

dxf 与 shp 转换中的问题研究

王强¹, 吴孟泉², 孙西兵³, 韩磊²

(1. 烟台市国土资源局福山分局, 山东烟台 265500; 2. 鲁东大学地理与规划学院, 山东烟台 264025; 3. 烟台市国土资源局开发区分局, 山东烟台 264000)

摘要:地理信息系统(GIS)具有强大的分析、管理、使用空间数据的能力,近几年发展很快。但大部分空间数据都是以CAD的格式存在的,将AutoCAD数据转化为GIS支持的文件成为必然。该文介绍了AutoCAD的dxf格式数据与ArcGIS的shp格式数据的区别与联系,两种格式数据转换涉及的要素及可能出现的问题,并提出了相应的解决办法。

关键词: dxf; shp; 数据转换; 地理信息系统

中图分类号: P208 **文献标识码:** A

地理信息系统(GIS, Geographic Information System)具有空间数据的采集、管理、处理、分析、建模和显示的功能,但很多空间数据都是以CAD的格式存在的,由于CAD软件在数据存储、图元定义、管理风格等方面同GIS软件有很大差别,数据转换不可避免地存在着各种问题,如何解决这些问题是当前GIS人员面临的主要的问题^[1-3]。

1 CAD 与 GIS 的格式区别

计算机辅助设计(CAD, Computer - Aided Design)通过计算机辅助设计人员进行设计,提高设计的自动化程度,节省人力和时间。其数据格式dxf(Drawing Exchange File)文件即标准图形交换文件,是AutoCAD软件的一种ASC II格式的文件。dxf文件由若干节组成,包括:①标题节:储存图形的一般信息。②类节:定义图形中的类,其实例在块、实体和对象节中表现,其属性不变。③表节:包括块参照表、图层表、线形表等一系列表。④块节:定义块和构成块的实体。⑤实体节:包括图形中的图形对象或实体,以及被看作插入实体的块参照。⑥对象节:包括图形中的非图形对象;⑦图形预览节:图形的预览数据^[3]。

ArcGIS中的shp(ESRI Shapefile)格式数据是可以建立元数据的数据类型。元数据是对数据进行描述和定义的数据,包括与空间数据相关的很多有用信息,如数据属性全名、原始数据比例尺、定位精度和投影等^[4-7],具体区别有:

(1)操作对象的不同。CAD研究对象为人造对象、规则几何图形及其组合。GIS处理的数据大多来自现实世界,比之人造对象复杂,数据量更大,数据采集的方式更趋于多样化,包括空间位置、投影方式、地理要素间的关系等。

(2)CAD的图形功能特别是三维图形功能强,一般很少涉及属性的内容。GIS有完整的空间和属性数据的管理特性,可以处理要素的空间定位特征,并将其空间和属性信息有机结合起来,从空间和属性2个方面对要素进行查询、检索和分析。

(3)CAD中的拓扑关系较为简单或没有拓扑,GIS强调对空间数据的分析,图形属性交互使用频繁。CAD一般采用几何坐标系,而GIS采用地理坐标系。

(4)AutoCAD以图层组织图形数据,不同类型的图素(如点、线、多边形等)可以共层,一个dxf文件可以包含多个图层。而在GIS中,每个视图只包

* 收稿日期:2009-02-23;修订日期:2009-04-29;编辑:曹丽丽

基金来源:高校博士科研基金资助(LY20073201)

作者简介:王强(1976—),男,新疆玛纳斯人,工程师,主要从事测绘及地理信息化工作。

含一个专题,同类型、不同专业的空间数据(如点、线、多边形等)应分层存放,不能共存于同一个专题(图1)。

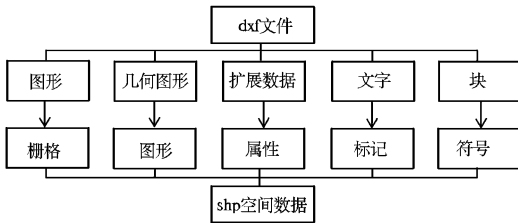


图1 dxf与shp数据格式的图元对应图

2 数据格式转换中产生的问题

由于CAD软件在数据存储、图元定义、管理风格等方面同GIS软件的差别,因此数据转换就不可避免地存在一些问题^[6]。

(1)要素丢失。由于CAD与GIS软件的图形实体数据结构不同或者转换前后各层数据的对应关系没处理好,有的内容没有转换过来,转换后造成数据丢失。

(2)要素变形。由于CAD与GIS软件的图形实体数据结构定义不一致(如CAD中的圆、弧等在GIS中可能变成多边形、线),或者原来的CAD数据在采集过程中作业不规范(如线不连续、采点太多、间隔太密等),造成要素变形。

