



# (12)发明专利申请

(10)申请公布号 CN 106547842 A

(43)申请公布日 2017.03.29

(21)申请号 201610898824.4

(22)申请日 2016.10.14

(71)申请人 华东师范大学

地址 200241 上海市闵行区东川路500号

(72)发明人 王忠樑 朱良峰

(74)专利代理机构 上海蓝迪专利商标事务所

(普通合伙) 31215

代理人 徐筱梅 张翔

(51)Int.Cl.

G06F 17/30(2006.01)

G06F 17/27(2006.01)

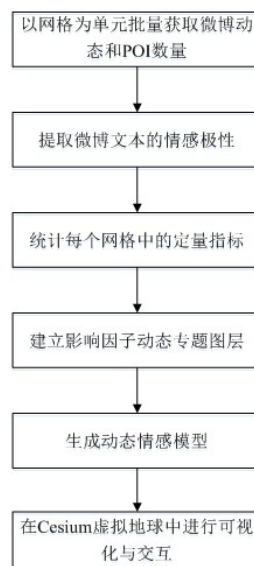
权利要求书2页 说明书13页 附图4页

## (54)发明名称

一种在虚拟地球平台上可视化基于位置的情感的方法

## (57)摘要

本发明公开了一种在虚拟地球平台上可视化基于位置的情感的方法,该方法首先利用新浪微博API获取微博动态和POI数量,从中提取微博文本情感极性和情感影响因子的定量指标;然后以网格为统计单元统计定量指标,在此基础上利用CZML语言建立动态的专题图层和情感可视化模型;最后将这些模型加载到Cesium平台中进行动态三维可视化及浏览、交互操作。本发明生成的微博情感模型可以表达时间和空间尺度上的动态变化,同时能够一定程度上反映微博情感与其影响因子的变化关系,并且能够在互联网上进行集成与共享。



1. 一种在虚拟地球平台上可视化基于位置的情感的方法,该方法包括以下具体步骤:

步骤1、以网格为单元批量获取微博动态和POI数量

建立三张空的数据表格,分别命名为网格单元信息表、微博动态信息表及用户信息表;

将可视化地区划分成四边形网格单元,将每一个网格的编号、网格中心点经度、网格中心点纬度、网格边长和搜索半径存入网格单元信息表;其中搜索半径为网格最小外接圆的半径;

用网格单元信息表中每一个网格的中心经纬度为圆心,搜索半径为圆的半径,定义圆形区域,将该圆形区域作为搜索范围;

用新浪微博API获取搜索范围内带地理标签的微博动态,得到微博动态文本信息和用户信息;分别存入微博动态信息表 and 用户信息表;

用新浪微博API获取搜索范围内的POI数量,将结果存入网格单元信息表中的POI数量字段;

步骤2、提取微博文本的情感极性

用基于语料库的情感分析方法对收集到的每一条微博进行文本情感极性评分,然后将分数转化成取值范围在0到10之间的情感极性分数,0代表消极情感的极端,10代表积极情感的极端;最后将结果存入微博动态信息表中的情感值EV字段;

步骤3、统计每个网格中的定量指标

计算网格内的所有微博的平均情感分数AEV;AEV取值范围在0到10之间,0代表消极情感的极端,10代表积极情感的极端;AEV的计算见下式,其中EmoValue代表网格内一条微博的情感值,n代表网格内的微博总数;

$$AEV = \frac{\sum EmoValue}{n}$$

计算网格内微博数WC和用户人数UC;如下式,其中w代表网格中的一条微博,u代表网格中的一个用户;

$$WC = \sum w$$

$$UC = \sum u$$

计算网格中的平均好友数AFC,如下式,其中f代表网格中一个用户的好友数,UC代表网格中的用户人数;

$$AFC = \frac{\sum f}{UC}$$

步骤4、建立影响因子动态专题图层

用色彩作为POI数量的视觉变量,通过网格色彩的透明度差异表达POI数量的变化,透明度越高,色彩越暗淡,代表POI数量越少;反之则越多;POI数量属于静态数据,用GeoJSON格式编码,将POI数量嵌入GeoJSON的properties属性中;

用色彩亮度和饱和度作为微博数量WC、平均好友数AFC和用户人数UC的视觉变量,饱和度和亮度越大,代表该定量指标的数值越大;用CZML标记语言动态模拟这三个定量指标:将这三个定量指标分别以CZML文件的格式进行编码;用rectangle标签模拟四边形网格,用material中的solidColor子属性和Color子属性设置;在Color的rgba子属性中记录具体时刻和该时刻对应的颜色RGBA值,让网格在不同时刻拥有不同颜色样式;

#### 步骤5、生成动态情感模型

建立一个侧面镂空的动态直四棱柱模型；其中直四棱柱的顶面与网格四边形的面积和形状都相同；直四棱柱突起的高度代表影响因子定量指标的数值大小；顶面和侧棱的颜色代表情感的正负程度，色彩方案如下：用红色和蓝色两种颜色表达情感的正负极性，红色代表正面积积极的倾向，蓝色代表负面消极的倾向；色彩的饱和度表示正负倾向的程度，最饱和的红色代表正面情感的极点，此时AEV等于10，最饱和的蓝色代表负面情感的极点，此时AEV等于0，红色和蓝色之间的过渡色代表10与0之间的值，此时AEV的值在0到10之间；

直四棱柱模型由两个CZML文件组合而成；一个是完全镂空的直四棱柱文件，用于存储完全镂空直四棱柱；用rectangle标签创建一个四边形要素，用extrudedHeight属性实现柱形高度随时间变化，再利用outlineColor属性实现侧棱颜色随时间变化；

另一个是直四棱柱的顶面文件，用于存储覆盖在直四棱柱之上的顶面四边形；顶面文件利用rectangle标签创建一个四边形片状要素，用material中的solidColor子属性设定四边形顶面颜色的变化；另外，用height属性控制四边形平面所在的高度，与镂空直四棱柱文件中对应柱形的extrudedHeight数值保持相同；

#### 步骤6、在Cesium虚拟地球中进行可视化与交互

将生成的专题图层和情感模型加载到Cesium地球平台中，进行可视化与交互设计；POI数量图层以GeoJSON格式载入Cesium，利用GeoJSON文件内存储的属性数据和Cesium提供的函数生成POI数量的色彩视觉变量；动态专题图层和动态情感模型以CZML的格式载入Cesium，能够直接进行动态可视化。

