**分 类 号 学号 Mxxxxxxxx**

**学校代码 10487 密级**



**硕士学位论文**

**基于xxxxxxx分析工具的研发**

|  |  |  |
| --- | --- | --- |
| **学位申请人** | **：** | **XX** |
| **学科专业** | **：** | **XXXX** |
| **指导教师** | **：** | **XXXX** |
| **答辩日期** | **：** | **2016年5月21日** |

**A Thesis Submitted in Partial Fulfillment of the Requirements**

**for the Degree for the Master of Engineering**

**An Analysis Tool on xxxx: Design and Implementation**

|  |  |  |
| --- | --- | --- |
| **Candidate** | **：** | **XX** |
| **Major** | **：** | **XXXX** |
| **Supervisor** | **：** | **XXXX** |

**Huazhong University of Science & Technology**

**Wuhan 430074, P.R.China**

**May, 2016**

**独创性声明**

本人声明所呈交的学位论文是我个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，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：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电子版，允许论文被查阅和借阅。本人授权华中科技大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密□， 在 年解密后适用本授权书。

本论文属于

不保密□。

（请在以上方框内打“√”）

学位论文作者签名： 指导教师签名：

日期： 年 月 日 日期： 年 月 日

摘 要

这是一个华科硕士论文模板，也是一个华科毕业狗的论文写作经验总结。我编写此模板的目的是希望这个模板能够给以后毕业的学弟学妹们在论文写作方面提供一点点帮助。第一次写这么重要的文档也是战战兢兢，如果有疑问，欢迎联系QQ：461453258。

具体格式上，我在参考《华中科技大学博士、硕士学位论文撰写规定》的基础上，还参考了历年的Word论文模板和Latex论文模板的格式。在章节的标题和正文的格式上，我把所有文中出现的内容都分门别类编辑成了应用样式，也就是说小伙伴们不需要在直接调节字、居中等等问题，直接根据不同的内容刷应用样式就可以了。对于公式、表和图片的题注，我是采用引用中“插入题注”的功能，在应用题注时采用的是“交叉引用”的功能，这样的好处是你再也不需要来回的计算表x.x了，也不需要因为删除一张图片而对每张图片的编号大动干戈。还有，最头疼的参考文献，我是采用NoteExpress插件做参考文献的插入的，它的用法稍微有一点点麻烦，但是绝对是终极精华总结，说不定你会爱上他呢，之后会详细介绍啦。

**关键词：**华科，硕士毕业论文，Word，模板

Abstract

英文摘要，应用格式“宋体小四”

**Key words：**HUST, Thesis, Word, Template

目 录

[摘 要 I](#_Toc451965219)

[Abstract II](#_Toc451965220)

[目 录 III](#_Toc451965221)

[1 绪论 1](#_Toc451965222)

[1.1 研究背景及意义 1](#_Toc451965223)

[1.2 国内外研究现状 2](#_Toc451965224)

[1.3 本文主要研究内容 5](#_Toc451965225)

[1.4 论文组织结构 6](#_Toc451965226)

[2 相关技术概述 7](#_Toc451965227)

[2.1 网络测量参数介绍 7](#_Toc451965228)

[2.2 话单定位算法介绍 9](#_Toc451965229)

[2.3 本章小结 11](#_Toc451965230)

[3 分析工具的设计 12](#_Toc451965231)

[3.1 需求分析和系统框架 12](#_Toc451965232)

[3.2 数据的获取 13](#_Toc451965233)

[3.3 显示模块设计 16](#_Toc451965234)

[3.4 信号覆盖模块设计 18](#_Toc451965235)

[3.5 话务负载模块设计 21](#_Toc451965236)

[3.6 基于地理的覆盖性能模块设计 23](#_Toc451965237)

[3.7 本章小结 25](#_Toc451965238)

[4 分析工具的实现 27](#_Toc451965239)

[4.1 交互流程设计 27](#_Toc451965240)

[4.2 显示模块实现 28](#_Toc451965241)

[4.3 信号覆盖模块实现 31](#_Toc451965242)

[4.4 话务负载模块实现 33](#_Toc451965243)

[4.5 基于地理的覆盖性能模块实现 33](#_Toc451965244)

[4.6 本章小结 36](#_Toc451965245)

[5 分析工具的测试 37](#_Toc451965246)

[5.1 测试平台与实验数据 37](#_Toc451965247)

[5.2 信号覆盖的统计与分析 38](#_Toc451965248)

[5.3 话务负载的统计与分析 42](#_Toc451965249)

[5.4 基于地理的覆盖性能统计与分析 43](#_Toc451965250)

[5.5 本章小结 51](#_Toc451965251)

[6 总结与展望 52](#_Toc451965252)

[6.1 总结 52](#_Toc451965253)

[6.2 展望 53](#_Toc451965254)

[致谢 54](#_Toc451965255)

[参考文献 55](#_Toc451965256)

# 一级标题（所有标题都是自动编号）

## 二级标题

### 三级标题

#### 四级标题

这是正文。

为了保持格式的一致性，强烈建议取消word程序的“格式跟踪”功能。设置方法：点击菜单“文件|选项”，调出“选项”对话框，在“高级|编辑选项”页，取消勾选“保持格式跟踪”。

另外，建议设置文件自动保存。设置方法：点击菜单“文件|保存”，调出“选项”对话框，在“保存”页，设置自动缓存文档间隔1分钟。



图 1‑1 应用样式列表

对于所有的标题样式直接点击图 1‑1，其中一级标题的样式是“标题1”，二级标题的样式是“标题2”，以此类推。

对于正文的样式，直接选择正文，然后点“宋体小四”，这样样式就变成正文的样式。

下面介绍下其他的样式：

* Java代码、代码样式：提供java代码、伪代码常用的样式
* 题注样式：这是给题注用的样式
* 图片居中：图片使用此样式可以使图片居中显示
* 无编号标题：这是给诸如摘要、Abstract、目录、致谢等标题使用的样式，该样式没有编号

## 图片、表格、公式的操作

对于图片、表格、公式比较烦人的就是题注的标注和交叉引用了。

### 图片

放置完图片后，点“图片居中”样式，图片居中显示。

点击菜单“引用”中的“插入题注”就可以为图片添加题注了。题注有专门的应用样式“题注”。如图 1‑2，点击“引用”中的“交叉引用”可以完成对图片的引用。这样引用的好处是，如果图片的编号发生变化了，全选文字然后按F9更新域，所有的编号都会自动调整。



图 1‑2 这是题注

如何在一行同时添加三张图片呢？

如图 1‑3，我的做法是：

1. 插入一张2行3列的表格
2. 在第一行的表格中放入图片，在第二行的表格中放入对应的题注
3. 隐去边框

|  |  |  |
| --- | --- | --- |
| C:\Users\Jasper\Desktop\chapter03\1point.png | C:\Users\Jasper\Desktop\chapter03\2point.png | C:\Users\Jasper\Desktop\chapter03\123point.png |
| (a) 图片1 | (b) 图片2 | (c) 图片3 |

图 1‑3 一行同时添加三张图片

### 插入表格和公式

表 1.1 表格的题注

|  |  |
| --- | --- |
| 在应用中点“插入题注”添加题注 | 如果没有“表”的标签，则需要自己创建 |
|  |  |

(1.1)

表格如表 1.1，公式如公式(1.1)。

## 关于参考文献的引用

我是使用工具NoteExpress添加参考文献的，NE可以在华中科技大学图书馆网站下载，网址<http://www.lib.hust.edu.cn/DBView.aspx?id=40&Tab=1>。

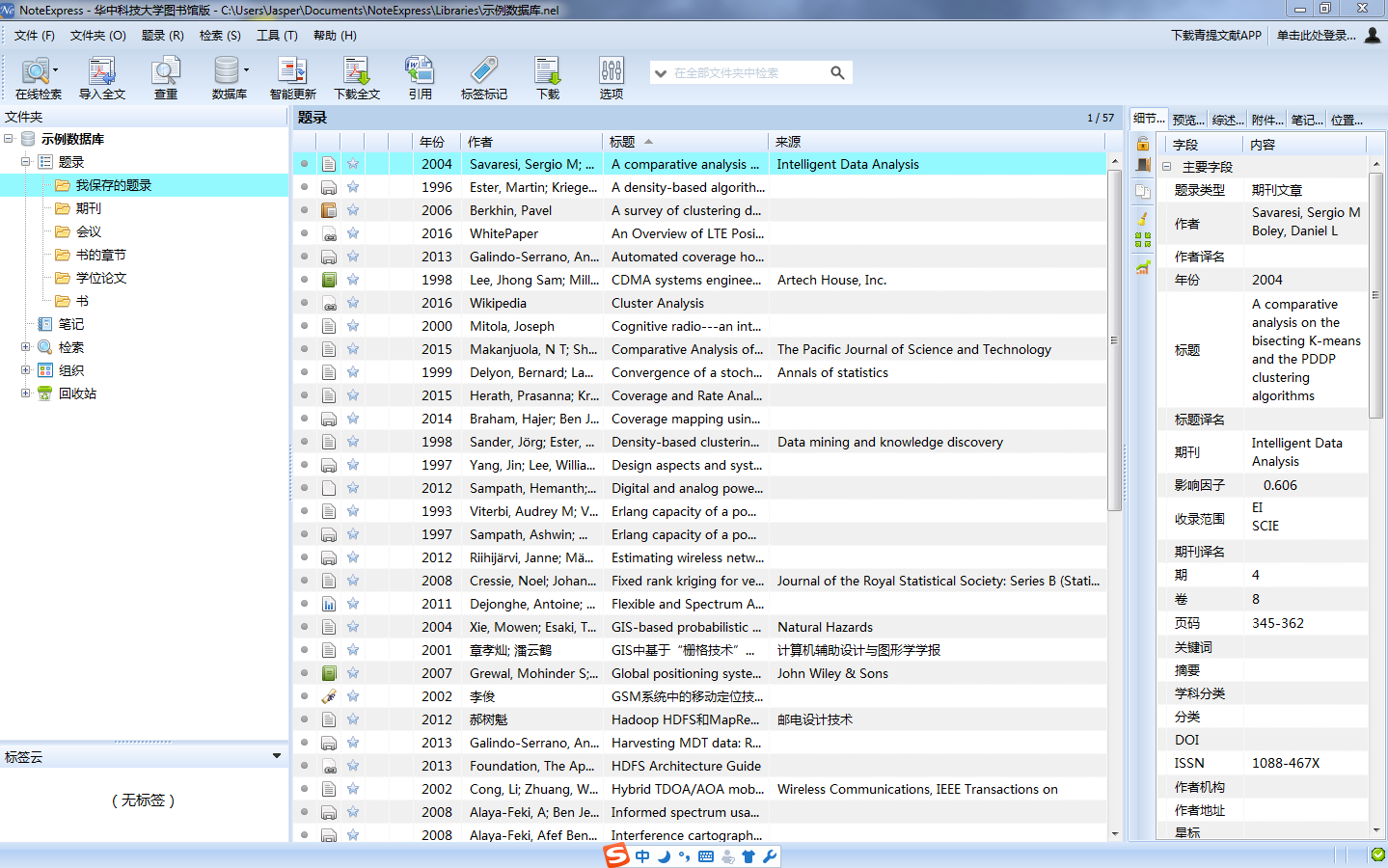


图 1‑4 NE界面

NE的界面如图 1‑4，首先要做的工作就是把你引用的参考文献都添加进来。其中中文的参考文献添加很方便啦，直接使用界面中的“在线检索”中的知网数据库搜索，之后直接添加相应文献的题录。英文参考文献会稍微麻烦点，首先要点开谷歌学术，收到相应的文章点击“引用”，保存EndNote格式文件。



图 1‑5 在谷歌学术上找到相应的文献点击“引用”

图 1‑6 保存EndNote文件

然后在NE中导入该文件到你之前添加参考文献的题录中。你引用的参考文献就在你的NE中了。

下一步，在论文需要添加参考文献的地方，点击如图 1‑7“插入引文”，然后你会发现参考文献就会神奇的出现在你的文章中了。这时，你会觉得高兴，可是过了一会儿，你可能又会发现不对劲，发现参考文献的样式和学校要求的不一样。那参考文献的样式可不可以修改，当然可以了，点击NE“工具”中的“样式”中的“样式管理器”导入样式。我已经根据学校参考文献的规定编辑好相应的样式规范，导入即可。

设置相应的快捷键，会更加方便操作。点击如图 1‑7中的设置可以设置“插入引文”功能的快捷键Alt+1。

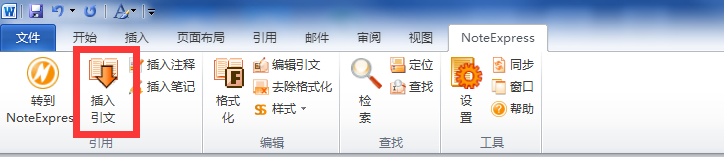


图 1‑7 点击“插入引文”

2015年5月，国际电信联盟（International Telecommunication Union，ITU）公布了《2015年信息与通信技术数据》，报告称3G网络覆盖率在过去5年间快速提高，由2011年的45%到2015年的69%，城市人口覆盖率达到89%。中国电信每个月都会在其官网上公布其移动用户数。2015年，电信全国3G/4G 用户数由11863万增至14313 万，增幅为20.3%。其中4G 用户数由708万增至5846 万，增幅为725.7%。以上数据说明，移动通信技术经历了前两代的发展，3G/4G业务正在快速增长。

随着3G/4G业务在我国的迅速普及，人们对于网络质量的要求也越来越高。网络覆盖和网络质量是移动通信网络运营中的重要部分，网络的优化质量直接决定了移动通信网络的好坏，因此移动通信网络运营商非常重视网络优化工作。做好网络优化工作可以维持或者提高网络质量，进而在用户数和周围环境不断变化的情况下，持续提供更好的用户体验，来保证在移动通信竞争环境中保持住有力地位。

移动通信网络优化是通过对入网的设备进行数据采集和分析，找到网络中存在的问题和造成问题的原因，然后通过对网络设备参数或者通信系统参数的优化和调整，使网络性能最佳[1]。通常网络优化包括两方面要求：解决问题和提高性能。解决问题主要是针对网络中发现的问题和故障提出有效方法加以解决和排除。提高性能是对网络中可能正常运行但是指标较差的情况，通过调整参数优化配置等方式来提高网络性能[2]。

