，将从数据库中按照一定的格式取出有用的数据，并经过融合处理后，将数据以一定的格式又存储到数据库中，以便其他应用调用。数据融合过程中的消息流向如图 3-9 所示。

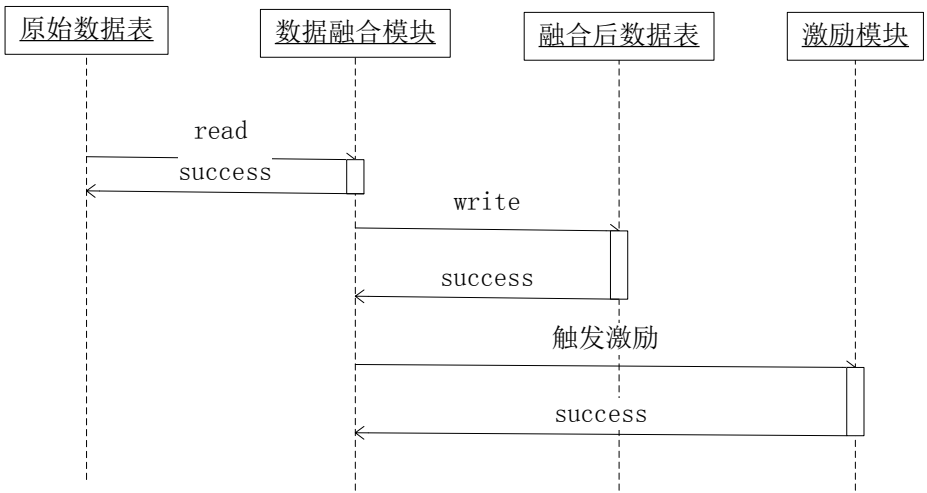


图 3-9 移动协作感知数据融合的消息图

图 3-9 展示了数据融合模块和数据库模块、激励模块的通信。应用场景为：从原始数据表 3-1 中读取融合所需要的采集数据，然后进行数据融合处理，如果在数据总量小于阈值或者数据存在缺失，则会触发激励模块进行下发激励，并同

时将融合的结果存储到融合数据表 3-3 中。

### 3.3.1 数据融合相关表的设计

数据融合模块相关表是根据收集数据的性质和数据融合的需求进行的。数据的存储设计对于数据操作来说非常重要，必须方便数据的存取。对于数据融合模块所涉及的数据表有三种，分别是：原始数据表、网格表、融合后的信息表。

#### (1) 原始数据信息表

该数据表中存储的是原始收集上来的数据，也就是未经过处理的数据。表 3-1 介绍了原始数据的各个数据项。

表 3-1 原始数据信息表

字段	类型	说明
id	int	记录的唯一标示
uid	int	用户的唯一标示
x	double	经度
y	double	维度
time	varchar	时间
humidity	int	湿度
temperature	int	温度
light	int	光线强度

#### (2) 网格表

通过对每一个融合的区域进行网格划分，进而对每一个网格内的数据进行融合处理。具体的融合策略在下文中介绍，具体你的网格表就如图 3-2 所示。

表 3-2 原始数据信息表

字段	类型	说明
id	int	记录的唯一标示
gridid	int	每个网格的唯一标示
grid_x	int	每个网格的 x 坐标
grid_y	int	每个网格的 y 坐标

#### (3) 融合之后的数据表

对每个网格内的数据进行融合之后，通过网格标示和网格内的具体数据项进行融合之后数据的记录，表 3-3 呈现了融合之后数据表的设计。

表 3-3 融合之后的数据信息表

字段	类型	说明
id	int	记录的唯一标示
gridid	int	网格的唯一标示
light	int	融合之后的光照值
temperature	int	融合之后的温度值
humidity	int	融合之后的湿度值
time	int	时间

3.3.2 数据融合模块接口设计

数据格式定义对于数据处理模块和各部分数据接口来说非常重要，在该移动协作感知平台中采用的是分布式数据库存储，由于这块数据存储不是由我来负责，因此了解取数据的格式和存入的数据格式对于这两个模块的接口通信就会变得很重要，因此在明确我取数据格式和存入数据的格式之后，这个接口函数就会由其他同学封装好，只提供给我们接口调用即可。因此明确模块之间调用所涉及到的数据内容以及数据格式，对于数据操作者和数据库管理者来说将会变得非常重要。

对数据处理之前，通过调用封装好的读取数据函数，将数据按照一定的格式从数据库中取出，也就是在数据融合处理过程中所用到的字段，然后融合处理之后，将数据按照一定的组织存入到数据库中。在数据融合过程中会存在数据缺失的情况，这就要启动激励机制。具体的和数据库模块和激励模块的接口设计如下：

(1) 从数据库中取出数据的接口

- 功能：返回一段时间内收集的数据
- 参数：（起始时间，结束时间）
- 返回值：如下表 3-4 所示

表 3-4 读取数据的格式

数据字段	字段含义	备注
Longitude	经度	
Latitude	纬度	
Time	时间	
Temperature	温度	

(2) 存储数据的数据接口

在数据融合过程中，采用小格式的方式进行融合，主要就是将区域分



为若干个小格子，然后将每个小格子内的数据进行融合处理，作为该格子的评估值。具体的格子横坐标  $X\_ID$  和纵坐标  $Y\_ID$  在第二部分会有介绍，也介绍了小格子的划分策略。经过数据融合后的数据存入数据库接口如下定义。

➤功能：将融合后的数据存入数据库

➤参数：如下表 3-5 所示

表 3-5 数据融合后的数据格式

数据字段	字段含义	备注
$X\_ID$	格子横坐标标示	
$Y\_ID$	格子纵坐标标示	
Time	时间	
Temperature	温度值	

➤返回值：存入成功返回 `true`，失败返回 `false`。

### (3) 和激励模块的数据接口

当对每个小格子内的数据进行数据融合时，一旦存在小格子内的数据缺失或者存在的数据量少与一定的阈值，这时就要触发激励模块，告诉激励模块哪个小格子区域内缺失数据，以便激励模块下发激励。

➤功能：触发激励机制，鼓励用户上传数据。

➤参数：如下表 3-6 所示

表 3-6 数据融合后的数据格式

数据字段	字段含义	备注
$X\_ID$	格子横坐标标示	
$Y\_ID$	格子纵坐标标示	
type	类型：声音，光照	

➤返回值：无

### 3.3.3 网格划分策略

对于一个较大的区域范围内，采集的数据量特别大，并且同一时间点上，该区域相差较远的位置确实测量值也相差较大。这种情况下，一个大区域内的数据彼此之间是有很大大偏差的，基于这样情况的思考，采用划分网格的策略对于解决该问题将会是一个很好的办法，也就是将区域再进行细分。在实际情况中，将会是对收集的数据进行分析，然后再决定网格的大小。但在该算法的实现中，就是人为的将北京邮电大学的区域分为  $M \times N$  的小格子。这种人为的选取，主要就是

借鉴这种方法的使用。因为实际情况下，北京邮电大学的整个学校区域本身就不是很大，完全可以不划分网格，整个区域都可以作为一个整体来进行数据处理。但在将来如果不仅仅收集北京邮电大学这个范围内，而是整个北京的数据时，可以采用这个网格的概念和思想进行实现。下面就将介绍一下划分北京邮电大学为几个网格的具体方法。

对于北京邮电大学的位置范围，取最大矩形作为该区域范围，然后取对角的两个点  $(x_1, y_1)$  和  $(x_2, y_2)$  作为描述边界范围，将该区域划分成  $M \times N$  个小格子，具体划分成的小格子用格子坐标来表示。对于每一个格子用格子坐标  $(X\_ID, Y\_ID)$  来表示，计算过程如下：对于该区域中的任意一点  $(x, y)$ ，首先检测该点是否位于该区域中，如果在，则计算该点所属的区域  $(X\_ID, Y\_ID)$ ，计算过程如下：

$$X\_ID = \frac{(x - x_1)}{M}, \quad Y\_ID = \frac{(y - y_1)}{N}, \quad \text{并且所得的两个值都向上舍入取整。}$$

这样就可以将每一个观测点都找到了所属的小格子区域内，当然也可以对所属于同一小格子内的数据进行获取，在通过设定时间阈值的限制，就可以获取同一小格子内一定时间范围内的数据，最后就可以对所属于同一小格子内的数据进行融合。

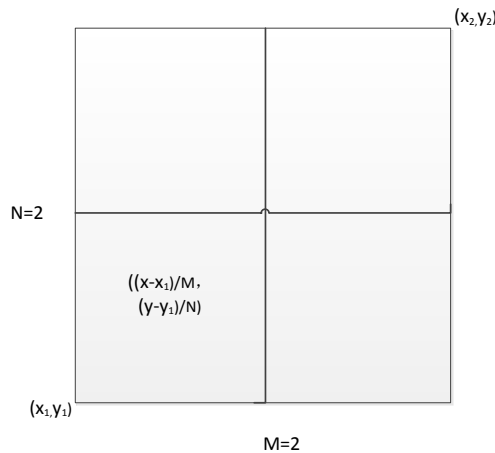


图 3-10 小格子具体划分策略

小格子的使用方便了大范围数据的融合过程。该方法的借鉴和使用为将来大范围收集数据打下了基础。

### 3.4本章小节

本章主要采用软件工程的设计流程进行整个需求分析和设计阶段的任务，主要介绍了移动协作感知平台的需求分析和设计，以及该平台中数据融合模块的需

求分析和设计，移动协作感知平台的设计，主要由服务器端和客户端组成，其中客户端包括：数据采集模块、用户 UI 界面模块、数据上传模块、PM2.5 处理模块、通信模块。服务器端模块包括：数据库存储模块、数据融合模块、轨迹预测模块、激励机制模块、web 展示模块、PM2.5 数据抓取模块等；再接下来对数据融合模块的需求进行分析和整体内部结构进行设计，并也介绍了数据融合模块和其他模块的通信流程的设计；最后介绍了数据融合模块的内部设计，主要包括对数据存储的数据库设计、融合模块和其他模块的接口设计、融合过程中用到划分网格的策略等。

## 第四章 克里金算法改进和模拟数据生成

### 4.1 基于空间模型的克里金算法的分析

从第二章中对数据融合相关技术的分析可以了解到数据融合的基本方法，但是这些方法大都是应用在固定多传感器网络中，在移动协作感知系统中还没有得到应用。插值算法介绍了克里金算法被用在移动协作感知系统中解决数据冗余和数据缺失的问题，并且取得了很好的效果。该方法本质也是利用数据融合中加权平均法的思想，本文中拟打算采用克里金算法来解决数据融合的问题，下文中对克里金算法进行具体的分析。

1951 年由南非采矿工程师克里金提出了一种数据插值算法——克里金法<sup>[41]</sup>，该算作为地质统计学中的一种插值计算方法。该方法是一种线性求最优解的空间插值方法。通过对观测值之间相互位置关系分析后，对每一个观测值赋予一定的权重值，通过加权平均的方法最后得到估计值。和一般插值法相比，权重的计算方法不仅仅是根据测量点和估计点之间的关系来计算，还考虑到各测量点之间的空间相关程度，也就是半方差。这样计算得到的权重会更加的准确和无偏。该算法的主要核心公式就是半方差公式。算法具体计算步骤如下：

- (1) 首先根据假设有  $n$  个测量点  $z_x$  ( $x=1,2,\dots,n$ )，然后根据  $n$  个观测点的值和空间位置相关性计算半方差，其计算公式为：

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n (z(x_i) - z(x_i + h))^2 \quad (4-1)$$