## 一种在虚拟地球平台上可视化基于位置的情感的方法

### 技术领域

[0001] 本发明涉及地球空间信息技术领域中的数字地球技术和自然语言处理领域的情感分析技术,具体地说是一种利用VGI数据获取基于位置的情感,并可视化在虚拟地球平台上的方法,该方法生成的微博情感模型可以表达情感和其影响因子在时间和空间尺度上的动态变化,同时能够一定程度上反映微博情感与其影响因子的变化关系,并且能够在互联网上进行集成与共享。

### 背景技术

[0002] 基于位置的情感(location-based emotions,简称LBEs)是带有地理标签的情感数据。在基于位置的服务(location-based services,简称LBS)中,它作为一种用户主观认知层面的数据,被广泛应用。基于位置的情感研究是将情感分析与空间信息相结合的交叉研究领域。研究学者通过在地图上描绘情感要素的方法,直观的展示情感的空间分布,对研究集体情感和区域活动的成因、变化趋势和结果都有重要意义。目前,该领域有很多以个人为单位进行基于位置情感的获取和分析的案例。在2004年的Bio Mapping项目中,参与者通过手持GPS和生物传感器生成地理坐标和极端情绪反馈,这些数据由制图工作者统计并绘制成情感地图。EmBaGIS项目通过收集残疾人士的情感和位置信息为“无障碍规划”的城市规划目标提供帮助。此外,大量的情感集聚还可以为智能导航和LBS推荐系统提供数据支持。然而,依靠手持设备监测人体情感的方法,耗费人力物力和财力,使得基于位置的情感研究陷入了窘境。

[0003] 随着移动设备的普及和信息传输能力的提升,用户志愿提供的志愿地理信息(Volunteered Geographic Information,简称VGI)数据在社交网络、在线社区和点评网站上广为流行。作为一种带有地域性和主观性的海量地理数据,VGI为基于位置的情感研究提供了新的数据来源。Twitter公司利用带有地理标签的推文提取情感指标,绘制了Twitter mood地图。地图中用圆圈表示地点,圆圈的大小表示推特数量,圆圈的颜色代表不同的情感。此后,EmoMap项目用同样的方法获取基于位置的情感数据,为市议会提供决策帮助。EMMA项目通过图片社交软件Flickr上的标签,提取基于位置的情感,研究旅游的动机和预期。在众多网络社交工具中,微博是VGI数据重要的数据来源,它作为一种分享和交流平台,更注重时效性和随意性,能表达出用户每时每刻的思想和动态。在中国,研究人员基于新浪微博的文本和标签,获取用户所表达的情感倾向和地理位置,分析集体情感和热点事件。基于位置的情感研究在VGI的帮助下,在舆情监控、非常规事件预测、城市规划满意度调查和公众心理健康分析领域得到了广泛的应用。

[0004] 通过VGI获取的情感数据具有共享、海量和时空变化的特点,这给数据的获取、管理和表达带来了挑战。在大数据时代,各个行业对基于位置的情感数据需求日益增加,许多跨学科工作人员和非专业人士也被吸引到这个领域。他们不仅需要一种获取和存储数据的通用方法,而且对数据的共享和实时动态的数据表达提出了的要求。情感的研究学者可能需要同时对集体情感和它的影响因子进行分析;地理工作者也许更注重情感或影响因子本

身的时空变化;企业级用户可能对获取数据的方法和数据的组织形式更感兴趣;普通民众可能需要借助实时的情感分析结果选择更好的位置服务。这种跨学科跨领域的研究和应用亟需一个集成了海量数据提取、时空数据组织、动态表达和情感影响因子可视化的地理信息共享平台。

[0005] 虚拟地球是一种表达地理数据的三维可视化平台,其可视化范围大到全球尺度的地壳模型,小到一个城市的设施点。它的优点是具有多维的空间尺度和友好直观的三维可视化效果,适合向不同专业程度的人群展示和分发地理数据。Google Earth是当前最流行的虚拟地球平台。利用KML标记语言可以在Google Earth可视化多样的三维地理实体。有学者在GoogleEarth虚拟地球平台中利用时空立方体的方法表达用户移动电话使用情况的时空信息。不过相比静态的时空立方体,动态的显示时空信息对于非专业人士而言是更直观、易懂,并且更贴近现实的表现形式。然而无论是Google Earth还是其他虚拟地球平台,都不能很好地支持地理实体的动态变化的可视效果。

[0006] Cesium虚拟地球平台是由Analytical Graphics公司(AGI)开发的开源网页端地图引擎。与其他虚拟地球平台相比,Cesium支持一种名为CZML的数据文件格式,可以动态的模拟地理实体空间位置和形态随时间的变化,常用于人造卫星、飞行器和运动实体的轨迹模拟以及时空数据的动态可视化。但是目前在虚拟地球平台(包括Cesium)上,尚未有对基于位置的情感进行时空动态可视化的案例。

## 发明内容

[0007] 本发明的目的是提供一种更为友好直观的方式动态地描绘基于位置的情感和其影响因子的关系和时空变化,实现数据的批量获取和情感分析,并且在互联网上共享、展示和应用。

[0008] 本发明要解决的技术问题可分解为两个方面:要解决从微博文本和标签中批量提取基于位置的情感指标和影响因子的问题;要解决基于位置的情感在时空尺度上动态可视化表达的问题。

[0009] 本发明所采用的具体技术方案是:利用VGI数据获取基于位置的情感并可视化在虚拟地球平台上的方法,该方法首先利用新浪微博API获取微博动态和POI数量,从中提取微博文本情感极性和情感影响因子的定量指标;然后以网格为统计单元统计定量指标,在此基础上利用CZML语言建立动态的专题图层和情感可视化模型;最后将这些模型加载到Cesium平台中进行动态三维可视化及浏览、交互操作。

[0010] 本发明包括以下具体步骤:

[0011] 步骤1、以网格为单元批量获取微博动态和POI数量

[0012] 建立三张空的数据表格,分别命名为网格单元信息表、微博动态信息表及用户信息表;

[0013] 将可视化地区划分成四边形网格单元,将每一个网格的编号、网格中心点经度、网格中心点纬度、网格边长和搜索半径存入网格单元信息表;其中搜索半径为网格最小外接圆的半径;

[0014] 用网格单元信息表中每一个网格的中心经纬度为圆心,搜索半径为圆的半径,定义圆形区域,将该圆形区域作为搜索范围;

[0015] 用新浪微博API获取搜索范围内带地理标签的微博动态,得到微博动态文本信息和用户信息;分别存入微博动态信息表 and 用户信息表;

[0016] 用新浪微博API获取搜索范围内的POI数量,将结果存入网格单元信息表中的POI数量字段;