移动通信网络具有较强的地理空间特性，因此在进行网络优化时结合地理位置进行具体分析才有更多的现实意义。目前一些常规的网络指标分析软件与网络优化软件，大多是面向某个单一领域，没有将测试的指标数据与地理空间进行关联，更难以对网络性能指标在地理空间上做更高级的分析。而目前被广泛使用的地理信息系统（GIS）则为移动通信领域提供了一个新的视角，可以将移动通信的网络指标在空间和时间两个维度上进行结合，以直观的空间视图界面展示业务的地理空间分布和随时间变化趋势[3]。同时，通过在空间和时间上的联合分析，可以找到更多影响网络质量的细节，为最终网络优化的决策提供更加丰富的信息。

基站记录有用户详细的网络数据和用户话单记录，利用这些数据结合大数据计算进行相应的分析，可以比较实时高效地检测网络的运行状况，并及时发现问题和定位原因。云计算平台Hadoop可以用来解决大数据面临的存储和计算问题[4]，目前Hadoop主要应用于国外的Google、Facebook和国内的腾讯、阿里等互联网公司，但是在移动通信运营商中应用较少。

该论文所设计的移动通信网络覆盖性能测试工具，采用Hadoop云计算平台实现了对基站的话单数据的实时存储，并且结合相应的话单定位算法可以高效地完成对话单数据进行定位处理，得到每个话单发生的时间位置和信号质量信息。最后，结合地理信息系统，可以在地图上展示出信号质量和话务量在空间和时间维度上的分布，并通过后续的分析可以找出地图上的弱覆盖和高负载区域，这些信息对于移动通信网络优化具有重要的参考意义。

## 国内外研究现状

### 网络优化技术和工具

随着移动网络技术不断的升级换代，移动通信网络分析和优化技术也需要不断去创新以提高网络检测的效率。传统的网络分析方式非常依靠人力，例如路测需要用专业的车辆和设备去采集相关数据，并且需要相关专家对数据做专业的分析。所以路测是比较低效的解决方法。为此，3GPP（the 3rd Generation Partnership Project）在其第9版标准[5]中已经包含了MDT（Minimization of Drive Tests），因此使用此版本网络通信标准可以显著降低成本。MDT的基本概念是用户的终端（the User Equipment, UEs）可以根据操作请求向网络报告它们的地理位置。传统的UE的测量和MDT的主要区别在于：传统的UE 是基于小区基站得到的地理位置，而MDT是基于用户终端的GPS技术得到的地理位置[6]。运营维护人员可以直接利用这些数据进行网络操作管理和优化任务。同时，研究人员也可以利用这些定位数据进行网络预测，进而为用户提供更好的网络质量。

为了建立一个准确和可依赖的覆盖地图，文献[7]提到一种从统计学借鉴来的空间插值技术Kriging[8]。 这种技术依赖于所测的数据的相关性并且可以在感兴趣的区域绘制出一个完整地图。在一些论文中应用Kriging技术[9]去做覆盖地图预测。在许多文献[7, 10, 11]都研究Kriging和它的一些衍生技术来进行覆盖预测。

在文献[11]，Galindo-Serrano等提出无线环境地图（Raido Environment Maps, REM）的方法用来解决蜂窝网络覆盖盲区检测的问题。REM对那些有地理位置的数据应用了空间差值的技术来得到真实的地理数据。这种方式可以自动鉴别覆盖盲区的数量、位置和形状。REM 可以对覆盖盲区的检测和预测效率都会有提高。

Milola提出在文献[12]提出地理位置的数据信息需要位置/环境信息。REM存储了地理位置信息、移动通信网络测量数据、环境信息以及过去的数据。在文献[13, 14]中，作者提出一种网状的REM结构，IC（Interference Cartography），这种结构中的点是方形栅格（比如像素点）。该方法的主要观点是：

1）用收集得到的带有地理位置的测量数据来预测未测量位置的测量数据；

2）测试未测量位置来提高预测的质量。

然而算法的复杂性随着测量点数数目的增加成指数增长O(N3)，N是测量点的数目）。在文献[15]中提到的FRK （Fixed Rank Kriging）是Kriging的一种变型，它的算法复杂度为O(NR2)，R是由用户定义的“固定等级”。在文献[16, 17]，该算法被用于覆盖预测。在仿真和实际测试的性能评估中，FRK均被证明在算法复杂度和预测准确度之间实现了非常好的平衡。

但是之前的这些工作大多是建立在移动终端可以准确定位的理想情况下的。然而，文献[18]中指出使用GPS定位的误差在5m到30m范围内，而在文献[19]中说明基于无线网络度量的定位技术误差在50m到300m。这都说明定位的不确定性降低了覆盖预测的准确性。

在文献[20]中，A. Palaios等人提出通过采用多种测量数据来提高定位的准确度和减少定位的误差。这种定位的方式是通过适当地组合不同传感器测到的结果，这种方法虽然能够得到不多的结果，但是显然不适用于用MDT特性测量的情况。在文献[21]中，作者Braham等人提出了通过扩展FRK算法来处理定位不准确的问题。在预测和校准模型中，在FRK模型中定位的不准确性影响了函数的均值和协方差。该文献的主要贡献是1）通过在模型中引入定位的不确定性，作者比较和研究了最佳线性无偏预测值和条件期望预测；2）引入了SAEM（Stochastic Approximation EM）算法。SAEM结合了随机EM和Gibbs抽样程序来处理大量的计算[22]。Gibbs 算法用并行处理的方法解决了定位概率密度抽样的问题。

至于网络优化工具，国际上的优化系统厂商在全球的无线网络优化市场中占据着较大的市场份额，其中爱立信公司开发的Tems优化系统应该最为广泛[23]。在国内，例如华为、中兴等公司也在开发相应的网络优化软件，但是开发的优化软件虽然在最近几年已经取得了长足的进步，但是还是存在一定的问题，首先在网络优化的实践过程中仍然需要人工参与，缺少自动化操作；其次是采用较为传统的数据存储和计算方法，对与海量的通信数据的处理效率不高；最后，仅仅是针对某种特定类型的数据提供处理，优化系统通用性有待提高。

### 地理信息系统

地理信息系统是在计算机软硬件的支持下，采集、管理、检索、分析和描述与地球表面空间位置相关的数据计算机系统[24]。它起步于60 年代，是一门集合了计算机科学、地理学、空间科学、环境科学和遥感测绘学的学科，它采用的基本技术是地理空间数据库技术、地图可视化技术和地理空间分析技术[25]。

近年来，随着科技进步和社会发展，地理信息系统在各行各业已经得到了广泛的应用。在移动通信服务领域，基于GIS的移动空间定位服务已经被绝大多数用户所使用，例如用户使用百度地图APP可以完成定位、路径规划和导航等功能，结合互联网数据，用户通过百度地图还可以找到附近的餐馆、银行。在通信网络的建设和优化方面，地理信息系统也有很多用武之地。在文献[26]中，作者提到使用GIS中的数字高程模型（Digital Elevation Model，DEM）以及其他相关分析方法在多山区域找到最佳建站地址；同时可以借助GIS 的空间数据分析的特点，结合移动通信网络参数和信号的传播模型，可以绘制出信号覆盖的预测图。在文献[27]中，作者通过结合现有的无线电波传播模型和基站辐射理论，提出并实现一种基于GIS的基站选址方案，该方案能够在基站选址时反映基站对于小区、学校和医院的电磁辐射情况，为监管部门在基站审批时提供一个可参考的直观化依据。在文献[28]中，作者提到使用GeoDatabse技术来应对GIS的海量数据存储的问题，并且GeoDatabase还支持多用户并发访问、版本管理和数据动态更新等功能，比传统的文件形式存储和空间数据的表示方法有巨大的优势。

按照空间数据的组织形式，可以将地理信息系统分为两种类型，一种是基于栅格的，另一种是基于矢量的。基于栅格的GIS以栅格元为地理特征的最小单元，并用同一个数据结构存储地理特征的空间位置信息和属性信息；而基于矢量的GIS是以点和线组成的，首先由点构成地理特征的边界，然后由点和这些点连接成的线共同组成对于地理特征的描述[29]，并且在通用的数据库中存储其属性信息，通过数标建立空间位置信息与属性信息的联系[30]。

本论文所研究的工具是基于栅格的GIS的应用，栅格图是由排列整齐的栅格点组成的，每个栅格点都有自己对应的空间位置信息和属性信息。比如栅格图是由a\*b个栅格组成，每个栅格用(x,y)坐标来标示，然后每个栅格的属性值可以通过颜色或者灰度值值进行表示[30]。

## 本文主要研究内容

本文的研究目的是设计并实现一个基于地理栅格的蜂窝网络覆盖性能分析的工具。本文处理的数据是某运营商在某一地区一段时间内的PCMD （Per Call Measurement Data）数据。本文所做的主要工作如下：

1. 通过Hadoop分布式处理平台对PCMD数据进行存取，生成栅格数据。PCMD数据记录的是每天通话记录的情况，包含通话时间、信号质量、基站编号、往返时延等基本信息。其中对PCMD中的基站编号和往返时延采用定位算法进行计算，可以得到每条通话的具体位置。由于PCMD 话单数据量巨大，一个月某省就能生成几十亿条话单数据，采用Hadoop 平台可以高效地对这些数据进行存取和计算，生成的栅格数据。
2. 将蜂窝网络的信号覆盖和话务负载信息以栅格图的形式进行展示。要更好地基于地理位置展示蜂窝网络的信息，需要将地理空间细分为若干个100m\*100m 的栅格。Hadoop平台生成的栅格数据包含具体位置、信号质量、话务量、时间戳和日期等信息，具体位置指栅格的ID，信号质量指某一时间段内在该栅格发生的所有通话的信号质量的平均值，话务量指某一时间段内在该栅格中发生的所有通话的数量，时间戳和日期是标记某个具体时间段。根据之前已经计算好的每个栅格的信号质量和话务量信息，根据值的大小在底图进行染色，最后生成信号覆盖和话务负载的栅格图。
3. 对蜂窝网络的信号覆盖和话务负载信息单独进行处理和分析，分别找到分析地区在该时间段内的弱覆盖区域和高负载区域。弱覆盖区域是指信号质量低于某一分析阈值的区域，高负载区域是指话务量高于某一分析阈值的区域，两个阈值都允许用户手动输入。通过数据分析，找到栅格数据中信号质量低于某一阈值的栅格，调低染色的透明度，并调高正常区域的染色透明度，在栅格图中重点突出显示弱覆盖区域。同理，突出显示高负载区域。
4. 对蜂窝网络的信号覆盖和话务负载进行联合统计与分析，统计出同时存在弱覆盖和高负载栅格的比例，并且在栅格图上重点展示此类栅格的位置。通过分析这些栅格点与基站距离之间的关系对此类需要优化的栅格点进行分类，针对基站能覆盖到的栅格点，通过调整基站参数结构、天线倾角等方式来优化；针对基站覆盖不到的栅格点，通过增建基站或者放置直放站的方式来优化。

## 论文组织结构

本论文由六个章节组成，其组织结构如下所述：

第一章介绍的是本文的的研究背景和意义，介绍国内外有关网络优化常采用的分析方法和工具，之后又介绍了GIS在移动通信网络中的应用现状，并介绍论文的主要解决的问题，最后给出了论文组织结构。

第二章介绍的是网络测量的相关参数和有关话单定位的算法，为之后的章节做铺垫。

第三章首先提出的该分析工具的需求分析和系统框架，然后依次谈到数据获取的流程和各个模块的设计工作，并详细说明了每个模块设置的意义、具体要解决的问题以及具体的执行流程。

第四章首先给出了该工具的交互流程设计，之后展开谈到每个模块的具体实现。对重点环节给出了伪代码和效果图，并且重点阐述了基于地理的覆盖性能模块对于II类目标栅格聚类的具体实现思路。

第五章是对本文所设计和实现的工具进行测试，首先分别对信号覆盖和话务负载模块单独测试，之后再联合测试，并对每一个测试的结果进行结果展示和分析，对最终的测试结果进行评估。该章节验证了本工具的各功能模块达到了预期效果。

第六章对本文进行了总结和展望，首先总结了本文的工作成果，最后客观地分析了本文所实现的蜂窝网络优化方法的不足，同时给出了未来可以改进的方向。

# 相关技术概述

## 网络测量参数介绍

在移动通信系统中，衡量一个网络好坏最重要的要素是覆盖、容量、质量、频谱利用率和传输效率[1]。

### 网络覆盖测量参数

无线网络覆盖是指在一定区域内，无线信号的强度和质量能够为用户提供移动通信服务，称该地区存在无线网络覆盖。移动通信系统的覆盖率和覆盖质量是衡量移动通信网网络质量和服务质量的重要指标[31]。在网络初期的建设和后期的维护期间，网络运营者通过测试、数据采集和分析等手段，来检验网络是否满足覆盖要求。如果不能达到要求，则需要对网络进行覆盖方面的优化。判断移动通信网络是否在某一地点形成覆盖，需要依据一定的测量参数。

对于GSM系统，判断是否存在网络覆盖只需要测试一个参数，即移动台接收电平（Rx\_Power）。一般当Rx\_Power-90dBm视为该地点存在GSM网络覆盖。在CDMA 系统中，与网络覆盖与导频强度（Ec/Io）有关，一般要求Ec/Io-12dB且Rx\_Power-90dBm 且Tx\_Power15dBm，只要这三个条件全部满足，才视为该点存在CDMA 网络的覆盖[32]。

**Rx\_Power：**在GSM系统中，Rx\_Power表示在工作频点上接收的电平强度，该参数直接反映了GSM网络覆盖的水平[33, 34]。CDMA 系统中，Rx\_Power表示在整个1.2288MHz带宽上总的接收功率[35]。该接收功率不仅包括服务小区的信号功率，也包括其他小区的信号功率、接收机热噪声及外界干扰。

**Ec/Io：**Yang在文献[36]介绍Ec/Io指每码片辛哈能量与总功率密度之比。在CDMA 系统中，Ec/Io反映了信号的干扰水平。移动台或者基站均要求信号的Ec/Io必须超过一定阈值才能正确解调。Ec/Io在-10dB 以上时，网络质量可以保持非常良好的水平；当Ec/Io-15dB时，则移动台无法接入网络。一般将Ec/Io-12dB作为满足网络的覆盖的标准。在CDMA系统中，Ec/Io和Rx\_Power 是评价网络前向覆盖能力的指标。

**Tx\_Power：**在CDMA系统中，移动台发射电平Tx\_Power反映网络的反向覆盖能力。CDMA系统对方向链路采用快速功率控制。在反向传播链路不好的情况下，会迅速提高移动台的发射功率，因此移动台的发射功率Tx\_Power可以衡量反向覆盖的水平。当Tx\_Power15dBm时，可认为该地点达到反向覆盖的要求[37]。