其中  $n$  为带测量值的点的数量， $h$  为各个测量点之间的距离， $z$  是测量点的测量值。

- (2) 根据上述计算得出不同距离  $h$  的半方差值  $\gamma(h)$ ，然后绘制  $h$ - $\gamma(h)$  的半方差图，半方差图中有三个参数：块金、基台值和变程。其中块金表示距离  $h$  为零时的半方差，也就是拟合曲线和  $\gamma(h)$  轴的交点；基台值表示  $\gamma(h)$  基本达到恒定的半方差值取值范围；变程表示一个取值范围，在该范围内半方差随距离  $h$  的增加而增加，超过该区域范围，半方差值将会趋于恒定。利用得出的点坐标  $(h, \gamma(h))$  用 MATLAB 工具编程拟合出半方差模型和公式，以球状模型为例。

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + c \left( \frac{3h}{2a} - \frac{h^3}{2a^3} \right), & 0 < h \leq a \\ c_0 + c, & h > a \end{cases} \quad (4-2)$$

(3) 用拟合的模型计算出三个参数。例如球状模型中块金为  $c_0$ ，变程为  $a$ ，基台值为  $c$ 。

(4) 然后根据半方差公式计算权重  $w_i$ ，假设有三个测量点。

$$\begin{bmatrix} C(h_{11})C(h_{12})C(h_{13}) \\ C(h_{21})C(h_{22})C(h_{23}) \\ C(h_{31})C(h_{32})C(h_{33}) \\ 1101 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} C(h_{10}) \\ C(h_{20}) \\ C(h_{30}) \\ 1 \end{bmatrix} \quad (4-3)$$

其中： $\gamma(h) = C(0) - C(h)$ ， $C(0) = c + c_0$ ，则  $C(h_{12}) = C(0) - \gamma(h_{12}) = c + c_0 - \gamma(h_{12})$ ，

(5) 利用拟合的模型来估算未知点的属性值，具体方程为：

$$z_0 = \sum_{i=1}^n z_i w_i \quad (4-4)$$

以上就是对克里金算法的整体描述，克里金算法的主要思想就是通过周围的测量点和观测值，利用空间位置之间的相关性来计算每个测量点所占的权重值，进而运用加权平均就可以得到最终的估测值。克里金算法主要是应用在地质学中，例如分析地质的整体温度情况、地下水含量情况等，该方法是属于网格化的方法，对分析的地带进行网格划分，然后再进行分析，具体的划分网格的策略视具体情况而定。该方法主要是考虑了空间各个位置点观测对象的大小、形状、温度等具体信息，并以线性无偏和最小估计方差的方法进行估计，该方法是一种光滑的数据插值方法，在数据量较大，并且数据点分布均匀时，估计结果会变得更加的准确。

## 4.2 基于空间模型的克里金算法的改进

通过以上对克里金算法的了解，我们可以知道克里金算法在计算各个观测点所占权重的时候，只根据各个观测点的空间位置关系以及观测点和估计点的空间位置关系来计算，基于这样只考虑空间位置的关系，而忽略了时间因素的方法，在实际应用中会存在很大的偏差。

在移动协作感知系统的实际应用过程中，由于节点的移动性和疏密程度的不

同,相比于固定传感器网络的固定节点来说,采集数据的时间不可确定,并且数据上传的频率也是不可确定的,这些因素就导致数据没有固定的收集频率,收集的数据会存在时间点上的杂乱无章,也会存在位置点上的无序,而且数据会在一个小区域范围内存在很多数据,并且也会存在无数据的位置点。在这样的数据背景下,利用这样的数据集,采用克里金算法对环境指数去进行评估就会变得很困难。对于只考虑空间维度的克里金算法而言,当用观测点去评估某一点时候,如果周围的点中有的空间位置距离较小,但时间间隔很大,其所占的权重应该比较小,或者就不应该被考虑进来。但是在考虑空间位置关系的克里金插值算法中,即使时间间隔比较长,但是空间位置离得较近的点所占的权重依然是很大的,所以这与实际的情况有点不大相符合。

基于以上存在问题的分析,时间因素对于克里金算法也是一个非常关键的因素。只有空间位置的考虑就会具有片面性和单一性,因此把时间因素考虑进来将会是一个很大的突破。在这里我们引入一个 **product** 模型<sup>[42]</sup>,该模型是基于克里金算法的空间维度的思考,并把时间考虑进来的方法。该模型主要思想就是给出了基于时间和空间维度的协方差和半方差公式,由上文对克里金算法的了解可知,通过半方差公式就可以实现该算法,并计算出权重,进而得到估测值。具体的 **product** 模型公式如下文。

#### 4.2.1 Product 模型介绍

##### (1) 基本公式

**Product** 模型主要就是针对时间和空间维度来考虑克里金算法,其中主要给出了时间和空间结合的混合协方差公式,时间的协方差和半方差公式可以按照空间的协方差和半方差的计算方法计算得到,详情可参照上文克里金算法的介绍,时间和空间结合的协方差公式如 4-5 所示。

$$C_{st}(h_s, h_t) = C_s(h_s) \bullet C_t(h_t) \quad (4-5)$$

其中  $C_{st}(h_s, h_t)$  是时间和空间的协方差,  $C_s(h_s)$  是空间协方差,  $C_t(h_t)$  是时间协方差。其中协方差  $C_{st}(h_s, h_t)$  和半方差  $\gamma_{st}(h_s, h_t)$  的关系如下:

$$\gamma_{st}(h_s, h_t) = C_t(0)\gamma_s(h_s) + C_s(0)\gamma_t(h_t) - C_{st}(h_s, h_t) \quad (4-6)$$

其中  $\gamma_s(h_s)$  和  $\gamma_t(h_t)$  分别为空间和时间的半方差。

##### (2) 计算方法

基本公式如下:  $z_0 = \sum_{i=1}^n z_x w_x$ , 其中计算权重值是该算法的重点和关键。权重公式如公式 4-7 所示, 其中  $[W]$  为各个观测点所占权重构成的矩阵,  $[K]$  为任意两个测量点之间的协方差构成的矩阵,  $[M]$  为任意一个测量点和估测点的协方差。

$$[W] = [K]^{-1} [M] \quad (4-7)$$

其中：

$$W = \begin{bmatrix} w_1 \\ w_1 \\ \dots \\ w_n \\ \lambda \end{bmatrix}, \quad K = \begin{bmatrix} C_{11}C_{12}\dots C_{1n}1 \\ C_{21}C_{22}\dots C_{2n}1 \\ \dots\dots\dots \\ C_{n1}C_{n2}\dots C_{nn}1 \\ 11\dots 10 \end{bmatrix}, \quad M = [C_{01}C_{02}\dots C_{0n}1]^T, \quad C_{ij} \text{ 表示的是第 } i \text{ 个点}$$

和第  $j$  个点的时空协方差。

根据以上公式就可以计算出各个测量点所占的权重值，进而得到估测点的值。以上公式就介绍了 **product** 模型的整体使用情况，主要就是计算各个观测值所占的权重。通过以上的步骤就可以得到估测点的值，进而可以实现时间和空间结合的数据融合算法。

该模型的引进可以在不同维度出发综合起来去计算各个测量点在估测点位置的权重值，增加一个时间维度，也为克里金算法的扩展打下了基础。

### 4.3 模拟数据集生产

#### 4.3.1 GeoLife 数据集描述

GeoLife<sup>[43-45]</sup> 是基于微软 Virtual Earth 的由 GPS 数据驱动的社交网络。其目标是建立一个可以进行个人管理的、可视化的网站，在该网站上可以根据收集自己的 GPS 轨迹信息来了解他们自己的历史记录。并且通过彼此间的相似的轨迹信息，建立互相之间的各种关系。

目前 GPS 位置信息采集设备的流行使拥有大量用户的位置信息成为可能，通过 GPS 可以记录用户的历史位置，这些位置可以体现用户的活跃地点。GeoLife 的出发点是通过每个用户上传的历史位置序列信息，通过信息的挖掘可以找到周围有共同兴趣爱好的朋友。拥有这样的应用，用户就可以很方便地根据系统的推荐结果发出邀请，一起参加活动，如徒步、自行车出游或驾车旅行等。这样的活动会带来更多用户的数据信息，丰富数据资源，并且可以为更多的用户提供更好的服务。

GeoLife 的数据集是通过测试他们的网站应用而采集的用户轨迹信息，这些数据也是数据挖掘的基础，进而也是推荐朋友等功能的基础。数据集的具体信息如下：该数据是由 182 个用户花费三年多的时间来收集的（2007.4-2012.8）。每一个 GPS 轨迹都是由一序列连续时间的位置点组成的，每个点都包含着经度、纬度、海拔等信息。整个数据集的位置在采集北京市内的位置点，包含 17621 条轨迹，大约总距离在 120 万公路，总时间约为 48000 多小时。数据的采集频率为

(1s-5s)/点, 或者 (5m-10m)/点。

该数据集记录了用户一连串的户外活动, 不仅仅包括用户的日常行为, 例如: 回家, 上班。而且也包括用户的一些娱乐活动和运动活动, 例如: 购物, 游览, 设宴, 徒步旅行等。该数据集可以被许多领域所使用, 例如用户行为识别, 基于位置的社交网络, 位置隐私和位置推荐等方面。

为了模拟移动协作感知平台收集的数据集, 我们从 GeoLife 数据集中截取了位于北京邮电大学的 1670 条数据, 模拟移动协作感知平台上的用户的轨迹, 其中包括经度、纬度、时间等信息, 主要就是北京邮电大学内的若干条轨迹信息, 具体的位置点分布情况如图 4-4 所示, 其中数据主要分布在北京邮电大学校园内的主干道上, 并且轨迹点分布比较集中。具体的北京邮电大学范围的数据表中的数据信息如下图 4-1 所示。

id	uid	tid	x	y	time
1	0	168	39.970766000000000	116.363640000000000	2008-12-10 02:35:03
2	0	168	39.970679000000000	116.363620000000000	2008-12-10 02:35:08
3	0	168	39.970598000000000	116.363672000000000	2008-12-10 02:35:13
4	0	168	39.970302000000000	116.363696000000000	2008-12-10 02:35:18
5	0	168	39.969881000000000	116.363696000000000	2008-12-10 02:35:23
6	0	168	39.969405000000000	116.363716000000000	2008-12-10 02:35:28
7	0	168	39.968896000000000	116.363735000000000	2008-12-10 02:35:33
8	0	168	39.968310000000000	116.363753000000000	2008-12-10 02:35:38
9	0	168	39.967693000000000	116.363762000000000	2008-12-10 02:35:43
10	0	168	39.967089000000000	116.363760000000000	2008-12-10 02:35:48
11	0	168	39.966830000000000	116.363791000000000	2008-12-10 02:35:53
12	0	168	39.966857000000000	116.363734000000000	2008-12-10 02:35:58
13	0	168	39.966850000000000	116.363734000000000	2008-12-10 02:36:03

图 4-1 北京邮电大学的轨迹信息

#### 4.3.2 GreenOrbs 数据集描述

GreenOrbs<sup>[46]</sup> 绿野千传项目, 是无锡清华信息科学与技术国家实验室物联网技术中心发起的面向长期大规模无线传感器网络开展研究、实验、部署和应用的科研项目。该项目主要致力于在林业碳汇、城市碳排放和空气质量监测等领域同时展开跨学科研究实践, 分别在浙江临安天目山自然保护区和江苏无锡高新技术开发区部署了当时世界上可持续稳定运行的规模最大的无线传感器网络系统。

在 2008 年的下半年“绿野千传”项目开始进入研发阶段。系统主要是用来监测森林生态环境, 并且使用传感器收集环境中的各项指标, 其中具体包括湿度、温度、二氧化碳浓度和光照强度, 采集到的多种数据可以为其他应用提供数据支撑, 例如森林监测、火灾风险评估、野外救援等方面。该项目在无线传感器网络领域建立长期大规模的监测系统很具有前瞻研究与探索性。

GreenOrbs 项目实现全年的森林生态监督, 收集各种传感器信息, 其中包括:



温度、湿度、光照强度、二氧化碳浓度等。目前部署的固定传感器节点只有 267 个，节点的采集频率为每 10 分钟采集一次信息，其中收集到的数据信息包括：节点 ID 编号、经度信息、纬度信息、光照强度、温度信息、湿度信息。整个数据集就是 265 个固定点提供的固定时间间隔（10 分钟）的信息值，其中有两个固定传感器节点没有上传过数据，因此属于无效节点，不具有参考价值，可以直接剔除，因此就只剩下 263 条有效的数据记录。

为了模拟移动协作感知平台收集的数据集，第一章中介绍了从 GeoLife 数据集中提取的北京邮电大学的数据集可以得到模拟移动协作感知系统的用户轨迹信息，但是缺少实际的环境观测值。对此我们从 GreenOrbs 数据集中提取 10 分钟内各个传感器收集到的数据，这些数据就可以模拟移动协作感知平台收集数据的过程。其中 GreenOrbs 数据的具体分布情况如图 4-4 所示。主要是一些散乱的固定位置点和该点上的数据值。在后续的第三节中我们有详细介绍模拟数据集的生产过程。下图 4-2 显示了 greenorbs\_data 数据表的具体信息，图 4-3 显示了 greenorbs\_location 数据表的具体信息。

id	location_id	temperature	light	humidity	time
151	1003	32.0500	2	6590	2011:08:04:00:07:16
192	1004	30.9500	3	6602	2011:08:04:00:08:34
171	1006	31.2400	3	6569	2011:08:04:00:07:53
143	1007	30.6100	2	6616	2011:08:04:00:07:11
150	1008	31.1700	0	6580	2011:08:04:00:07:15
123	1010	28.7700	2	6622	2011:08:04:00:06:32
74	1011	31.4100	3	6575	2011:08:04:00:04:37
79	1014	31.6900	3	6581	2011:08:04:00:04:51
21	1016	29.7500	3	6580	2011:08:04:00:02:39
181	1018	30.4000	3	6630	2011:08:04:00:08:08
132	1019	32.0300	14	6566	2011:08:04:00:06:53
213	1020	30.0600	1	65535	2011:08:04:00:09:24
225	1021	32.2900	3	6588	2011:08:04:00:09:56
164	1022	30.9200	2	6610	2011:08:04:00:07:42
11	1023	30.9300	4	6622	2011:08:04:00:02:21
84	1024	40.9500	3	9809	2011:08:04:00:05:09
114	1030	31.8600	2	6602	2011:08:04:00:06:21

图 4-2 GreenOrbs 数据表的信息数据

图 4-2 展示了数据信息值，其中 id 代表记录的唯一标示，location\_id 是位置点标示，是外键，light 是光照信息，temperature 是温度信息，humidity 是湿度信息，并且 light，humidity 在实际应用过程中需要转换成实际值。

id	x	y	location_id
1	31.48343833333330	120.37292333333300	1001
2	31.48324000000000	120.37285000000000	1003
3	31.48432666666670	120.37276833333300	1004
4	31.48297666666670	120.37117166666670	1006
5	31.48371500000000	120.37308000000000	1007
6	31.48373666666670	120.37268000000000	1008
7	31.48331000000000	120.37264833333300	1010
8	31.48316666666670	120.37235666666670	1011
9	31.48396166666670	120.37223666666670	1013
10	31.48432833333330	120.37312000000000	1014
11	31.48322833333330	120.37313500000000	1015
12	31.48334500000000	120.37301833333300	1016
13	31.48418333333330	120.37315833333300	1018
14	31.48387000000000	120.37227000000000	1019
15	31.48332500000000	120.37292166666670	1020
16	31.48347500000000	120.37225333333300	1021
17	31.48308333333330	120.37208333333300	1022
18	31.48351000000000	120.37226666666670	1023
19	31.48413500000000	120.37224500000000	1024
20	31.48327000000000	120.37217166666670	1026
21	31.48431500000000	120.37288333333300	1030
22	31.48319500000000	120.37306166666670	1033

图 4-3 GreenOrbs 数据表内的位置信息

图 4-3 展示了数据采集位置点的信息,其中 x 是经度,y 代表纬度,location\_id 是轨迹点唯一标示,并且和图 4-2 中 location\_id 相对应。

### 4.3.3 模拟数据集的生成

在移动协作感知平台上,对于该平台上数据融合算法的研究离不开该平台上的数据集。在没有该数据集的情况下必须先通过一定的方法生产模拟数据集。在本文中模拟数据的生产将会基于以上两种数据集,其位置范围是在北京邮电大学校园范围内,并采用克里金插值的方法来生成。

根据以上描述的 GeoLife 和 GreenOrbs 两个的数据集,我们将生产移动协作感知系统的数据集。首先根据 GreenOrbs 的数据集和 GeoLife 数据集的分布情况,我们采用平移的方法将 GreenOrbs 的坐标移到 GeoLife 中北京邮电大学的坐标上,这样得到了北京邮电大学的数据分布情况,总体数据的分布趋势和 GreenOrbs 的分布趋势一致。然后再根据 GreenOrbs 的数据集拟合出模型,最后通过该模型插值得到北京邮电大学的数据集。具体过程如下:

- (1) GreenOrbs 数据集: 共有 267 个传感器点来每十分钟采集一次数据,其中有两个传感器节点没有上传数据,因此节点一共有 (265) 条。我们从中

提取任意 10 分钟的数据，保证这 10 分钟内的数据节点不重复，且都包含其中。具体数据信息包括：经度、纬度、光照信息、温度、湿度。并且数据的分布情况如下图 4-4 所示。

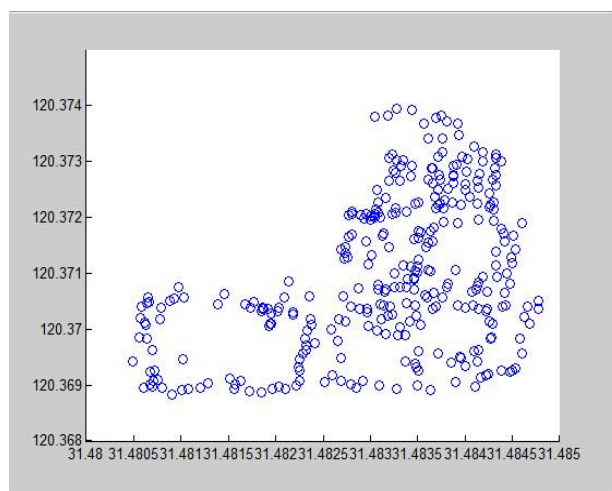


图 4-4 GreenOrbs 数据集分布

- (2) 北京邮电大学数据点位置：北京邮电大学的数据采集点的位置是来源于微软 GeoLife 数据集中截取的北京邮电大学的 1670 条数据，其中包括若干条完整的轨迹信息。具体数据分布比较集中，几乎都是分布在北京邮电大学主干道上的数据。其中包括：经度、纬度、用户 id、轨迹 id。具体分布如下图 4-5 所示。

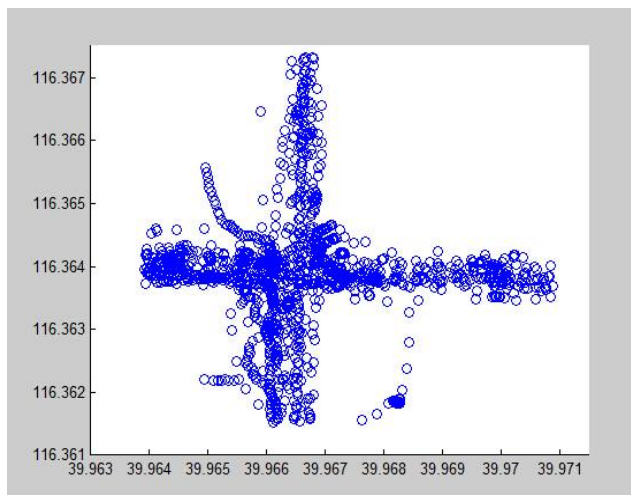


图 4-5 北京邮电大学 (GeoLife 数据集) 分布

- (3) 通过北京邮电大学数据分布范围和 GreenOrbs 数据分布范围的比较：北京邮电大学数据分布范围：经度范围为：39.964—39.971，经度差值：0.007，纬度范围：116.361—116.367，纬度差值：0.006。GreenOrbs 数据分布范围：经度范围：31.48—31.485，经度差值：0.005，纬度范围：120.368—120.374，纬度差值：0.006。北京邮电大学的数据范围比 GreenOrbs 的数据集的分布范围大一点，通过平移的方式将 GreenOrbs 数据集平移到北京

邮电大学的数据范围内，因为 GreenOrbs 数据集的经度和纬度相比于北京邮电大学的分别是多 4.007 和少 8.484。则具体的平移过程为：将 GreenOrbs 上的所有数据点的经度坐标减去 4.007，纬度坐标加上 8.484，这样就将数据进行了位置的平移，平移算法的流程如图 4-6 所示，平移后的数据分布如下图 4-7 所示，平移之后的部分数据的数据存储结构如图 4-8 所示。

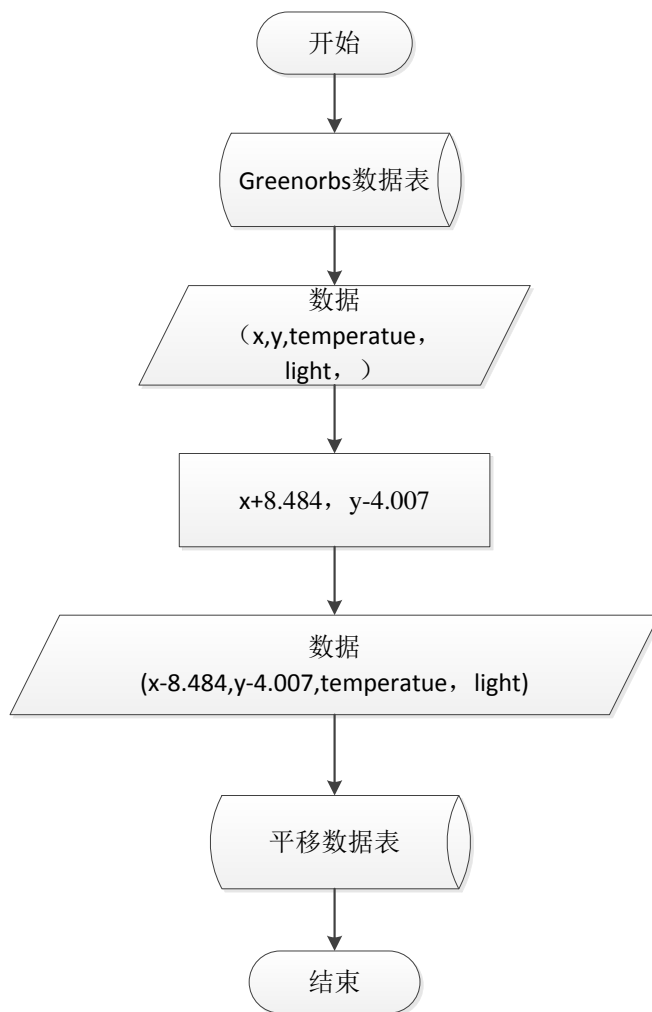


图 4-6 平移算法的流程

图 4-6 展示了平移算法的流程，因为北京邮电大学范围与 GreenOrbs 的地理范围几乎接近，就没有将二者平移到一个公共坐标系进行转换，而是直接将 Greenorbs 的数据直接平移到北京邮电大学范围内，减少了平移数据的工作量。