[0017] 步骤2、提取微博文本的情感极性

[0018] 用基于语料库的情感分析方法对收集到的每一条微博进行文本情感极性评分,然后将分数转化成取值范围在0到10之间的情感极性分数,0代表消极情感的极端,10代表积极情感的极端;最后将结果存入微博动态信息表中的情感值 (EmoValue) 字段;

[0019] 步骤3、统计每个网格中的定量指标

[0020] 计算网格内的所有微博的平均情感分数 (average emotion value,简称AEV);AEV取值范围在0到10之间,0代表消极情感的极端,10代表积极情感的极端;AEV的计算见下式,其中EmoValue代表网格内一条微博的情感值,n代表网格内的微博总数;

$$[0021] \quad AEV = \frac{\sum EmoValue}{n}$$

[0022] 计算网格内微博数 (WC) 和用户人数 (UC);如下式,其中w代表网格中的一条微博,u代表网格中的一个用户;

$$[0023] \quad WC = \sum w$$

$$[0024] \quad UC = \sum u$$

[0025] 计算网格中的平均好友数 (average friends count,简称AFC),如下式,其中f代表网格中一个用户的好友数,UC代表网格中的用户人数;

$$[0026] \quad AFC = \frac{\sum f}{UC}$$

[0027] 步骤4、建立影响因子动态专题图层

[0028] 用色彩作为POI数量的视觉变量,通过网格色彩的透明度差异表达POI数量的变化,透明度越高,色彩越暗淡,代表POI数量越少;反之则越多;POI数量属于静态数据,用GeoJSON格式编码,将POI数量嵌入GeoJSON的properties属性中;

[0029] 用色彩亮度和饱和度作为微博数量 (WC)、平均好友数 (AFC) 和用户人数 (UC) 的视觉变量,饱和度和亮度越大,代表该定量指标的数值越大;用CZML标记语言动态模拟这三个定量指标:将这三个定量指标分别以CZML文件的格式进行编码;用rectangle标签模拟四边形网格,用material中的solidColor子属性和Color子属性设置;在Color的rgba子属性中记录具体时刻和该时刻对应的颜色RGBA值,让网格在不同时刻拥有不同颜色样式;

[0030] 步骤5、生成动态情感模型

[0031] 建立一个侧面镂空的动态直四棱柱模型;其中直四棱柱的顶面与网格四边形的面积和形状都相同;直四棱柱突起的高度代表影响因子定量指标的数值大小;顶面和侧棱的颜色代表情感的正负程度,色彩方案如下:用红色和蓝色两种颜色表达情感的正负极性,红色代表正面积积极的倾向,蓝色代表负面消极的倾向;色彩的饱和度表示正负倾向的程度,最饱和的红色代表正面情感的极点,此时AEV等于10,最饱和的蓝色代表负面情感的极点,此时AEV等于0,红色和蓝色之间的过渡色代表10与0之间的值,此时AEV的值在0到10之间;

[0032] 直四棱柱模型由两个CZML文件组合而成;一个是完全镂空的直四棱柱文件,用于

存储完全镂空直四棱柱;用rectangle标签创建一个四边形要素,用extrudedHeight属性实现柱形高度随时间变化,再利用outlineColor属性实现侧棱颜色随时间变化;

[0033] 另一个是直四棱柱的顶面文件,用于存储覆盖在直四棱柱之上的顶面四边形;顶面文件利用rectangle标签创建一个四边形片状要素,用material中的solidColor子属性设定四边形顶面颜色的变化;另外,用height属性控制四边形平面所在的高度,与镂空直四棱柱文件中对应柱形的extrudedHeight数值保持相同;

[0034] 步骤6、在Cesium虚拟地球中进行可视化与交互

[0035] 将生成的专题图层和情感模型加载到Cesium地球平台中,进行可视化与交互设计;POI数量图层以GeoJSON格式载入Cesium,利用GeoJSON文件内存储的属性数据和Cesium提供的函数生成POI数量的色彩视觉变量;动态专题图层和动态情感模型以CZML的格式载入Cesium,能够直接进行动态可视化。

[0036] 本发明的有益效果

[0037] 本发明具有步骤简明、自动性高的优势,具有较高的实用性和较强的扩展性与可复用性。使用本发明生成的微博情感模型可以表达时间和空间尺度上的动态变化,同时能够一定程度上反映微博情感与其影响因子的变化关系,并且能够在互联网上进行集成与共享。本发明为开发一个集成海量数据提取、时空数据组织、动态表达和情感影响因子可视化的微博情感信息共享平台提供了总体框架,为基于位置的情感的跨学科跨领域的研究和应用奠定了基础。

## 附图说明

[0038] 图1是本发明流程示意图;

[0039] 图2是本发明使用的直四棱柱情感模型示意图,其中横轴代表平均情感分数AEV,纵轴代表一种影响因子的定量指标,在实施例是微博数,图中(1)、(2)、(3)依次为红色、白色和蓝色的直四棱柱;

[0040] 图3是本发明实施例上海市中心城区的1000×1000米网格示意图;

[0041] 图4是本发明实施例实现的基于位置情感模型动态可视化平台——EmoGrids的网页在计算机上展示的效果示意图;

[0042] 图5是本发明实施例实现的多维度、多视角切换视图功能图,(a)为三维视角下的侧视图,(b)为二维视角下的顶视图;

[0043] 图6是本发明实施例实现的影响因子专题图层显示的效果示意图,(a)为餐馆POI数量图层,(b)为公园POI数量图层,(c)为平均好友数图层,(d)为用户人数图层。

## 具体实施方式

[0044] 下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0045] 参阅图1所示,本发明首先利用新浪微博API获取微博动态和POI数量,从中提取微博文本情感极性和情感影响因子的定量指标;然后以网格为统计单元统计定量指标,在此基础上利用CZML语言建立动态的专题图层和情感可视化模型;最后将这些模型加载到Cesium平台中进行动态三维可视化及浏览、交互操作。本发明可在微机平台下采用PHP,Python语言开发实现,具体实施方式如下:

## [0046] 数据来源

[0047] 本发明对基于位置的情感和情感的影响因子进行可视化研究。数据来源是新浪微博。新浪微博是全中国最主流,最具人气的微博产品,截至2014年3月,微博月活跃用户1.438亿,日活跃用户6660万。微博情感常常表现在微博文本和表情上。通过文本分析可以提取情感指标。表达情感常用的指标是情感的极性,也就是情感的正负面倾向。这一指标构造的情感模型将情感的正负程度放在一维的线段上表达,两个端点分别表示积极和消极情感的极点,是最简单易懂的构造方法。因此,本发明选择情感极性作为微博情感的定量指标。