(3)数据冗余。CAD中分层不合适(如母线和符号未区分好,符号也当成母线转换过来),或一条线段上点太多太密,造成数据冗余。

(4)属性信息不足。各种CAD数据都很少有属性信息,造成转换后信息不足。属性信息追加起来费时费力,容易产生错漏,而且不易检查修改。在AutoCAD数据转换为GIS数据格式时总会存在部分数据的丢失,增加了数据编辑的工作量,很难保证数据的质量和精度。

3 问题解决方法及流程

3.1 问题解决的主要方法

转换后的数据检查是确保数据正确与否的一道重要工序,检查内容包括数据精度、图形信息、拓扑信息、属性信息4个方面,数据检查一般采用如下方法^[7]:

(1)叠合比较法。该法是空间数据正确与否的

最佳检核方法,把原图定位导入GIS软件中,与数字地图进行观察和比较。一般可以马上发现空间数据位置的不完整和不准确。

(2)目视检查法。指在屏幕上用目视检查的方法,检查一些较明显的的数据误差和错误,包括由于线段的过长或过短导致的交叉、不闭合、悬挂线、多边形的重叠压盖和不闭合等。

(3)逻辑检查法。基于数据拓扑的一致性进行检查,对属性数据的检查一般也最先采用这种方法,检查属性数据的值是否超过其取值范围。属性数据之间或属性数据与地理实体之间是否有荒谬的组合。对于空间数据的不完整或位置的误差,主要是利用GIS的图形编辑功能如删除、修改(平移、拷贝、连接、分裂、合并、整饰)、插入等进行处理。

3.2 问题解决的流程

3.2.1 预处理

建立对应关系,将转换前后的图形实体建立一对一或多对一(如CAD中分层较详细的几个图层转换时可以合并为一层)的对应关系,防止丢失要素,某一类要素出现问题容易追本溯源,找出原因所在。建立转换对照表^[6,7],尤其是图元参数对应表,程序执行时只需按照对照表即可将复杂的图元替换成GIS可以识别管理的地理实体。充分利用CAD的图形信息自动生成GIS的属性信息、图层信息可以转换成要素类型编码;注记可以生成许多属性信息,如由名称注记可以得到要素名称属性,可以大大减少属性编辑的工作量(图2,表1)。

3.2.2 规范化处理

(1)图形元素的整合。ArcGIS中的每一个图形元素都有其具体的物理意义和属性信息,因此,对AutoCAD中对应的图形元素要根据其代表的地物或物理意义进行整理或合并。例如,在AutoCAD中,一条支渠可能由数条线段组成,这并不影响其出图效果,但转到ArcGIS中,它就变成数条支渠。一个灌溉区域,如果没有对构成边界的多线段进行闭合,则转化后的图形只能为线状图形,无法形成面。

(2)图形元素的分层。ArcGIS的shp文件只能代表一类图形元素,这与AutoCAD中的图层有明显的区别。虽然图层是两者管理图形数据最为接近的方式,能自动按实体类型进行图层元素的分配。但是转换后的图中将包含所有图层中的同一类型数据,如果类型相同而性质属性不同的数据归于一个

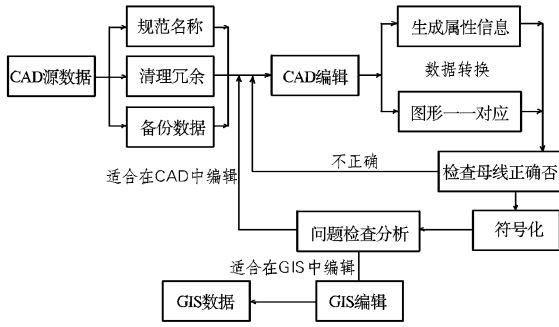


图 2 数据转换编辑流程

表 1 dxf 与 shp 格式转换的图层对应

所表示的对象	AutoCAD	ArcGIS
地名、设施、设备等	点	点
建筑物、行政区域、河流等	多段线	面
道路、楼栋附属物、轨道交通线等	线、多段线	多段线
光缆、电缆	线、多段线	多段线
各类名称、数值等	文本	注记
属性	注记	标注

图层或同一性质和类型的数据归于不同的图层，都将给转换后的图形管理造成混乱。应将 AutoCAD 的图形元素按性质进行分层，使每一图层与 ArcGIS 的某一类相对应。

(3) 注记的属性化。当把 AutoCAD 数据添加到 ArcGIS 视图中时，ArcGIS 会自动识别源图形的类型，并根据实体类型进行分配。如果注记为块的属性，就会在对应的主题属性表中自动建立包含相应注记字符的属性字段，可以方便地转换为 Shape 数据文件；如果注记为文字实体，则只能通过建立联接主题，注记字符串可缺省保存在“TEXT”属性字段中，显示为不可编辑的图形字符，但在转换为 Shape 数据文件时，原文字实体转化为线对象，表现为注记丢失，只有在原图形对象对应的主题属性表中重新建立属性字段或进行属性表联接。因此，对于需要显示为点主题的图形对象，例如控制点位置，其注记信息最好定义为同名属性，并把同一类型的图形元素修改为含属性的块对象。

