**南京大学软件学院研究生学位论文中期检查报告格式**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **南京大学软件工程硕士学位论文中期检查报告** | | | | | |
| 导师1姓名 | 刘海涛 | 研究生姓名  （学号） | 韩淳 MF1732038 | 方向 | 软件工程方向 |
| 导师2姓名 |  |
| 论文题目 | 基于优先点树和地图瓦片的地理轨迹数据服务的设计和实现 | | | | |
| 论文选题来源及研究的目的和意义（500字左右）：  在移动互联网、卫星定位技术、LBS技术高速发展的背景下，无时无刻不在产生轨迹数据，轨迹数据包括交通数据、人类移动数据、动物迁移数据和自然现象轨迹数据等。海量的轨迹数据潜在性地暴露了个人的行为特征、兴趣爱好和社会关系等信息。这种细节信息的暴露，很大程度上是由于轨迹数据本身存在位置特征和时空特征上的关联性。这种关联性是很多高价值商业场景的运行基础。而在所有这些的关联性中，最直观，最有利用价值的，就是轨迹的相似性。  与轨迹相似有关的高价值应用场景很多，例如基于轨迹相似的用户分类，交通路线预测，犯罪同伙分析等等。因此，一个稳定高效的，实现轨迹数据相似检索和数据更新的轨迹数据系统是符合商业发展要求的必需产品，也是行业发展的一个重要方向。 | | | | | |
| 该方向的研究现状或技术进展综述（2000字左右）ArcGis这两个系统。  目前国内外轨迹相关系统所提供的轨迹分析功能着重于对速度变化和停靠点处理两方面，本文主要调查了百度鹰眼和ArcGis这两个系统。百度鹰眼是一套集轨迹追踪、存储、运算、查询的完整轨迹开放服务，可帮助开发者管理多达100万人/车轨迹。百度鹰眼支持持续的轨迹追踪，鹰眼SDK可以实时地采集终端的地理位置，持续回传轨迹。其采集回传的频率一般在2s到5min这个区间内。百度鹰眼支持轨迹存储，提供数据访问隔离和分布式存储，保证数据安全。其存储的内容包括坐标，速度，图片，视频和用户自定义字段。在数据存储量上，鹰眼目前支持100万终端，储存1年的轨迹数据。百度鹰眼支持轨迹查询和展示，提供历史轨迹查询服务，开发者可以毫无延时地查询终端的实时位置，并回放轨迹。百度鹰眼支持轨迹数据分析。目前已提供的轨迹分析包括驾驶行为分析（急速，超速判断），停留点分析（是否违法停车）等。  ArcGis是由ESRI公司开发的地理信息系统系列软件，其ArcGis1.0是世界上第一个现代意义上的GIS软件，第一个商品化的GIS软件。在轨迹数据服务方面ArcGis提供了路径图层创建，障碍创建，停靠点编辑，轨迹运动方向生成，运动轨迹展示，轨迹运动分析。其中轨迹运动分析包括所有停靠点最佳访问方式路径生成和展示。  TileServer-GL是一个针对矢量瓦片的开源地图服务器。它能够在服务器端使用MapBox GL内置引擎对矢量瓦片进行栅格化，进而为web应用和移动应用提供提供地图数据。它支持Mapbox GL JS,Android SDK,IOS SDK,Leaflet, OpenLayers, HighDPI/Retina, GIS via WMTS等众多前端库的数据调用。  TileServer-GL不仅能够提供瓦片数据，还提供了基于Mapbox GL Style的地图渲染。用户只要提供了有效的Mapbox GL style文件，tileServer就能够按照指定风格渲染地图数据，并返回给浏览器或移动端。  尽管TileServer-GL对外服务有良好而良好的适应性，但是它在商业应用领域存在以下两方面明显的短板。  首先，它的数据是保存在mbtiles文件中，而mbtiles是sqlite数据库的一种文件格式。TileServer-GL强耦合了这种文件格式使得其对不同数据源的扩展性几乎为零，面对那些数据保存在传统关系型数据库或是列数据库中的用户，TileServer-GL将无能为力。  其次，TileServer-GL只能提供瓦片读取服务，而不能提供地图数据的实时更新。而某些商业场景下，地图数据发生更新变化的可能性是非常大的。对于这种有更新要求的商业场景,TileServer也无法胜任。  vp-tree(vantage point tree),中文名称，优先点树，正是上述思路的一种实现。vp-tree从原理上说，是基于三角不等式进行递归分解的剪枝技术，其核心思想奠基了两种情况下的正确性。第一种，是在检索过程中，对于那些远远超出搜索范围的分支，就不需要进行搜索了。第二种，是当搜索目标点显然在某一个范围内的时候，外部的其他分支就都不必搜索了。基于这两个原则，搜索点的数量和点之间距离计算的次数都被大幅度地减少，从而显著提升了性能。  vp-tree的基本思路就是对点集合进行空间划分。第一步，要选择一个点作为vantage point，也就是优先点。第二步，集合中的所有点要计算自己与vp的距离。第三步，根据距离值的大小将点集合均分为两支，距离小于等于中值的为left/inside子集合，距离大于等于中值为right/outside子集合。  第四步，以left/inside集合作为左子树的根节点，right/outside集合为右子树的根节点，再针对这两棵子树分别递归地进行上述划分，从而形成一颗平衡的二叉树。vp-tree实际上实现了整个点集合内部的一个球状分割。而在整个数据空间中，大量的数据点集合被以不同的优先点为中心划分成了大量的相互交错的球型子空间。 | | | | | |