## [0048] 情感指标与影响因子选取

[0049] 情感的影响因子也在本发明可视化的范畴。影响情感的因子有内在的心理因素,也有外部环境和社会关联等因子。本发明中更关注与空间相关联的外部影响因子。因此,可以将本发明中涉及的影响因子分为两类,一类是环境因子,另一类是社会关系。在城市化加速的大都市,城市环境对居民情感的影响不可忽视,天气、噪音、基础设施、社会密集度、开放空间、周边环境和周围绿化都能成为环境因子。社会关系反映人与人之间的联通性,在微博中指的就是用户与他人的关联程度。有研究显示,与他人关联更紧密的用户在社交网络中所发的状态更积极向上,这说明社会关系对情感的影响在社交网络中同样适用。本发明的影响因子用定量化的指标表达:环境因子中,选择用户人数和微博条数表达实时的社会密集度,用餐饮POI作为周边环境的定量指标;公园POI数量作为开放空间和绿地的定量指标;社会关系则用微博用户好友数来量化。

## [0050] 表1影响因子和定量指标

[0051]

影响因子	定量指标
<b>环境因子</b>	
社会密集度	微博用户人数、微博条数
周边环境	餐饮 POI 数量
开放空间和绿地	公园 POI 数量
<b>社会关系</b>	微博用户好友数

[0052] 因为微博情感和其影响因子的数据量巨大,以单个坐标点作为可视化单元会占用大量的内存资源,而且可视化效果不佳。所以本发明在数据收集、统计和可视化时,选择以四边形网格为单元。

## [0053] 第一步 以网格为单元批量获取微博动态和POI数量

[0054] 首先,将可视化地区划分成四边形网格单元,然后利用新浪微博API获取各个网格单元内的微博动态、用户信息和POI数量,为后续的模型生成与可视化提供初始数据。

[0055] 以边长为L米的四边形网格单元为例,首先要建立一张存放网格空间信息的数据表“网格单元信息表”,包含的字段有网格编号、网格中心点经度、网格中心点纬度和搜索范围。之后的搜索、统计和处理都以网格单元信息表中的网格为基本单元。

[0056] 获取微博动态的工作原理是搜索给定中心点附近一定范围,获取此范围内带地理标签的微博动态。在此例中,搜索的中心点就是网格中心点,搜索范围设置为网格最小外接



圆的半径,约等于 $\sqrt{2}L$ 米。得到的结果(包含微博动态文本信息和用户信息)存入“微博动态信息表”和“用户信息表”。需要注意的是,有一些信息属于广告性质,需要基于微博标签进行筛选和剔除。

[0057] POI数量获取的工作原理与微博动态获取相同,搜索对象从微博动态变为POI点,返回的信息中直接包含搜索范围内POI数量。需要设置的参数除了中心点和搜索范围,还需要设置查询类型,也就是POI分类类型,比如“餐饮”或者“公园”类型的POI。结果存入网格单元信息表中。

[0058] 表2网格单元信息表

字段名	数据类型	说明
gridID	INT(11)	网格 ID
long	FLOAT	中心点经度
lat	FLOAT	中心点纬度
numCaterPOI	INT(11)	网格内餐饮服务 POI 数量
numParkPOI	INT(11)	网格内公园 POI 数量

[0060] 表3微博动态信息表

字段名	数据类型	说明
id	BIGINT(64)	微博动态的 ID
gridID	INT(11)	所在网格 ID
userID	BIGINT(64)	用户 ID
create_at	DATETIME	发表时间
long	FLOAT	所在经度
lat	FLOAT	所在纬度
text	TEXT	微博文本
emoValue	FLOAT	情感极性分数, 在提取微博文本情感极性之后计算获取

[0062] 表4用户信息表

[0063]

字段名	数据类型	说明
userID	BIGINT(64)	用户 ID
gender	TINYTEXT	用户性别
age	INT(11)	用户年龄
friendsCount	INT(11)	用户好友数

[0064] 第二步提取微博文本的情感极性

[0065] 情感极性是情感正负面的倾向和程度。评价微博文本中用户态度的积极和消极程度,量化成情感极性。本发明采用自然语言处理技术中基于语料库的情感分析方法提取文本情感极性。判断文本情感正负面程度的基本流程如下:1.对文本进行分词操作;2.基于情感知识的语料库对每一个词进行正负面的判断;3.统计文本中情感正负面词的数量;4.对文本进行定量的情感极性评分。

[0066] 本发明使用自然语言处理与信息检索(NLPIR)共享平台提供的汉语分词工具和情

感组件对收集到的每一条微博进行文本情感极性评分,然后将分数转化成取值范围在0到10之间的情感极性分数,0代表消极情感的极端,10代表积极情感的极端。最后将结果存入微博动态信息表中的情感值 (EmoValue) 字段。

[0067] 第三步 统计每个网格中的定量指标

[0068] 定量指标中POI数量是在数据获取时就根据网格统计完成的,其他的定量指标都需要基于网格进行统计,获得能代表网格单元的单一数值。

[0069] 网格的情感极性分数是该区域集体情感的定量指标,用网格内的单条微博的平均情感分数 (average emotion value,简称AEV) 表示。理论上,AEV和EmoValue的取值范围相同,都在0到10之间,0代表消极情感的极端,10代表积极情感的极端。AEV的计算见下式,其中EmoValue代表网格内一条微博的情感分数,n代表网格内的微博总数。

$$[0070] \quad AEV = \frac{\sum EmoValue}{n}$$

[0071] 网格的微博数 (WC) 和用户人数 (UC) 反映网格内的社会密集程度,计算方法是分别求网格内两项指标的总和,如下式,其中w代表网格中的一条微博,u代表网格中的一个用户;

$$[0072] \quad WC = \sum w$$

$$[0073] \quad UC = \sum u$$

[0074] 好友数代表网格内社会关系的强弱程度,为了消除社会密集程度对统计结果的影响,本发明计算网格中的平均好友数 (average friends count,简称AFC),如下式,其中f代表网格中一个用户的好友数,UC代表网格中的用户人数;

$$[0075] \quad AFC = \frac{\sum f}{UC}$$

[0076] 第四步 建立影响因子动态专题图层

[0077] 根据不同数据的特征为情感的影响因子设计视觉变量,建立影响因子的动态专题图层。数据需要通过地图符号系统在虚拟地球上进行可视化,其本质上是视觉变量的参数设置。尺寸、色相、方位、形状、位置、纹理以及密度都属于视觉变量的范畴。

