引文格式:朱庆,付萧.多模态时空大数据可视分析方法综述[J].测绘学报,2017,46(10):1672-1677. DOI:10.11947/j.AGCS.2017.20170286. ZHU Qing,FU Xiao.The Review of Visual Analysis Methods of Multi-modal Spatio-temporal Big Data [J]. Acta Geodaetica et Cartographica Sinica,2017,46(10):1672-1677. DOI:10.11947/j.AGCS.2017.20170286.

多模态时空大数据可视分析方法综述

朱 庆,付 萧

西南交通大学地球科学与环境工程学院,四川 成都 611756

The Review of Visual Analysis Methods of Multi-modal Spatio-temporal Big Data ZHU Qing, FU Xiao

Faculty of Geoscience and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China

Abstract: The visual analysis of spatio-temporal big data is not only the state-of-art research direction of both big data analysis and data visualization, but also the core module of pan-spatial information system. This paper reviews existing visual analysis methods at three levels: descriptive visual analysis, explanatory visual analysis and exploratory visual analysis, focusing on spatio-temporal big data's characteristics of multi-source, multi-granularity, multi-modal and complex association. The technical difficulties and development tendencies of multi-modal feature selection, innovative human-computer interaction analysis and exploratory visual reasoning in the visual analysis of spatio-temporal big data were discussed. Research shows that the study of descriptive visual analysis for data visualizationis is relatively mature. The explanatory visual analysis has become the focus of the big data analysis, which is mainly based on interactive data mining in a visual environment to diagnose implicit reason of problem. And the exploratory visual analysis method needs a major break-through.

Key words: spatio-temporal big data; visual analysis; pan-spatial information system; multi-modal data Foundation support: The National Natural Science Foundation of China(No. 41471320); National Key R&D Program of China(No. 2016YFB0502303)

摘 要:时空大数据可视分析是近年国际大数据分析与数据可视化领域研究的热点前沿,也是全空间信息系统的核心研究内容之一。本文针对时空大数据多源、多粒度、多模态和时空复杂关联的特点,按照描述性、解释性和探索性3个层次分类归纳了现有时空大数据可视分析方法,论述了时空大数据可视分析中多模态特征筛选、新型人机交互分析以及探索性可视推理等技术难点和主要发展动态。研究表明,以数据可视化为主的描述性可视分析研究相对成熟,以可视环境下交互式挖掘分析实现问题诊断为主的解释性可视分析已成为当前大数据分析的焦点,而面向复杂问题协同决策的探索性可视分析方法则是大数据分析有待突破的重要发展方向。

关键词:时空大数据:可视分析:全空间信息系统:多模态数据

中图分类号:P208 文献标识码:A 文章编号:1001-1595(2017)10-1672-06

基金项目:国家自然科学基金(41471320);国家重点研发计划(2016YFB0502303)

大数据时代,传感网、物联网和媒体/自媒体社交网络等信息通信技术日新月异,使得描述与记录人类社会、计算机世界和物质世界复杂事物的时空数据迅猛增长,时空数据规模越来越庞大,数据语义越来越丰富,因而催生了科学研究的第四范式—数据密集型科学^[1]。如何从海量、高维、动态的多模态时空大数据中挖掘有价值的、潜在

的、复杂时空语义关联,综合感知时空大数据反映的发展态势,进行科学合理的探索推理预测成为 地理信息科学迫切需要解决的关键问题。

可视分析综合人脑感知、假设、推理的优势与 计算机对海量数据高速、准确计算的能力,变"信 息过载"问题为机遇,已成为当下大数据分析的研 究热点[2-6]。本文从人类空间认知规律出发,针对 多模态时空大数据特点,按照描述性、解释性和探索性3个层次分类归纳了现有可视分析方法的优缺点,指出了未来时空大数据可视分析的发展方向。

1 概 述

时空大数据由于其所属空间从宏观的宇宙空间到地表室内空间以及更微观的空间,其时间、空间和属性 3 个方面的固有特征呈现出时空紧耦合、数据高维、多源异构、动态演化、复杂语义关联的特点^[7-9]。传感网、互联网、物联网及其多层耦合网络成为探索人类活动和事物发展密切相关的时空大数据来源,时刻产生着与时空位置相关的时空大数据来源,时刻产生着与时空大数据还具有的,现频等多种模态,因此时空大数据分析的难点在于如何从不同来源、不同粒度、不同类型的多据,还为类态的特点。目前时空大数据分析的发展,多模态的特点。目前时空大数据分析的难点在于如何从不同来源、不同粒度、不同类型的多据融合,进而实现数据特征降维、聚类分析,关联分析及分类预测。