### 网络容量测量参数

对于一个移动通信系统而言，除保障覆盖和质量要求外，容量的保障也至关重要。容量优化是网络优化工作中必不可少的一部分，移动给用户数量的增长、用户通信行为的变化、新业务的使用及设备故障灯，都会引起网络的可用容量降低，导致用户尝试接入网络时可用资源分配。因此，为保证移动网络高质量地运行，必须实时监控网络的容量变化，及时发现网络容量不足的问题，并采取相应增扩容量的措施，以满足用户通信的要求。

同时，作为移动通信运营商，在满足用户的话务需求的同时，也要充分合理利用既有网络设备和频谱资源，实现资源利用率最大化，从而达到提供优质服务、节省投资的目的。这也是网络优化的目标和原则。

无线网络的容量体现了移动通信网络提供服务的能力。移动通信系统中，使用无线电波作为信息的传输载体，完成用户终端与基站之间的信息传送。由于无线频谱的资源限制，使得无线网络容量的瓶颈不同于固定网络。固定网络的容量“瓶颈”主要在于交换机的交换能力，而移动通信网络的容量瓶颈则取决于无线频谱的宽度、频率复用方式和调制方式等。

在通信系统中，无论是固定通信系统还是移动通信系统，都用话务量来反映通信业务量的大小，也用来反映网络容量的大小和通信设备的负荷。

在文献[38, 39]提到，话务量的单位是爱尔兰（Erlang，简写为Erl），是为了纪念话务理论的创始人，丹麦学者A. K. Erlang 而命名的。爱尔兰定义为单位时间内信道被占用的时长。如果一个信道在1小时内被全部连续占用，称此时的话务量为1Erl；如果一个信道在1小时内被占用30分钟，称此时的话务量为0.5Erl。

话务量的计算方法为：单位时间内发生的呼叫次数和每次呼叫所占用的时间的乘积[40]。话务量一般用A表示，即

(2.1)

其中，为单位时间内的呼叫次数，为每次呼叫的保持时间。

从话务量的公式可以看出，话务量受两个因素的影响：一个是单位时间内的呼叫次数，用户越多或者呼叫越频繁，话务量就越高；另一个是呼叫保持时间，用户的通话保持时间越长，话务量就会越高。

对于有多个用户的系统，系统的总话务量等于所有用户话务量之和：

(2.2)

其中，中系统的总话务量，为系统中总的用户数量，为平均每个用户的话务量。

值得注意的是，通信网络中各个时段的话务量往往并不是平均分配的，话务量最高的一个小时称为忙时。忙时这一个小时的话务量称为忙时话务量。忙时话务量与全天话务量之比，称为忙时集中系数。忙时集中系数一般在10%~15%之间。

因此，忙时话务量可由下式计算，即

(2.3)

式中，为全天话务量，为忙时集中系数。

在进行网络规划和优化时，一般以满足网络忙时的话务量需求作为规划和优化的目标。

## 话单定位算法介绍

在PCMD话单数据中提取用于话单定位的数据，针对每条话单连接基站数目的不同分别采用不同的定位算法。

**COO(Cell of Origin)**定位算法的基本原理是根据移动终端距离基站的距离和扇区的编号来确定移动终端的位置[41]。该论文中使用COO 定位算法对只连接一个基站的话单进行定位。

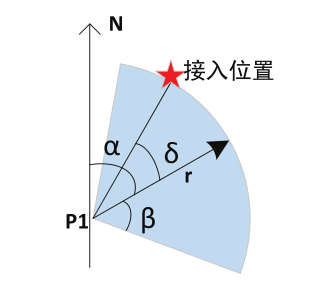


图 2‑1单点定位示意图

如图 2‑1，P1为小区位置，是小区方位角，是小区半波瓣角60，为小区传播时延换算的接入距离，介于[-,+]的随机数。接入位置以P1为原点，以 方向，半径为r处的位置。

**AOA(Angle of Arrival)**定位算法是基于方向角的两基站定位算法。文献[42]提到，其基本原理是利用移动终端到达两个基站的距离和所在的扇区的编号来计算其具体位置。如图2-2：

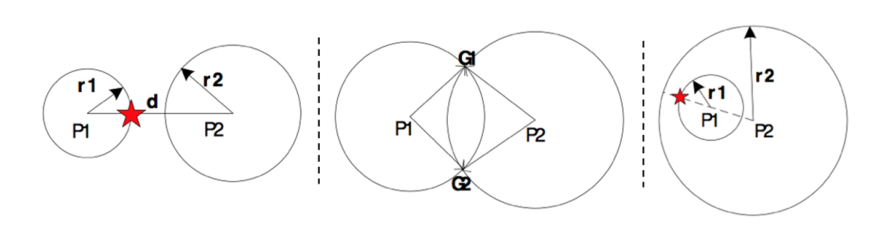


图 2‑2两点定位示意图

* 两圆相离(d>r1+r2)：d为两小区之间直线距离，以P1为原点，P1、P2连线方向，r1为接入距离；
* 两圆相含(d+r1<r2)：d为两小区之间直线距离，以P1为原点，P1、P2连线方向，r1为接入距离；
* 两圆相交(d<r1+r2)：求出两交点G\_1、G\_2；小区P1、P2的方向角为 和；G1和P1、P2 连线的方位角分别为、；G2和P1、P2连线的方位角分别为、；取角度偏差小的点，即如果，则接入位置为G1，否则为G2。

**TOA(Time of Arrival)/TDOA(Time Difference of Arrival)**定位算法基本思想是找到移动终端距离三个移动终端的交叉点即为移动终端的位置[43]，该方法要求移动终端和基站的时间精确同步。为了测量移动终端的发射信号的到达时间,需要在每个基站处设置一个位置测量单元，为了避免定位点的模糊性，该论文中使用TOA/TDOA 定位算法对连接至少三个基站的话单进行定位，如下图 2‑3所示：

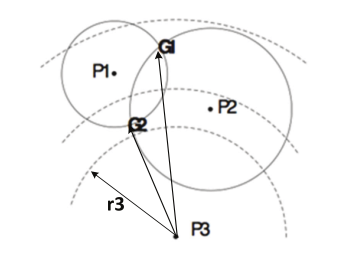


图 2‑3三点及三点以上定位示意图

* 两圆相离、相含：处理方法同AOA，不再考虑第三点。
* 两圆相交：求出P1、P2圆的交点G1 和G2;求出P3到G1、G2的距离分别为d31 和d32；用P3的接入距离r3进行选取，如果|d31-r3|>|d{32-r3|, 则接入距离为G2，否则为G1。

## 本章小结

本章介绍了网络覆盖分析最新的一些技术以及网络覆盖和网络容量经常用到的一些测量参数，这些参数中部分会在之后的章节中出现。在这之后本章又介绍了地理信息系统的一些概念，重点讲解了下地理栅格产生的背景。最后，本文概要地介绍了下话单的定位算法。

# 分析工具的设计

## 需求分析和系统框架

移动通信网络是一个动态的网络，在网络持续运营的过程中，会出现一些在网络规划设计中无法周全考虑的一些变化，例如传播环境的变化、用户业务量的变动以及业务质量的改变。这都需要对移动通信网络进行相应的优化，以更适应实际的变化。网络优化就是解决网络运行中存在的问题，优化资源配置，最大限度地发挥网络和设备的效能。相较于网络建设，网络优化是一个持续的过程[32]。

针对网络优化，目前已经存在一些分析工具，这些工具主要是通过采集路测数据、用户投诉数据进行分析，然后再结合现网的运行和工程情况制定出适宜的优化调整方案。虽然对采集的样本数据和对用户投诉数据进行分析得到的结果具有一定的可信性，但这样的做法比较耗费人力成本并且得到的数据偏少不能完整反映某一地区所有区域的网络覆盖情况。而基站中的话单数据完整地记录了每一次通话的网络覆盖情况，某一地区的所有基站则记录了该地区所有通话的网络情况，对某一地区的所有话单数据进行整体大数据分析，就能简单直观地反映出该地区信号覆盖和话务负载情况，并且进一步分析可以得出更加可信的网络优化方案。

该工具通过对某运营商某地区一定时间段内的 PCMD数据进行大数据处理，以100m\*100m栅格为粒度将该地区分成若干个小栅格，计算出该地区每个栅格的信号覆盖质量和话务量信息。针对每条 PCMD数据连接不同数量基站的情况，分别采取 COO、AOA、TOA/TDOA算法进行定位处理。之后对获得的每个栅格的信号覆盖质量和话务量信息进行统计和处理，以栅格图的形式直观展现信号质量和话务量覆盖情况。并对处理后的信号覆盖质量和话务负载联合统计，找出该地区中需要优化的目标栅格，根据栅格距离基站的距离是否在基站有效覆盖半径之内，对目标栅格进行分类。最后对于不同类别的目标栅格采取不同的改善策略。

通过对移动通信网络信号覆盖质量和话务量的分析，发现网络覆盖中话务量较高但是信号覆盖质量较差的区域，为移动通信网络的优化提供有价值的参考信息。其中得到的基站优化排序和增建基站区域排序，再结合实际情况，可以作为网络优化方案的一部分。

该工具是从 PCMD话单数据中提取出栅格话单数据，来分析蜂窝网络信号覆盖和话务负载特性，并通过栅格图、统计图表等方式直观展示分析结果，最终提出网络优化方案的。

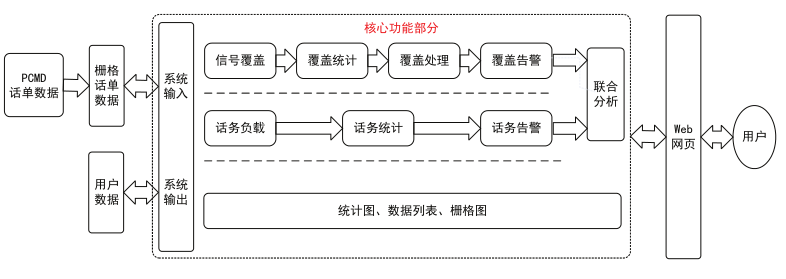


图 3‑1系统总体框架

该分析工具主要框架如图 3-1所示，分为数据存取部分、核心功能部分和界面展示部分。其中数据获取部分是从 PCMD话单中提取栅格话单数据，其中 PCMD话单来源于某省的某运营商传输局，该部分主要运用了 Hadoop分布式处理平台和 COO、AOA、 TOA/TDOA三种定位算法。界面展示部分是 Web网页，使用 HTML、CSS和JavaScript语言编辑实现，用户可以通过该部分输入相应参数，最终结果也会在这里呈现出来。该工具的关键部分是核心功能部分，包括信号覆盖特性的统计和分析模块、话务负载的统计和分析模块、两者的联合分析模块以及制图显示模块等。信号覆盖模块会利用栅格话单数据中的 Ec/Io和栅格 ID信息，统计 Ec/Io在数值空间上的分布，并分析出信号覆盖弱覆盖区域；话务负载模块会利用栅格话单数据中的话务量和栅格 ID信息，统计话务量在数值和空间上的分布，并分析出话务高负载区域；最后使用基于地理的覆盖性能模块，联合分析找出弱覆盖目标栅格，并最终分析出具体的目标栅格的改善策略。

## 数据的获取

### 原始数据介绍

该课题采集的 PCMD话单数据时间长度为 4个周，覆盖范围为某省全省，采集共计几十亿条的 PCMD话单数据。这些数据按照地级市进行划分，每个地级市一个文件夹。每个文件夹内记录着这个地级市内的话单记录，话单数据按照小时以地级市为划分生成，平均每小时的产生量为数GB左右。

在论文中，为了方便，统一将某地级市称为某地区，某地区下的某县区镇或者截取的某一固定大小的地方则统一称为某区域。通常，该工具是以某地区为单位进行计算的，但是最后的呈现可以按照用户需求显示某一区域的特性分析情况。

该工具要对PCMD进行提取，计算出以 100m\*100m栅格为粒度的蜂窝网络信号覆盖质量和话务量信息。其中会用到的PCMD字段如下表3.1所示：

表 3.1从PCMD中提取到的字段表

|  |  |  |
| --- | --- | --- |
| 序号 | 字段名称 | 说明 |
| 1 | *CallStartTime* | 电话开始时间 |
| 2 | *CallType* | 呼叫类型 |
| 3 | *CallEstablished* | 通话是否已建立 |
| 4 | *init\_CellSite* | 参考基站 |
| 5 | *init\_ref\_CellPAF* | 参考扇区 |
| 6 | *init\_ref\_PilotStrength* | 参考导频强度 |
| 7 | *init\_ref\_RoundTripDelay* | 参考往返延迟 |
| 8 | *init\_non\_ref\_ CellSite1* | 非参考基站1 |
| 9 | *init\_non\_ref \_CellPaf1* | 非参考扇区1 |
| 10 | *init\_non\_ref\_ PilotStrength1* | 非参考导频强度1 |
| 11 | *init\_non\_ref\_PrimaryCellRoundTripDelay1* | 非参考基站往返时延1 |
| 12 | *init\_non\_ref\_CellSite2* | 非参考基站2 |
| 13 | *init\_non\_ref \_CellPaf2* | 非参考扇区2 |
| 14 | *init\_non\_ref\_PilotStrength2* | 非参考导频强度2 |
| 15 | *init\_non\_ref \_PrimaryCellRoundTripDelay2* | 非参考基站往返时延2 |

### Hadoop分布式处理平台

Hadoop是一个在机群上使用简单的编程模型来处理超大数据集的软件框架。它可以支持从一台服务器到上千台有本地计算和存储功能的机器处理数据。并且它能自发检测和处理应用层的故障从而摆脱对硬件高可靠性的依赖，来提供高可靠性的服务[44]。

Hadoop主要由 HDFS(Hadoop Distributed File System)和MapReduce两部分组成。最底层是 HDFS，它存储 Hadoop集群中所有存储节点上的文件， HDFS的上一层是MapReduce引擎，用于处理大规模数据集[45]。

HDFS是一个运行在服务器上的分布式文件系统。与其他分布式文件系统不同的是，它[46]

•具有高容错性，可以部署在价格低廉的服务器上；

•提供高吞吐量来获取应用数据，适合于那些有着超大数据集的应用程序；

•放宽一些POSIX的要求，支持以流式数据访问文件系统数据。

MapReduce是一个软件框架，它可以让用户更加容易地编写程序来在大规模服务器机群（千台量级）上以一种可靠、高容错性的方式并行处理大规模数据（ TB量级）。MapReduce工作原理是将任务分解为成百上千个小任务，然后发送到计算机集群中。每台计算机再传送自己那部分信息， MapReduce则迅速整合这些反馈并形成答案。简单来说,就是任务的分解和结果的合成[47]。

图 3‑2 基于Hadoop平台的数据获取流程

如图 3‑2所示，在该课题中MapReduce主要应用于：

* 处理话单数据，定位每条话单经纬度；
* 并行导入数据到Hbase数据库；
* 计算全省各地区的地图栅格信息；
* 计算每个栅格内的平均Ec/Io值和话务量，生成TXT文件。