[0078] 在现实中,不同的数据在不同的时间尺度下回表现出动态和静态两种特性:比如,某地区的POI数量在一周内是静态不变的数据,但是如果以一个月为采样间隔,每个采样时间点的POI数量就可能发生动态变化。为了研究一种同时适用于静态和动态数据的通用可视化方法,本发明将定量指标分为两个类型,一类是只在空间尺度上变化的空间数据,包括餐饮和公园POI数量;一类是在时空尺度上变化的时空数据,包括情感分数、微博数、用户人数和平均好友数。

[0079] 针对不同类型的数据需要设计不同样式的视觉变量,达到最优化可视效果。POI数量属于空间数据,表现形式是静态的专题图层。本发明选用色彩作为POI数量的视觉变量,通过网格色彩的透明度差异表达POI数量的空间变化,透明度越高,色彩越暗淡,代表POI数量越少;反之则越多。

[0080] 微博数量、平均好友数和用户人数以色彩亮度和饱和度作为视觉变量,饱和度和亮度越大,代表的数值越大。这三类数据属于时空数据,需要进行动态的模拟。在模拟数据动态变化之前,首先需要明确每一个关键时刻各个定量指标的具体数值,并转换为对应视

觉变量的参数。

[0081] 为了实现网格颜色随时间的动态变化,需要将数据以CZML文件的格式进行编码。用rectangle标签模拟四边形网格,用material中的solidColor子属性设定网格颜色。Color属性中的rgba可以在RGBA值之前标注具体时刻,让网格在不同时刻拥有不同颜色样式。每一个关键帧时刻的Color属性都是一串数组,排列方式是 $[T_1, R_1, G_1, B_1, \dots, T_n, R_n, G_n, B_n]$ 。其中 $T_n$ 代表第n个时刻的时间,以国际标准ISO 8601日期时间表示法书写; $R_n, G_n, B_n$ 代表该时刻实体的颜色RGBA值。

[0082] 使用rectangle标签定义网格单元随时间改变颜色的CZML代码如下:

[0083]

```
{
    "availability": "2015-12-09T00:00:00Z/2015-12-16T00:00:00Z", //设置动态显示时间
    段(set animation time range)
    "id": "0", //网格编号
    "rectangle": { //定义四边形网格对象
        "material": { //四边形材质属性
            "solidColor": { //设置材质为颜色填充
                "color": { //设置具体的颜色
                    "rgba": [ //设置特定时刻的网格颜色, 格式: 时间+RGBA 值
                        "2015-12-09T00:00:00Z",84,84,84,150,
                        "2015-12-10T00:00:00Z",0,0,0,150,
                        "2015-12-12T00:00:00Z",56,56,56,150,
                        "2015-12-14T00:00:00Z",84,84,84,150,
                        "2015-12-15T00:00:00Z",56,56,56,150,
                        "2015-12-16T00:00:00Z",28,28,28,150]
                }
            }
        }
    }
}
```

```

[0084]
        }
    }
},
"outline": 0,    //四边形边框不显示
"coordinates": { //设置四边形位置
    "wsenDegrees": [121.328,31.1309,121.338,31.1409] //西南东北边界的经纬
    度
},
"fill": 1    //填充四边形
},
"name": "grid0" //定义四边形名称
}

```

[0085] 第五步 生成动态情感模型

[0086] 本发明的情感模型是一种综合的可视化表达。该模型通过三维地图符号的视觉变量,不仅可以表达质量特征(正负极性),数量特征(正负程度),还可以描述关系特征(与影响因子的关联)和动态特征(随时间的变化)。

[0087] 情感的质量特征是指情感的正负极性,数量特征是指情感极性的正负程度,也就是AEV。为了同时可视化质量和数量特征,本发明选择在两种饱和颜色(红色与蓝色)过渡的方案,用红蓝两种颜色表达质量特征,用颜色的饱和度表达情感的数量特征。在情感极性中,用红色代表正面倾向,蓝色代表负面倾向,色彩的饱和度表示正负倾向的程度。最饱和的红色代表正面情感的极点(AEV=10),最饱和的蓝色代表负面情感的极点(AEV=0),中间的过渡色代表10与0之间的值。

[0088] 关系特征描述的是情感与其影响因子的关联,本发明将关系表达在一个侧面镂空的直四棱柱上(如图2)。直四棱柱的顶面与网格四边形的面积和形状都相同。顶面和侧棱的颜色代表情感的正负程度,色彩参数与上文相同,柱形突起的高度代表影响因子的数值大小。这种方法利用了三维空间、色彩要素的特性,表达了数据在多个维度的属性特征。

[0089] 动态特征是情感和影响因子随时间变化的表现。与动态专题图层相似,需要将数据以CZML文件的格式进行编码。为了建立直四棱柱模型,我们需要两个CZML文件,一个是完全镂空的直四棱柱文件,一个是直四棱柱的顶面文件,分别存储网格之上突起的镂空直四棱柱和覆盖在柱形之上的顶面四边形。

[0090] 镂空直四棱柱文件用rectangle标签创建一个四边形要素,用extrudedHeight属性实现柱形高度随时间变化,再利用outlineColor属性实现侧棱颜色随时间变化,以下CZML代码块示范了直四棱柱文件的单个实体的语法结构:

```
{
  "name": "grid0",
  "availability": "2015-12-09T00:00:00Z/2015-12-16T00:00:00Z",
  "rectangle": {
    "outline": 1,      //设置边框显示
    "extrudedHeight": {      //高度突起设置
      "number": [      //特定时间下突起高度的值
        "2015-12-09T00:00:00Z",1500,
        "2015-12-10T00:00:00Z",1100,
        "2015-12-12T00:00:00Z",1300,
        "2015-12-14T00:00:00Z",1400,
        "2015-12-15T00:00:00Z",1400,
        "2015-12-16T00:00:00Z",1200]
      },
    [0091] "coordinates": {
      "wsenDegrees": [121.328,31.1309, 121.338, 31.1409]
    },
    "outlineWidth": 4,      //设置边框宽度
    "fill": 0,              //设置各个面透明，镂空效果
    "outlineColor": {      //边框颜色设置
      "rgba": [            //特定时间下边框的颜色
        "2015-12-09T00:00:00Z",138,174,193,255,
        "2015-12-10T00:00:00Z",22,94,131,255,
        "2015-12-12T00:00:00Z",255,255,255,255,
        "2015-12-14T00:00:00Z",236,96,83,255,
        "2015-12-15T00:00:00Z",244,174,167,255,
        "2015-12-16T00:00:00Z",248,213,209,255]
      }
    },
  },
}
```

```

    "id": "0"
  [0092] }