人脑是天然的大数据处理引擎,有较强的容 错性以及对听觉、视觉、嗅觉、味觉信息等多模态 特征的高度概括及综合能力[10-11]。模拟人脑神 经网络构建的神经网络方法,在单一模态数据的 自动处理方面取得了明显进展,如语音识别、图像 特征处理。关于多模态数据的智能处理,有特征 级和语义级两个层次的跨域多模态数据融合方 法,特征级融合和语义级融合多依赖机器学习方 法,如深层神经网络 DNN、迁移学习、多核学习 等[12]。允许不同模态数据之间相互作用、彼此增 益是实现多模态数据融合分析面临的挑战,仅依 靠机器学习等数据挖掘方法使得用户难以对数据 和算法过程本身直观地进行理解、探索和优 化[13-16]。基于可视化方法辅助多模态数据融合、分 析和预测,充分结合了人脑接收处理多模态数据时 的高度抽象概括能力、联想推理能力和计算机自动 计算能力的优势[17]。可视分析方法通过可视交互 界面,将人的智慧,特别是"只可意会,不能言传"的 人类知识和个性化经验可视地融入整个数据分析 和推理决策过程中成为最有潜力的方向[18]。

2 现有的可视分析方法

可视分析经历了 3 个阶段的发展: 可视表达、 交互式可视化、可视化推理。这个发展过程十分明 显地表明越来越多的人类智慧被引入到数据处理过程之中[19]。可视表达即数据的可视化、可视表征,帮助分析者以可视化的形式直观了解数据,属于对数据分布、异常、聚集等特征描述性质的可视分析;随着信息获取与社会交往的日益增进,传统描述性可视分析越来越难以满足人们对信息理解与交流的动态需求。交互式可视化即实现在可视化环境下交互地浏览数据和挖掘过程,属于对数据背后模式、规律的解释性质的可视分析;可视化推理是以可视表征辅助人进行假设推理,包含探索发现数据间潜在的、有价值的关系和基于此构建模型进行假设推理和验证。本文以下按照描述性、解释性和探索性对现有的可视分析方法进行归纳和对比。

2.1 描述性可视分析

描述性可视分析主要为数据驱动,通过不同类型时空数据(场景数据、时序数据、轨迹数据、多变元网络数据等)符号化或真实感的专题地图或逼真动态场景等形式表达,直观展现多模态时空数据的分布、异常、聚集、演化等多维动态特征^[20]。实时接入的多模态动态感知数据的融合分析是目前研究的焦点。典型的有:

- (1) 时序数据可视分析方法。可以分为需要坐标轴或视觉变量展示时间变化的静态表征方法和将时间变化体现于动画的动态表征方法^[21-22]。传统静态表征方法如扇形图(sector graph)、堆积饼图(stacked graph)、圆环图(circle graph)和折线图等,针对数据高维多变元特点的静态表征方法如螺旋图(spiral graph)时序平行坐标轴、梳形图(multi comb)时间轮图(time wheel)^[23-24]。动态表征是以动画形式通过颜色、形状、大小、纹理等视觉变量的变化来反映时序数据。
- (2) 轨迹数据可视分析方法。轨迹数据具有时间和空间属性,传统的轨迹数据可视分析方法以时空立方体为代表,时空密度图、聚集和汇总、马赛克镶嵌图、出发地-目的地数据矩阵等可视化方法适用于大规模、多变元的轨迹数据可视化[25-26]。
- (3) 网络可视分析方法。网络是以节点代表对象、弧段代表对象间关系的数据结构模型,被广泛运用于具有多维度、复杂关联关系的时空属性数据的表达^[27-28]。节点-边的复杂关系使得网络数据可视表征可认知程度降低。现有的按照网络节点的布局方法可分为力导引布局(FDA)、圆形布局、相对空间布局、聚类布局、层布局、基于时间布局、手工布局和随机布局^[29]。

