

**研究生学位论文开题报告登记表**

学 院： 资源与环境科学学院

专 业： 测绘工程

学 号： 2014282050150

姓 名： 李冬琳

导师姓名： 李精忠

导师职称： 副教授

2015年 12 月 31 日

**武汉大学关于研究生学位论文开题报告的有关规定**

根据《中华人民共和国学位条例》及其《暂行实施办法》和《武汉大学学位授予工作细则》的精神，为做好研究生学位论文的开题报告，保证学位论文质量，特作如下规定：

**第一条** 学位论文开题报告是研究生写作论文的必经过程，所有研究生（含：博士生、硕士生）在修完学位课程，写作学位论文之间都必须作开题报告。

**第二条** 开题报告主要检验研究生对专业知识的独立驾驭能力和研究能力，考察写作论文准备工作是否深入细致，包括选题是否恰当，资料占有是否翔实、全面，对国内外的研究现状是否了解，本人的研究是否具有开拓、创新性。

**第三条** 学位论文开题报告前，研究生必须根据专业培养目标，结合导师、教研室（或研究室）所承担的国家、省部委等有关部门下达的研究项目或课题以及本人的研究特长，与导师协商，确定选题，广泛查阅文献，深入调研，收集资料，制定学术研究方案，在此基础上撰写开题报告。

**第四条** 研究生进行开题报告，必须提交“开题报告”的书面材料，内容包括：（1）论文选题的理由或意义；（2）国内外关于该课题的研究现状及趋势；（3）本人的研究计划，包括研究目标、内容、拟突破的难题或攻克的难关、自己的创新或特色、实验方案或写作计划等；（4）主要参考文献目录。开题报告的书面材料不得少于3000字。

**第五条** 研究生进行学位论文开题报告要向导师提出申请，申请获准后，博士生在博士生指导小组范围内作开题报告，硕士生在导师所在教研室或教学小组作开题报告。参加开题报告的教师，包括导师在内，一般不得少于3人。无论博士生还是硕士生，在作开题报告时，本学科专业的研究生一般必须参加，跨学科或相近专业的研究生亦可旁听。

**第六条** 参加研究生学位论文开题报告的教师应当对开题报告进行评议，主要评议论文的选题是否恰当，研究设想是否合理、可行，研究内容与方法是否具有开拓性、创新性，研究生是否可以开始进行论文写作等。评议结果分“合格”与“不合格”二种。评议结束后，由研究生指导教师在《研究生学位论文开题报告登记表》“评语”栏中填写评语。学位论文开题报告通过后，研究生方可进行论文撰写工作。