### 数据结构

通过 Hadoop平台得到的 TXT文件的每条记录都包含着栅格 ID编号、经纬度、 Ec/Io、话务量、时间戳等信息，在该论文中统一称为栅格数据。其中通过单点定位算法计算得来的栅格数据称为一点栅格数据，同理有两点栅格数据和三点栅格数据。每条栅格数据包含的字段如下表 3.2，其中时间戳是以 2个小时为粒度标注的，时间戳的格式是 0:00-2:00为1，2:00-4:00为2，以此类推至22:00-24:00为12。

表 3.2每条栅格数据的字段表

|  |  |  |
| --- | --- | --- |
| 字段编号 | 字段名称 | 说明 |
| 1 | *grid\_id* | 栅格的ID编号，如648ˆ738 |
| 2 | *longitude* | 栅格对应的经度 |
| 3 | *latitude* | 栅格对应的纬度 |
| 4 | *Ec/Io* | 该栅格的信号质量Ec/Io |
| 5 | *call\_count* | 该栅格的话务量 |
| 6 | *call\_count* | 时间戳 |
| 7 | *data* | 日期 |

比如表 3.3，代表在 2015年5月3日2:00-4:00时间段内，在编号为 1000ˆ1003、经纬度为(30.85763,114.72652)的栅格中通话的Ec/Io的平均值为-9.79，总话务量为90。

表 3.3数据库中的栅格数

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *grid\_id* | *longitude* | *latitude* | *Ec/Io* | *call\_count* | *call\_count* | *data* |
| 1000ˆ1003 | 104.72652 | 34.85763 | -9.79 | 90 | 2 | 150503 |

## 显示模块设计

本论文研究的是对网络信号覆盖和话务负载进行分析的工具，核心功能模块都需要与数据打交道，但是该工具在处理之后得到的数据本身并不能够给用户带来直观的展示。为了能够把最终展示的结果直接地呈现给用户，以便用户根据处理结果来调整参数和评估最后得到的方案，该工具提供了两种表示方式：数据统计图和栅格图。

### 数据统计图

数据统计图允许用户通过输入统计区域和统计时间来查看话单数据中的某一地区某个时间段内的Ec/Io 或者话务量在数值区间上或者时间上的分布情况。

图 3‑3数据统计模块流程

如图3-3所示，用户在Web页面输入要统计的区域和时间段，该模块从某地区的TXT文件中获得要统计区域的栅格数据进行统计，最后将数据通过ajax请求以JSON格式发送给前台，前台调用Highcharts库进行统计图的渲染，最后在Web页面展示。

### 栅格图

通过将全省划分成 100m\*100m的栅格并对栅格进行染色处理，栅格图可以将某区域的Ec/Io或话务量基于地理位置进行直观地展示。

该工具采用经纬度每相隔 0.001度为一个栅格。经度每相差 0.001度，平均直线距离为97m；纬度每相差 0.001度，平均直线距离为 111m。按经纬度 0.001度为单位划分栅格，基本可满足百米精度要求，可以近似将栅格大小看成 100m\*100m。据此，某省全省可划分为约2400万个栅格。每个栅格能够按两个字段（Ec/Io和话务量）的进行颜色填充。



图 3‑4绘制栅格图流程

如图 3‑4所示，绘制栅格首先要对 Ec/Io或者话务量进行排序，找到每种染色的颜色对应的数值边界，后要计算好每个栅格在地图上的对应的像素值，最后染色，输出 PNG图片。

## 信号覆盖模块设计

信号覆盖模块是对栅格话单中Ec/Io 进行统计、处理和分析的模块，它可以帮助用户更好的了解Ec/Io在数值和空间上的分布。用户可以根据需求自行配置参数，经过模块计算，最终结果以比例图和栅格图的形式呈现。

### 信号覆盖模块需求分析

|  |  |  |
| --- | --- | --- |
| C:\Users\Jasper\Desktop\chapter03\1point.png | C:\Users\Jasper\Desktop\chapter03\2point.png | C:\Users\Jasper\Desktop\chapter03\123point.png |
| (a)一点栅格图有圆环示意图 | (b) 栅格图栅格有缺失示意图 | (c) 栅格图可读性差示意图 |

图 3‑5 信号覆盖模块需求分析

通过之前介绍的定位算法，我们已经明确知道通过该算法定位的话单会在一个的圆环上，我们采用的是随机在该圆环取一点作为该话单的位置。所以单点定位的栅格数据会有圆环，可信度较低，如图 3‑5(a)。 全省的栅格数据大约有605是单点栅格数据，所以如果完全舍弃单点栅格数据会造成很过栅格数据缺失。所以需要对单点和多点栅格数据进行融合，使得融合后的数据能够避免单点定位的圆环现象，同时减少出现栅格缺失的情况。

图 3‑5 (b)所示，栅格图栅格有缺失。这是由于话单统计时间过短或者该栅格处恰好没有通话产生，所以会造成栅格图中某些栅格由于数值缺失导致不能染色，对用户利用栅格图进行信号覆盖情况分析会造成影响，这会影响用户体验。这需要通过参考周围栅格的数值对缺失栅格进行合理的填补。

如图 3‑5 (c)所示，栅格图有杂点，这会对栅格图的可读性带来一些障碍。可以给用户设置一些可调整的平滑参数，让用户可以自行调节参数来达到平滑栅格图中颜色的效果，以便用户更好地观察某区域信号覆盖的整体情况。同样的，用户还需要可以设置相应阈值，过滤出信号质量差的栅格，绘制出弱覆盖区域。

### 信号覆盖模块功能设计

经过考虑，该模块设计了四个子模块来满足如上的需求，分别是数据融合、栅格增补、栅格平滑和寻找弱覆盖区域。



图 3‑6信号覆盖模块设计流程

信号覆盖模块的设计流程如图 3‑6。在数据融合子模块中，对单点定位的栅格数据和两三点定位的栅格数据按一定比例进行融合；在栅格增补子模块中，通过计算周边栅格的数据对确实的栅格进行填补。在数据处理完毕之后，可以使用栅格平滑和寻找弱覆盖区域两个子模块对数据进行分析。在栅格平滑子模块中，对栅格数据进行平滑处理，来显示整体信号质量的覆盖态势；在寻找弱覆盖区域子模块中，设定弱覆盖阈值，重点显示局部的弱覆盖区域。下面详细讲一些每个模块的具体设计：

数据融合子模块是为了解决话单单点定位不准确而多点定位的栅格数据又偏少的问题。基于的基本原理是，在邻近的栅格范围内，Ec/Io值不会发生突变。所以每个栅格的Ec/Io值，对于邻近栅格都具有参考意义，这是融合算法的基础。

该子模块的数据融合算法流程如图 3‑7所示，首先要遍历该区域的每一个栅格，判断该栅格处都有几种类型定位的栅格数据，之后对于不同类型定位方式采取不同的处理方法，根据输入参数进行计算输出融合后的栅格数据。用户需要输入的参数有：当融合定位数据都有时的单点比例*one\_point\_percent*、多点比例 *several\_points\_percent*；当只有单点定位数据时的参考半径*radius*、参考比例*refer\_percent*。例如用户输入 (0,1,3,0.7)，表示当单点和多点定位栅格数据都存在时，只取多点定位数据，舍弃掉单点数据；当只有单点定位栅格数据时，参考横坐标在 [x-3,x+3]、纵坐标在 [y-3,y+3]共计48个栅格的 Ec/Io值，然后按照70%的权重与30%权重的单点定位数据相加求和。



图 3‑7数据融合算法流程图

栅格增补子模块是通过参考缺失栅格周围的栅格数据来对该缺失栅格进行增补。用户需要输入的参数为参考半径radius，可以计算参考栅格数据的平均值作为该缺失栅格值。例如，参数为3，假设该缺失栅格坐标为(x,y)，则要计算横坐标在 [x-3,x+3]、纵坐标在[y-3,y+3]共计48个栅格的Ec/Io的平均值。

栅格平滑的主要功能是通过用户自行调节参数能够取出栅格杂点的干扰，更加方便的观察到某区域覆盖的整体情况。

该子模块的栅格平滑算法流程如图 3‑8所示，用户需要输入的参数有阈值*gap*、参考半径*radius*、参考比例*refer\_portion*。栅格平滑算法判断一个栅格是否为需要平滑的主要依据是该栅格的Ec/Io值与以其为中心的周围 个栅格 Ec/Io平均值的差值是否大于*gap*，如果大于，则判断该点需要平滑；类似于增补算法，平滑算法平滑也是根据*refer\_portion*参考周围栅格数据的，如果*refer\_portion*为0.8，则表示该栅格取周围 个栅格Ec/Io平均值的80%与本身Ec/Io值的20%作为该栅格平滑后的Ec/Io值。



图 3‑8栅格平滑算法流程图

寻找弱覆盖区域子模块的功能与栅格平滑子模块相对应，栅格平滑子模块展示的是信号覆盖整体态势图，而它展示的是局部的弱覆盖区域的栅格图。它的算法类似于话务负载模块中的寻找高负载区域子模块的算法，将在下文谈到。

## 话务负载模块设计

话务负载模块是对栅格话单中话务量进行统计、分析的模块，从而使话务量可以在数值和空间上呈现。最终反映的结果可以帮助用户更好的把握整个区域的网络负载情况，从而及时作出网络优化调整。

首先，需要有话务负载的栅格展示功能，可以让用户看到话务量在空间上的呈现。另外，该模块要允许用户设定阈值，可以过滤出超出阈值的话务高的栅格，绘制话务高负载区域。

如图 3‑9，话务负载模块需要有数据融合、栅格平滑和寻找高负载区域三个子模块来满足需求。数据融合子模块把单点栅格数据和多点栅格数据中的话单进行融合；栅格平滑子模块通过对栅格数值的平滑处理，来显示话务负载的整体态势；在寻找高负载区域子模块中，设定高负载阈值，通过数据统计和显示模块，重点显示局部的高负载区域。



图 3‑9话务负载模块设计流程

其中数据融合子模块和栅格平滑子模块与信号覆盖模块中的相应子模块类似，这里不再多做解释。



图 3‑10寻找高负载告警算法流程图

如图 3‑10是寻找高负载告警算法流程图，该算法重点是要判断该栅格数据与周围参考栅格的平均值是否大于阈值。用户可以输入的参数有：参考半径*radius*、高负载分析阈值*threshold*。参考半径为计算某个栅格话务量时的参考范围，使用该子模块计算以该栅格为中心、参考半径范围内的栅格区域内的栅格平均话务量。高负载分析阈值为该栅格区域的话务负载报警阈值，如果该区域话务量超出阈值，则判别该栅格为高负载栅格。

## 基于地理的覆盖性能模块设计

### 覆盖性能模块需求分析

前文主要是讲到从信号覆盖和话务负载单方面分析的功能设计，单方面分析虽然能够检测出蜂窝网络的运行状况，但是对于网络优化提出解决方案帮助不大。所以需要通过对两者进行联合分析，进一步定位出网络中出现问题的原因并给出具体的解决措施。

基于地理的覆盖性能模块是要对信号覆盖和话务负载进行基于地理栅格的联合统计与分析。首先要通过联合统计子模块找到同时存在弱覆盖和高负载的“目标栅格”，具体弱覆盖和高负载的阈值应允许用户自行设定，并通过栅格图的形式直观展示目标栅格的分布。同时对目标栅格进行分类，比如分为可以通过调整基站参数可以改善的目标栅格和需要增建基站才能改善的目标栅格，最后对不同类别的栅格进行分析，给出优先需要调整的基站和优先需要增建基站的区域等。

### 覆盖性能模块功能设计

如图 3‑11，覆盖性能模块应有信号覆盖与话务负载联合统计、覆盖性能分析两个子模块来满足需求。

联合统计子模块是用来对某区域的信号覆盖和话务负载做联合统计的子模块，比如它可以统计出在 Ec/Io在某一数值区间的栅格的话务量的分布比例。用户需要输入 Ec/Io的数值区间、话务量分档的数值区间，最后得到反映话务量比例的饼形图。



图 3‑11覆盖性能模块设计流程

通过设置合适的参数，统计子模块可以筛选出用户需要的栅格。设置参数，经过统计子模块，可以得到信号差且话务量多的区域，该区域通常是需要进行网络优化的，在论文中称该区域为目标区域，称该区域中的栅格为目标栅格。

覆盖性能分析子模块可以对目标栅格进行分类，对得到的不同类别栅格采取不同的优化策略，最后输出给出待改善基站的排名和增建基站区域。

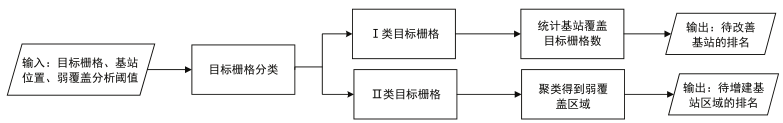


图 3‑12覆盖性能分析算法流程图

如图 3-12是该子模块的覆盖性能分析算法流程图。用户需要自行输入弱覆盖分析阈值，在本论文中作为基站的有效覆盖半径使用。虽然真实情况是基站在各个方向的有效覆盖半径受到周围环境、天线朝向功率等各方面条件的影响，但是应需求方要求，假设基站在各个方向的覆盖半径相同可以更方便实验和测试，所以本论文中所有基站在各方向的覆盖半径统一为。同时输入还有经过联合统计子模块得到的目标栅格数据和从运营商处获得的基站位置数据。



图 3‑13 目标栅格分类

参照基站位置数据，对目标栅格做分类，具体分类示意图如图3-13。在基站范围内的目标栅格，这一类栅格是可以通过调整基站参数、优化基站配置或者调整天线等措施改善的，在该论文中我们统一称这一类目标栅格为Ⅰ类目标栅格。对于在基站范围外的目标栅格，这一类栅格可能因位置比较偏远而不能被有效覆盖到，主要通过增建基站或者放置直放站的方式来改善，我们统一称这一类为Ⅱ类目标栅格。

为了优化Ⅰ类目标栅格，首先对所有基站按照覆盖Ⅰ类目标栅格个数排序，其中排名靠前的基站可以作为优化的重点来解决Ⅰ类目标栅格的覆盖性能问题。而对于Ⅱ类目标栅格，首先对其进行聚类分析，然后对聚类后得到的目标区域采取增建基站的方式来解决其覆盖性能的问题。聚类分析子使用Ⅱ类目标栅格数据作为输入源，通过基本 K均值算法、 DBSCAN等聚类算法，对栅格进行聚类分析，将目标区域按照位置区分成小的目标区域，并对每个区域按照统一的弱覆盖区域评价准则评价排序，其中排名靠前的弱覆盖区域应是优先考虑建站的区域。最后得到的待改善基站排名和增建基站区域排名可以为该地区的网络优化提供有价值的信息参考。