2.2 解释性可视分析

解释性可视分析是数据驱动与模型驱动结 合,通过示意性与沉浸式协同的增强现实可视化 表达,进行深度挖掘分析,以实现对数据蕴含的规 律和模式的分析与解释,如可视化方式实现数据 挖掘过程及空间分析过程中参数调整和特征选择 等。可视化驱动的数据挖掘方法是目前时空大数 据分析的热点。交互式降维分析方法突破传统降 维分析在可视化输出结果后调整参数重新计算再 可视化的缺陷,对计算中每次迭代的中间过程进 行可视化并实时动态调整参数[30]。关联分析是 常用的数据挖掘方法,然而关联规则集难以直观 理解,规则之间的相关性会被淹没在规则集中。 多层次多模态混合的时空可视化能促进关联规则 结构的深度洞察,交互操作、高亮强调相关性高和 有意义的规则,如采用马赛克图对关联规则进行 可视化[31-32]。聚类分析依据数据特性,能够对其 进行有效的分类和分组,其可视化形式多样,能够 直观反映数据的聚集模式。传统的聚类方法难以 处理,许多聚类分析的可视化方法交互性差且只

简单考虑图形布局问题。星形坐标、VISTA 和 HOV³ 针对上述问题进行了改进^[33-35]。

2.3 探索性可视分析

探索性可视分析方法是交互驱动、模型驱动 与数据驱动的有机结合,通过将抽象事物如数据、 空间布局等进行增强现实表达,结合人的思维能 力进行感知认知、联想假设,探索发现多源多模态 数据间隐含的关联关系等知识规律,揭示智慧城 市运行规律、驱动力、影响因子以及相互作用机 制,如多粒度时空对象复杂的关联关系网络可视 化以及可视推理。网络可视分析是分析和推理异 构数据内在关系的最重要方法[18]。多粒度时空对 象不同的粒度划分之间有层次结构,表征对象时空 属性的多源数据之间存在着关联关系,构成多层次 的复杂关联关系网络,其可视化以及基于此的可视 推理是目前探索性可视分析的重要研究内容[36]。

2.4 可视分析方法对比

表1按照描述性可视分析、解释性可视分析 和探索性可视分析 3 个层次分类归纳了现有可视 分析方法的特点。

表 1 现有可视分析方法对比

Tab.1 Comparison of existing visual analysis methods

分类 驱动因素及主要任务 适用范围 实例

描述性可视 数据驱动,可视表达时空数据的分布、异常和聚集 等一般性特征

时序数据可视分析 轨迹数据可视分析 高维数据可视分析 场景可视分析



场景特征统计

分析

解释性可视 模型驱动+数据驱动,深度挖掘分析多模态时空 数据隐含的规律和模式

可视化关联分析 可视化聚类分析 可视化降维分析



房价与可视域的关联关系分析

分析

探索性可视 交互驱动+模型驱动+数据驱动,电脑与人脑协 同进行联想、假设与推理

多粒度时空对象复杂关联关系 网络可视推理 复杂时空过程模拟预测



复杂建筑物室内外物理环境分析

3 可视分析方法的主要发展趋势

针对时空大数据多源、多粒度、多模态和时空 复杂关联的特点,为满足探索时空大数据潜在关联

关系、综合感知时空数据反映的态势并进行科学合 理的推理预测与决策需求,本文提出时空大数据多 层次可视分析体系,如图1所示,人类左脑侧重逻 辑技术性思维,右脑侧重空间形象性思维,相互协 同、不可分割。分析和可视化是全空间信息系统的核心功能,4 个层次(描述、诊断、预测、处方)的分析功能与 3 个层次(展示、分析、探索)的可视化功能相互融合与协同,构成从描述性可视分析到解释性可视分析和探索性可视分析的多层次可视分析体系,为快速有效地从多模态时空大数据中发现价值,诊断问题,检验预测以及探索未知规律提供以空间思维为中枢的"超级大脑"。