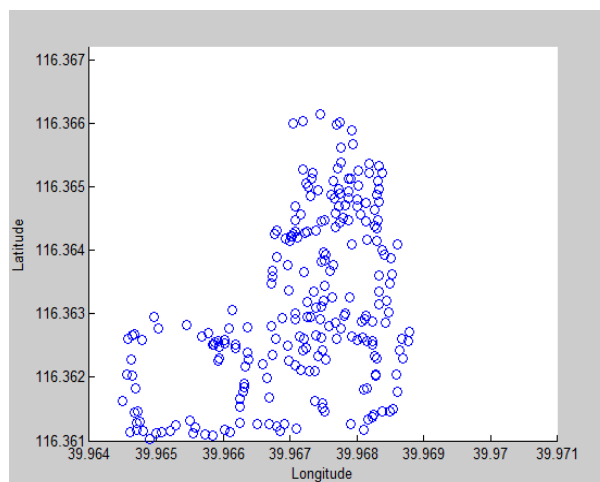


图 4-7 北京邮电大学数据分布

x	y	temperature	light	humidity
39.96724000000000	116.36585000000000	32.0500	2	6590
39.96832666666670	116.36576833333300	30.9500	3	6602
39.96697666666670	116.36417166666699	31.2400	3	6569
39.96771500000000	116.36608000000000	30.6100	2	6616
39.96773666666670	116.36568000000000	31.1700	0	6580
39.96731000000000	116.36564833333300	28.7700	2	6622
39.96716666666670	116.36535666666700	31.4100	3	6575
39.96832833333330	116.36612000000000	31.6900	3	6581
39.96734500000000	116.36601833333299	29.7500	3	6580
39.96818333333330	116.36615833333299	30.4000	3	6630
39.96787000000000	116.36527000000000	32.0300	14	6566
39.96732500000000	116.36592166666699	30.0600	1	65535
39.96747500000000	116.36525333333300	32.2900	3	6588
39.96708333333330	116.36508333333299	30.9200	2	6610
39.96751000000000	116.36526666666700	30.9300	4	6622

图 4-8 平移到北京邮电大学的数据存储结构

- (4) 然后使用图 4-8 所示平移数据的经度、纬度、温度三个数据项，利用公式 4-1 编程实现计算出  $r(h)$ ，计算  $r(h)$  的过程为：计算任何两个点之间距离在  $h$  范围内的数据值之差平方，然后统计出总和，再除以 2 倍的差值个数，就得到了  $r(h)$ ，用这种方法再计算  $r(2h)$ 、 $r(3h)$ ……。然后再用 MATLAB 对  $((h, r(h)), (2h, r(2h)), (3h, r(3h)))$  这样的数据点拟合，得出半方差公式，在这里是拟合球状模型公式，模型如下图 4-8 所示，公式如 4-8 所示，计算半方差流程如图 4-9 所示。

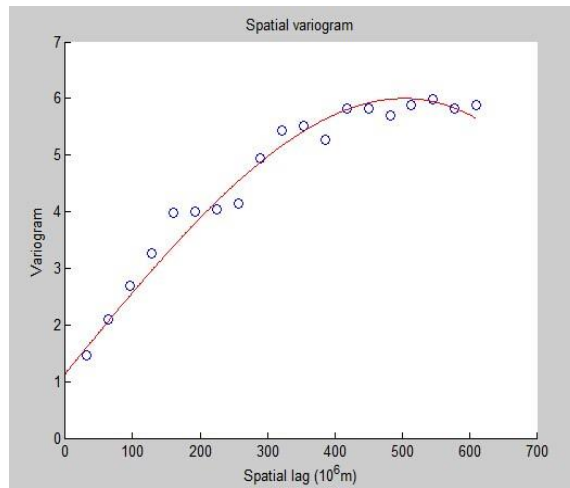


图 4-8 平移数据的空间位置球状模型  
球状模型的具体公式如下 4-8:

$$\gamma(h_s) = \begin{cases} 0, & h_s = 0 \\ c_0 + c \left( \frac{3h_s}{2a} - \frac{h_s^3}{2a^3} \right), & 0 < h_s \leq a \\ c_0 + c, & h_s > a \end{cases} \quad \text{其中 } c_0 = 1.15, c = 4.85, a = 503.65 \quad (4-8)$$

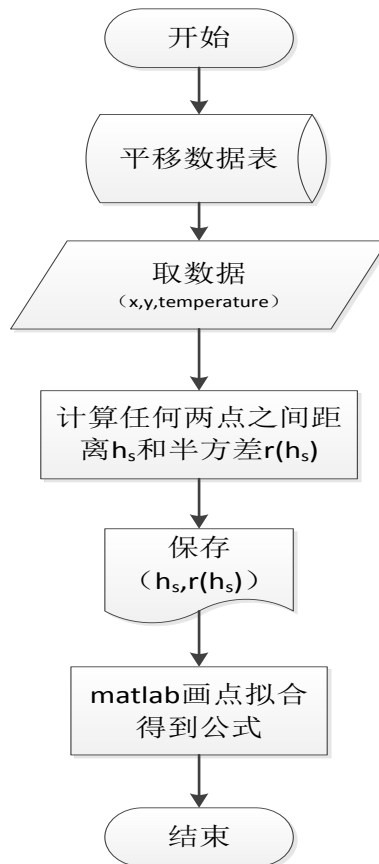


图 4-9 计算半方差的流程

(5) 再利用半方差公式 4-8 和权重计算公式 4-3 和 4-4 进行北京邮电大学

(GeoLife 数据集) 数据点插值计算, 最终生成了 GeoLife 数据集上北京邮电大学的数据 (带有数据值)。计算过程为: 选取图 4-1 所示 bupt 表中的一个数据点, 然后在图 4-7 的平移数据集中选取以这个位置点为圆心, 半径为 0.001 的数据点, 根据公式 4-3 中的  $\gamma(h) = C(0) - C(h)$ , 其中  $C(0) = c_0 + c$ , 就可以计算得出  $C(h)$ , 然后套用公式 4-3, 就可以算出每个点所占的权重, 再根据公式 4-4 就得到 bupt 表中位置点的数据值, 依次循环选取 bupt 数据表中的其余各点, 就能够将 bupt 数据表中没有的数据信息进行插值得出。生成的模拟数据分布如图 4-12 所示, 算法的步骤如图 4-11 所示, 模拟数据的存储数据表 product\_bupt 的结构如图 4-10 所示, 包括唯一标示 id、轨迹标示 tid、用户标示 uid、经度 y、纬度 x、温度信息 temperature。

id	uid	tid	x	y	temperature	time
1	0	168	39.97076600000000	116.36364000000000	29.47	2008-12-10 02:35:03
2	0	168	39.97067900000000	116.36362000000000	30.83	2008-12-10 02:35:08
3	0	168	39.97059800000000	116.36367200000000	29.27	2008-12-10 02:35:13
4	0	168	39.97030200000000	116.36369600000000	31.06	2008-12-10 02:35:18
5	0	168	39.96988100000000	116.36369600000000	30	2008-12-10 02:35:23
6	0	168	39.96940500000000	116.36371600000000	31.23	2008-12-10 02:35:28
7	0	168	39.96889600000000	116.36373500000000	29.55	2008-12-10 02:35:33
8	0	168	39.96831000000000	116.36375300000000	32.06	2008-12-10 02:35:38
9	0	168	39.96769300000000	116.36376200000000	30.61	2008-12-10 02:35:43
10	0	168	39.96708900000000	116.36376000000000	30.51	2008-12-10 02:35:48
11	0	168	39.96683000000000	116.36379100000000	30.93	2008-12-10 02:35:53
12	0	168	39.96685700000000	116.36373400000000	31.15	2008-12-10 02:35:58
13	0	168	39.96685000000000	116.36373400000000	30.23	2008-12-10 02:36:03
14	0	168	39.96684600000000	116.36373500000000	31.74	2008-12-10 02:36:08
15	0	168	39.96684300000000	116.36373600000000	30.71	2008-12-10 02:36:13
16	0	168	39.96683800000000	116.36373600000000	29.81	2008-12-10 02:36:18
17	0	168	39.96683500000000	116.36373600000000	30.63	2008-12-10 02:36:23
18	0	168	39.96683500000000	116.36373600000000	31.89	2008-12-10 02:36:28
19	0	168	39.96683300000000	116.36373500000000	31.32	2008-12-10 02:36:33

图 4-10 模拟数据表的结构

图 4-10 展示了模拟数据的存储结构, 包括经度、纬度、时间、温度值等。对于模拟数据集的温度信息分布情况和整体温度趋势,



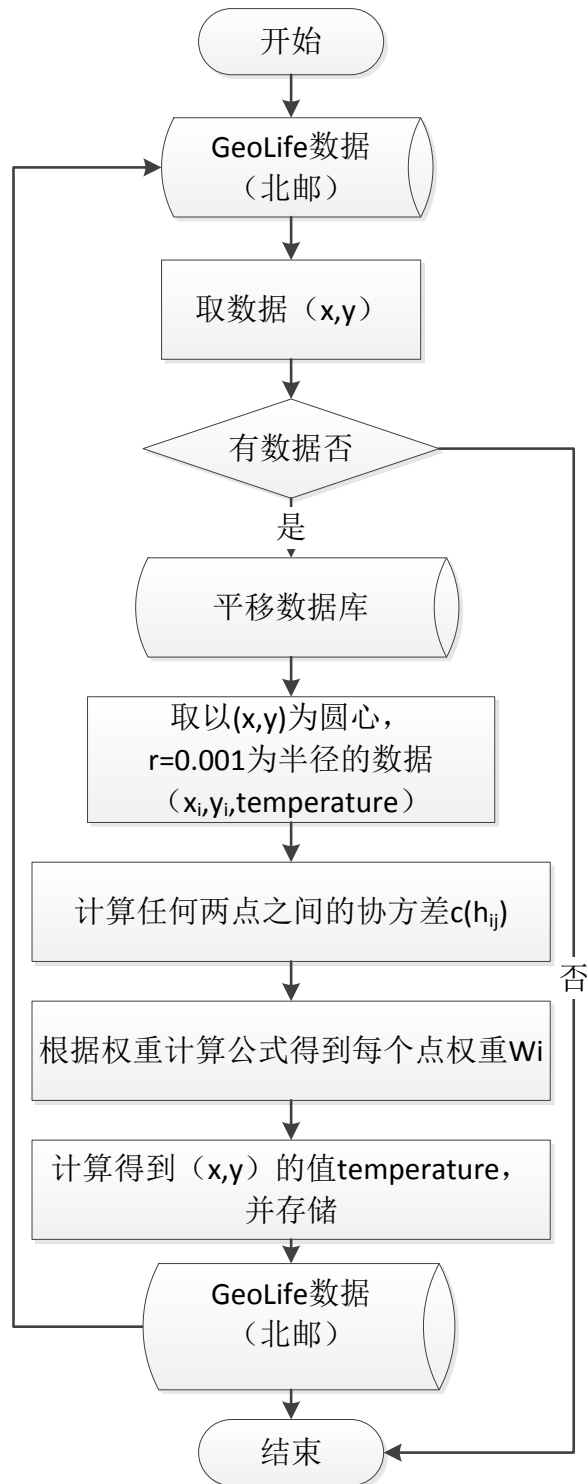


图 4-11 插值算法的流程

图 4-11 展示了插值生成模拟数据的算法流程图，主要就是根据平移的数据集的整体走势对北京邮电大学范围内的模拟数据进行插值生成，最终得到了模拟移动协作感知平台收集的数据集。



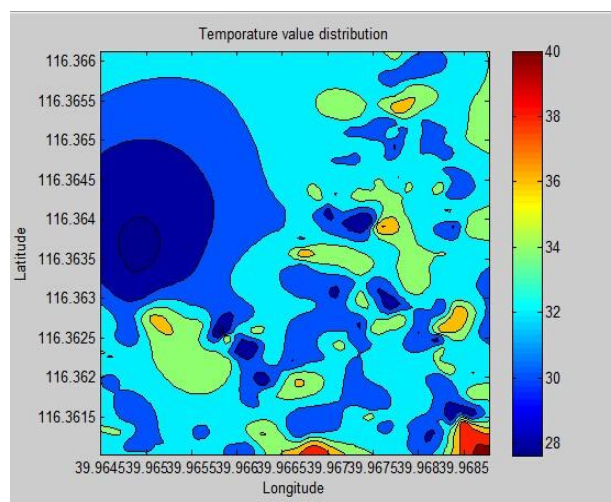


图 4-12 模拟数据集的分布

图 4-12 展示了模拟数据数据值的整体分布情况，从图中可以看到局部区域的数据值比较小，有个别地点的数据值比较大，但是大量的数据还是呈现比较平稳的状态的，并没有特别大的差值，还是一个比较理想的数据集。通过 MATLAB 对图 4-12 的数据进行走势分析，图 4-13 展示了生成模拟数据集的整体走势。