为了对弱覆盖区域进行评价，判断该区域是否有必要增建基站，本文设计了一个弱覆盖区域评价准则。

弱覆盖区域评价准则：

其中表示对聚类后的第个区域的评估值，表示该区域中第个目标栅格所对于栅格的信号质量，小于 0；表示该区域中第个目标栅格所对应栅格的话务量，是用户输入的参数，用来调整栅格信号质量和话务量权重的系数，、是常数，分别是-12dB、10。信号质量越差，越小，此时值越大，代表该区域增建基站的需求越大；话务量越大，越大，此时值越大，同样也代表该区域增建基站的需求越大。

## 本章小结

本章首先给出了该工具的需求分析，详细说明了该工具的功能需求。其次介绍了分析工具的整体框架设计，并详细介绍了各功能模块。接下来介绍了本文的数据的获取。最后对各功能模块的设计进行详细说明分析。

首先详细叙述了显示模块的设计，该模块作为本工具的核心功能之一，能够将统计 或者处理之后的数据以直观地数据统计图表和栅格图的形式展示。然后设计了信号覆盖模块和话务负载模块，信号覆盖模块主要用于处理与信号质量相关的数据，话务负载则用于处理与话务量相关的数据，最后处理的结果通过显示模块做出相应展示，展现出数据在数值区间、时间和空间上的分布关系。最重要的是，根据用户需求，完成了对基于地理的覆盖性能模块的设计，该模块是对之前模块功能的总结和升华，采用联合统计的方式，找到目标栅格，再结合目标区域的基站数据对目标栅格进行分类。对于不同分类的目标栅格采取不同的优化方案。本章介绍各模块方法的设计思想，为后续实现章节提供指导。

# 分析工具的实现

本工具主要包括四个模块：显示模块、信号覆盖模块、话务负载模块、基于地理的覆盖性能分析模块。显示模块主要完成对信号话务的统计和分布信息展示，主要展示方法有数据统计图和栅格图。信号覆盖模块包括信号覆盖的数据融合、栅格增补、栅格平滑、寻找弱覆盖区域四个子模块。话务负载模块包含有数据融合、栅格平滑和寻找高负载区域。基于地理的覆盖性能分析模块有信号覆盖与话务负载联合统计、覆盖性能分析两个子模块。下面将详细说明各模块的实现方法。

## 交互流程设计

该课题的主要任务是提供蜂窝网络信号覆盖和话务特性的分析工具，工具的人机交互力求简洁好用。本文的前端展示页面通过Web 页面实现，Web页面具有跨系统的特性，无论用户工作在哪一个系统下，都可以方便使用该工具。后台使用Java语言实现。良好的页面交互设计是设计该工具的基础，交互过程需要根据功能的需求设计。从需求分析中可知，该工具给用户提供了一系列信号特性分析工具，不同的分析工具应对不同的使用场景。针对同一处理算法，配置不同的参数，建模结果可能不同，因此需要提供参数配置界面，以便用户调节参数，获取最终理想的网络优化。而且为了能够直观展示每一步处理的结果，该工具提供了包含统计图表和栅格图等展示方式。综上所述，工具集的交互流程设计框架如图 4‑1所示:



图 4‑1 交互流程图

图中所示的交互流程主要包括四个界面：数据选择界面、数据统计界面、数据处理界面和数据分析界面。首先用户在数据选择界面通过配置地区和时间参数来选择HBase数据库中不同区间的话单数据，通过Hadoop 分布式处理平台获取所需格式的话单数据。用户通过数据统计页面获取统计信息。同时，用作统计的数据也是后续数据处理和数据分析的数据源。由于用户需求的不同，在数据处理界面中，用户在对处理的参数自行进行配置时，采用不同参数处理之后的结果相差可能也会比较大，通过该界面中的栅格图可以将数据处理结 果直观地展示给用户，用户根据处理结果来调整处理参数从而得到最优结果。最终在数据分析界面中，用户继续配置相应参数进行分析，最后以栅格图的形式展示后台计算完成后的结果。

## 显示模块实现

工具的显示模块很重要，因为它能以图形化的形式直观的显示该工具中间运算的过程和最后的结果。显示模块主要包括两部分：数据统计图和栅格图。

### 数据统计图

数据统计图用于显示信号特性的统计信息，是调用名为Highcharts的JS库对统计数据在Web页面进行渲染并展示。

下面分别展现了曲线图和饼图的实现效果。如图4-2，显示的是某地区Ec/Io 随时间的变化曲线，可以看到08年7月有明显的波谷，猜测此时网络应该需要进行网络优化。



图 4-2 曲线图示例

如图4-3，利用饼形图我们可以直观看到Ec/Io值或者话务量中每个数值区间的比例，其中低于-12的Ec/Io值只有5%，说明该地区覆盖质量较好。

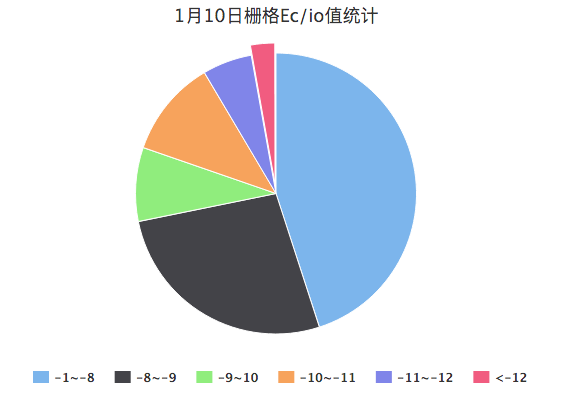


图 4-3 饼形图示例

### 栅格图

栅格图是显示信号特性基于位置的分布信息，以Ec/Io栅格图为例，首先根据待绘制区域的经纬度范围在某地区底图上截取相应的地图，并取出与之对应的栅格数据，根据某地区的栅格数据排序后确定Ec/Io数值区间的染色颜色，最后根据每条数据的栅格编号确定染色的位置，遍历每条数据对区域内的栅格进行染色，最终绘制出Ec/Io栅格图。

input:

dataCity;

dataSet;

regionPoint1;//待绘制区域左下角经纬度坐标

regionPoint2;//待绘制区域右上角经纬度坐标

originPoint;//原点经纬度坐标

output:

gridPng;

begin:

//截取某区域地图

distancePixels = LonglatToPixel(regionPoint1- originPoint);//将坐标转换成像素

areaPixels = LonglatToPixel (regionPoint2- regionPoint1);//将坐标转换成像素

regionImg = cutImage(sourceImg, distancePixels, areaPixels);//截取

gridImg = newImgae(width,height);//新建图片

//确定每种颜色对应的边界

dataTemp = dataCity.sort();

length = dataCity.length();

threshold[1] = dataTemp[length\*0.1];

threshold[2] = dataTemp[length\*0.2];

...

threshold[7] = dataTemp[length\*0.9];

//遍历栅格数据进行染色

n = dataSet.length();

for i = 1 to n

gridPixel = IdToPixel(dataSet[i].gridId);//将栅格数据的ID编号转换成像素

gridColor = GetColor(dataset[i].ecio, threshold);//根据之前计算的边界获得栅格颜色

draw(gridImg, regionImg, gridPixel, gridColor);//对栅格染色,生成新图片

endfor

end

绘制栅格首先要对某地级市某时间段内的Ec/Io或者话务量进行排序，分别按照数据量占比10%、20%、 35%、50%、65%、80%、90% 取出7 个数据作为栅格填充颜色的阈值，从而可以将栅格图染成八种颜色，如图4-4

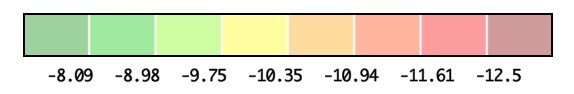


图 4-4 统计某地区栅格数据后得到的信号覆盖栅格图例

之后需要计算好每个栅格对应地图上的像素值，在截取好的某区域地图上按照已经处理好的栅格数据值进行染色，最后生成PNG 图片，在Web 页面展示给用户。如图4-5。



图 4-5 栅格图示例

## 信号覆盖模块实现

信号覆盖模块的实现是对栅格数据进行融合、增补的处理子模块和栅格平滑、寻找弱覆盖区域分析子模块的实现。数据处理是在进行数据分析前通过数据清洗时提高数据可信度的方式。

数据处理模块的难点在于很难找到一种完美的方法可以处理不同场景下的栅格数据，所以该课题在开发处理模块时，从提高该模块的适应度出发，为用户提供多种可调节的方式，保证最后经过参数调节能出现可信度较高的栅格数据。在数据融合子模块中，对单点定位的栅格数据和两三点定位的栅格数据按一定比例进行融合。数据融合算法具体实现的伪代码如下：

input:

dataSet1;//待处理单点定位的栅格数据

dataSet2;//待处理多点定位的栅格数据

gridSet;//该区域所有栅格

ontPointPercent;//当单点和多点定位都有时，单点比例

severalPointsPercent;//当单点和多点定位都有时，多点比例

radius;//当只有单点定位时，参考半径

referPercent;//当只有单点定位时，参考比例

output:

dataProcess;//融合后的栅格数据

begin:

n = gridSet.length();

hashMap1.put(dataSet1.gridId, dataSet1.ecio);

hashMap2.put(dataSet2.gridId, dataSet2.ecio);

for i = 1 to n //遍历该区域的每一个栅格

if(hashMap2.containsKey(i))

if(hashMap1.containsKey(i))) //有单点和多点定位，根据比例确定

gridEcio = hashMap1.get(i)\* ontPointPercent + hashMap2.get(i)\* severalPointsPercent;

else //只有多点定位，只取多点定位值

gridEcio = hashMap2.get(i);

endif

else if(hashMap1.containsKey(i))) //只有单点定位，参考周边

for j = radisu\*(-1) to radius

ecioSum = hashMap2.get(j);

num++;

ecioAverage= ecioSum / num;

gridEcio = ecioAverage\*referPercent + hashMap1.get(i)\*(1-referPercent);

endfor

endif

dataProcess[i] = gridEcio;

endfor

end

栅格增补的主要功能是通过参考缺失栅格周围的栅格数据来对该缺失栅格进行增补，该算法具体实现的伪代码与融合算法中的部分代码类似，这里就不重复介绍了。

栅格平滑的主要功能是通过用户自行调节参数能够更加方便的观察到某区域覆盖的整体情况。具体的实现的伪代码：

input:

dataSet;//待分析的栅格数据

radius;//参考半径

gap;//差值

refer\_portion;//参考比例

output:

dataSmooth;//输出的报警栅格位置

begin:

n = dataSet.length();

for i = 1 to n

ecio = dataSet.get(i).ecio;

for j = radius\*(-1) to radius

ecioCount += dataSet.get(j).ecio;

num++;

endfor

ecioAver = ecioCount/num;

if(math.abs(ecioAver-ecio)>gap)//栅格与周围栅格的Ec/Io差值超过gap，需要平滑

ecio = ecioAver\*refer\_portion+ecio\*(1-refer\_portion);

endif

dataSmooth[i] = ecio;

endfor

end

## 话务负载模块实现

话务负载告警子模块是要计算栅格内的话务量超过某个阈值的栅格位置。与用户沟通得知，用户在分析话务量告警时，需要的常常不是某个告警栅格，更多的某块告警区域，所以计算栅格话务量时通常还参考周围的栅格的话务量。具体实现的伪代码如下：

input:

dataSet;//待分析的栅格数据

radius;//参考范围

threshold;//阈值

output:

callAlert;//输出的报警栅格位置

begin:

n = dataSet.length();

for i = 1 to n

for j = radius\*(-1) to radius

callCount += dataSet.get(j).calls;

num++;

endfor

if(callNum>threshold\*num)

callAlert.add(dataSet.get(i)); //记录超过阈值的栅格

endif

num=0;

endfor

end

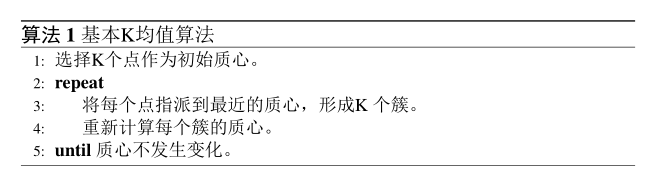
## 基于地理的覆盖性能模块实现

基于地理的覆盖性能模块包含联合统计和覆盖性能分析两个子模块。该模块首先使用信号覆盖和话务负载的联合统计模型找到需要进行网络优化的目标栅格，之后结合该区域的基站位置数据，计算出每个目标栅格距离最近基站的距离，按照用户输入的基站覆盖半径将目标栅格分为有基站覆盖和无基站覆盖的目标点，之后在对不同类别的目标栅格进行不同的分析和优化。覆盖性能分析子模块是本工具的重点，它的好坏直接影响到本工具最后输出的具体优化方案的质量，下面重点讲解覆盖性能分析子模块的实现。

用户在自行输入弱覆盖分析阈值Rw后，覆盖性能分析子模块参照目标栅格与基站之间的距离把目标栅格分为Ⅰ类目标栅格和Ⅱ类目标栅格。

为了优化Ⅰ类目标栅格，首先对所有基站按照覆盖Ⅰ类目标栅格个数排序，其中排名靠前的基站可以作为优化的重点来解决Ⅰ类目标栅格的覆盖性能问题。而对于Ⅱ类目标栅格，首先对其进行聚类分析，然后对聚类后得到的目标区域采取增建基站的方式来解决其覆盖性能的问题。聚类分析子使用Ⅱ类目标栅格数据作为输入源，通过基本K均值算法、DBSCAN等聚类算法，对栅格进行聚类分析，将目标区域按照位置区分成小的目标区域，并对每个区域按照统一的弱覆盖区域评价准则评价排序，其中排名靠前的弱覆盖区域应是优先考虑建站的区域。最后得到的待改善基站排名和增建基站区域排名可以为为该地区的网络优化提供有价值的信息参考。

聚类分析（cluster analysis，亦称为群集分析）是对于统计数据分析的一门技术，在许多领域受到广泛应用，包括机器学习，数据挖掘，模式识别，图像分析以及生物信息[48]。聚类是把相似的对象通过静态分类的方法分成不同的组别或者更多的子集（cluster），这样让在同一个子集中的成员对象都有相似的一些属性[49]，常见的包括在坐标系中更加短的空间距离等\citeup{cluster}。 基本K均值是一种基于中心的聚类技术，表示以空间中k个点为中心进行聚类，对最靠近它们的对象归类[50]。K 均值具体描述形式参见算法1[51, 52]。