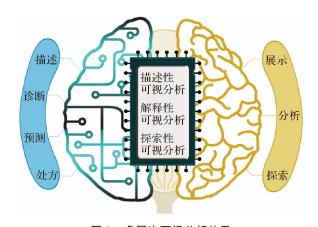


图 1 多层次可视分析体系

Fig.1 Multi-level visual analysis framework

如图 2 所示,描述性分析是对多模态时空对

象的时空参照、位置、空间形态、属性、行为、认知、 组成结构和关联关系等实际状况的形式化表达与 量化描述,如实时路况。诊断性分析旨在验证或 建立涵盖多模态时空对象的分布态势、关联关系 及异常表现等方面的假设,发现时空现象演化过 程中隐含的模式与规律,典型的有异常分析和演 化分析等功能,如道路拥堵分析。预测性分析是 在掌握多模态时空对象现有特征、模式与规律的 基础上,对其发展变化的状态、趋势与过程进行预 测,包括空间预测、时空预测、数据流预测、轨迹预 测等功能,如短时的路况预测。处方性分析针对 具体的任务,综合多模态时空对象的特征、规律与 演化态势,为决策者提供优化、调控与干预的可行 方案,包括优化分析和智能决策等功能,如交通路 口信号灯的智能控制。描述性分析和诊断性分析 通常是一种延时处理,随着传感网、物联网和社交 网等实时动态数据的在线接入处理,实时分析计 算成为多模态时空数据处理的典型特征。预测性 分析和处方性分析为前瞻性分析,结合历史数据 和实时动态数据对发展趋势进行预测并为决策提 供依据。

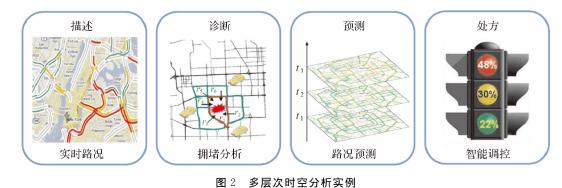


Fig.2 Example of multi-level spatio-temporal analysis

展示性可视化主要以多模态时空数据、信息和知识的高效表达与传递为基本目标,未来将发展离散-连续、动-静、真实感-抽象、精细-概略场景相宜的自适应表达方法以及与真实场景高度融合的协同可视化表达方法,基于语义级视觉变量映射的全空间增强现实场景动态生成方法。分析性可视化则要充分准确地表达复杂计算分析获取的时空数据所隐含的信息与知识,突出任务所关注的特征,未来将发展如多焦点群集对象变形可视化。探索性可视化将发展语义级视觉变量和新型人机交互界面,通过场景聚焦、变形、选择、突出和

简化等全空间增强现实表达,实现数据、人脑、机器智能和应用场景 4 方面的有机耦合,进行假设验证、知识归纳和推理论断等深度关联分析,将研究适合复杂环境的多机多用户协同交互模式和位置敏感的新型人机界面,研究面向大规模多粒度时空对象复杂数据分析的可视化探索、筛选、映射和布局等方法。

4 结束语

随着智慧城市建设的快速推进,人机物三元空间的深度融合对多模态时空大数据可视分析提

出了前所未有的新挑战,也为全空间信息系统的发展提供了新机遇。大数据时代的空间信息系统无疑将成为一个"超级大脑",高效灵活地综合处理不同来源不同性质的时空数据,为智慧城市精准化与智能化管理提供强大的全空间全信息时空关联、由浅入深的多层次可视分析能力。

致谢:感谢中南大学邓敏教授、电子科技大学 周艳副教授和邵俊明教授、浙江大学郑文庭副教 授对论文提出的建议,感谢西南交通大学博士生 李赟、刘铭崴、冯斌、张骏骁协助制图及对此文作 出的贡献。

参考文献:

- [1] HEY T, TANSLEY S, TOLLE K. The Fourth Paradigm: Data-intensive Scientific Discovery[M]. Redmond, Washington; Microsoft Research, 2009.
- [2] ANDRIENKO G, ANDRIENKO N, DEMSAR U, et al. Space, Time and Visual Analytics[J]. International Journal of Geographical Information Science, 2010, 24(10): 1577–1600.
- [3] KEIM D A, MANSMANN F, THOMAS J. Visual Analytics: How Much Visualization and How Much Analytics? [J]. ACM SIGKDD Explorations Newsletter, 2009, 11(2): 5-8.
- [4] KEIM DA, MANSMANN F, SCHNEIDEWIND J, et al. Visual Analytics: Scope and Challenges [M] // SIMOFF S J, BÖHLEN MH, MAZEIKA A. Visual Data Mining. Berlin Heidelberg: Springer-Verlag, 2008.
- [5] MENG L, ZIPF A, WINTER S. Map-based Mobile Services[M]. Berlin: Springer, 2005.
- [6] BERTONE A, BURGHARDT D. A Survey on Visual Analytics for the Spatio-temporal Exploration of Microblogging Content[J]. Journal of Geovisualization and Spatial Analysis, 2017, 1(1-2): 2.
- [7] 周成虎. 全空间地理信息系统展望[J]. 地理科学进展, 2015, 34(2): 129-131.