**第七条** 开题报告结束后，研究生应将登记表复印一份连同登记表原件和开题报告等一并交所在院、系研究生干事将登记表复印件加盖公章后报送研究生院培养教育处，其他材料留存院、系查备查。研究生院培养教育处将不定期抽查研究生开题报告材料。

**第八条** 本规定自2008年级研究生开始实行。

**第九条** 本规定由研究生培养教育处负责解释。

武 汉 大 学 研 究 生 院

**研究生学位论文开题报告表**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 姓 名 | | | 李冬琳 | | | | 院、系（所） | | 资源与环境科学学院 | |
| 学 科 专 业 | | | 测绘工程 | | | | 攻读学位 | | 硕士 | |
| 研 究 方 向 | | | 空间数据可视化 | | | | 指导教师 | | 李精忠 | |
| 拟定学位论文题目：网络VGI数据可视化方法研究 | | | | | | | | | | |
| 参加开题报告教师人数 | | | | | 3 | | | 参加旁听学生人数 | | 17 |
| 开  题  报  告  组  成  人  员 | | 姓 名 | | | 职 称 | 所 在 工 作 单 位 | | | | |
| 李精忠 | | | 副教授 | 资源与环境科学学院 | | | | |
| 江文萍 | | | 副教授 | 资源与环境科学学院 | | | | |
| 蔡忠亮 | | | 教授 | 资源与环境科学学院 | | | | |
|  | | |  |  | | | | |
|  | | |  |  | | | | |
|  | | |  |  | | | | |
|  | | |  |  | | | | |
|  | | |  |  | | | | |
|  | | |  |  | | | | |
|  | | |  |  | | | | |
|  | | |  |  | | | | |
|  | | |  |  | | | | |
| 一、研究生开题报告及指导教师所提问题回答的内容记录：   1. 论文选题的理由及其意义   **1论文选题的背景及其意义**  随着信息技术的发展，我们进入到web2.0的时代，网络的技术和终端技术的发展极大地丰富了人类获取信息的手段和内容， 使任何人在任何地方都能通过泛在网络获取需要的任何信息，同时也可以产生大量的信息，用户现在不仅是信息的消费者同时也是生产者。 这些由非专业的普通自愿者用户参与创建及共享的地理信息统称为自发地理信息（volunteered geographic information,VGI）。由于各类传感器日益普及，通讯技术的飞跃以及网络基础设施的高速发展，VGI数据量也以指数级增长，过去两年所产生的数据量为有史以来所有数据量的90%[1]。而据统计这些信息中80%带有空间信息，所以挖掘网络环境下的VGI数据是一项有重大意义的研究课题。  当前数据挖掘中，可视分析是一项重要的分析手段，可视分析结合了可视化、人机交互和自动分析，并使数据分析过程透明化。用户通过人机交互技术评价、修改和改进自动分析模型， 从而得到新的自动分析结果。由人来定义分析任务和识别复杂的模式， 由机器来存储和分析大量的数据，这样就让人和机器实现优势互补，能够非常有效直观的挖掘出数据背后的价值，发现数据背后隐藏的规律。  VGI数据根据其主要特征不同可以分为三类：时间特征数据、空间特征数据、语义特征数据。本文将针对这三类数据。选取具有代表性的出租车轨迹数据、POI数据、微博数据进行可视化方法研究。出租车轨迹数据是随着GPS技术的发展而产生的一种VGI数据，是其中非常典型的一种时间数据，能够反映出城市空间中的人口流动、城市交通信息。POI数据则是VGI数据中最常见的空间数据，其地理空间位置信息中往往蕴含着城市空间内的建筑分布等信息；微博数据是带有地理空间信息的文本语义信息数据，往往能从中发现微博用户的行为特征，是VGI数据中最典型的语义特征数据；  **2． 国内外研究现状及趋势**  **2.1 时间特征可视化**  由于各种传感器、互联网的发展，单个地理目标的空间位置信息的连续收集成为可能，也因此产生了大量的带有时序信息的空间数据，如何对这种数据进行有效的可视化，国内外也有大量的学者进行研究。。Krüger等人开发的轨迹可视化系统允许用户交互式定义轨迹的起止点，实现轨迹可视化[14]。Zeng等人计算了乘客从一个公交或地铁站点出发到达其他站点的时间花费，并实现了多种公共交通状况专题的可视化形式，比如交通可达性等专题[15]。Tominski等人利用时空立方体技术，将轨迹折线扩展成彩色条带, 并在高度方向将不同轨迹的条带堆叠起来, 以方便轨迹间的比较，不同的条带代表不同的属性，实现了轨迹的多属性三维可视化[16]。Buchin等人利用边捆绑技术，用Flow Map对美国人口移动事件进行时空可视化[17]。Slingsby等人利用多维平行坐标的方法，多具有多指标多维度特性的的人口统计数据进行了可视化，能非常直观的发现和了解人口统计数据中各组指标之间的相互关系[18]。袁晓如团队利用OD-Wheel方法对城市内的出租车轨迹数据进行可视化，可以从中发现出行模式和异常交通拥堵[19]。陈为团队开发了一个交互式可视化系统，能基于出租车轨迹数据实时对城市道路交通流量、道路交叉口交通状况进行模拟可视分析[20]。  **2.2 空间特征可视化**  网络环境下的VGI数据中最常见的就是空间位置数据，比如OpenstreetMap、Google Map Maker等平台，志愿者可以在上面自由的编辑、发布空间位置信息。针对空间位置数据的可视化，也有非常多专家学者在进行相关研究。禹文豪等人基于网络核密度法对网络空间POI点进行三维可视化分析[12]。杨敏等人基于服务器端多层次结构化组织与客户端冲突移位相结合的在线式综合策略，实现了POI点的多尺度可视化[13]。  **2.3 语义特征可视化**  在网络环境下，用户可以在各种终端上的博客、微博、微信朋友圈等平台发布信息，这些信息中大量的数据是文本数据，如何对文本数据进行有效的可视化，国内外学者都相关内容进行了很多研究。Thom等人利用Twitter用户发布的信息用点标注的方法来实现可视化，并利用算法将局部信息和全球信息进行对比来探测异常事件[2]。Marcus等人通过算法对Twitter数据进行聚合，实现了交互式事件可视化，帮助用户梳理整个事件的走向[3]。Zhao等人将文本的叙述结构语义以树的形式进行可视化，同时展现了相似度统计、修辞结构、以及相应的文本内容[4]。  周霞娟等人开发了一个用户驱动的微博可视化搜索系统，通过对用户、行为特征等建立模型，实现了更加高效的微博搜索[5]。袁晓如团队开发了一套微博事件可视化系统，能够对微博事件的传播链路进行网络可视化[6]。陈思明等人开发了一套交互式微博可视化系统，针对抽样微博用户发布微博的位置移动，挖掘用户的出行模式[7]。赵华等人通过分析用户发表微博特点，利用3D标签云实现用户兴趣可视化[8]。王松等人开发了一套微博可视化系统，能对微博热点、发布数量、地理信息、时序等信息进行可视化[9]。武泽旭利用融合FN的多层级图可视化算法对微博数据进行网络可视化[10]。陈宏飞等人基于微博进行了西安市夜间时空分布可视化、交通拥堵状况时空分布研究[11]。  三、本人的研究计划，包括研究目标、内容、拟突破的难题或攻克的难关、自己的创新或特色、实验方案或写作计划等；  **3.1研究目标**  本文的研究目标是针对网络VGI数据中的时间数据、空间位置数据、文本属性数据进行可视化方法研究，提出几种分别适合于三类数据的可视化方法，并用出租车轨迹数据、POI数据、微博数据作为具体实验对象进行可视化，验证方法的有效性。  针对时间特征数据，以出租车轨迹数据为例，拟进行多尺度动态可视化来表现其时间特征，发现居民的出行规律并尝试进行拥堵预测。  针对空间特征数据，以POI数据为例，拟实现一种空间特征数据的多尺度可视化方法。解决小尺度下空间特征数据的压盖、拥挤等问题。  针对语义特征数据，以微博数据为例，采用合适的可视化方法，表达其语义特征，挖掘语义信息。  **3.2研究内容**  基于以上对研究背景和国内外研究现状的分析，结合本文的研究目标，本文的研究内容主要包括：  **（1）针对时间特征数据的可视化**  如果对整个城市的轨迹数据直接可视化，那结果必然是杂乱无章的。所以面对整个问题，本文拟对出租车轨迹进行预处理，先在空间、时间尺度上进行综合，根据不同的属性进行分类、分级。最终对轨迹数据进行多尺度动态可视化。  **（2）针对空间特征数据的可视化**  目前常见的POI可视化方法比如标注、聚类、热图都各有优劣，在使用的时候也都是针对各自的优点分开使用，本文将尝试将这几种方法的优点进行融合，同时和交互技术相结合，研发出一种新的多尺度POI可视化方法，能够自适应于底图和数据。  **（2）针对语义特征数据的可视化**  微博数据除了用户发布的文字内容中的文本信息外，大量微博它还带有地理位置信息。目前常见的微博可视化更多的将这两种信息分开分别进行可视化，本文将尝试将这两种信息融合进行可视化。  对话题文本进行分词后得到的标签和空间数据可视化中的聚类方法相结合，将文本标签在地理视角下进行可视化。同时针对同一个话题在全国不同城市的热度信息，尝试进行三维可视化，从一个新的角度对这种信息进行可视化。针对有着事件属性的事件，将尝试在进行空间信息+时间维度的可视化，探究事件的生命周期。  **3.3 拟突破的技术难点与解决方案**  本选题拟突破的技术难点如下：   1. **时间特征数据的处理及可视化技术**   时间特征数据同时拥有时序信息和空间信息，并且会有很多异常数据，如何处理这种异常数据，得到适合于可视化的数据是一个难点，并且空间信息也会遇到小尺寸空间和大数据表达的矛盾，如何实现在时间尺度和空间尺度下的可视化是一个难点。   1. **空间特征数据的处理及多尺度可视化技术。**   空间位置数据小比例尺的可视化中会遇到拥挤、压盖等冲突现象，会严重影响用户的认知过程，不能有效的展现空间位置数据的信息。所以需要合适的数据处理方法和多尺度可视化技术，解决小尺寸空间和大数据表达的矛盾，实现多尺度可视化。   1. **语义特征数据的关键信息提取技术、分词其可视化技术。**   文本数据往往都是大量的文本，为了提取其中的有效信息，其中就涉及到自然语言处理技术和分词技术。针对处理后的文本数据，需要有合适的可视化技术来展现文本数据的信息，并能从可视化的结果中发现有价值的信息。  **实验方案**  为验证本文提出的对文本数据、空间数据、时空数据的可视化方法，本文拟开发一个网络VGI数据可视化系统，整个系统主要分为三层：数据采集层、数据处理及输出接口层、可视化层。  **3.3.1 数据采集层**  针对VGI数据主要分布在互联网中，本文拟采用Python语言开发一个高效的网络爬虫，持续爬取互联网中的VGI数据。并在爬取的过程中对一些异构数据作预处理。  由于VGI数据复杂、异构、多源、多维、数据量大的特征，本文拟采用非关系型数据库MongoDB，MongoDB作为一个开源的、分布式的、面向文档存储的非关系数据库，具有自动分片、模式自由的特性。MongoDB模式自由的特性使得它非常适合存储各种异构数据，且内置的水平扩展机制提供了从百万到十亿级别的数据量处理能力，完全可以满足海量数据的存储。此外，MongoDB还支持空间索引。  **3.3.2 数据处理及输出层**  因为VGI数据复杂、多源的特征，在数据库中的数据和可视化工具之间必须有一个中间层，这个中间层负责进行对数据库中的数据进行预处理，并针对不同可视化工具实现数据定制化输入输出，满足可视化工具的需求。本文拟采用Node.js开发一个高速的数据处理及输出层。Node.js最大的特性就是异步无堵塞，当面对可视化工具需要频繁的请求数据时，这个特性将能非常好的满足这个需求，而且和可视化层所使用的是同一种语言，易于后期开发和维护。  **3.3.3 可视化层**  本文拟采用OpenLayers作为基础类库来进行可视化工具的开发工作，OpenLayers是一个专为Web GIS客户端开发提供的JavaScript类库包，除了可以实现标准格式发布的地图数据访问。还能提供很多比如数据的加载、处理等其他操作。本文还将利用ECharts、D3.js等成熟的可视化类库进行先关可视化工具的开发。  因为VGI数据的多源性特性，可视化平台必须要能有高度的可扩展性、低耦合性。因此开发将严格按照模块化开发，采用ECMA2015标准进行模块化开发，每个可视化工具都将抽象成一个独立的可视化模块，做到可以“随意插拔”，整个可视化层将由这些独立的模块组合而成。  **3.4 论文的创新之处**  本文创新点：  1．针对时间特征数据，用出租车轨迹数据作为实验对象，将出租车轨迹在综合、分类、分级后，实现动态多尺度可视化。  2. 针对空间特征数据，用POI数据作为实验对象，将常见的可视化方法融合后，和交互技术想结合，提出新的可视化方法，实现多尺度可视化。  3. 针对语义特征数据，以微博数据为例，创新性的将文本信息和空间信息融合，在地理空间内对文本信息进行分词聚类，实现地理视角的微博文本二维、三维可视化。同时利用时间维度的信息，探究微博事件的生命周期。  **3.5论文章节安排**  **第一章 绪论**  **1.1． 研究目的与意义**  **1.2. 国内外研究现状**  **1.3. 本论文主要研究内容和方法**  **第二章VGI数据、可视化技术及理论**  **2.1. VGI数据**  **2.2. 可视分析**  **2.3. 当前主流可视化技术**  **第三章 时间特征数据可视化**  **3.1 出租车轨迹数据的预处理**  **3.2 出租车轨迹数据综合、分类、分级**  **3.3 出租车轨迹数据可视化**  **第四章 空间特征数据可视化**  **4.1 空间特征数据特征、可视化需求**  **5.3 自适应可视化理论与技术**  **6.2 POI点自适应可视化**  **第五章 语义特征数据可视化**  **5.1 微博语义数据的特点**  **5.2 自然语言处理技术**  **5.3 事件时空可视化**  **5.4 事件热度三维可视化**  **第六章 总结与展望**  **参考文献**  **致谢**  主要参考文献  [1] Krüger R, Thom D, Wörner M, et al. TrajectoryLenses–A Set‐based Filtering and Exploration Technique for Long‐term Trajectory Data[C]//Computer Graphics Forum. Blackwell Publishing Ltd, 2013, 32(3pt4): 451-460.  [2] Zeng W, Fu C W, Arisona S M, et al. Visualizing mobility of public transportation system[J]. Visualization and Computer Graphics, IEEE Transactions on, 2014, 20(12): 1833-1842.  [3] Tominski C, Schumann H, Andrienko G, et al. Stacking-based visualization of trajectory attribute data[J]. Visualization and Computer Graphics, IEEE Transactions on, 2012, 18(12): 2565-2574.  [4] Buchin K, Speckmann B, Verbeek K. Flow map layout via spiral trees[J]. Visualization and Computer Graphics, IEEE Transactions on, 2011, 17(12): 2536-2544.  [5] Slingsby A, Dykes J, Wood J. Exploring uncertainty in geodemographics with interactive graphics[J]. Visualization and Computer Graphics, IEEE Transactions on, 2011, 17(12): 2545-2554.  [6] Donghao Ren, Xin Zhang, Zhenhuang Wang, Jing Li, and Xiaoru Yuan. *Proceedings of IEEE Pacific Visualization Symposium (PacificVis 2014),*pages 330-334, (Notes Paper), Yokohama, Japan, Mar. 4-7, 2014.  [7] Wang F, Chen W, Wu F, et al. A visual reasoning approach for data-driven transport assessment on urban roads[C]//Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on. IEEE, 2014: 103-112.  [8] 禹文豪,艾廷华. 核密度估计法支持下的网络空间POI点可视化与分析[J]. 测绘学报,2015,01:82-90.  [9] 杨敏,艾廷华,卢威,成晓强,周启. 自发地理信息兴趣点数据在线综合与多尺度可视化方法[J]. 测绘学报,2015,02:228-234.  [10] 李清泉,李德仁. 大数据GIS[J]. 武汉大学学报(信息科学版),2014,06:641-644+666.  [11] Thom D, Bosch H, Koch S, et al. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages[C]//Pacific visualization symposium (PacificVis), 2012 IEEE. IEEE, 2012: 41-48.  [12] Marcus A, Bernstein M S, Badar O, et al. Twitinfo: aggregating and visualizing microblogs for event exploration[C]//Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2011: 227-236.  [13] Zhao J, Chevalier F, Collins C, et al. Facilitating discourse analysis with interactive visualization[J]. Visualization and Computer Graphics, IEEE Transactions on, 2012, 18(12): 2639-2648.  [14] 周霞娟,汪飞,金玲,陈为,王章野. 用户驱动的微博可视化搜索[J]. 中国图象图形学报,2015,05:715-723.  [15] 王祖超,袁晓如.轨迹数据可视分析研究[J].计算机辅助设计与图形学学报,2015,01:9-25.  [16] Chen S, Yuan X, Wang Z, et al. Interactive Visual Discovering of Movement Patterns from Sparsely Sampled Geo-tagged Social Media Data[J]. Visualization and Computer Graphics, IEEE Transactions on, 2016, 22(1): 270-279.  [17] 赵华,纪晓文,曾庆田,郝春燕. 基于话题相关空间的微博用户兴趣识别及可视化方法[J]. 计算机科学,2015,S1:500-502+509.  [18] 王松,吴亚东,李秋生,蒋宏宇,邹勇刚. 基于时空分析的微博演化可视化[J]. 西南科技大学学报,2014,03:68-75.  [19] 武泽旭. 社交网络海量数据的分析与可视化[D].北京邮电大学,2014.  [20] 陈宏飞,张心萍,赵艳慧,刘广,孙九林. 基于微博的西安市交通拥堵状况时空分布研究[J]. 陕西师范大学学报(自然科学版),2015,06:83-88. | | | | | | | | | | |
| 研究生开题报告及指导教师所提问题回答的内容记录： | | | | | | | | | | |
| 开题报告记录人签名：  年 月 日 | | | | | | | | | | |
| 指导教师意见：  指导教师签名：  年 月 日 | | | | | | | | | | |
| 开  题  报  告  评  语 | 评议结果 | | |  | | | | | | |
| 参加开题报告的教师（3~5人）签名：  年 月 日 | | | | | | | | | |

注：评议结果分“合格”或“不合格”。