基本K均值算法主要是为了能够大的目标区域划分成k个小的目标区域，用户在输入参数点击运行后，该子模块按照K均值算法对目标栅格进行聚类，最后输出k个聚类区域。

覆盖性能分析子模块使用基本K均值算法对II类目标栅格进行聚类分析，找到k个簇，即为k个小的目标区域；并且通过一定的弱覆盖区域评价准则对每个目标区域的建站需求做评估。

用户可以输入的参数有：初始质心个数k。k个初始质心代表经过该子模块计算，最后会将所有目标栅格划分成k个目标区域。具体实现的伪代码如下：

input:

dataSet;//待分析的栅格数据

k;//初始质心个数

output:

cluster;//K均值模型处理后的簇集

centerEcio;//簇集中心

begin:

//初始化k个质心

for i = 1 to k

temp = random.nextInt(dataSet.length());

center[i] = dataSet.get(temp);

endfor

while(true)

clusterInit(); //初始化k个簇集

clusterSet(); //将当前栅格放到最小距离中心相关的簇中

m++;

errorSquare[m] = countRule(); //计算误差平方和

if(errorSquare[m]== errorSquare[m-1])

break;

endif

//根据新计算的簇集计算新的簇质心

for i = 1 to k

n = cluster.get(i).length();

for j = 1 to n

newCenterSum += cluster.get(i).get(j);

center[i] = newCenterSum / n;

//考虑每个栅格Ecio值，重新计算质心

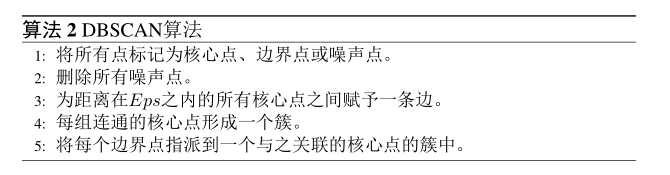
centerEcio = getCenterEcio(center);

endfor

endfor

end

基于密度的聚类寻找被低密度区域分离的高密度区域。DBSCAN算法是一种简单、有效的基于密度的聚类算法[50]。DBSCAN 算法具体描述形式参见算法2[53]。



在通过DBSCAN算法聚类，分离得到小的目标区域后，类似基本K均值，按照对每个目标区域中目标栅格数目的排序，取出排名前k个目标区域，计算这k个目标区域的中心。DBSCAN算法相对来说具有一定的抗噪声能力，能够抵抗离群点的干扰，并且能够处理任意大小的和形状的目标区域，但是如果目标区域中目标栅格的密度变化过大时，可能得到的结果相对较差。

对于DBSCAN算法，较重要的是DBSCAN 本身参数Eps和MinPts的确定问题。Eps、MinPts主要用于寻找核心点，如果某点在Eps距离内点的个数超过阈值MinPts，则判断该点为核心点。通常，DBSCAN算法取MinPts为4。计算所有点到第4个最近邻点的距离，以递增次序将他们排序，然后绘制排序后的值，则我们会看到4-距离的急剧变化，合适的Eps值对应于其急剧变化的4-距离点。

## 本章小结

本章主要探讨了蜂窝网络覆盖性能分析工具中各个模块的具体实现，并且通过伪代码的形式对每个技术难点做了详细的说明。

其中交互流程设计从宏观上说明了该工具的整体交互流程，显示模块中统计图的实现是通过调用Highcharts库在Web页面渲染实现的，栅格图则是确定每个栅格在图片像素中的位置，对底图的每个像素描点染色实现的。信号覆盖处理子模块的实现主要是通过参考周围栅格对每一个栅格数据进行重新评估的过程。话务负载告警子模块不仅要考虑栅格本身的话务量，还要考虑周围栅格的话务量。基于地理的覆盖性能需求分析模块需要对基站位置做栅格化处理，根据基站与目标栅格的距离对目标栅格做分类；其中对于Ⅱ类目标栅格采用聚类算法得到弱覆盖区域，同时设计一个弱覆盖区域评价准则对得到的弱覆盖区域评价排名，排名靠前的应优先考虑建站。

# 分析工具的测试

## 测试平台与实验数据

本实验采用的测试平台配置如下表5.1所示：测试平台由12台机器组成，其中10台机器组成话单处理集群，1台机器作为数据备份服务器，一台客户机作为测试客户端。

表 5.1 测试平台配置表

该组实验采集到的某电信运营商的话单数据，时间长度为2 个周，覆盖范围为某个省级市，经Hadoop处理平台计算后得到846505条共计13MB 的单点定位栅格数据和621824条共计9.6MB 的多点定位栅格数据。同时采集到的数据还有该省级市的基站位置信息。

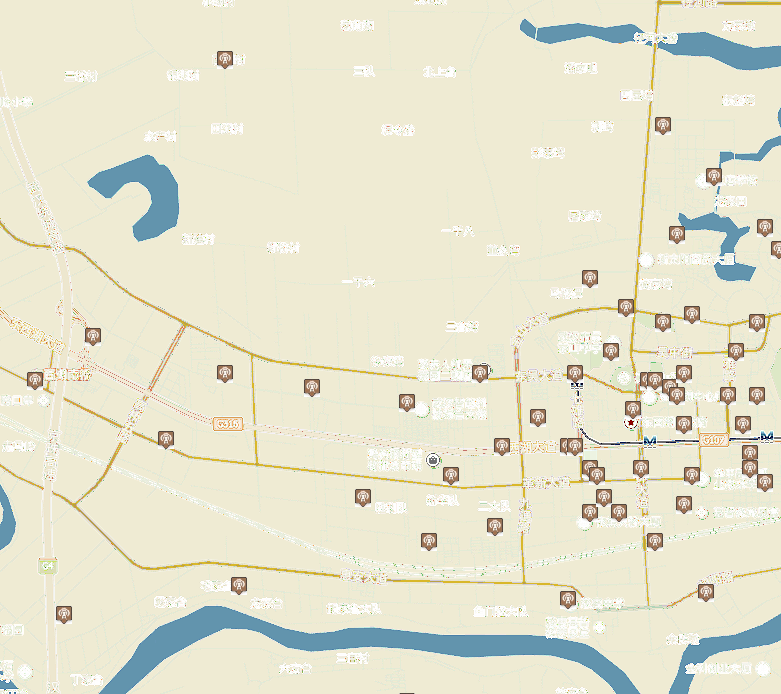


图 5-1 样本区域基站图

因为省市地区的面积过大，不利于描述，该实验接下来部分选择该地区中一块区域作为样本区域继续分析，如图5-1本区域包含有10476个栅格，建有55个基站，其中图中褐色图标标注的是该区域的基站，基站呈现左疏右密的形势。该区域比较典型，其中图片右侧有商业广场、区政府和地铁站，人流量密集；底部有河流和学校；左上侧有村庄，人流量较少。

## 信号覆盖的统计与分析

信号覆盖的统计与分析模块是用来统计、处理、分析和显示信号覆盖在时间和空间上分布的模块。该模块的测试主要测试以下四个功能：数据融合、栅格增补、栅格平滑、寻找弱覆盖区域。实验的输入数据是之前采集到的话单数据计算生成的单点和多点定位数据，输出以栅格图的形式呈现。

### 数据融合

信号覆盖统计模型中的栅格图可以反映信号覆盖在区域中位置分布。在统计该地区全部栅格Ec/Io 数值比例之后，可以得到信号覆盖栅格图例如图5-2。前端的Web页面上输入相关时间和区域参数后，经过后台的数据获取和栅格图绘制，可以得到如图5-3的栅格图。

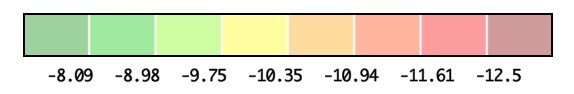


图 5-2 信号覆盖栅格图例

数据融合的目的在于可以把单点定位和多点定位的数据融合在一起，从而避免单基站定位算法不准确性和多点定位栅格数据不足的情况。如在本文第三章中所述一样，我们会发现单点点定位栅格图在市郊存在明显的圆环现象，如图5-3(a)所示；会在多点定位栅格图中发现栅格缺失的现象，如图5-3(b)所示。

我们采用数据融合算法对两种栅格数据进行数据融合，采用的参数为(0,1,3,0.7)。参数的具体意义请参考第三章。融合后的得到的栅格图如图5-3(c)所示。比较图5-3三张图，该模型已经很好地避免了单点的圆环现象和多点的栅格缺失现象。

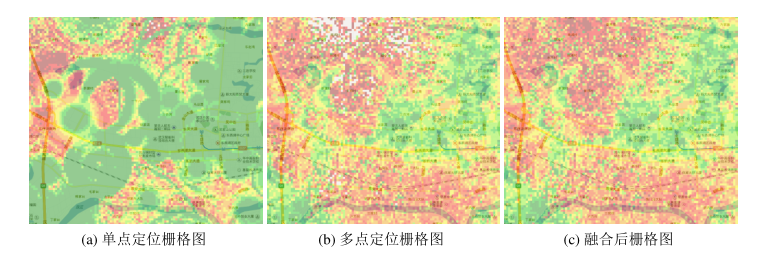


图 5-3 数据融合结果展示

### 栅格增补

栅格增补是对经过数据融合算法融合后的栅格数据进行补齐。如图5-3(c) 所示，该区域栅格缺失很少，只有2 个缺失栅格。经过栅格增补，该区域全部10476 个栅格都已经有数据了。其中一个栅格的增补结果如图5-4，图(a)圈中白色栅格代表此栅格缺失，图(b) 中白色栅格已经通过栅格增补算法补齐。

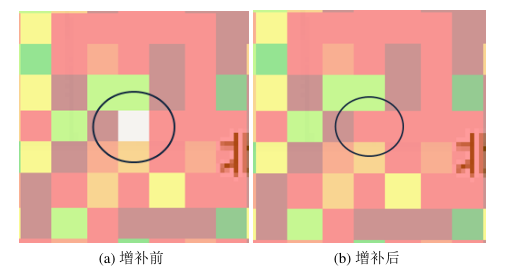


图 5-4 栅格增补结果展示

### 栅格平滑

栅格平滑可以通过调整平滑参数，达到平滑栅格图中颜色的效果，这样可以更好地观察某区域信号覆盖的整体情况。为了能够更好地测试平滑的效果，我们设计了3组实验。



图5-5 阈值对平滑效果的影响

第一组实验参数：(0,3,1)、(1,3,1)、(2,3,1)，参数的意义分别是阈值gap、 参考半径Rw、 参考比例refer\_portion。其中第一个参数gap 是该栅格的Ec/Io 值与以其为中心的周围(radius+1)^2-1个栅格Ec/Io平均值的差值，如果大于gap，则判断该栅格需要调用平滑算法进行平滑；反之则不需要。类似于增补算法，平滑算法也是根据refer\_portion的值参考周围栅格数据的。得到的效果如图5-5，其他参数不变，随着阈值的减少，平滑算法能够平滑到更多的栅格。到阈值为0时可以平滑到所有栅格，这时能够更加明显地反映信号覆盖的整体形势。

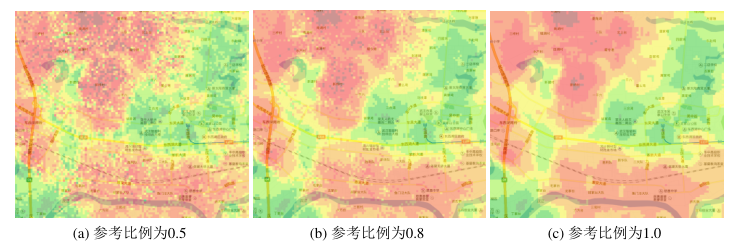


图 5-6 参考比例对平滑效果的影响

第二组实验参数：(0,3,0.5)、(0,3,0.8)、(0,3,1.0)，得到的效果如图5-6，其他参数不变，随着参考比例的提升，栅格图中的杂点越来越少，视线干扰会变少。当参考比例为1.0时，每个栅格完全参考周围栅格的平均值作为该栅格的Ec/Io 值，此时在栅格图中几乎看不到杂点。

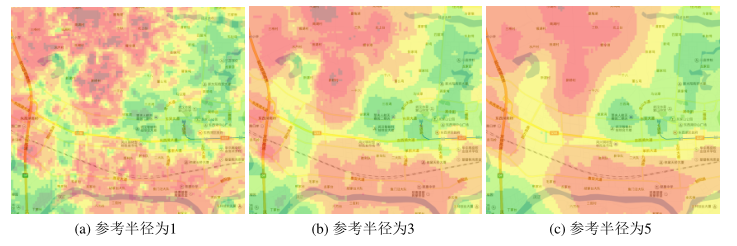


图 5-7 参考半径对平滑效果的影响

第三组实验参数：(0,1,1)、(0,3,1)、(0,7,1)，得到的效果如图5-7，参考半径变大，其他参数不变，会使得色块的边界更加圆滑，更容易辨识出不同数值区间的Ec/Io 值的分布范围。当参考半径为7时，每种颜色所代表的区域连为一体，Ec/Io值的分布范围能够更加明显地识别出来。

### 寻找弱覆盖区域

寻找弱覆盖区域模型 可以反映信号质量差的区域，如图5-8 (a)，该图中Ec/Io 低于-12 的区域主要集中在左上部，可以看到左上部只有一个基站，而在基站密集的右边则信号质量偏好；如图5-8 (b)，Ec/Io低于-13的栅格比-12的栅格已经少很多了；如图5-8 (c)，Ec/Io低于-14的栅格已经很少了，主要在左上部稀疏分布。

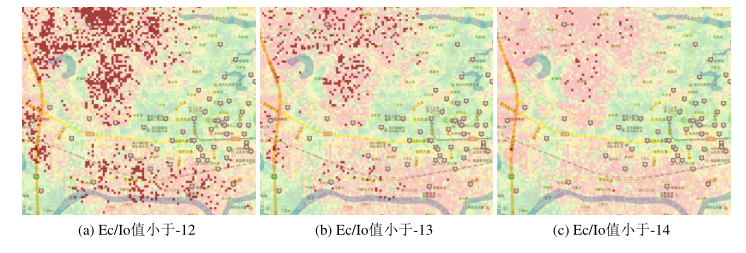


图 5-8 寻找弱覆盖区域栅格图

## 话务负载的统计与分析

### 话务负载栅格展示

根据整个地区的话务量在数值上的分布比例，按照数据量占比10%、20%、35%、50%、65%、80%、90% 取出7个数据作为话务负载的栅格填充颜色的阈值，话务负载栅格图例如图5-9。