 ZHOU Chenghu. Prospects on Pan-spatial Information System [J]. Progress in Geography, 2015, 34(2): 129-131.
- [8] 华一新. 全空间信息系统的核心问题和关键技术[J]. 测绘科学技术学报, 2016, 33(4): 331-335.

 HUA Yixin. The Core Problems and Key Technologies of Pan-spatial Information System[J]. Journal of Geomatics Science and Technology, 2016, 33(4): 331-335.
- [9] 李德仁. 展望大数据时代的地球空间信息学[J]. 测绘学报, 2016, 45(4): 379-384.

 LI Deren. Towards Geo-spatial Information Science in Big Data Era[J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(4): 379-384.
- [10] 章毅, 郭泉, 王建勇. 大数据分析的神经网络方法[J]. 工程科学与技术, 2017, 49(1): 9-18.

 ZHANG Yi, GUO Quan. WANG Jianyong. Big Data Analysis

- Using Neural Networks[J]. Advanced Engineering Sciences, 2017, 49(1): 9-18.
- [11] MCGURK H, MACDONALD J. Hearing Lips and Seeing Voices[J]. Nature, 1976, 246(5588): 746-748.
- [12] ZHENG Yu. Methodologies for Cross-Domain Data Fusion: An Overview[J]. IEEE Transactions on Big Data, 2015, 1 (1): 16-34.
- [13] LAHAT D, ADALI T, JUTTEN C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects
 [J]. Proceedings of the IEEE, 2015, 103(9): 1449-1477.
- [14] KHALEGHI B, KHAMIS A, KARRAY F O, et al. Multi-sensor Data Fusion: A Review of the State-of-the-Art[J]. Information Fusion, 2013, 14(1): 28-44.
- [15] TURK M. Multimodal Interaction: A Review[J]. Pattern Recognition Letters, 2014(36): 189-195.
- [16] 马昱欣,曹震东,陈为. 可视化驱动的交互式数据挖掘方法综述[J]. 计算机辅助设计与图形学学报,2016,28(1): 1-8.

 MA Yuxin, CAO Zhendong, CHEN Wei. A Survey of
 - Visualization-driven Interactive Data Mining Approaches [J]. Journal of Computer-Aided Design and Computer Graphics, 2016, 28(1): 1-8.
- [17] DE OLIVEIRA M C F, LEVKOWITZ H. From Visual Data Exploration to Visual Data Mining: A Survey[J]. IEEE Transactions on Visualization and Computer Graphics, 2003, 9(3): 378-394.
- [18] 陈为. 大数据可视化与可视分析[J]. 金融电子化, 2015 (11): 62-65.

 CHEN Wei. Big Data Visualization and Visual Analysis
 [J]. Financial Computerizing, 2015(11): 62-65.
- ZHUANG Yueting, WU Fei, CHEN Chun, et al. Challenges and Opportunities: From Big Data to Knowledge in AI 2.0
 [J]. Frontiers of Information Technology and Electronic Engineering, 2017, 18(1): 3-14.
- [20] 朱庆. 三维 GIS 及其在智慧城市中的应用[J]. 地球信息科学学报, 2014, 16(2): 151-157.