```

[0093] 顶面文件利用rectangle标签创建一个四边形片状要素,用material中的solidColor子属性设定四边形顶面颜色,方法与动态专题图层相似。另外,用height属性控制四边形平面所在的高度,与镂空直四棱柱文件中对应柱形的extrudedHeight相同,以下CZML代码示范了顶面文件单个实体的代码结构:

```

  {
    "name": "grid0",
    "availability": "2015-12-09T00:00:00Z/2015-12-16T00:00:00Z",
    "rectangle": {
      "outline": 0,    //设置边框为透明
      "material": {
        "solidColor": {
          "color": {
            "rgba": [    //在特定时间下的顶面颜色
              "2015-12-09T00:00:00Z",138,174,193,255,
              "2015-12-10T00:00:00Z",22,94,131,255,
              "2015-12-12T00:00:00Z",255,255,255,255,
[0094] "2015-12-14T00:00:00Z",236,96,83,255,
              "2015-12-15T00:00:00Z",244,174,167,255,
              "2015-12-16T00:00:00Z",248,213,209,255]
            }
          }
        }
      }
    },
    "coordinates": {
      "wsenDegrees": [121.328,31.1309,121.338,31.1409]
    },
    "height": {    //在特定时间下的顶面所在高度,
      "number": ["2015-12-09T00:00:00Z",1500,
        "2015-12-10T00:00:00Z",1100,
        "2015-12-12T00:00:00Z",1300,

```

```
        "2015-12-14T00:00:00Z",1400,  
        "2015-12-15T00:00:00Z",1400,  
        "2015-12-16T00:00:00Z",1200]  
    },  
[0095]    "outlineWidth": 4,  
        "fill": 1  
    },  
    "id": "0"  
}
```

[0096] 为了保证柱形和顶面组合在一起形成一个整体，直四棱柱文件中的extrudedHeight属性与顶面文件中的height属性在任何时刻都应该保持数值相等。同时，直四棱柱文件中的outlineColor属性也要和顶面文件中的solidColor属性保持一致。

[0097] 可以看出，利用视觉变量的组合和三维空间的特性，可以在同一视图、同一空间中叠加多种类型的地理信息。结合CZML格式对时空数据的组织能力和Cesium平台的动态可视化效果，能够实现多个维度的时空数据可视化，形成一种直观、友好的数据组合表达方式，并且一定程度上表达数据之间的关系。

[0098] 第六步 在Cesium虚拟地球中进行可视化与交互

[0099] 将生成的专题图层和情感模型加载到Cesium地球平台中，进行可视化与交互设计。POI数量图层以GeoJSON格式载入Cesium，利用GeoJSON文件内存储的属性数据和Cesium提供的函数生成POI数量的视觉变量。动态专题图层和动态情感模型以CZML的格式载入Cesium，可以直接进行动态可视化。

[0100] 利用Cesium平台不仅可以将情感和影响因子的时空变化直接展示在地球表面，还可以进行多种交互操作。使用鼠标可以在三维地球空间中选择特定的网格单元，查询与网格对应的属性信息；从不同的视角（侧视图和顶视图）不同的维度（2D, 2.5D, 3D）观察数据的分布和变化；查询特定时间各个数据的空间分布情况；控制专题图层的显示、隐藏和叠加；控制动态显示的速率。

[0101] 实施例

[0102] 为实现本发明提出的方法，本发明以上海市中心城区为可视化区域进行了实例研究。上海中心城区指的是上海外环线以内的区域，面积约660平方公里，是全上海市人口最密集的地区。本发明用752个覆盖上海市中心城区的1000×1000米网格作为统计单元（图3），收集了从2015年12月9日到12月16日一周时间内公开的微博数据（共计47757条微博），进行可视化方法的应用。

[0103] 为了让数据收集和处理的办法能够扩展和复用，实施例中编写了微博数据的批量获取、情感评分和CZML生成（将数据库中的表格数据转换成CZML文件）的程序。其中，微博数据获取程序是基于PHP编写，通过访问新浪微博API获取数据，然后存入数据库表格中；情感评分程序集成了NLPIR的情感组件，对数据库中的微博文本逐条进行情感评分；CZML生成程

序利用Python编写,能够基于前文描述的可视化方法将表格数据转换成CZML文件。这三个程序能快速、批量地处理大量数据,并生成可以导入Cesium的动态三维模型文件。

[0104] 为了满足不同领域用户对数据共享和浏览的需求以及不同学科之间的合作研究和交流,实施例基于Cesium地图引擎设计了一个网页端的公众可视化平台EmoGrids,并将原始的微博数据、统计数据和CZML文件存放在了服务器上进行共享和分发。

[0105] 如图4所示,EmoGrids的可视化交互界面由控件和可视化窗口组成。任何支持WebGL的浏览器都可以通过国际互联网访问这个网页,自由浏览和查询微博情感和其影响因子的时空变化。

[0106] 用户打开EmoGrids可视化界面时,首先是侧视图的视角,呈现在眼前的是由红蓝直四棱柱组成的动态情感模型,随着浏览器下部时间条的推移,情感模型的色彩和柱形的高度也会随之改变,模拟微博情感和微博数的动态变化,可以辅助情感和影响因素关系的研究。用户可以通过网页上部的视角切换按钮切换成顶视图,或者用右上角的视图切换工具切换成二维视图(2D)(图5),这种视角的切换可以在正上方俯视整个网格区域,直观清晰地将一种定量指标在空间上的动态变化展示给用户,可以帮助研究情感和影响因素在空间和时间上的分布和变化规律。

[0107] 右下方的图层控制工具用于专题图层的控制,用户可以通过简单的操作切换不同的影响因素专题图层(图6)和情感模型。拖动下部的时间滑块可以定位到某一特定时刻,方便用户查询特定时间点数据的空间分布。