话务负载统计模型的栅格图可以反映话务量在位置上的分布情况，类似信号覆盖统计模型，在输入时间和区域参数后，得到的栅格图如图5-10。



图 5-9 话务负载栅格图例

### 寻找高负载区域

在话务负载告警模型中，本实验测试了3组参数(3,500)、(3,1500)、(3,2500)。 得到的高负载栅格图如图5-11。我们可以观察到话务量大的区域主要集中在图片右侧，右侧区域有地铁站、广场和市政府等人流量密集场所，并且右侧还建有大量基站来应对应大话务量，这都与话务负载分析的结果相吻合。通过统计，我们也可以得到表5.2所示为各数值区间话务量的占比。

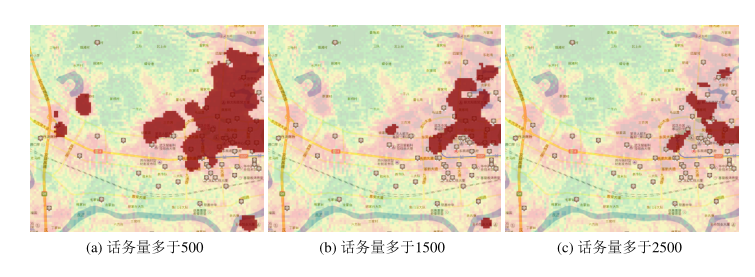
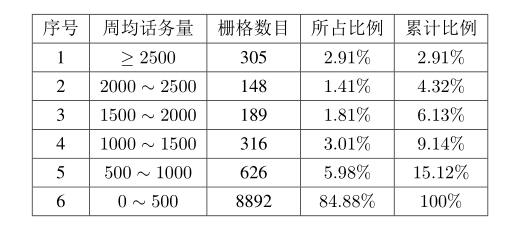


图 5-11 话务负载告警栅格图

表 5.2 各数值区间话务量的分布表  


## 基于地理的覆盖性能统计与分析

经过数据融合和栅格增补模型的处理，该区域栅格数据的圆环、缺失的问题已经得到了很好的解决。实验测试的下一步工作是要将经过处理后的数据导入基于地理的覆盖性能分析模块。

### 信号覆盖与话务负载的联合统计

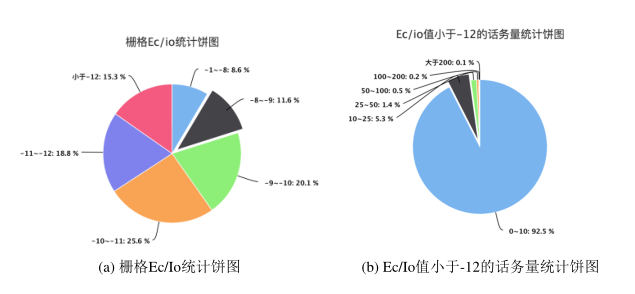


图 5-12 统计饼图

本实验使用联合统计模型对经过覆盖分析和话务分析后的栅格数据进行联合统计。首先我们要对整个地区Ec/Io在每个数值区间的分布情况有所了解。如图5-12 (a)所示，整个地区的Ec/Io值在[-10,-11]占的比例最高为25.6%，另外有15.3%的栅格Ec/Io值小于-12dB。根据经验，移动通信网络覆盖率考核值是大于-12dB，为了能够了解这部分区域的通话量组成，我们继续对该区域的栅格数据进行话务量统计，统计其每个栅格的话务量的比重。如图5-12 (b)所示，我们可以发现该区域有92.5%的栅格话务量小于10个，也就是说两周内在这其中每个栅格的通话数量小于10 个。值得注意的是，仍然有7.5%的栅格话务量是高于10 个的，这一部分栅格应该作为网络优化的重点。在本实验中，我们称这些Ec/Io值小于-12dB但话务量多于10个的栅格为目标栅格，称这些栅格所围成的区域为目标区域。

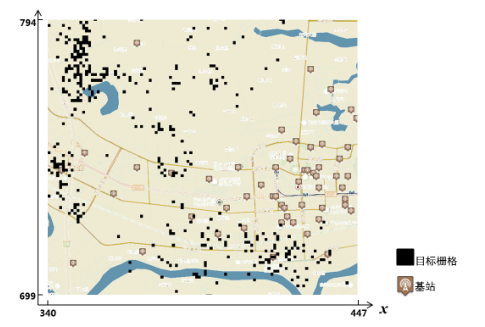


图 5-13 “目标栅格”位置

为了能够了解目标区域的具体位置，我们继续使用绘制栅格功能绘制栅格图。如图5-13，该栅格图标注出了Ec/Io 小于-12dB且话务量大于10的7.5% 的目标栅格在样本区域中的位置。

### 基于地理的覆盖性能分析

该子模块对上文中得到的目标栅格继续进行分析，目的是通过对目标栅格的分类找到针对不同类型目标栅格的优化策略。用户设置不同的Rw，会得到不同的结果。为了测试该子模块性能，该实验分别输入Rw为1km、2km、3km。

#### 基站的弱覆盖分析阈值为1km

设置Rw为1km，经过计算，得到目标的分类结果如表5.3所示，在总数为378 的目标栅格中，其中Ⅰ类目标栅格个数为165，占总数的43.7%，Ⅱ类目标栅格个数为213，占总数56.3%。

表 5.3 Rw为1km 时，“目标栅格”分类表



目标栅格在地理空间中的位置如图5-14 (a)，其中棕色的图标表示基站的位置，棕色的栅格点表示Ⅰ类目标栅格，而黑色的栅格点表示Ⅱ类目标栅格。从图中可以看出Ⅱ类目标栅格主要集中在图5-14(a)的左上方和左边的中间区域。下面主要是对于Ⅱ类目标栅格的实验分析。

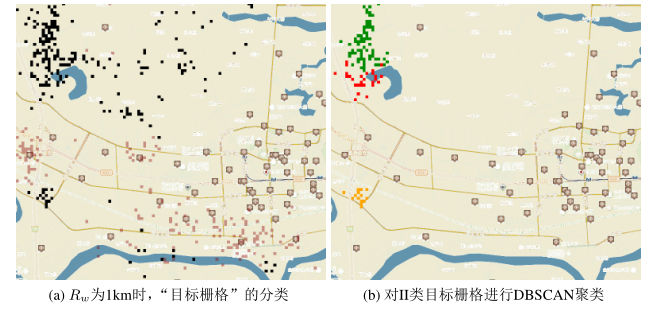


图 5-14 radius = 1km 时，“目标栅格”分析

如图5-14 (a)所示，目标栅格在栅格图上呈现出稀疏不同的分布，如果采用基本K 均值聚类得到的弱覆盖区域会因为离群点影响最终的结果。DBSCAN 算法是一种简单、有效的基于密度的算法，可以比较有效排除离群点的干扰，根据目标栅格的分布，选用DBSCAN 模型可能会使一个不错的选择。

表 5.4 R w 为1km 时，三个弱覆盖区域基本信息列表



选取DBSCAN参数Eps=4和MinPts=4，经过计算，得到3 个弱覆盖区域基本信息统计表如表5.4，覆盖目标栅格个数代表在半径为1km 范围内目标栅格的个数，它反映了每个中心如果建设基站所能够覆盖到的目标栅格的数目；同理话务量。由于此时基站覆盖面积为2.1\*2.1km，对应C0 取1，该模型导入弱覆盖区域评价准则计算得到每个弱覆盖区域的评估值。评估值是通过如下的弱覆盖区域评价准则得到的，它综合了目标栅格的信号强强度和话务量，反映了每个弱覆盖区域增建基站的迫切程度。通过比较发现，虽然1号弱覆盖区域覆盖目标栅格个数少于2号区域，但是由于1号的话务量多，1 号的评估值仍大于2号，因此可以判断在1号区域内增建基站优化周围目标栅格的需求更加迫切。

#### 基站的弱覆盖分析阈值为2km

设置Rw为2km，该模型首先根据基站的覆盖半径对所有的目标栅格进行分类。如表5.5所示，在总数为378的目标栅格中，其中Ⅰ类目标栅格个数为314，占总数的83.1%，Ⅱ类目标栅格个数为64，占总数16.9%。

表 5.5 R w 为2km 时，“目标栅格”分类表



目标栅格在地理空间中的位置如图5-15，其中棕色的图标表示基站的位置，棕色的栅格点表示有基站覆盖目标栅格，而黑色的栅格点表示Ⅱ类目标栅格。从图中可以看出Ⅱ类目标栅格主要集中在图5-15的左上方和上方的中间区域，在这里基站分布很少，而在基站分布密集的下方的目标栅格都能被基站覆盖到。

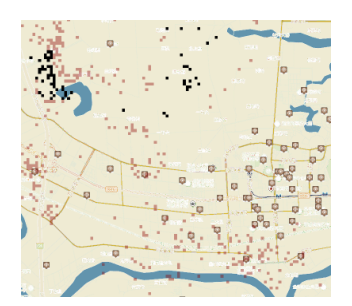
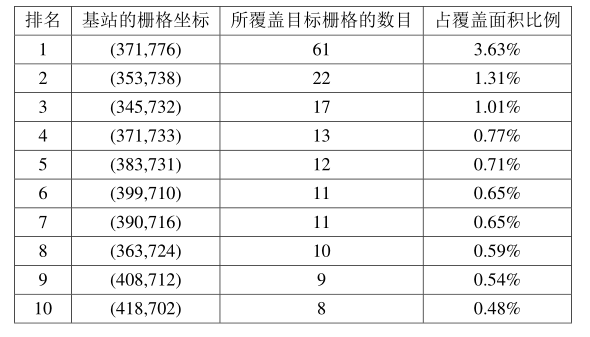


图 5-15 R w 为2km时，“目标栅格”的分类

对于所有Ⅰ类目标栅格，该模型通过统计每个基站所能覆盖到的目标栅格的数目，在区域中所分布的55个基站中，找到覆盖目标栅格个数最多的Top10 基站。得到的数据如表5.6，其中排名第1 的基站所覆盖的目标栅格个数为61，由于每个目标栅格都是100m\*100m 的栅格，因此可以在4.1km\*4.1km的覆盖面积中计算得到所有覆盖到的目标栅格占基站覆盖面积的比例，该基站的目标栅格占比为3.63%。

利用显示模块可以绘制出Top10 基站的分布图，如图5-16 (a)所示，橙色图标表示Top10基站的位置。在网络优化理论中，这些Ⅰ类目标栅格之所以会有弱覆盖问题，通常是由于周围基站的影响、天线方向倾角的问题、接收机灵敏度、周围环境的影响等等。一般情况下，可以通过调整天线方向角、增大基站发射功率、提高天线倾角、增加第四扇区等等手段去解决。

表 5.6 R w 为2km 时，覆盖目标栅格个数最多的Top10 基站



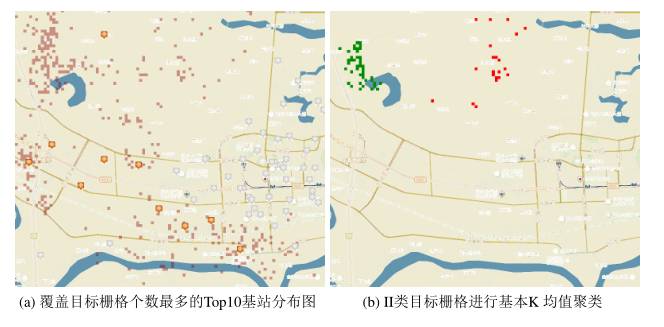


图 5-16 R w 为2km时，“目标栅格”分析

对于Ⅱ类目标栅格，这些通常是属于位置比较偏远基站覆盖不到的地方，如果区域面积比较大，可以通过增建基站的方式解决；而如果面积较少且通话量不是很多，可以考虑建设直放站的方式解决。如图5-16(b)所示，这是所有基站未覆盖到的目标栅格分布图，通过基本K均值算法对这些目标栅格进行聚类。

基本K均值算法适用于基于中心分布的目标栅格的区域，所以每个区域的点到区域中心的距离会比到任何其他区域的中心距离更近。K均值算法的参数k 代表有k个初始质心，即最后会聚类形成k个区域。如图5-16(b)所示，取参数K为2，通过聚类，得到2 块弱覆盖区域。

表 5.7 R w 为2km 时，两个弱覆盖区域基本信息列表



经过计算，得到2个弱覆盖区域基本信息统计表如表5.7，其中中心点代表每个中心点的栅格ID；目标栅格个数代表在半径为2km 范围内目标栅格的个数，它反映了每个中心如果建设基站所能够覆盖到的目标栅格的数目；话务量代表2周时间长度内在半径为2km范围内每个栅格的平均话务量，它反映了每个中心基站所覆盖范围的负载程度，话务量越高，代表在此中心建站所能够影响到的通话越多；通过比较发现，在序号为1的弱覆盖区域即图中红色栅格区域的评估值高于序号为2的弱覆盖区域，说明在1号区域增建基站更加迫切。

#### 基站的弱覆盖分析阈值为3km

设置基站的覆盖半径为3km，经过计算，得到目标的分类结果如表5.8所示，其中有基站覆盖目标栅格个数为377，占总数的99.74%，Ⅱ类目标栅格个数为1，占总数0.26%。

表 5.8 Rw为3km 时，“目标栅格”分类表



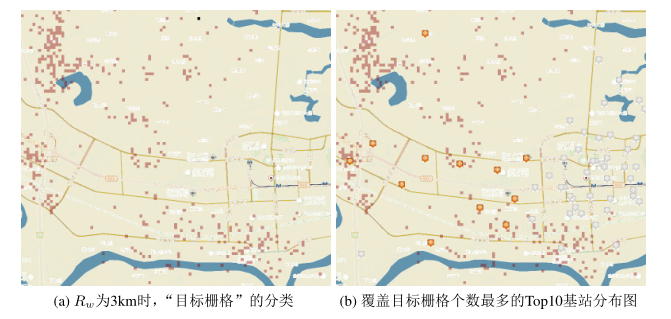
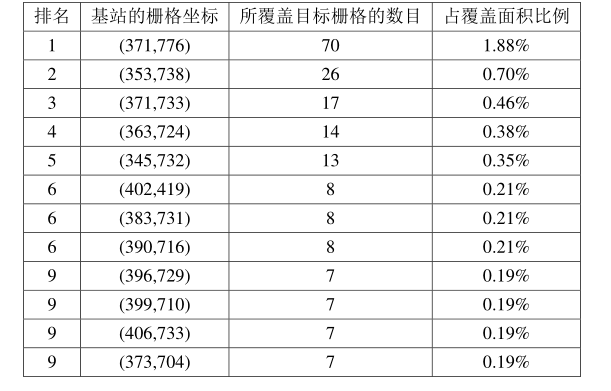


