doi:10.3969/j.issn.1672-4623.2017.10.014

融合社交媒体大数据的城市三维模型构建

郭正扬[1]，许捍卫[1]
（1. 河海大学地球科学与工程学院，江苏 南京 211100）

摘　要：通过 C# 代码编写微博大数据抽取系统（定时自动抽取），自动抽取有关新街口地区消费者对商户的海量最新评论数据，从网页上获取南京新街口的 GIS 数据与微博数据一起形成数据库，再将数据库导入 Cityengine 中，通过规则建模，形成具有情感信息的三维城市。该方法能够直观地看出南京市新街口地区的情感信息，便于选择评价好的区域，实现 GIS 服务于人。

关键词：CityEngine；微博大数据；三维；情感信息
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2017）10-0046-03

对微博大数据而言，如何使消费者对商户的评价信息以三维形式展现并给人直观的视觉感受显得尤为重要[1]。本文对微博大数据的自动抽取进行研究，设计开发微博大数据抽取系统，对获取的微博大数据进行分析，将分析后的微博大数据进行三维建模显示，将大数据技术与三维建模技术结合起来，帮助用户直观地理解数据，发现规律，有利于为消费者提供高质量的商户信息，同时也有利于促进商户不断改善其服务水平与商品质量，对实际生产生活具有重要意义。

1　系统设计

本系统的主要模块有数据库设计模块和前端三维系统设计模块两大部分。数据库设计模块包括微博大数据的获取与预处理、地图数据的获取与预处理两部分，如图 1 所示。

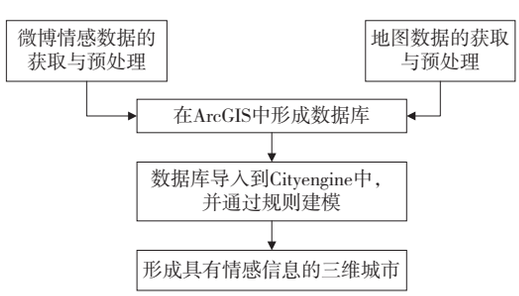

图 1　系统架构示意图

2　数据库设计

本系统数据库设计需要使用到 ArcGIS 数据库 Geodatabase 和 SQL Server 数据库。利用 SQL Server 数据库强大的数据聚合和处理能力来存储和处理微博情感数据，Geodatabse 数据库则用来存储和处理地图数据。最终将处理好的微博数据和地图数据导入到一个数据库中。本系统选用的数据库是 Geodatabase，能与 CityEngine（以下简称 CE）无缝结合，形成完整数据库，为下一步前端系统设计提供数据支持。

2.1　微博数据的获取与预处理

2.1.1　微博数据的获取

对于微博数据的获取模块，首先聚合新浪微博和腾讯微博及大众点评网开放平台 API 于开发的数据抽取系统中。对新街口地区特定的点、话题以及特定时间对特定点等进行时空数据抽取。通过发送 http 请求到开放平台，返回 JSON 数据，对 JSON 数据进行解析，将抽取到的南京市新街口地区的商户点评信息以及与南京市新街口地区有关的微博搜索数据入库。对文本数据采用网络上开源分词工具盘古分词来进行分词，提取存在的情感词及情感值。表 1 为抽取到的部分微博数据条数。

表 1　抽取到的微博数据条数

位置	接口	新浪微博	腾讯微博
8 个特定的 POI 点	周边的微博	243 660 条	71 354 条
	话题搜索 / 搜索微博	无	91 844 条 /14 354 条
南京市新街口地区	周边的微博	237 174 条	173 642 条
	话题搜索 / 搜索微博	无	46 976 条 /58 588 条

2.1.2　微博数据的预处理

由于通过数据抽取系统抽取到的微博数据数量非常大并且情感信息复杂，要有效地运用这些微博数据，必须进行详细的数据预处理，由此获得方便在 CE 中运用的微博数据。以新街口地区微博数据的预处理为例进行说明。