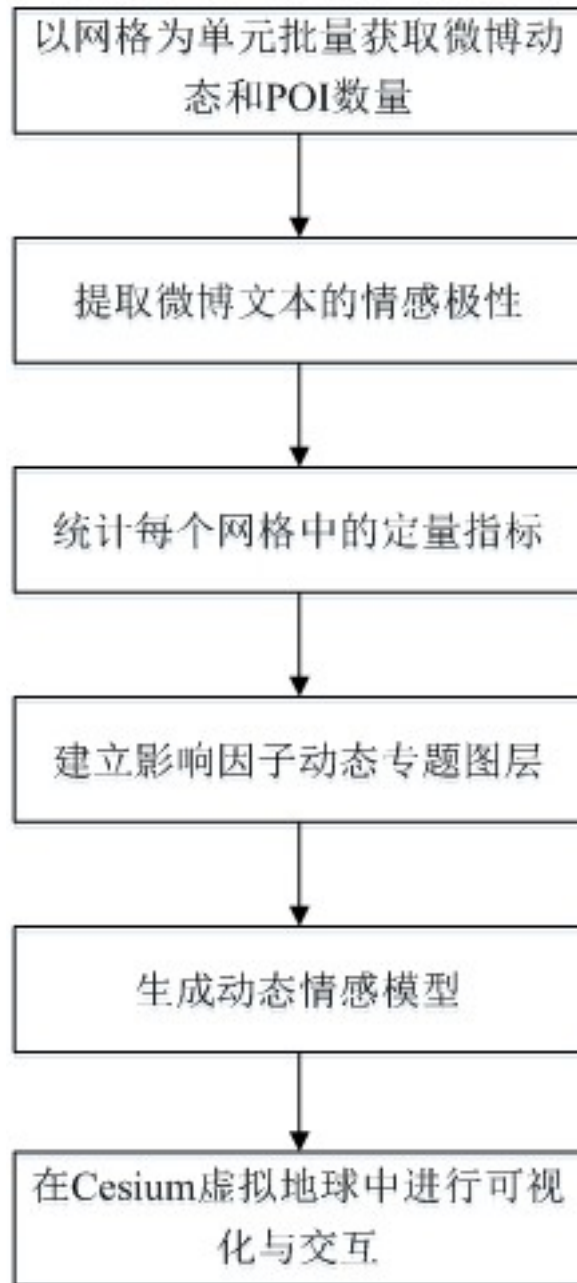


图1

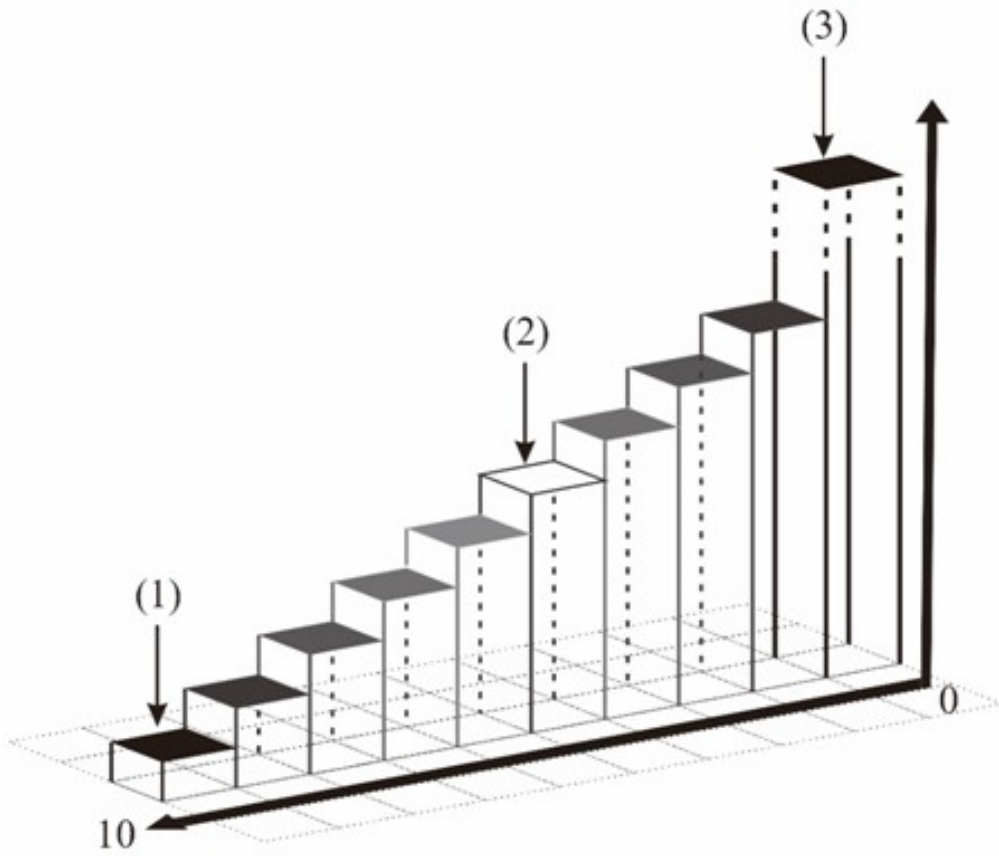


图2

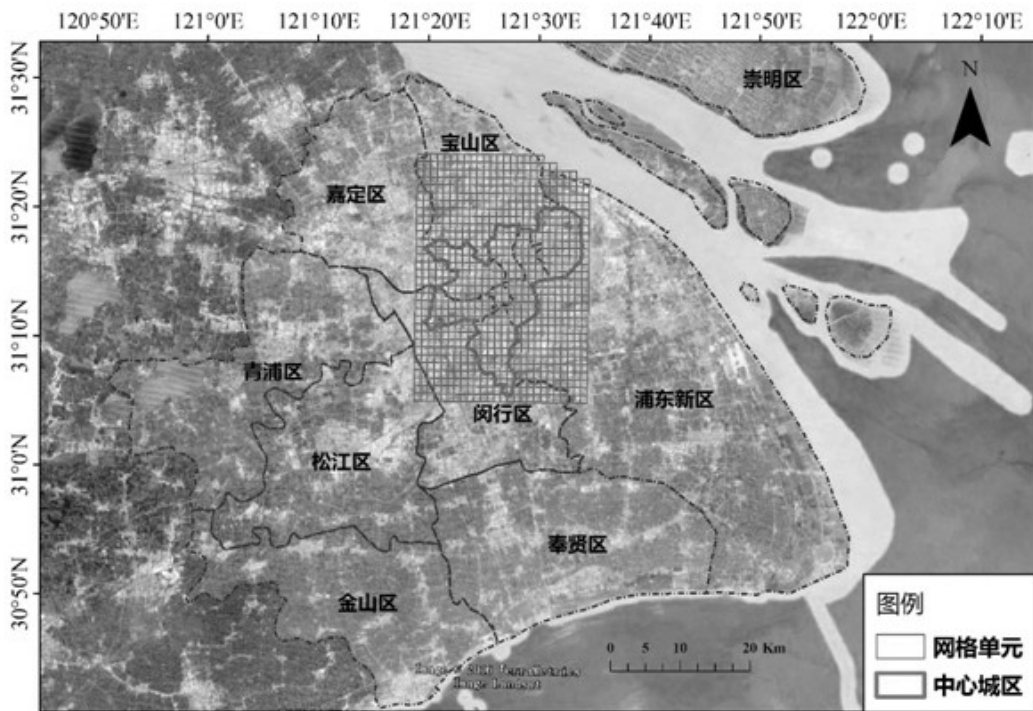


图3