图 5-17 “目标栅格”分析

目标栅格在地理空间中的位置如图5-17(a)，从图中可以看出几乎所有目标栅格都被基站覆盖到。下面主要是对于有基站覆盖目标点的实验分析。

对于所有Ⅰ类目标栅格，该模型通过统计每个基站所能覆盖到的目标栅格的数目，在区域中所分布的55个基站中，找到覆盖目标栅格个数最多的Top10 基站，其基站的地理分布如图5-17(b)。Top10 基站具体数据如表5.9，与表5.6相比，在Top10基站中有8个相同，这些基站可以作为网络优化的重点去处理。

表 5.9 R w 为3km 时，覆盖目标栅格个数最多的Top10 基站



## 本章小结

本章通过实验测试的方式对工具集中的功能进行了测试，并且根据不同的模块采用不同的方式进行评估。在信号覆盖的统计与分析模块中，着重测试了覆盖处理模型中，讨论了不同参数对于结果影响，同时针对不同的栅格数据具体问题具体分析，采用合适的参数才能获得最好的结果，最后对处理结果进行统计，印证模块功能的实现情况。在话务负载的统计与分析模块中，则具体测试了话务负载的统计和分析模型，并且结合实际对分析的结果做出判断。基于地理的覆盖性能的统计与分析模块是该实验章节测试的重点。首先在联合统计子模块中，着重统计整个地区Ec/Io 和话务量在不同数值区间的整体分布情况，并且采用Ec/Io 和话务量联合查询的方式，定位网络优化的重点——目标栅格；只有在覆盖性能分析子模块中，分别设置Rw为1km、2km 和3km，测试不同Rw值对最后度最终得到的优化方案的影响。

测试结果表明，各模块均能通过测试，用户可以针对自己的需求设计参数，得到预期的结果，基本满足了设计需要，说明本工具具有一定的实用价值。

# 总结与展望

## 总结

本文首先对移动通信网络的发展现状做了简要的概述，提出随着4G用户数的不断增多和用户行为的变化等，移动运营商在网络优化方面存在不小的压力，提出了有关网络分析和优化的创新方法。本论文取得的进展如下：

1) 本文实现了对于话单数据的存取和计算的功能。在Hadoop平台上，通过定位算法，可以每月对几十亿条话单记录做计算生成栅格数据。

2) 本文设计并实现了一种蜂窝网络信号覆盖和话务负载基于地理位置的展示方法。使用这种方法生成的栅格图可以帮助用户直观地分析信号覆盖和话务负载。同时，用户通过设置不同的参数，可以通过较低栅格染色透明度的方式着重突出显示某些栅格。

3) 本文设计并实现了对于蜂窝网络特性分析的功能。该工具集既有对信号覆盖的分析，也有对话务负载的分析，同时也可以对两者联合分析。工具集的功能较为全面，同时允许用户通过前端界面设置相应的参数对分析过程进行控制，可以满足用户的不同需求，具有良好的实用性。

4) 本文按照需求方的要求设计并实现了根据栅格数据提供蜂窝网络覆盖性能改善建议的功能。对于联合分析查找到的目标栅格，根据距离基站的远近采取不同的优化改善策略。对于基站不能有效覆盖的II 类目标栅格，通过聚类分析的方法对栅格进行聚类找到需建基站的弱覆盖区域，并且设计了弱覆盖区域评价准则用于对增建基站的需求进行评估；对于基站能够有效覆盖到的I类目标栅格，统计该区域基站覆盖栅格数目的排名，定位引起网络弱覆盖的基站。用户可以设置不同的弱覆盖分析阈值，来控制对于目标栅格的分类。

综上所述，本工具设计并实现了一种基于地理栅格的蜂窝网络覆盖性能的分析工具，用户可以方便的输入参数对分析过程进行控制，并且最终结果通过栅格图和统计图的形式直观展示，方便了用户网络分析和优化的流程，最终得到的结果和数据可以作为最终网络优化的参考。

## 展望

由于本人的精力以及研究水平有限，另外受限于所获得的实验数据，本文中的一些研究工作还有许多不足，具体如下：

本论文虽然完成了蜂窝网络覆盖性能在空间维度上的分析，并且对不同类别的目标栅格数据提供了不同的优化建议。但是如果能够结合地理栅格在时间维度上的分析结果，本文所提出的优化建议就具有了更高的可信性。

另外本文在考虑基站距离对于目标栅格的影响时，假设基站在各个方向上的有效覆盖距离具有各向同性以及不同位置的基站的有效覆盖半径相同。其实受限于具体的地形地貌、天线的高度倾角以及话务负载、越区覆盖等问题，实际情况与实验的理想情况会存在不同。如果能够从运营商那拿到每个基站在各个方向的覆盖半径，则能够对目标栅格进行更加准确的分类。

目前，本文仅从网络信号覆盖和话务负载两个方面分析给出了相应的网络优化建议，并没有从移动台的接入性能和移动网络的保持性能去考虑网络的优化问题，同时也缺少对于网络数据传输速率方面的分析和优化。这些都是该工具以后可以改进和提高的方向。

致谢

三年时光匆匆，光阴像武汉边上的长江水一直不断地流淌，无论你是否注意到它。再过一个月，无论我有多少的不舍，都要离开华科和武汉了，离开春天烦人的梧桐雨和中心操场刺眼的日光灯，离开光谷施工的日夜轰鸣和森林公园的鸟语花香，离开篮球场上开心的奔跑和刘老师语重心长的教导……感谢华科三年时光带给我的挫折、成长、感动和快乐！

感谢我的导师许炜老师，许老师给我提供了一个非常好的学习成长环境，在南一楼323实验室学习工作的半年时间里，他让我时刻充满着紧迫感。特别感谢刘威老师给予我的指导和帮助，刘老师为人正直、工作严谨、待人宽厚，能够在研三期间蒙您指导，真是人生一件幸事，同时也感谢徐晶老师在日常生活中给予的帮助和建议。

感谢同级的小伙伴张鸿宇、卢希和陈骏，在写论文时候我们互相鼓励、互帮互助。感谢同在一个电信项目组、一起并肩完成无数挑战的师弟易江南、杨磊、李文杰，除了在实验室，我们一起打球的时光也同样难忘。感谢实验室的所有师弟师妹们，正是我们一起，才使得WST大家庭如此这般温暖。

最后，感谢我的家人在我求学过程中给予我的爱与支持，未来我会创造更多的奇迹！

参考文献

[1] 邹铁刚, 刘建民, 张明臣. 移动通信网络优化技术与实践. 第1版. 北京: 清华大学出版社, 2015.

[2] 吴松, 何照东, 葛海平. 面向3G的无线网络优化体系. 邮电设计技术, 2007, 11: 10-14

[3] 熊华伟. 基于Gis的移动通信网络质量监控与分析系统研究. 信息通信, 2012, 04: 129-130

[4] 黎宏剑, 刘恒, 黄广文 et al. 基于Hadoop的海量电信数据云计算平台研究. 电信科学, 2012, 08: 80-85

[5] Network T R A. Study On Minimization of Drive-Tests in Next Generation Networks;(Release 9). 2015.

[6] Holma H, Toskala A. LTE Advanced: 3GPP Solution for IMT-Advanced. 第John Wiley & Sons, 2012.

[7] Riihijärvi J, Mähönen P. Estimating Wireless Network Properties with Spatial Statistics and Models. In, ed. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2012 10th International Symposium on. IEEE, 2012. 331-336

[8] Cressie N. Statistics for Spatial Data. 第John Wiley & Sons, 2015.

[9] Dejonghe A, Van Wesemael P, Pavloski M et al. Flexible and Spectrum Aware Radio Access through Measurements and Modelling in Cognitive Radio Systems. 2011.

[10] Galindo-Serrano A, Sayrac B, Ben Jemaa S et al. Automated Coverage Hole Detection for Cellular Networks Using Radio Environment Maps. In, ed. Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt), 2013 11th International Symposium on. IEEE, 2013. 35-40

[11] Galindo-Serrano A, Sayrac B, Ben Jemaa S et al. Harvesting MDT Data: Radio Environment Maps for Coverage Analysis in Cellular Networks. In, ed. Cognitive Radio Oriented Wireless Networks (CROWNCOM), 2013 8th International Conference on. IEEE, 2013. 37-42

[12] Mitola J. Cognitive Radio---An Integrated Agent Architecture for Software Defined Radio. 2000,

[13] Alaya-Feki A, Ben Jemaa S, Sayrac B et al. Informed Spectrum Usage in Cognitive Radio Networks: Interference Cartography. In, ed. Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on. IEEE, 2008. 1-5

[14] Alaya-Feki A B H, Sayrac B, Jemaa S B et al. Interference Cartography for Hierarchical Dynamic Spectrum Access. In, ed. New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on. IEEE, 2008. 1-5

[15] Cressie N, Johannesson G. Fixed Rank Kriging for Very Large Spatial Data Sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2008, 1: 209-226

[16] Braham H, Ben Jemaa S, Sayrac B et al. Coverage Mapping Using Spatial Interpolation with Field Measurements. In, ed. Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on. IEEE, 2014. 1743-1747

[17] Braham H, Ben Jemaa S, Sayrac B et al. Low Complexity Spatial Interpolation for Cellular Coverage Analysis. In, ed. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on. IEEE, 2014. 188-195

[18] Grewal M S, Weill L R, Andrews A P. Global Positioning Systems, Inertial Navigation, and Integration. 第John Wiley & Sons, 2007.

[19] Whitepaper. An Overview of LTE Positioning. In, ed2016.

[20] Palaios A, Jagadeesan S, Perpinias N et al. Studying and Mitigating the Impact of GPS Localization Error On Radio Environment Map Construction. In, ed. Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on. IEEE, 2014. 258-263

[21] Braham H, Jemaa S B, Fort G et al. Spatial Prediction Under Location Uncertainty in Cellular Networks. arXiv preprint arXiv:1510.03638, 2015,

[22] Delyon B, Lavielle M, Moulines E. Convergence of a Stochastic Approximation Version of the EM Algorithm. Ann Stat, 1999, 94-128

[23] 郑伟. 栅格化数据在移动通信网络优化中的应用研究:[ 硕士学位论文]. 北京邮电大学, 2015.

[24] 方裕, 周成虎, 景贵飞 et al. 第四代Gis软件研究. 中国图象图形学报, 2001, 09: 5-11

[25] 刘学锋, 李先华, 何幼斌. 地理信息系统在通信领域的应用研究进展. 上海大学学报(自然科学版), 2007, 04: 389-393

[26] 周成国, 范玉山, 赵修涛. 基于Gis技术的沂蒙山区无线通信基站选址与网络优化. 城市勘测, 2005, 06: 13-15

[27] 宋海营. 城市移动通信基站的建站分析和选址方案设计. 中国无线电, 2013, 10: 55-58

[28] 方黎, 于海波. 基于GeoDatabase的管线数据库建库若干问题研究. 城市勘测, 2005, 06: 22-25

[29] Xie M, Esaki T, Zhou G. GIS-based Probabilistic Mapping of Landslide Hazard Using a Three-Dimensional Deterministic Model. Nat Hazards, 2004, 2: 265-282

[30] 章孝灿, 潘云鹤. Gis中基于“栅格技术”的栅格数据矢量化技术. 计算机辅助设计与图形学学报, 2001, 10: 895-900

[31] Herath P, Krzymien W, Tellambura C. Coverage and Rate Analysis for Limited Information Cell Association in Stochastic-Layout Cellular Networks. 2015,

[32] 陈威兵, 何松华, 彭曙光. 移动通信系统. 第1版. 北京: 清华大学出版社, 2010.

[33] Makanjuola N T, Shoewu O O, Akinyemi L A et al. Comparative Analysis of GSM Network and IS-95 CDMA Network Using Signal Strength. The Pacific Journal of Science and Technology, 2015,

[34] Ruohonen J. Method for Reducing Power Consumption of a Mobile Station and a Mobile Station. In, edGoogle Patents, 2002.

[35] Sampath H, Gore D A, Teague E H. Digital and Analog Power Control for an OFDMA/CDMA Access Terminal. In, edGoogle Patents, 2012.

[36] Yang J, Lee W C, Shin S. Design Aspects and System Evaluations of IS-95 Based CDMA Systems. In, ed. Universal Personal Communications Record, 1997. Conference Record., 1997 IEEE 6th International Conference on. IEEE, 1997. 381-385

[37] Gilhousen K S, Padovani R, Wheatley Iii C E. Method and Apparatus for Controlling Transmission Power in a CDMA Cellular Mobile Telephone System. In, edGoogle Patents, 1991.

[38] Viterbi A M, Viterbi A J. Erlang Capacity of a Power Controlled CDMA System. 1993,

[39] Sampath A, Mandayam N B, Holtzman J M. Erlang Capacity of a Power Controlled Integrated Voice and Data CDMA System. In, ed. Vehicular Technology Conference, 1997, IEEE 47th. IEEE, 1997. 1557-1561

[40] Lee J S, Miller L E. CDMA Systems Engineering Handbook. 第Artech House, Inc., 1998.

[41] 李俊. Gsm系统中的移动定位技术研究:[ 硕士学位论文]. 国防科学技术大学, 2002.

[42] Cong L, Zhuang W. Hybrid TDOA/AOA Mobile User Location for Wideband CDMA Cellular Systems. Wireless Communications, IEEE Transactions on, 2002, 3: 439-447

[43] 刘全胜. 基于Gsm系统的移动台定位方法研究:[ 硕士学位论文]. 大连海事大学, 2004.

[44] Foundation T A S. What is Apache Hadoop. In, ed2016.

[45] 郝树魁. Hadoop HDFS和MapReduce架构浅析. 邮电设计技术, 2012, 07: 37-42

[46] Foundation T A S. HDFS Architecture Guide. In, ed2013.

[47] Foundation T A S. MapReduce Tutorial. In, ed2013.

[48] Berkhin P. A Survey of Clustering Data Mining Techniques. In, ed. Grouping multidimensional data. Springer, 2006. 25-71

[49] Boley D. Principal Direction Divisive Partitioning. Data Min Knowl Disc, 1998, 4: 325-344

[50] Tan P, Steinbach M, Kumar V. Introduction to Data Mining. 第Pearson Addison Wesley Boston, 2006.

[51] Savaresi S M, Boley D L. A Comparative Analysis On the Bisecting K-means and the PDDP Clustering Algorithms. Intell Data Anal, 2004, 4: 345-362

[52] Bradley P S, Fayyad U M. Refining Initial Points for K-Means Clustering. In, ed. ICML. Citeseer, 1998. 91-99

[53] Sander J, Ester M, Kriegel H et al. Density-Based Clustering in Spatial Databases: The Algorithm Gdbscan and its Applications. Data Min Knowl Disc, 1998, 2: 169-194