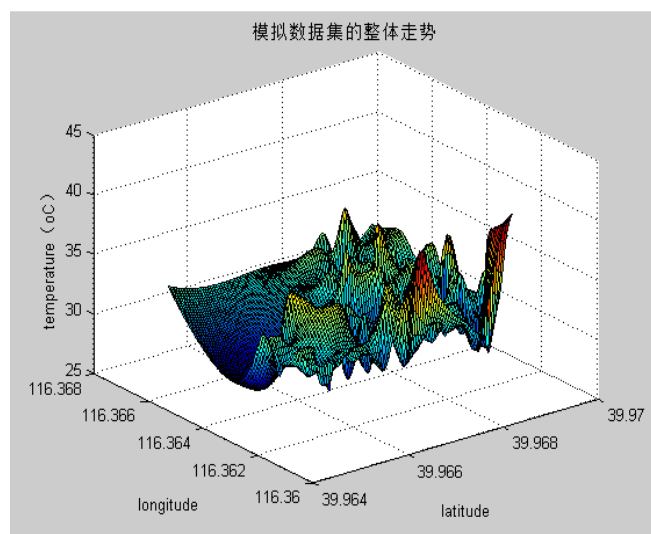


图 4-13 模拟数据集的分布

从图 4-13 中可以看到模拟数据的数据值范围比较广，可以看到最高的可以达到 36 到 40 之间，但是在实际数据表中并没有数值这么高的数据，这张图是由有限个数据点和其上的数据值进行拟合得到的走势图，是由局部上升趋势自动延展开而导致，除个别点之外，其余的数据情况还是接近事实的，因此数据本身也存在着误差。

#### 4.4本章小结

本章主要介绍了在移动协作感知系统中拟使用的数据融合算法、算法的改进以及模拟数据集的生成。首先对拟使用的克里金算法进行介绍，并对这个算法的分析之后，对算法的不足进行了改进，以便适应移动协作感知平台。然后通过对 GeoLife 和 GreenOrbs 两个现有数据集进行描述，其中 GeoLife 数据集包括经度、纬度、时间等信息，提供了一些连续点的轨迹信息，但是无具体的各种环境信息值。而 GreenOrbs 数据集包括了经度、纬度、时间、光照，温度，湿度、时间等信息，提供了环境信息，但是没有提供轨迹信息，仅仅是一些散乱的点。基于这样的两种不同类型的数据集的研究和分析，我们截取 GeoLife 中北京邮电大学的数据范围，然后将 GreenOrbs 数据集的环境监测值平移到北京邮电大学范围内，并用 MATLAB 工具来拟合 GreenOrbs 数据集的球状模型公式，这样北京邮电大学内的数据分布和走势与 GreenOrbs 数据集成一个趋势，我们可以通过这个模型公式，对北京邮电大学轨迹数据点进行插值，最终得到了模拟数据集，并且该数据集模拟了移动协作感知平台的数据收集。

## 第五章 基于时间和空间模型的克里金算法的验证与实现

在第四章中已经介绍了基于时间和空间模型的克里金算法,对于该算法的好坏目前还是无法判断,必须通过实验验证来评估算法的准确性和适用性。第四章也介绍了生成的模拟数据集,因此本章将利用模拟的数据集进行算法的验证和实现。首先建立数据模型,根据生成的模拟数据集分别拟合出时间和空间维度的半方差球状模型公式,并根据 **product** 模型计算出基于时间和空间的半方差模型,然后用 **MATLAB** 工具进行仿真验证,最后使用 **JAVA** 语言,在该平台上实现算法。

### 5.1 建立数据模型

从生成的模拟数据集图 4-10 中取出 (time,temperature),图 5-1 展示了取出的数据结构。针对上文中所介绍的基于时间和空间的克里金方法,进行实际数据的建模分析,进而进行实验验证和实现。

time	temperature
2008-12-10 02:37:18	30.34
2008-12-10 02:37:23	32.83
2008-12-10 02:37:28	31.45
2008-12-10 02:37:33	28.95
2008-12-10 02:37:38	29.87
2008-12-10 02:37:43	30.98
2008-12-10 02:37:48	30.51
2008-12-10 02:37:53	30.23
2008-12-10 02:37:58	31.27
2008-12-10 02:38:03	29.01
2008-12-12 04:34:40	31.19
2008-12-12 04:34:45	31.69
2008-12-12 04:34:50	29.73
2008-12-12 04:34:55	30.64
2008-12-12 04:35:00	30.15
2008-12-12 04:35:05	32.09
2008-12-12 04:35:10	29.99
2008-12-12 04:35:15	29.81
2008-12-12 04:35:20	30.37

图 5-1 计算时间协方差的数据

- (1) 由于生成的模拟数据集的走势和原始的 **GreenOrbs** 数据集的走势相同,前文中得到 **GreenOrbs** 数据集的空间维度的半方差球状模型和公式,则模拟

数据集的半方差模型和公式具体模型和公式如图 4-8 和公式（4-8）所示。

- (2) 按照上文中图 4-9 所示的计算半方差流程，利用移动协作感知系统的模拟数据集，通过编程实现进行数据处理和计算，再通过 MATLAB 仿真拟合，最终得到了模拟数据集在时间维度上的半方差球状模型和公式，具体模型和公式如图 5-2 和公式（5-1）所示。

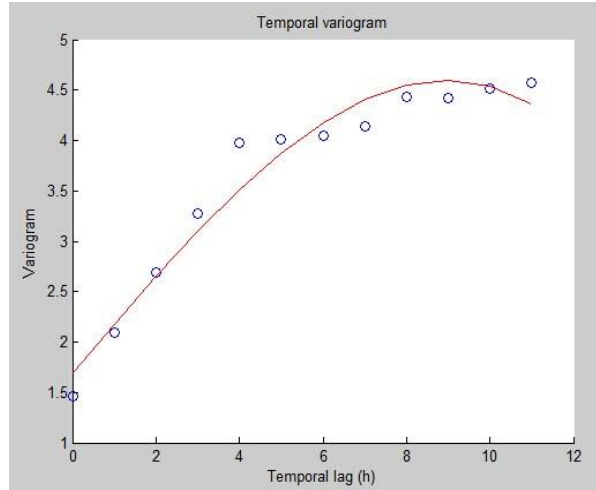


图 5-2 时间半方差球状模型

$$\gamma(h_t) = \begin{cases} 0, & h_t = 0 \\ c_0 + c \left( \frac{3h_t}{2a} - \frac{h_t^3}{2a^3} \right), & 0 < h_t \leq a \\ c_0 + c, & h_t > a \end{cases} \quad \text{其中 } c_0 = 1.7, c = 2.9, a = 9.0 \quad (5-1)$$

主要计算过程为：就是只考虑时间关系而忽略空间关系，提取图 4-9 中的数据包括温度值、时间这两项，然后计算所有时间差  $h_t$  的所有数据值之差的平方和，然后根据公式 4-1 得到了  $r(h_t)$ ，然后用 MATLAB 拟合得到公式 5-1，具体的流程图如图 5-3 所示。

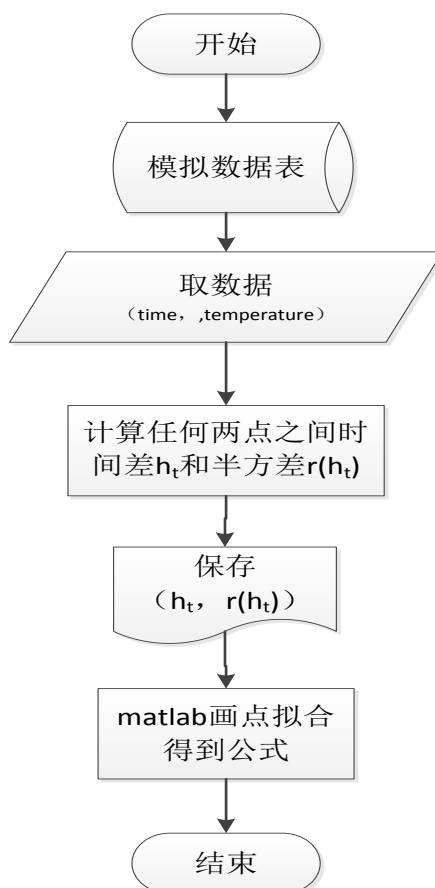


图 5-3 时间半方差计算过程

- (3) 根据 product 模型公式，结合以上的时间半方差模型公式 (5-1) 和空间半方差模型公式 (4-8) 根据 (4-5) 和 (4-6) 得到基于时间和空间结合的半方差模型公式 (5-2)，具体的时空半方差模型如图 5-4 所示。

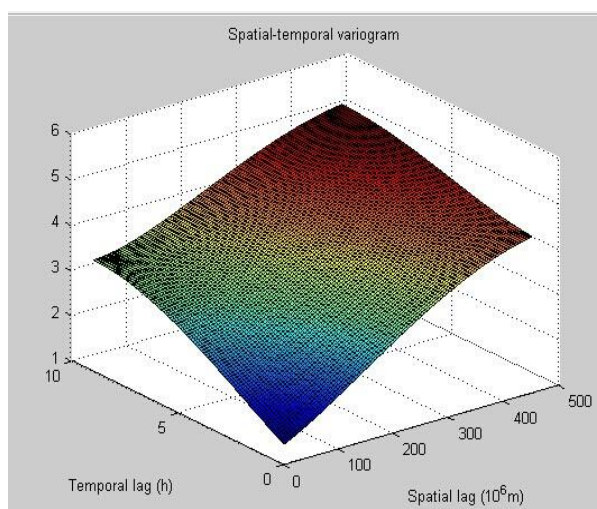


图 5-4 时间和空间半方差

$$\gamma_{st}(h_s, h_t) = C_t(0)\gamma_s(h_s) + C_s(0)\gamma_t(h_t) - C_{st}(h_s, h_t) \quad (5-2)$$

其中：

$$\gamma(h_t) = \begin{cases} 0, & h_t = 0 \\ 1.7 + 2.9 \left( \frac{3h_t}{2 \times 9} - \frac{h_t^3}{2 \times 9^3} \right), & 0 < h_t \leq a \\ 1.7 + 2.9, & h_t > a \end{cases}$$

$$\gamma(h_s) = \begin{cases} 0, & h_s = 0 \\ 1.15 + 4.85 \left( \frac{3h_s}{2 \times 503.65} - \frac{h_s^3}{2 \times 503.65^3} \right), & 0 < h_s \leq a \\ 1.15 + 4.85, & h_s > a \end{cases}$$

公式 5-2 是根据 product 模型套用得到的，根据时间半方差和空间半方差公式就可以得到时空结合的半方差公式，然后再用 MATLAB 仿真得到该公式的图形。