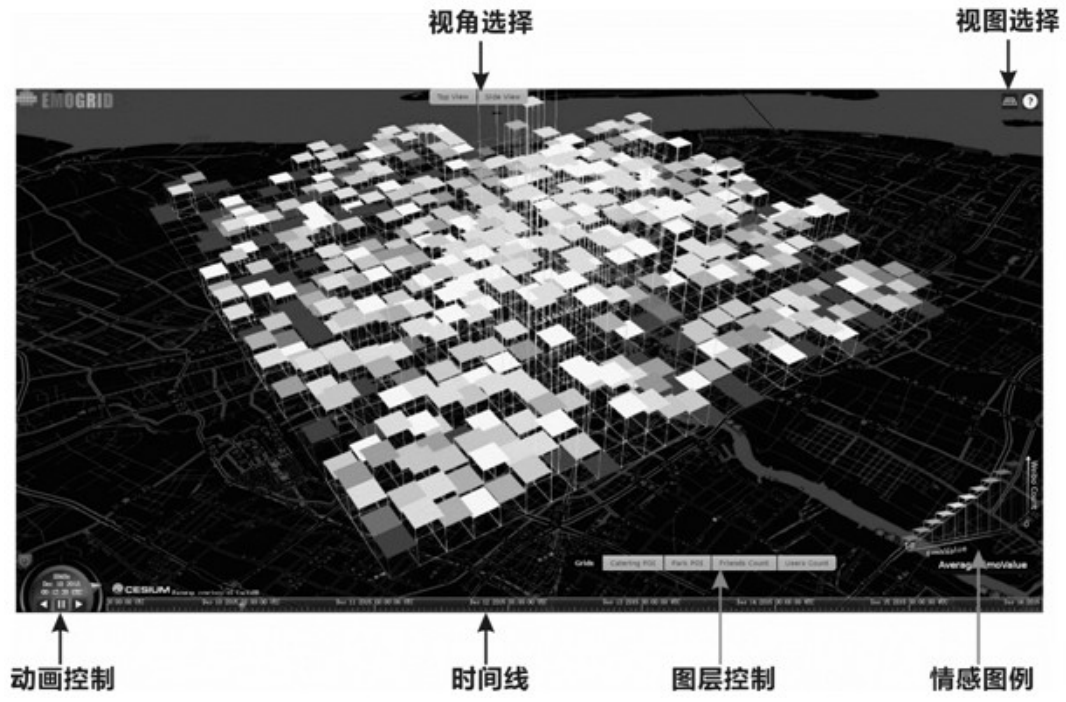


图4

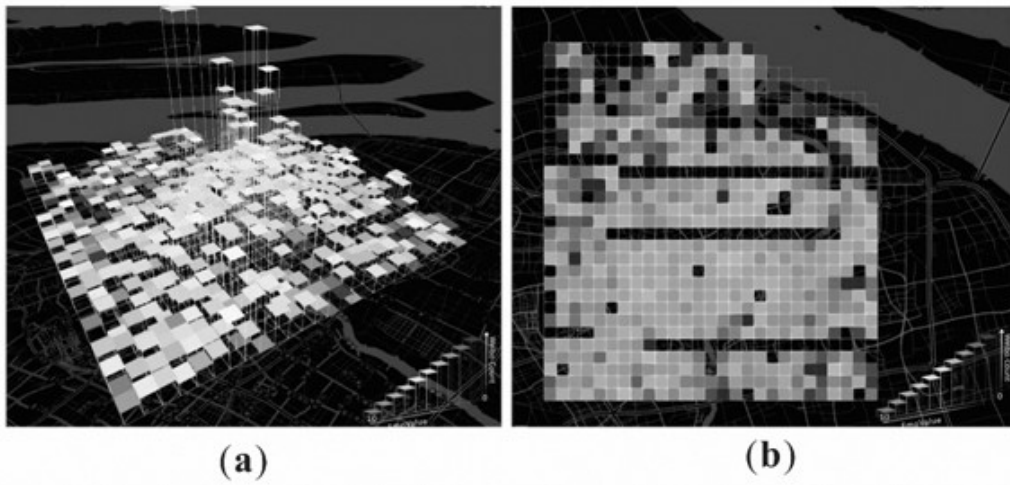


图5

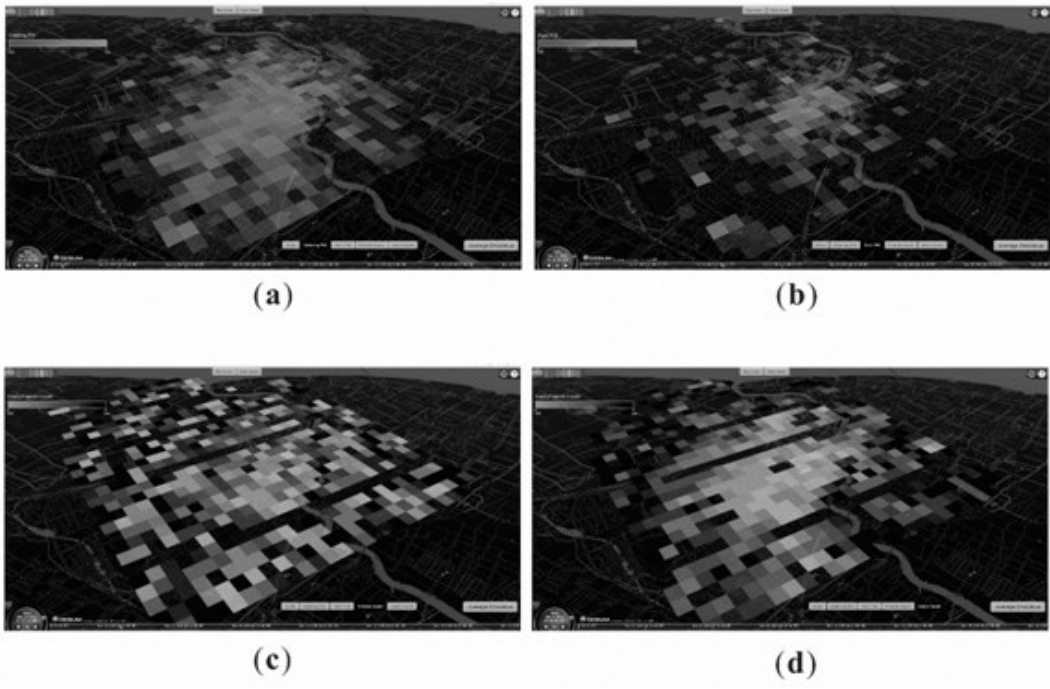


图6