| 论文的主要技术路线、研究思路和实现方法；相关项目应用前景：（重点说明变更部分）：  本文从轨迹数据的本质特征出发，认识到轨迹数据本身其实是一种几何图形，而几何图形作为空间中某些坐标点的集合。由此可以想到，如果把轨迹与轨迹之间的距离，作为轨迹相似的评定标准，也就是距离越小，相似度越高，距离越大，相似性越小。那么这样就可以构建一个轨迹相似的衡量标准，进而以此标准作为优先点树的构建规则和搜索规则。沿着这个思路，就可以实现轨迹检索功能。  而对于瓦片数据服务，其基本思路是对地图坐标系到瓦片坐标系的转换。只要正确地实现这种转换规则，就能对瓦片数据的读取和更新功能加以实现。  本项目非常适合应用于火车，船舶，飞机以及犯罪嫌疑人的运行轨迹分析，通过相似性检索，发现可能的跟踪，伴随交通工具或者犯罪同伙。对于交通事业和侦察分析行业具有和现实的意义和作用。 | | | | | |
| 本人在相关项目中的扮演的角色和承担的工作（重点说明变更部分）：  本人是此项目的设计者和实现者，承担了系统技术调研，需求开发，架构设计，详细设计，代码实现，功能和性能测试的工作。 | | | | | |
| 论文的主要工作（500字左右）：  本文在原生优先点树的理论基础上，设计并实现了多路优先点树的Lucene索引结构。并针对多路优先点树结构开发了批量初始建树，插入新数据点和相似轨迹检索等三个主要功能。同时，为了支持轨迹数据在界面端的展示功能，本文基于Mapbox Tile标准开发了独立的瓦片数据服务，实现了可扩展的，可动态更新的瓦片数据服务。然后再通过前端库，将轨迹检索功能和瓦片数据功能结合起来，将两者的数据共同展示在浏览器上，为用户提供完整，简洁的，方便的轨迹数据服务。 | | | | | |
| 论文三级大纲：  第一章引言  l.1 项目背景  1.2国内外相关系统的发展概况  1.2.1 国内外轨迹数据系统发展概况  1.2.2 国内地图瓦片数据系统发展况  1.3本文的主要作  1.4本文的组织结构  第二章 相关技术概念综述  2.1 优先点树  2.1.1 NN问题  2 1.2优先点树述.  2.1.3优先点树的基本原理  2.1.4最简单优先点树的结构和搜索过程  2.2豪斯多夫距离  2.3 Lucene简介  2.3.1 Lucene运行原理  2.3.2.Lucene核心数据结构  2.3.4 为什么选择lucene  2.4 地理相关技术概念介绍  2.5.地理坐标与投影法  第三章．系统需求分析与概要设计  3.1.GTDS 系统概述  3.2 轨迹数据服务需求分析  3.2.1轨迹数据服务的功能需求  3.2.2 轨迹数据服务的非功能需求  3.2.3 轨迹效据服务用例图  3.2.4 初始建立轨迹索引用例描述表  3 2 5插入新数据用例描述表  3.2.6 相似轨迹检索用例描述表  第四章．系统详细设计与实现  4.1轨迹检索服务设计概述  4.1.2.优先点树设计类图  4.1.3.优先点树节点结构  4.1.3.初始建树流程图  4.1.4.初始建树代码实现  4.1.5.新插入数据点流程图  4.1.6.新插入数据点代码实现  4.1.7. 相似轨迹检索流程图  4.1.8.相似轨迹检索代码实现  4.2 瓦片数据服务设计概述  4.2.1.瓦片更新功能流程图  4.2.2.瓦片更新功能代码实现  第五章．系统运行效果图  第六章．总结与展望 | | | | | |
| 论文和相关项目的当前进度：项目已经完成功能代码编写和功能测试。论文基本完成，尚在校正阶段。 | | | | | |
| 论文和相关项目进展过程中遇到的困难和问题，以及解决的措施：  在最初实现中，出现优先点树常驻内存使用过多问题。本文采用了坐标栈的方式，实现内存的大量复用，以避免冗余内存的使用，进而解决这一问题。 | | | | | |
| 主要参考文献：  [1] Gao Q, Zhang FL, Wang RJ, and Zhou F. Trajectory big data:a review of key  technologies in data processing. Ruan Jian Xue Bao, 28(4):28–34, 2016.  [2] BaiduYingyan System. http://lbsyun.baidu.com/trace.  [3] tileserver System. http://tileserver.org.  [4] Daniel Mateos-Garc´ ıa, Jorge Garc´ ıa-Gutiérrez, and José Cristóbal Riquelme San-tos. On the evolutionary weighting of neighbours and features in the k-nearest  neighbour rule. Neurocomputing, 326-327:54–60, 2019.  [5] Peter N. Yianilos. Data structures and algorithms for nearest neighbor search in  general metric spaces. pages 311–321, 1993.  [6] Keinosuke Fukunaga and Patrenahalli M. Narendra. A branch and bound algo-  rithms for computing k-nearest neighbors. IEEE Trans. Computers, 24(7):750–  753, 1975.  [7] David A. Forsyth, Philip H. S. Torr, and Andrew Zisserman, editors. Computer  Vision - ECCV 2008, 10th European Conference on Computer Vision, Marseille,  France, October12-18, 2008, Proceedings, PartII,volume5303ofLectureNotes  in Computer Science. Springer, 2008.  [8] Andrés Vargas and Johan Bogoya. A generalization of the averaged hausdorff  distance. Computación y Sistemas, 22(2), 2018.  [9] Laurence Hirsch and Teresa Brunsdon. A comparison of lucene search queries  evolved as text classifiers. Applied Artificial Intelligence, 32(7-8):768–784, 2018.  [10] Peilin Yang, Hui Fang, and Jimmy Lin. Anserini: Reproducible ranking baselines  using lucene. J. Data and Information Quality, 10(4):16:1–16:20, 2018.  [11] Michael McCandless, Erik Hatcher, and Otis Gospodnetic. Lucene in Action,  Second Edition: Covers Apache Lucene 3.0. Manning Publications, 2010.  56  [12] Fateh Boucenna, Omar Nouali, Samir Kechid, and M. Tahar Kechadi. Secure  inverted index based search over encrypted cloud data with user access rights  management. J. Comput. Sci. Technol., 34(1):133–154, 2019.  [13] Vyron Antoniou, Jeremy Morley, and Mordechai (Muki) Haklay. Tiled vectors:  A method for vector transmission over the web. In Web and Wireless Geographi-  cal Information Systems, 9th International Symposium, W2GIS 2009, Maynooth,  Ireland, December 7-8, 2009. Proceedings, pages 56–71, 2009.  [14] Wendy Osborn. A k-nearest neighbour query processing strategy using the mqr-  tree. In Advances in Network-Based Information Systems, The 20th International  Conference on Network-Based Information Systems, NBiS 2017, Ryerson Univer-  sity, Toronto, ON, Canada, August 24-26, 2017., pages 566–577, 2017.  [15] AdaWai-CheeFu, PollyMei-shuenChan, Yin-LingCheung, andYiuSangMoon.  Dynamic vp-tree indexing for n-nearest neighbor search given pair-wise dis-  tances. VLDB J., 9(2):154–173, 2000.  [16] Hugo Proenc ¸a and Jo˜ ao C. Neves. Fusing vantage point trees and linear discrim-  inants for fast feature classification. J. Classification, 34(1):85–107, 2017.  [17] Guohang Zeng, Qiaozhi Li, Huiming Jia, Xingliang Li, Yadi Cai, and Rui Mao.  An inclusion rule for vantage point tree range query processing. In Human Cen-  teredComputing-FirstInternationalConference, HCC2014, PhnomPenh, Cam-  bodia, November 27-29, 2014, Revised Selected Papers, pages 777–783, 2014. | | | | | |
| 导师意见： | | | | | |
| 学院备案意见：        年 月 日 | | | | | |