## 5.2 基于时间和空间模型的算法的实验评估

### 5.2.1 算法对比分析

利用以上描述的时间和空间结合的球状模型公式，根据生成的实验数据集进行验证，我们在一个小区内随机选取 40 个点，实验分成四组，我们分别选取 10 个点、20 个点、30 个点、40 个点，对于每一组实验数据中的每个点，我们都是通过该组中其余点通过空间克里金和时空结合的克里金算法计算来得到该数据点的数据值，然后把两种方法的计算结果再和这个点的实际测量值比较，实验数据如图 5-5 所示。

id	uid	tid	x	y	temperature	time
21	0	168	39.96680500000000	116.36373400000000	29.75	2008-12-10 02:36:43
22	0	168	39.96681000000000	116.36373420000000	31.38	2008-12-10 02:36:48
23	0	168	39.96679700000000	116.36373400000000	30.1	2008-12-10 02:36:53
24	0	168	39.96678900000000	116.36373400000000	29.74	2008-12-10 02:36:58
25	0	168	39.96678100000000	116.36373500000000	28.91	2008-12-10 02:37:03
26	0	168	39.96677600000000	116.36373500000000	30.05	2008-12-10 02:37:08
27	0	168	39.96677400000000	116.36373500000000	30.32	2008-12-10 02:37:13
28	0	168	39.96677400000000	116.36373500000000	30.34	2008-12-10 02:37:18
29	0	168	39.96677400000000	116.36373500000000	32.83	2008-12-10 02:37:23
30	0	168	39.96677800000000	116.36373500000000	31.45	2008-12-10 02:37:28
31	0	168	39.96679100000000	116.36373500000000	28.95	2008-12-10 02:37:33
32	0	168	39.96680300000000	116.36373500000000	29.87	2008-12-10 02:37:38
33	0	168	39.96681300000000	116.36373500000000	30.98	2008-12-10 02:37:43
34	0	168	39.96681400000000	116.36373500000000	30.51	2008-12-10 02:37:48
35	0	168	39.96681400000000	116.36373600000000	30.23	2008-12-10 02:37:53
36	0	168	39.96681400000000	116.36373600000000	31.27	2008-12-10 02:37:58
37	0	168	39.96681200000000	116.36373600000000	29.01	2008-12-10 02:38:03
38	0	168	39.96679200000000	116.36372000000000	31.19	2008-12-12 04:34:40
39	0	168	39.96656500000000	116.36372800000000	31.69	2008-12-12 04:34:45
40	0	168	39.96623800000000	116.36372700000000	29.73	2008-12-12 04:34:50

41	0	168	39.965883000000000	116.363825000000000	30.64	2008-12-12 04:34:55
42	0	168	39.965514000000000	116.363869000000000	30.15	2008-12-12 04:35:00
43	0	168	39.965129000000000	116.363884000000000	32.09	2008-12-12 04:35:05
44	0	168	39.964664000000000	116.363931000000000	29.99	2008-12-12 04:35:10
45	0	168	39.964341000000000	116.363980000000000	29.81	2008-12-12 04:35:15
46	0	168	39.964382000000000	116.363964000000000	30.37	2008-12-12 04:35:20
47	0	168	39.964332000000000	116.363966000000000	31.06	2008-12-12 04:35:25
48	0	168	39.964052000000000	116.364016000000000	30.55	2008-12-12 04:35:30
49	0	168	39.963948000000000	116.363728000000000	29.17	2008-12-12 04:35:35
50	0	168	39.964486000000000	116.363730000000000	30.2	2008-12-12 04:35:40
51	0	168	39.964932000000000	116.363714000000000	30.55	2008-12-12 04:35:45
52	0	168	39.965366000000000	116.363801000000000	29.08	2008-12-12 04:35:50
53	0	168	39.965568000000000	116.363798000000000	31.08	2008-12-12 04:35:55
54	0	168	39.965648000000000	116.363816000000000	29.17	2008-12-12 04:36:00
55	0	168	39.965668000000000	116.363844000000000	31.43	2008-12-12 04:36:05
56	0	168	39.965671000000000	116.363842000000000	30.84	2008-12-12 04:36:10
57	0	168	39.965671000000000	116.363842000000000	31.33	2008-12-12 04:36:15
58	0	168	39.965663000000000	116.363841000000000	30.61	2008-12-12 04:36:20
59	0	168	39.965651000000000	116.363845000000000	29.59	2008-12-12 04:36:25
60	0	168	39.965651000000000	116.363845000000000	31.18	2008-12-12 04:36:30

图 5-5 实验数据

图 5-5 展示了算法验证的实验数据，包括经度、纬度、温度和时间，本实验将从空间和时空结合两个方面对比验证。首先将该数据集取出 10 个点、20 个点、30 个点、40 个点，然后对于每个组中的每个点都用其余各点用基于空间模型和基于时空模型进行计算，然后再和真实值进行对比，对比结果如下各图所示。

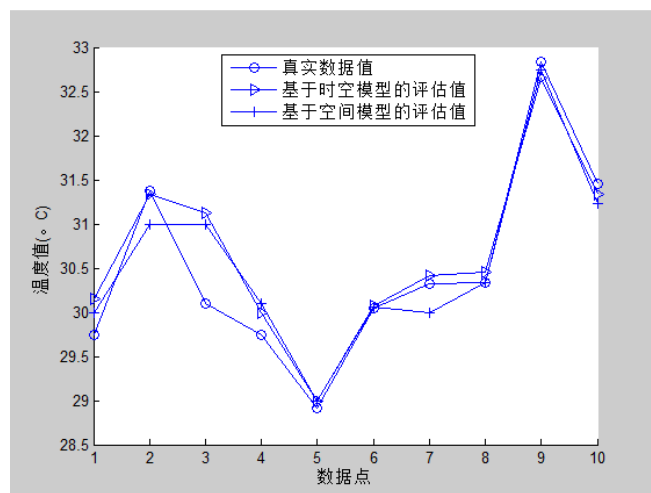


图 5-6 10 个数据点的实验结果

图 5-6 展示了 10 个点的数值，然后对于其中任何一个点，都用其余的 9 个点分别用空间模型和时空模型计算得到结果，然后再和真实值进行对比分析，下面各图也是按照这种思想进行对比。



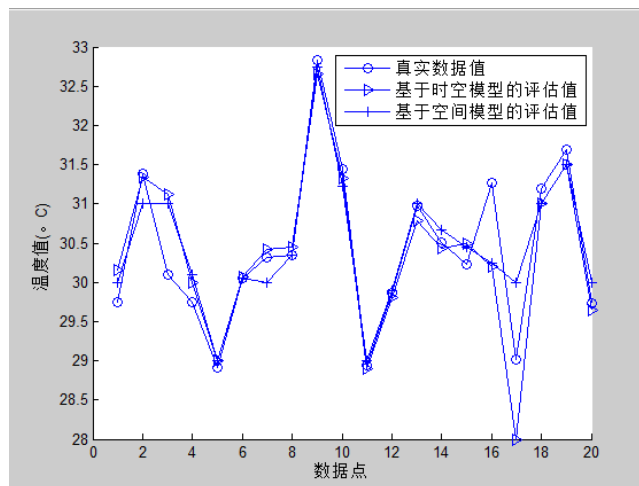


图 5-7 20 个数据点的实验结果

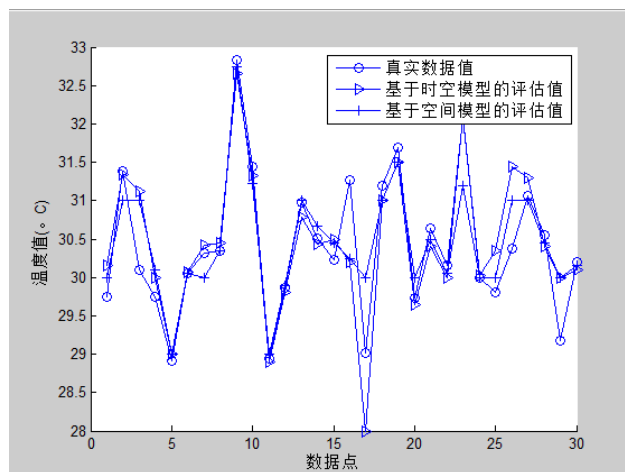


图 5-8 30 个数据点的实验结果

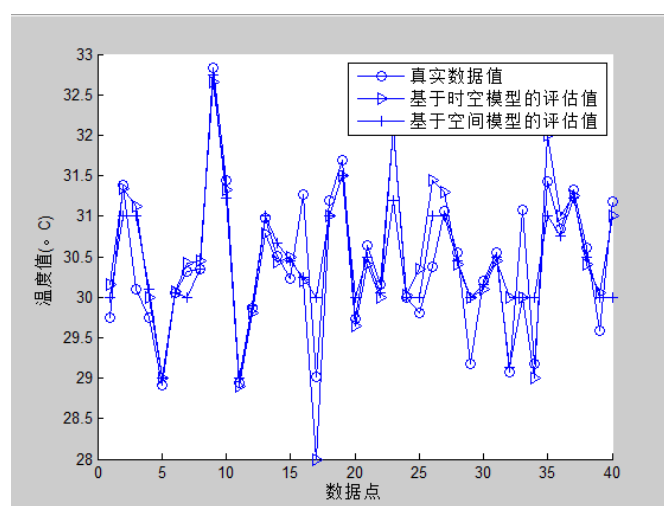


图 5-9 40 个数据点的实验结果

根据以上实验结果，我们可以进行误差分析和统计，表 5-1 展示了各组数据的误差实验结果。



表 5-1 实验结果误差分析

实验	基于时空结合的克里金算法 误差	基于空间维度的克里金算法 误差
10 个点	2.3400	2.6100
20 个点	5.5600	5.7600
30 个点	8.7700	8.9200
40 个点	12.8600	13.0900

由表 5-1 分析得出，随着实验数据点的增大，基于时空结合的克里金算法的误差总比基于空间维度的克里金算法的误差也要变大，因此基于时空维度的克里金算法比单独考虑空间因素的克里金算法的误差更加小一些，从整体的实验误差分析可以得出时间因素对于移动协作感知数据融合是至关重要的因素，引入时空模型的克里金算法的准确性有所提高。

### 5.2.2 算法准确率和性能比较

对于算法的准确性分析，从实验数据中随机取出一个点，然后用在一定时间范围和空间范围内、不用数量的数据对该点用基于时空模型和基于空间模型的算法分布进行计算，然后再进行对比得到如图 5-10 的数据结果。

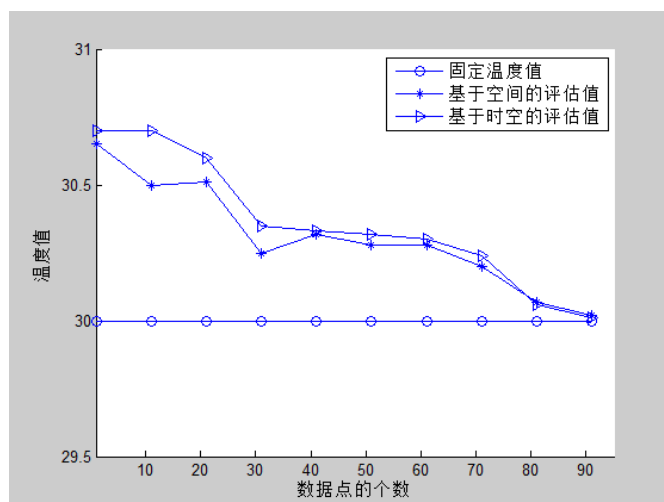


图 5-10 对比实验结果

从图 5-10 中可以看出随着数据量的增加，得到的数据结果更加的接近数据的真实值。并且基于时空模型的克里金算法比基于空间模型的克里金算法更加容易接近真实值。实验误差对比结果如图 5-11 所示。