1）首先对 ArcGIS Editor for OpenstreetMap 获取的房屋底面数据（以下简称 OSM 数据）获取中心点，并以这些中心点为圆心、50 m 为半径作缓冲区分析。对房屋底面数据的中心点作缓冲区分析的目的在于对微博数据进行抽稀处理，使落在缓冲区内的点被保留，落在缓冲区外的点被排除，与房屋底面数据相关的微博数据被保留，而与房屋底面数据关系不大的微博数据被排除，以此来对微博数据进行有效的利用。

收稿日期：2016-07-20。
项目来源：国家自然科学基金资助项目（41101374）；水利部公益性行业科研专项经费资助项目（201201025）。

2）由于在缓冲区内的每个微博点数据都有情感值，需要对这些情感值获取平均值，用这个平均值来表达此地区的情感信息。首先，需要获取每个缓冲区内微博点的情感值总分，除以缓冲区内的微博点个数，以此来获取缓冲区内微博点的情感值平均分。将获取的微博点情感平均分分别赋给房屋底面数据中心点。获取到的微博点平均分在 2~6 分之间，分数越高则表示此地区的情感值越高，即此地区评价越好。因此，需要将这些情感平均分分段，以便在 CE 中用不同的符号来表达。将这些情感分数分为（2~3）、（3~4）、（4~5）、（>5）4 个分段，在 CE 中分别用下雨、阴天、多云、晴天来表达。微博数据的预处理流程如图 2 所示。

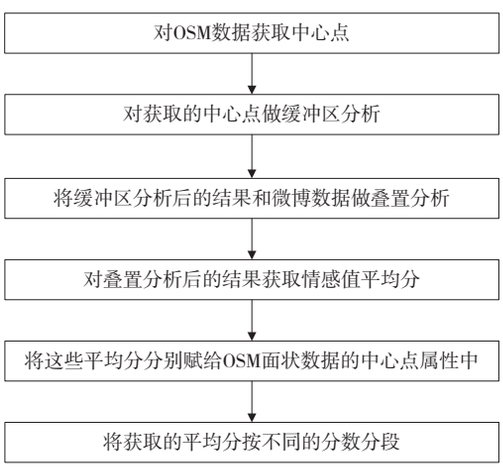

图 2　微博数据预处理过程

2.2　地图数据的获取与预处理

由于本系统所使用的数据要求是南京市新街口地区的实时数据，因此不能选用现成的南京市新街口的数据，需要通过一定的方法从网上获取所需要的数据。由于所需要的数据类型不同，所使用的方法也不尽相同。新街口地区的影像图是通过稻歌软件（http://www.daogle.com）截取的，即在稻歌软件中通过鼠标直接点取要截取区域的左上角和右下角坐标，设置好地图级别与地图类型，最后选择好输出文件夹点击输出即可；道路以及房屋底面数据使用 OpenStreetMap(http://www.openstreetmap.org/copyright）在网页上爬取，即在 OpenStreetMap 网页中输入研究区域的经纬度，点击导出即可。OpenStreetMap® 是开放数据，由 OpenStreetMap 基金会（OSMF）采用开放数据共享开放数据库许可协议（ODbL）授权；地铁数据通过 ArcGIS 中的插件 ArcTilerPlugin1.4.2（http://www.arctiler.com/index.html）获取；底图数据是通过之前获取的影像图在 ArcGIS 中经过重新配色，再经过图幅配准处理得到的 [2]。

通过以上方法即可得到南京市新街口地区的地图数据，将其导入到 ArcGIS 的数据库 Geodatabase 中，图 3 所示为获取并预处理之后的南京市新街口地区的二维地图数据。

