

# 北京邮电大学

## 本科毕业设计（论文）开题报告

学院	信息与通信工程	专业	通信工程	班级	2014211112
学生姓名	林文鼎	学号	2014210328	班内序号	07
指导教师姓名	别红霞	所在单位	信息与通信工程学院	职称	教授
设计（论文）	基于腾讯定位数据的异常事件检测算法				
题目	Anomalous Event Detection Based on Tencent Positioning Data				

### 1. 选题的背景和意义

随着 GPS 定位，传感器网络和无线通信等应用的日益普及，越来越多的定位数据被收集和保存在应用服务器，如何快速地从这些定位数据中挖掘出有效信息日益成为一个令人关注的研究课题。在地图的定位数据中，人流密度可以反映出当地的一些事件变化。在通常情况下，人流密度的分布应该会服从一个基于时间的规律变化。而在异常事件（如异常气象，交通管制等）出现时，定位数据较以往同时段的数值必然会有异常的波动，我们可以从这些波动中获取到异常事件发生时上述定位数据的异常特征，从而在相似事件发生时可以及时做好有关准备。

### 2. 研究的基本内容和拟解决的主要问题

通常在异常事件（如异常气象，交通管制等）出现时，定位数据较以往同时段的数值必然有异常的波动。本论文研究基于位置服务数据的异常事件检测算法，并开发实际应用。主要研究内容如下：

- (1) 首先针对当前定位数据的时代背景，提出异常检测的必要性和重要性。然后调查异常检测的研究概况与发展状况，对常用检测方法进行介绍。
- (2) 通过给定的腾讯定位数据，读入并分析在异常事件出现时的数据异常性特征，标出所要处理数据的异常。
- (3) 研究并实现基于位置服务数据的异常检测算法。
- (4) 研究并实现基于上述位置服务数据形成的图像异常检测算法。
- (5) 分析可能提高准确率的改进，集成上述算法开发实际应用。

基于上述研究内容，需要解决的主要问题如下：

- (1) 腾讯定位数据的读入，MATLAB 信号分析工具的使用，信号异常点特征的明确，数据的标定；
- (2) 信号异常检测的方法调查，输入曲线并识别出异常点的功能实现；
- (3) 基于上述位置服务数据形成图像，图像异常检测的算法调查，机器学习算法的运行框架选择。
- (4) 软件实现平台的调查，软件架构的设计，输入输出的标准化。

### 3. 研究方法及措施：

- (1) 针对原始的腾讯定位数据，首先需要了解 TIF 格式的数据样式（维度、有效信息），将数据读入 MATLAB 中作信号分析以便于观察异常点特征并标注数据。在此过程中，需要熟悉 MATLAB 中常用的信号分析工具，在此基础上调研相关信号异常检测算法并尝试实现 demo。由于常规的定位数据应大致符合某一个变化曲线，可以通过基于统计的方法展开调查；同时也可以从信号的频

谱特征或是聚类方法入手寻找可能解决问题的方法。当积累了一定的信号异常检测算法，可以对实际数据进行测试，汇总各个算法的优劣，准备中期检查。

- (2) 实现定位数据的曲线异常检测后，需要对数据进行图像的重建，从图像的角度分析异常。可以从以下两个角度进行处理：

①建立判别模型：和信号的处理方法类似，直接寻找异常点。基于这种角度可以通过主成分分析（PCA）与矩阵分解的方法将图像的高维数据映射到低维度空间，将异常点的特征最大化地表示在低维度空间中进行判断。

②建立生成模型：预测正常情况下的数据点分布并通过计算实际点差值来寻找异常点。近年来机器学习、神经网络在异常检测中的应用得到了广泛关注，递归神经网络（RNN）是一类神经网络，包括一层内的加权连接（与传统前馈网络相比，连接仅馈送到后续层）。因为 RNN 包含循环，所以它们可以在处理新输入的同时存储信息。这种记忆使它们非常适合处理必须考虑事先输入的任务（比如时序数据）。在充分调研该网络的理论以及最佳运行框架后，可以采用 RNN 去预测时序的定位数据并寻找异常点。

根据以上两种思路，寻找相应的算法应用到数据图像上，并对它们的有效性进行分析。

- (3) 实现上述相关算法后，查找并实现一些相对复杂的改进策略，分析其复杂度和有效性；同时，调研并设计应用，整合输入输出与上述异常检测算法。汇总所有资料并核对任务完成情况，完成论文撰写及答辩工作。

#### 4. 研究工作的步骤与进度

01/01-01/26 资料查阅，论文开题；

01/27-03/04 阅读综述论文，调研相关算法；

03/05-03/19 技术路线确定；

03/20-04/15 完成异常数据的标定以及实现异常数据的检测算法；

04/16-04/20 准备中期检查相关材料；

04/21-05/10 完成异常数据生成的图像检测算法并开发应用；

05/11-05/20 核对任务完成情况，毕设论文撰写；

05/21-06/04 汇总毕设相关材料，准备答辩。

#### 5. 主要参考文献

- [1] Patcha A, Park J M. An overview of anomaly detection techniques: Existing solutions and latest technological trends[J]. Computer Networks, 2007, 51(12):3448-3470.
- [2] Chan P K, Mahoney M V, Arshad M H. A machine learning approach to anomaly detection[J]. 2003.
- [3] Malhotra P, Vig L, Shroff G, et al. Long short term memory networks for anomaly detection in time series[C]//Proceedings. Presses universitaires de Louvain, 2015: 89.
- [4] Eskin E. Anomaly detection over noisy data using learned probability distributions[C]//In Proceedings of the International Conference on Machine Learning. 2000.
- [5] Ringberg H, Soule A, Rexford J, et al. Sensitivity of PCA for traffic anomaly detection[J]. ACM SIGMETRICS Performance Evaluation Review, 2007, 35(1): 109-120.

- [6] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. ACM computing surveys (CSUR), 2009, 41(3): 15.
- [7] Lee W, Xiang D. Information-theoretic measures for anomaly detection[C]//Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on. IEEE, 2001: 130-143.
- [8] Noble C C, Cook D J. Graph-based anomaly detection[C]//Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003: 631-636.

指导教师签字

日期

年 月 日