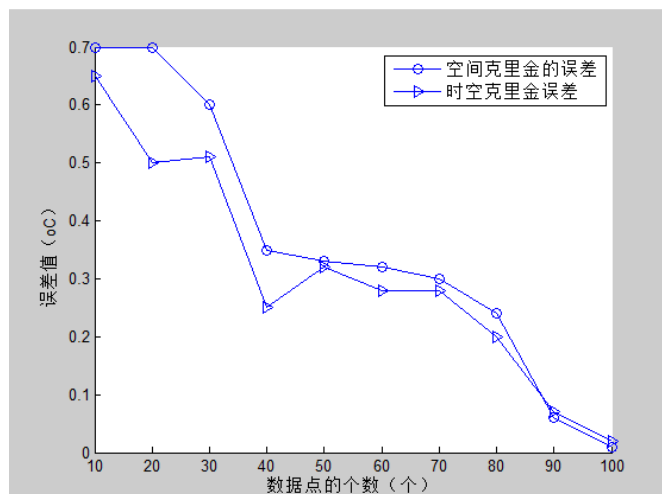


图 5-11 对比实验结果

从 5-11 就可以看出基于时空模型克里金算法得到的误差更小一些，具体的统计结果如表 5-2 所示。

表 5-2 误差对比分析

实验组	基于时空模型的克里金算法	基于空间模型克里金算法误差
1 组 (10 点)	0.6500	0.7000
2 组 (20 点)	0.5000	0.7000
3 组 (30 点)	0.5100	0.6000
4 组 (40 点)	0.2500	0.3500
5 组 (50 点)	0.3200	0.3300
6 组 (60 点)	0.2800	0.3200
7 组 (70 点)	0.2800	0.3000
8 组 (80 点)	0.2000	0.2400
9 组 (90 点)	0.0700	0.0600
10 组 (100 点)	0.0200	0.0100

### 5.2.3 算法的误差分析

针对以上的误差结果，我们将要对产生误差的原因进行分析。误差的原因主要以下两个方面：

#### (1) 实验数据集的误差

在进行算法验证的过程中，所使用的实验数据是通过两个原始数据集生成的。一方面原始数据集本身可能存在误差，另一方面用带误差的原始数据按照一定的策略生成实验数据集，这个过程中也会存在误差。这就从一定程度上影响了实验结果的精度。

#### (2) 数据拟合模型的问题

在用实验数据集进行数据建模过程中,最后得到一个比较合适的数据模型公式。没有任何办法去判断该模型公式就是最合适的,只是通过不断调整,得到一个比较接近的模型公式,这个过程中也存在着误差。

### 5.3 基于时间和空间的克里金算法的实现

通过对基于时间和空间结合的克里金算法的实验验证,了解到该算法在移动协作感知系统中具有很高的准确性和适用性,因此在该平台上实现该算法。算法实现的流程如图 5-12 所示。描述了从数据库中取出数据后,对数据进行的一系列计算处理,最终得到了融合结果。

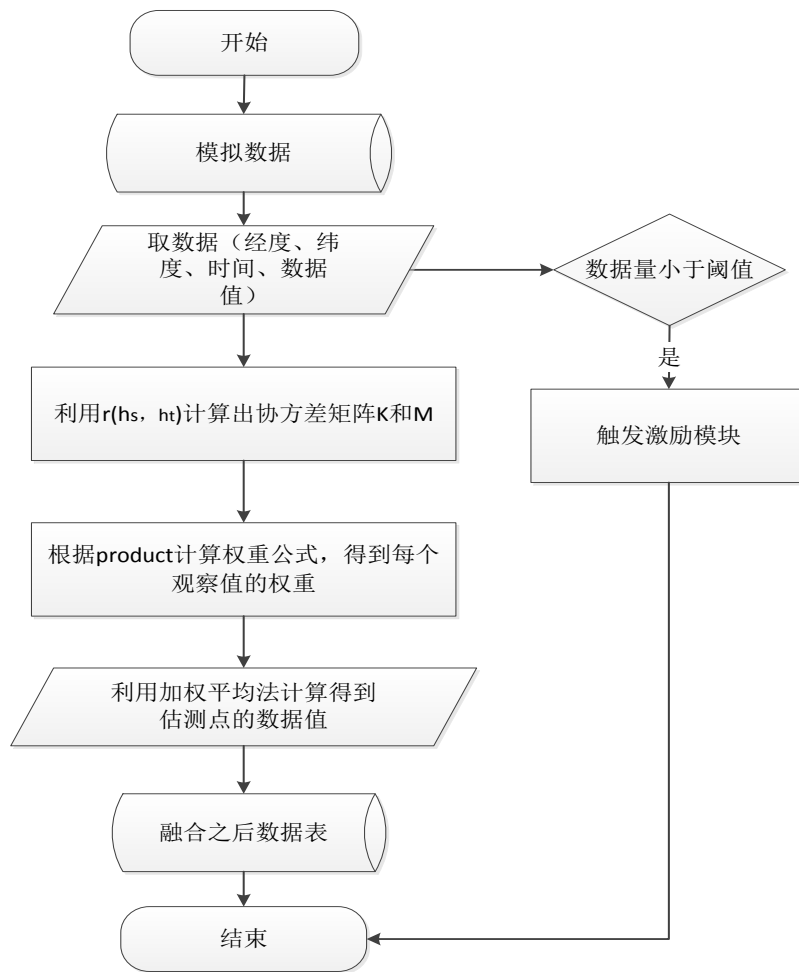


图 5-12 算法实现的流程图

图5-12展示了数据融合算法的流程图,首先从数据库中读出在一个网格内、一段时间范围内的所有数据,包括经度、纬度、时间和数据值,然后根据时间和空间结合的协方差公式计算出任意两个测量点之间的协方差  $C_{ij}$  构成的矩阵  $K$ , 和任意一个测量点和估测点的协方差  $M$ 。然后再利用  $[W] = [K]^{-1}[M]$  计算出权重,最终得到融合的计算结果。

在克里金算法实现过程中，所设计的类有四个，主要有 DataInfo、DBConnection、QueryOperation、Kriging 四个类，其中 DataInfo 主要是定义了数据的基本结构和基本的数据操作方法，方便数据的存储；DBConnection 类定义类连接数据库的一些列操作；QueryOperation 类定义了创建网格的操作；Kring 类定义了实现克里金算法的一些类方法。Kringing 类调用其他三个类来实现克里金算法，具体四个类的类关系如图 5-13 所示。

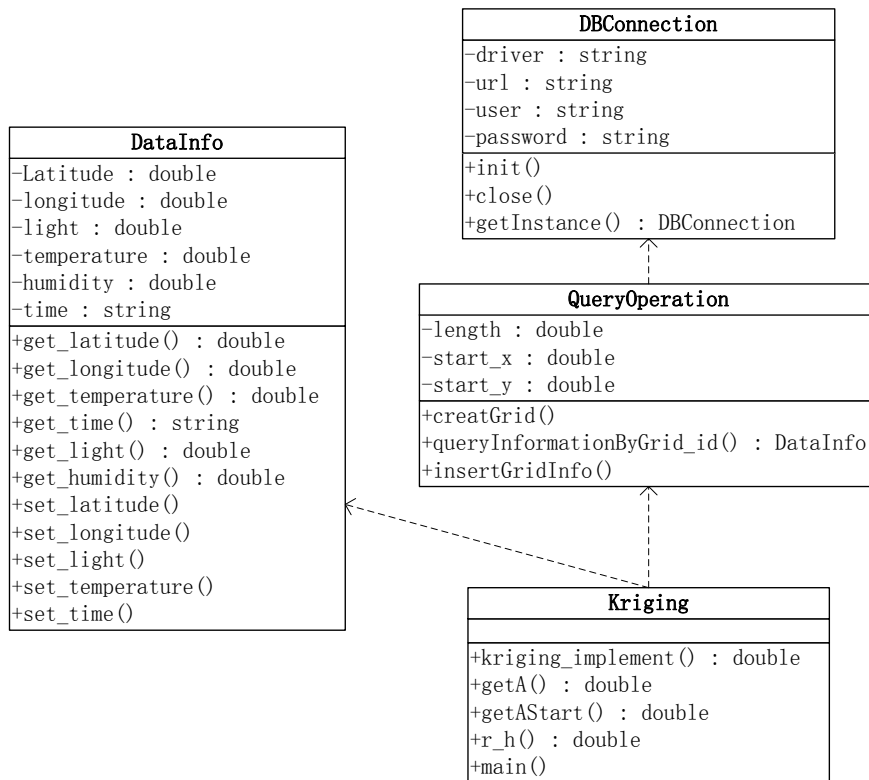


图 5-13 克里金算法实现的类图

通过算法的验证了解到该算法具有很高的准确性，按照需求分析的要求，最终在该平台上实现了数据融合算法，并且经过数据融合之后的数据的分布情况如图 5-14 所示。



图 5-14 融合之后的结果分布

图 5-14 展示了经过数据融合之后的数据分布情况，将每个网格内一段时间范围的数据都融合成一个点，用这个点的数据值来代表整个区域的数据评估。将整个区域面积分为  $4 \times 5$  个网格，对每个网格内的数据融合成一个值。如果存在缺失数据，则会启动激励模块进行下发激励。

## 5.4 本章小结

本章主要是基于第四章介绍的时间和空间结合的克里金算法进行数据建模和算法验证实现。首先利用生成的模拟数据利用 MATLAB 工具进行数据建模，并拟合出基于时间的克里金球状模型，再加上第四章拟合的基于空间的克里金球状模型，然后利用 product 模型计算出基于时间和空间的克里金的半方差模型，得到计算公式，利用该公式，并随机选取 40 个数据点，然后进行算法的实际验证，最终误差分析结果显示：基于时间和空间结合的克里金算法比基于空间维度的克里金算法更加有效和适用。最后用 JAVA 语言实现了该算法，得到融合结果，并用百度地图显示数据的分布图。

## 第六章 结论及展望

### 6.1 结论

本论文主要就是阐述移动协作感知平台上数据融合算法的研究与实现,其中主要是借鉴加权平均的数据融合方法,采用引入时间维度的克里金算法进行实现。首先是对数据融合的背景概念、融合层次,以及具体的融合方法进行了了解;然后对移动协作感知平台的整体设计进行描述,包括平台的整体设计,我所负责的数据融合模块的设计,数据存储的数据表的设计,和数据融合模块通信的数据接口的设计,还有融合过程中划分网格的策略等。由于平台没有搭建起来,还没有该平台上的数据集,接下来就通过 GeoLife 数据集和 GreenOrbs 项目的数据集这两个数据集去生成北京邮电大学的数据,并且该数据集模拟了移动协作感知平台上数据收集的方式。最后对该平台上使用的克里金算法进行介绍,并将引用的 product 模型进行引入,通过已生成的数据集,建立自己的数据模型,最终通过实验对比分析基于空间维度的克里金算法和基于时空模型的克里金算法这两种算法,通过对比实验和误差分析得出:基于 product 模型的时空克里金算法比基于空间模型的克里金算法在移动协作感知平台上更加具有适用性。

在本文移动协作感知平台上数据融合算法的研究和实现之后,未来还有很多工作有待改善,首先就是真实数据的实验验证方法,要通过移动协作感知平台收集的真实可靠数据去做算法的验证,这样的结果会更加具有说服力。其次在数据融合算法上,可以去继续探索新的适用的方法,也可以对原始数据融合方法进行改进使其能用在移动协作感知平台上。最后就是对比实验部分显得单薄,可能在后期的研究中会更加的完善。

### 6.2 展望

针对移动协作感知数据融合的研究,本文采用的是克里金插值算法,也是借鉴了加权平均的思想,在后续的研究中可以考虑将克里金插值算法和卡尔曼滤波算法结合起来实现数据融合。通过克里金算法可以得到网格内的测量值,在根据网格内上一时刻的估测值来推算现时刻的最优值,进而不断递归,就可以进行数据融合,卡尔曼算法的不断递归性会让系统更加趋于平稳和正常,可以去除过程中的噪声干扰,具有很好的应用前景。这只是一个初步的想法,具体的可行性分