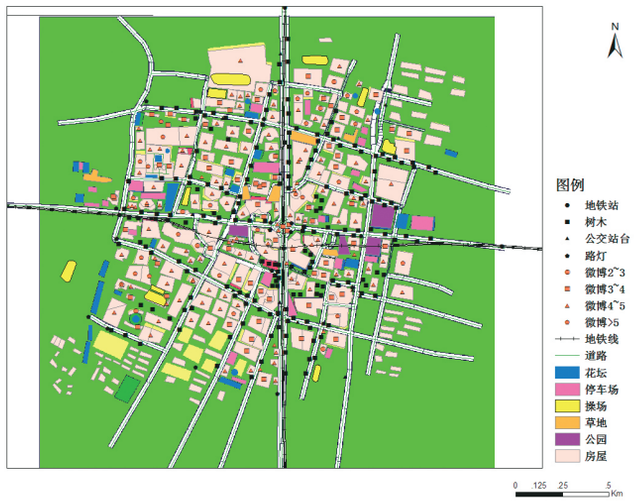

图 3　研究区二维地图数据

3　三维 GIS 实现

3.1　创建场景

要实现本系统的功能，需要在 CE 中进行前端系统设计。在 CE 中进行设计，首先必须创建场景。打开 CE，新建一个场景，选择投影坐标为墨卡托投影，将之前得到的底图转换成和数据库一致的投影（这里选择的是墨卡托投影），分别将底图和数据库加载到场景中，效果如图 4 所示。再在这个场景中进行规则建模，将二维的数据转化为三维的城市，达到三维城市建模的目的。

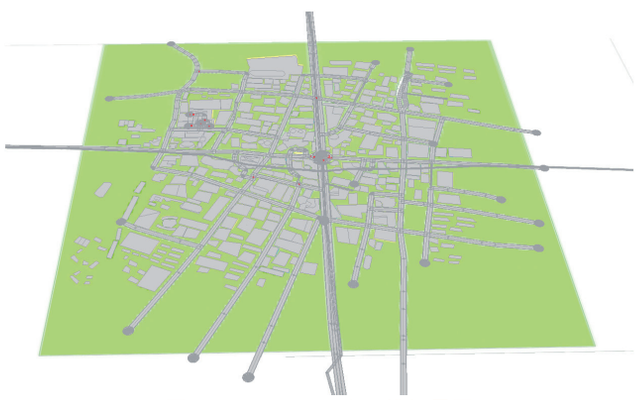

图 4　Cityengine 中二维场景图

3.2　基于规则的建模

规则是 CE 中自带的一种脚本语言，通过编写规则代码可以进行大批量的模型生成。规则定义一系列几何和纹理特征，决定模型如何生成。基于规则建模的思想是定义规则，反复优化设计，以创造更多细节。当有大量模型创造和设计时，基于规则建模可以节省大量时间和成本 [3-6]。

1）公共设施规则的编写思路。这里的公共设施包括操场、花坛、停车场、公园、草地、公交站台、交通灯等。由于本系统追求卡通化的效果，而这些公共设施的结构又极为复杂，故在规则当中使用 Dae 模型。Dae 模型是一种在 3DMax 或者 Sketchup 导出的格式，

能与 CE 无缝结合[7]。以下为公共设施建模的方法：①用 extrude 函数进行拉伸操作；②用 *i* 函数导入 dae 模型；③用 comp 函数分面，方便为每个面贴图；④用 setupProjection 函数设置纹理的大小，以及纹理对应的纹理坐标系图；⑤用 texture 函数导入图片并贴图；⑥用 ProjectUV 函数结束创建纹理。

2）房屋规则的编写思路。对于房屋可以不用 Dae 模型导入，只在 CE 中用代码来实现。本系统的房屋大体上分为南京市新街口中心的标志性建筑和新街口周边的民房两大类。首先，对于标志性建筑，需要先用 extrude 函数拉伸一定的高度，此高度可以随着属性进行调节；再用 comp 函数对建筑分为前、后、左、右、上 5 个面，对于每个面再运用 split(*x*) 函数沿 *x* 轴方向切分为不同的长度，对这些划分的长度运用 texture，projectUV，setupProjection 这三个函数进行贴图[8,9]。

对于南京市新街口周边的民房，与标志性建筑建模不同的地方有两个方面。首先，拉伸的高度不同；其次，南京市新街口地区的民房屋顶不是通过贴图来实现的，而是需要使用 roofGable 函数来具体设置屋顶的样式。