 ZHU Qing. Full Three-Dimensional GIS and Its Key Roles in Smart City[J]. Journal of Geo-Information Science, 2014, 16(2): 151-157.
- [21] AIGNER W, MIKSCH S, MÜLLER W, et al. Visualizing Time-Oriented Data: A Systematic View[J]. Computers and Graphics, 2007, 31(3): 401-409.
- [22] MÜLLER W, SCHUMANN H. Visualization Methods for Time-Dependent Data: An Overview [C] // Proceedings of 2003 Winter Simulation Conference. New Orleans, LA, USA: IEEE, 2003.
- [23] CARLIS J V, KONSTAN J A. Interactive Visualization of Serial Periodic Data[C] // Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology. San Francisco, California, USA: ACM, 1998: 29-38.
- [24] TOMINSKI C, ABELLO J, SCHUMANN H. Axes-based Visualizations with Radial Layouts [C] // Proceedings of

- the 2004 ACM Symposium on Applied Computing. Nicosia, Cyprus: ACM, 2004: 1242-1247.
- [25] DEMŠAR U, VIRRANTAUS K. Space-time Density of Trajectories: Exploring Spatio-temporal Patterns in Movement Data[J]. International Journal of Geographical Information Science, 2010, 24(10): 1527-1542.
- [26] ANDRIENKO G, ANDRIENKO N. Visual Analytics for Geographic Analysis, Exemplified by Different Types of Movement Data[M]//POPOVICH V V, CLARAMUNT C, SCHRENK M, et al. Information Fusion and Geographic Information Systems. Berlin; Springer, 2009.
- [27] 陈为,朱标,张宏鑫. BN-Mapping: 基于贝叶斯网络的地理空间数据可视分析[J]. 计算机学报, 2016, 39(7): 1281-1293.

 CHEN Wei, ZHU Biao, ZHANG Hongxin. BN-Mapping:

Visual Analysis of Geospatial Data with Bayesian Network[J]. Chinese Journal of Computers, 2016, 39(7): 1281-1293.

- [28] COSSALTER M, MENGSHOEL O, SELKER T. Visualizing and Understanding Large-scale Bayesian Networks [J]. Proceedings of the 25th AAAI Conference on Artificial Intelligence, AAAI Workshop on Scalable Integration of Analytics and Visualization. San Francisco, AAAI, 2011.
- [29] 孙扬, 蒋远翔, 赵翔, 等. 网络可视化研究综述[J]. 计算机科学, 2010, 37(2): 12-18, 30.

 SUN Yang, JIANG Yuanxiang, ZHAO Xiang, et al. Survey on the Research of Network Visualization[J]. Computer Science, 2010, 37(2): 12-18, 30.
- [30] DIAZ I, CUADRADO A A, PÉREZ D, et al. Interactive Dimensionality Reduction for Visual Analytics [C] // Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium, 2014.
- [31] BRUZZESE D, DAVINO C. Visual Mining of Association Rules[M]//SIMOFF S J, BÖHLEN M H, MAZEIKA A. Visual Data Mining. Berlin, Heidelberg: Springer, 2008: 103-122.

- [32] HOFMANN H, SIEBES A P J M, WILHELM A F X. Visualizing Association Rules with Interactive Mosaic Plots [C] // Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Boston, Massachusetts, USA: ACM, 2000; 227-235.
- [33] KANDOGAN E. Visualizing Multi-dimensional Clusters, Trends, and Outliers Using Star Coordinates [C] // Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California; ACM, 2001; 107-116.
- [34] CHEN Keke, LIU Ling. VISTA: Validating and Refining Clusters Via Visualization[J]. Information Visualization. 2004, 3(4): 257-270.
- [35] ZHANG Kebing, ORGUN M A, ZHANG Kang, et al. Hypothesis Oriented Cluster Analysis in Data Mining by Visualization[C]//Proceedings of the Working Conference on Advanced Visual Interfaces. Venezia, Italy: ACM, 2006: 254-257.
- [36] ROBINSON A C, DEMŠAR U, MOORE A B, et al. Geospatial Big Data and Cartography: Research Challenges and Opportunities for Making Maps that Matter [J]. International Journal of Cartography, 2017; 1-29. DOI: 10.1080/23729333.2016.1278151.

(责任编辑:张艳玲)

收稿日期: 2017-06-02 修回日期: 2017-07-24

第一作者简介: 朱庆(1966—),男,博士,长江学者特聘教授,博士生导师,研究方向为多维动态 GIS 与虚拟地理环境。

First author: ZHU Qing (1966—), male, PhD, professor, PhD Yangtze River Scholar distinguished supervisor, majors in dynamic GIS and VGE.

E-mail: zhuq66@263.net