析和实验验证还需要进一步的深入研究。

由于近几年移动协作感知平台的研究还处于初级阶段，还不是很成熟，所以也会存在着很多不足。首先移动协作感知平台可以增加移动智能终端之间的通信，更好的实现协作感知的想法，这样也会提高系统的性能。其次在能源节省上，可以监测用户移动终端的能源剩余，这样方便任务的均衡分配，也会方便用户对移动智能设备的日常行为操作。再次在数据融合算法上，应该对于数据缺失和数据冗余这块也同样可以和激励机制模块结合在一起，通过激励机制的相应策略去调整采集数据的分布，使数据更加的均匀化。然后在激励机制上，这块是很重要的，它的激励下发策略和能源节省也有很大的关系，通过对用户终端能源的监测，来合理选择用户，进行激励任务下发。最后是在隐私保护上，通过以上的分析可以了解到激励模块在整个系统的参与过程中，还是必不可少的，能源节省、数据融合等都和激励模块有很大的关系，当然隐私保护也和激励模块有关系，因为激励模块是对用户完成系统下发的任务会有奖励机制，这种奖励机制就是需要用户的身份，但是对于用户来说未必想要公开自己的身份，因此一个隐私策略就会变得尤为重要了。

## 参考文献

- [1] Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of International conference on World Wild Web (WWW 2009), Madrid Spain. ACM Press: 791-800.
- [2] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, Wei-Ying Ma. Understanding Mobility Based on GPS Data. In Proceedings of ACM conference on Ubiquitous Computing (UbiComp 2008), Seoul, Korea. ACM Press: 312-321.
- [3] Yu Zheng, Xing Xie, Wei-Ying Ma, GeoLife: A Collaborative Social Networking Service among User, location and trajectory. Invited paper, in IEEE Data Engineering Bulletin. 33, 2, 2010, pp. 32-40.
- [4] Mo L, He Y, Liu Y, et al. Canopy closure estimates with GreenOrbs: sustainable sensing in the forest[A]/Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems[C]. ACM, 2009: 99-112.
- [5] R. Bonney, C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk, "Citizen science: A developing tool for expanding science knowledge and scientific literacy", *BioScience*, 2009, 59(11): 977-984.
- [6] D. Cuff, M. Hansen, J. Kang. Urban sensing: out of the woods. *ACM Commun*, 2008, 51(3): 24-33.
- [7] A. Krause, E. Horvitz, A. Kansal, F. Zhao. Toward Community Sensing. In Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN), 2008:481-492.
- [8] S. Liu, J. Yang, B. Li, C. Fu. Volunteer Sensing: The New Paradigm of Social Sensing. In Proceedings of
- [9] D. Philipp, F. Durr, K. Rothermel. A Sensor Network Abstraction for Flexible Public Sensing Systems. In Proceedings of the 8th International Conference on Mobile Adhoc and Sensor Systems (MASS), 2011:460-469.
- [10] R.K. Ganti, F. Ye, H. Lei. Mobile crowdsensing: current state and future challenges. *Communications Magazine, IEEE*, 2011, 49(11):32-39.
- the 17th International Conference on Parallel and Distributed Systems, 2011:982-987.
- [11] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden. CarTel: A Distributed Mobile Sensor Computing System. In Proceedings of the 4th ACM International Conference on Embedded



Networked Sensor Systems (SenSys), 2006:125–138.

[12] D. Prabal, A. M. Paul, N. Kumar, M. Alan, M. Chris, W. Wesley, W. Allison. Common Sense: participatory urban sensing using a network of handheld air quality monitors. In Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, 2009:349-350.

[13] S. Poduri, A. Nimkar, and G. S. Sukhatme. Visibility monitoring using mobile phones. Architecture. Retrieved from <http://robotics.usc.edu/mobilesensing/visibility/MobileAirQualitySensing.pdf>, Dec 2009.

[14] E. Kanjo. NoiseSPY: A Real-Time Mobile Phone Platform for Urban Noise Monitoring and Mapping. Mobile Networks and Applications, 2010, 15(4):562-574.

[15] S. Eisenman, E. Miluzzo, N. Lane, R. Peterson, G. Ahn and A. Campbell. The Bikenet Mobile Sensing System for Cyclist Experience Mapping. In Proceedings of the 5th international conference on Embedded networked sensor systems (SenSys), 2007:87-101.

[16] S. Gaonkar, J. Li, R. Choudhury, L. Cox, and A. Schmidt. Micro-Blog: Sharing and Querying Content through Mobile Phones and Social Participation. In Proceedings of the 6th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), 2008:174-186.

[17] L. Deng and L. Cox. LiveCompare: Grocery Bargain Hunting through Participatory Sensing. In Proceedings of the 10th Workshop on Mobile Computing Systems and Applications (HotMobile), 2009:1-6.

[18] Y. Dong, S. Kanhere, C. Chou, and N. Bulusu. Automatic Collection of Fuel Prices from a Network of Mobile Cameras. In Proceedings of the 4th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS), 2008:140-156.

[19] Q. Huang, Y. Cui, S. Samarasekera, M. Greiffenhagen. Auto Cameraman Via Collaborative Sensing Agents. Computer Vision, 1997:418-425.

[20] F.Z. Qureshi. Collaborative Sensing via Local Negotiations in Ad Hoc Networks of Smart Cameras. Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras ICDSC 2010:190-197.

[21] X. Bao, R.R. Choudhury. MoVi: Mobile Phone based Video Highlights via Collaborative Sensing. In the Proceedings of the 8th international conference on Mobile systems, Applications, and services, MobiSys, 2010:357-370.

- [22] W.T. Luke Teacy, J. Nie. Collaborative Sensing by Unmanned Aerial Vehicles. In Proceedings of the 3rd International Workshop on Agent Technology for Sensor Networks, 2009:13-16.
- [23] 上海交通大学的智能车辆技术研究所, <http://cyberc3.sjtu.edu.cn/www/research.htm>. Accessed in March, 2013.
- [24] 刘云浩, 群智感知计算, 中国计算机学会通信, Vol. 8, No. 10, 2012, pp. 38-41.
- [25] 耀红. 数据融合理论与应用[M]. 西安电子科技大学出版社, 1997.
- [26] Hall D L, Llinas J. An introduction to multisensor data fusion[J]. Proceedings of the IEEE, 1997, 85(1): 6-23.
- [27] Steinberg A N, Bowman C L, White F E. Revisions to the JDL data fusion model[C]//AeroSense'99. International Society for Optics and Photonics, 1999: 430-441.
- [28] Llinas J, Bowman C, Rogova G et al. Revisiting the JDL data fusion model II[R]. SPACE AND NAVAL WARFARE SYSTEMS COMMAND SAN DIEGO CA, 2004.
- [29] 宪尧. 数据融合技术与船舶自动化[M]. 大连海事大学出版社, 2003.
- [30] 王耀南, 李树涛. 多传感器信息融合及其应用综述[J]. 控制与决策, 2001, 16(5): 518-522.
- [31] 蒋庆全. 数据融合在雷达目标识别中的应用[J]. 兵工自动化, 2001, 20(2): 14-18.
- [32] 永元, 洪钺, 自动化技术研究者, 等. 卡尔曼滤波与组合导航原理[M]. 西北工业大学出版社, 1998.
- [33] 诗松, 茆. 贝叶斯统计[M]. 中国统计出版社, 1999.
- [34] 朱大奇, 于盛林. 基于 DS 证据理论的数据融合算法及其在电路故障诊断中的应用[J]. 电子学报, 2002, 30(2): 221-223.
- [35] 李成. 神经网络系统理论[M]. 西安电子科技大学出版社, 1990.
- [36] 伍斯龙. 群智感知中语音识别系统的研究与实现[D]. 华南理工大学, 2013.
- [37] 李向阳. 面向情境感知的数据融合研究[D]. 南京: 南京邮电大学, 硕士学位论文, 2012.
- [38] 孙晓燕, 杜景林, 周杰. 参与感知在气象信息服务系统中的应用研究[J]. 计算机应用与软件, 2014, 31(5): 71-75.
- [39] 杨静. 基于参与式感知的个体移动特征识别研究[D]. 燕山大学, 2013.
- [40] Mendez D, Labrador M, Ramachandran K. Data interpolation for participatory sensing systems[J]. Pervasive and Mobile Computing, 2013, 9(1): 132-148.

- [41] Oliver M A, Webster R. Kriging: a method of interpolation for geographical information systems[J]. *International Journal of Geographical Information System*, 1990, 4(3): 313-332.
- [42] Myers D E. Space-time correlation models and contaminant plumes[J]. *Environmetrics*, 2002, 13(5 - 6): 535-553.
- [43] Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of International conference on World Wild Web (WWW 2009)*, Madrid Spain. ACM Press: 791-800.
- [44] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, Wei-Ying Ma. Understanding Mobility Based on GPS Data. In *Proceedings of ACM conference on Ubiquitous Computing (UbiComp 2008)*, Seoul, Korea. ACM Press: 312-321.
- [45] Yu Zheng, Xing Xie, Wei-Ying Ma, GeoLife: A Collaborative Social Networking Service among User, location and trajectory. Invited paper, in *IEEE Data Engineering Bulletin*. 33, 2, 2010, pp. 32-40.
- [46] Liu Y, He Y, Li M, et al. Does wireless sensor network scale? A measurement study on GreenOrbs[J]. *Parallel and Distributed Systems, IEEE Transactions on*, 2013, 24(10): 1983-1993.

## 致 谢

岁月如歌，光阴似箭，两年半的研究生求学生活即将结束，站在毕业的门槛上，回首往昔，奋斗和辛劳成为丝丝的记忆，甜美与欢笑也都尘埃落定。经历了找工作的残酷与坎坷，我深深体会到了学校生活的那份宁静和平和。北京邮电大学以其优良的学习风气、严谨的科研氛围教我求学，以其博大包容的情怀胸襟、浪漫充实的校园生活育我成人。值此毕业论文完成之际，我谨向所有关心、爱护、帮助我的人们表示最诚挚的感谢与最美好的祝愿。

首先我要特别的感谢我的导师王文东老师，感谢他收我为他的学生，感谢他给我读研究生的机会，感谢他给予我的鼓励和支持，感谢老师每一次对我热情的指导。在我攻读硕士研究生期间，也深深受益于实验室的阙喜戎导师和龚向阳导师的关心、爱护和谆谆教导。他们作为老师，点拨迷津，让人如沐春风；作为长辈，关怀备至，让人感念至深。能师从各位老师，我为自己感到荣幸和幸福。在此谨向实验室的各位老师表示我最诚挚的敬意和感谢！

在做毕业设计的这段日子里，有实验室三位导师的热情指导和关心，内心更是充满了感激与感动。导师们虽然教研工作非常繁忙，但仍会从项目研发中抽出时间来了解我们的论文进度情况及所碰到的问题，给予了我精心的指导和帮助。导师们渊博的知识、严谨的科学态度、精益求精的工作作风、诲人不倦的高尚师德、强烈的事业心和一丝不苟的工作作风以及活跃的学术思想，给我留下了深刻的印象，为我树立了学习的榜样。导师不仅授我以文，而且教我做人，虽历时三载，却赋予我终生受益无穷之道。在此对他们表示由衷的感谢。

感谢一直关心与支持我的同学、朋友们和家人！感谢你们的鼓励和帮助。我还要感谢的是我寝室的姐妹们，两年来，我们朝夕相处，共同进步，感谢你们给予我的所有关心和帮助，感谢你们带给我的视野，感谢你们在我无助时给我的帮助。同窗之谊，我将永远珍惜！在这里我还要感谢宋峥师兄、高慧师兄以及张波师兄给我的指导和帮助。

还要感谢父母在我求学生涯中给与我无微不至的关怀和照顾，一如既往地支持我、鼓励我，父母的养育之恩无以为报，他们是我十多年求学路上的坚强后盾，在我面临人生选择的迷茫之际，为我排忧解难，他们对我无私的爱与照顾是我不断前进的动力。最后感谢北京邮电大学，带给我宽阔的学习平台，感谢北京邮电大学的老师们，感谢我的同学们，感谢我的室友们，千言万语一句话：谢谢你们。

## 攻读学位期间发表的学术论文目录

- [1] 周雪松. Android 平台上旋转积木游戏的设计与实现[EB/OL]. 中国科技论文在线 .2014.12