4 数据采集建议

由于 AutoCAD 的广泛应用，大量的数据以 CAD 的格式存在，所以实现 dxf 文件的转换具有极大的实用价值。既提高了数据的采集效率、降低了 GIS

的建设成本，又缩短了 GIS 的建设周期。为避免在转换时浪费大量的人力与时间，特在数据采集的步骤上提出几点数据采集标准方面的建议。

(1) 点状要素采集原则。点状要素一般在一定位点上画一个固定的、不依比例尺的地物要素，这类要素形状和尺寸固定，例如变电室、水轮泵和消防栓等。此类要素可利用形文件来定义要素，然后采集其空间位置和属性。对复杂点状要素应以固定的线型提取其边界，而且应该特别注意线的连续性和整体性。

(2) 线状要素采集原则。简单的线状要素应以“形”方式采集，保证线要素的整体性和属性连续性。例如一条乡村道路被分成若干条等距的线段，虽然满足了地图制图的要求，但与 GIS 完全相背，不能为 GIS 服务。其次，在线要素采集过程中应注意线实体的方向性，这对于 AutoCAD 完全无碍，但用于 GIS 的分析往往产生严重的后果。例如流域分析、水库库容分析等其往往产生相反的结果。复杂线状要素应以“组”方式采集，按这种方式采集数据可保证数据的完整性和属性逻辑一致性，也利于要素的修改。

(3) 面状要素采集原则。面要素的采集相对点和线较简单些，可直接用多段线采集。但应注意面边界的采集方向和闭合性，因为这影响到空间要素的拓扑关系的建立。其次飞地和岛等问题也要注意，要保证其属性一致，可用“组”采集。

(4) 属性数据采集原则。AutoCAD 有非常强大的图形编辑功能，但是与 GIS 系统软件相比，其属性编辑功能相对较弱。在 AutoCAD 中，数据库中仅记录表示图形元素的几何位置、形状、大小等的几何坐标和表示图形元素性质（如颜色、层号等）的辅助数据，远不能达到 GIS 的要求。如果能将图形元素的其他属性数据添加到数据库中，将大大提高 AutoCAD 的内部属性数据库的功能。

参考文献：

[1] 党安荣, 贾海峰, 易善楨, 刘钊. ArcGIS 8 Desktop 地理信息系统应用指南[Z]. 北京: 清华大学出版社, 2003.
 [2] 黄杏元, 马劲松, 汤勤. 地理信息系统概论[M]. 北京: 高等教育出版社, 2007.
 [3] 任沂军. AutoCAD 中 dxf 文件的坐标转换分析与实现[J]. 测绘通报, 2006, (7): 23-26.
 [4] 申胜利, 李华. 基于 ArcEngine 的 ArcGIS 与 AutoCAD 数据转换研究[J]. 测绘通报, 2007, (2): 41-42.

- [5] 张叶,孙毅中,陈年松. CAD城市基础数据到GIS转换的有关问题探讨[J]. 测绘与空间地理信息系统,2007,(1):94-95. [J]. 测绘通报,2007,(12):46-47.
- [6] 胡胜华,何宗宜,陶利佳. AutoCAD与GIS数据转换的研究[J]. 测绘通报,2005,(8):11-12. [7] 陈能,施蓓琦. AutoCAD地形图数据转换为GIS空间数据的技术研究与应用[J]. 测绘通报,2005,(8):11-12.

Study on Datas Conversion between dxf and shp

WANG Qiang¹, WU Meng-quan², SUN Xi-bing³, HAN Lei²

(1. Fushan Branch Bureau of Yantai Bureau of Land and Resources, Shandong Yantai 265500, China; 2. Geography and Planning College of Ludong University, Shandong Yantai 264025, China; 3. Development Zone Branch Bureau of Yantai Bureau of Land and Resources, Shandong Yanta 264000, China)

Abstract: Geographic Information System (GIS) with its powerful analysis, management and use of space data has developed quickly. But now, most datas are the CAD format. Thus, it is very necessary to transfer AutoCAD to datas supported by GIS technology. datas with dxf format supported by AutoCAD technology and shp format datas supported by ArcGIS are introduced, relations and distinction between these two datas format, and elements and problem involved in datas format conversion are introduced in this paper. Countermeasures for solving problems occurred in datas conversion are put forward as well.

Key words: dxf; shp; Data conversion; GIS