3）道路规则的编写思路。首先将道路分为路灯、人行道、树木、人和交通工具 5 大模块。然后对这些模块分别编写规则，其中对于树木、人、交通工具和人行道运用 *i* 函数引入 dae 模型来实现，对于路灯则采用之前获取的交通灯的 dae 模型来实现。

4）地铁规则编写思路。地铁需要对其进行拉伸负值来实现。拉伸负值后会发现，在地上有一部分可见，需要运用 split(*y*) 函数沿 *y* 方向进行分割。将地上分为 2 个单位，地下分为 18 个单位，通过 NIL 函数将地上部分隐藏。由于本系统是卡通风格，对地下部分可用 color 函数进行绿色染色。

5）微博大数据的表达。对于之前获取的微博数据，需要用不同的模型来表达。将微博数据分为 4 个分段，即（2~3）、（3~4）、（4~5）、（>5），运用下雨、阴天、多云、晴天来表现。首先，用 extrude 函数拉伸一定的高度，此高度必须比微博数据相对应的建筑物的高度高。建筑物的高度信息是在网页上获取的，将这些高度信息加入到建筑物图层属性中，通过属性设定建筑物的高度，进而建筑物上的微博情感点高度也可通过属性获得。将微博数据点的高度设置为比与之对应的建筑物高出 20 m。微博数据点的高度确定之后，用 *i* 函数导入在 sketchup 中制作好的 dae 模型，再用 *s* 函数定义为点状要素，即可完成微博情感数据的表达。具体实现方法为：①用 Attr 函数定义常量；②用 case 和 else 函数分情况讨论旋转角度和尺寸大小；③用 *s* 函数设置模型的尺寸、*r* 函数设置旋转角度、*t* 函数设置宽度的大小；④用 *i* 函数导入 dae 模型。

3.3　规划设计以及系统测试

由于系统中一些数据通过手动矢量化得到，其位置可能会与实际有些偏差，需要进行后期的位置调整和规划设计。对于地图数据，通过修改 CGA 规则的相关属性来修改模型的高度等特征，以动态的方式来调整整体显示效果，从而增加 CGA 规则的灵活性。而对于一些公共设施，则需要通过手动调节来调整其位置。经过调整后的效果如图 5 所示。

图 5　整体效果图

4　结　语

对微博大数据的自动抽取进行研究，设计开发微博大数据抽取系统，对获取的微博大数据进行分析，并且将分析后的微博大数据进行三维建模显示，将大数据技术与三维建模技术结合起来，帮助用户直观地理解数据，发现规律，有利于为消费者提供高质量的商户信息，同时也有利于促进商户不断改善其服务水平与商品质量，对实际生产生活具有重要意义。

参考文献

[1] 王军，刘金辉．大数据的国内外研究现状及发展动态分析 [J]. 电子技术与软件工程，2015(23):200
[2] 李学祥．GIS 数据动态获取方法研究 [J]. 地理信息世界，2010(5):77−82
[3] 徐汝坤．Esri CityEngine: 开启三维规则建模新篇章 [M]. 北京：中国信息技术有限公司，2012
[4] 董晓非．Esri CityEngine 中文教程 [M]. 北京：Esri 中国信息技术有限公司，2012
[5] 董晓非．Generating a 3D City[M]. 北京：Esri 中国信息技术有限公司，2012
[6] 李德仁．从数字地球到智慧地球 [J]. 武汉大学学报（信息科学版），2010,35(2):127−132
[7] 吕永来，李晓莉．基于 CityEngine 的三维建筑模型研究 [D]. 合肥：合肥工业大学，2012
[8] 花利忠，王赵兵．基于 CityEngine 与 ArcGIS Flex API 的校园 WebGIS 系统 [D]. 合肥：合肥工业大学，2013
[9] 吕永来，李晓莉．基于 CityEngine 平台的高速铁路建模方法的研究与实现 [D]. 合肥：合肥工业大学，2012

第一作者简介：郭正扬，硕士，研究方向为 GIS 系统开发与应用。