

中国矢量数据交换格式的应用研究

陈泽民¹

(1 南京大学城市与资源学系 南京市汉口路 22 号, 210093)

摘要: 在研究中国《地球空间数据交换格式》标准的矢量数据交换格式的基础上, 提出了矢量空间数据的两种属性分类概念, 并就标准的功能扩充作了简要的论述和建议。

关键词: 矢量数据; 定位数据; 非图形属性

中图法分类号: P283. 1; P208

目前, 人们常使用的 CAD 和 GIS 软件有数 10 种之多, 而这些软件大都采用不同的数据结构和格式, 其格式有公开、半公开或不公开的, 用户在不同图形平台或 GIS 平台上使用同一数据受到了极大的限制, 同时由于数据标准、使用的数据平台和对数据模型理解的不同, 都大大减缓了空间信息共享的进程, 使我国存在着大量的重复生产和开发等浪费现象。中国《地球空间数据交换格式》(CNSDTF)标准的制定和推出, 较好地解决了 GIS 行业发展和数据交换的瓶颈问题。然而, 目前的中国《地球空间数据交换格式》标准在实际应用中尚存在一定的局限和不足, 因此也影响了本标准的应用和推广。本文中的地理空间信息主要是指用几何图形元素对空间实体对象的描述即矢量数据, 而几何图形也仅仅是指用于描述地理空间现象的空间对象。

通常, 空间矢量数据由几何图形和属性两部分组成, 图形部分是空间矢量数据的可视部分, 而属性部分则是与空间矢量数据相关的数据库信息(以下简称非图形属性)。当然, 也可以将图形部分再分解成图形的基本定位部分, 即图形的基本定位数据(以下简称定位数据)和图形的表现属性(以下简称图形表现属性)两部分。前者确定图形的类别(点、线、面或注记等)和定位信息(坐标(串)), 这是 GIS 交换不可缺少的部分, 而后者则提供图形的表现属性(图层、颜色、线型、线宽、显示比例、旋转方向和是否显示、填充与否、线型方向等), 这在 GIS 的应用中是部分需要或不太需要的部分, 但对于地

图制图(包括一切形式的显示)却是不可或缺的。

图形表现属性和非图形属性同为空间实体对象的属性, 但却具有不同的特点、实现方法和连接手段。图形表现属性对于同一类对象具有完全一致的设定或数值, 而非图形属性对于每一个对象都可以有不同的数值。在连接方法上, 图形表现属性应该在“几何图形数据”段内的“层名”位置上加挂图形表现属性表的连接, 非图形属性则可以在“属性数据”段内, 根据对象的 ID 号进行挂接(现有的 CNSDTF 挂接方法)。在存储位置上, 图形表现属性可以不与数据库发生关系(如果仅仅是电子地图), 而直接保存在图形文件里, 而非图形属性常常需要挂接到关系型数据库上。

因此可以认为, 空间矢量数据是由图形的定位数据、图形表现属性和非图形属性 3 个部分组成。当空间矢量数据仅包含图形的定位数据和图形表现属性两部分时, 能基本满足地图制图数据的转换要求, 即电子地图的交换需求; 当空间矢量数据包含图形的定位数据和非图形属性时, 则基本可以为 GIS 数据的交换服务即通常所说的建库数据的交换; 当空间矢量数据包含全部 3 个部分数据时, 则既可以为 GIS 数据的交换服务, 又可以满足地图制图数据的转换要求。

《地球空间数据交换格式》里已经提供了较完善的 ID 号对照机制, 主要有“要素类型编码”、“目标标识码”等, “属性表名”也属于此范畴。尽管如此, CNSDTF 仍使用了较多的“绝对值”方法, 即指定某名称或数值的方法。使用该方法可

以保证数据的“不走样”，但却不容易做到“兼容”，即有可能损伤 CNSDTF 设计中的一项主要原则——完备性原则。

1 使用现有标准的困难

现有的《地球空间数据交换格式》在单纯满足 GIS 数据的交换上较合理，但在进行电子地图数据的转换时，由于其对图形表现属性方面结构描述的薄弱，还存在一些困难。

1.1 RGB 颜色定义的缺陷

在《地球空间数据交换格式》里，颜色值是使用 RGB 格式即 24bit 长度的整型量表示的。在计算机中使用 RGB 作为颜色值，可以表示一种唯一的颜色。而在实际的应用中，直接使用 RGB 有其不便之处，甚至无法避免出错。目前，相当多的图形平台使用的是颜色号，而人们习惯上也乐意使用颜色号，它具有简单、易用等特点，如在 Microstation 中，限制用户仅能使用 0~254 号颜色，至于每一种颜色使用的 RGB 数值则是通过定义不同的颜色表来实现。在该软件中，所有的操作都是针对颜色号而非 RGB 颜色值进行的。这样就带来一个问题，对于 CNSDTF 中的同一种 RGB 颜色，不同人或不同软件的读入就可能会对应于不同的颜色号。

1.2 层名的丢失

《地球空间数据交换格式》规定，无属性的要素类型编码和层名为 Unkonwn。这里的无属性其实是指无非图形属性，然而，层名其实是属于图形表现属性范畴的属性。显然，此项规定对于交换纯电子地图的应用是不可接受的，因为经过如此转换，所有的图形元素将使用同一个图层，即 Unkonwn 图层，而不管转换之前电子地图的分层情况。

1.3 关于模拟图形的问题

模拟图形是现实生活中用到很多的一种图形，在这类图形里，数据点是可以不通过显示图形本身的，但可以通过一个函数得到图形上任意点的坐标，同时，图形上的点是连续的，有无穷多个。

《地球空间数据交换格式》支持部分模拟图形，如圆弧、圆、椭圆和 B 样条曲线等，并将其统统归入线状要素类。事实上，模拟图形也可以分线和面两大类，圆弧、B 样条曲线属于线的范畴，而圆、椭圆则应该属于面类。

1.3.1 圆、椭圆类型定义的不足

圆、椭圆按类型来说属于模拟图形，即用函数来表示的图形范畴，在表现形式上和实际应用中，都可以作为面类型来处理。因为它具有全部的面

的特征，诸如封闭、具有唯一的面积值、可以普染颜色等。同时，它可以没有线的特征，如没有端点。因此，把圆、椭圆按线类型来归类是并不科学的。

根据模拟图形的特点，图形的组成不是一系列点的坐标，而应该是一组数学表达式：

$$P(x, y) = f(a, b, c, \dots)$$

根据这个函数式，可以非常方便地计算各点的位置坐标。如圆的函数关系为：

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

式中， x_0, y_0 为圆心坐标； r 为圆的半径； x, y 的取值范围分别为 $(x - x_0) \pm r$ 和 $(y - y_0) \pm r$ 。

通常，根据函数式来确定图形上点的坐标很方便，且是唯一的；反之，则非常困难，也没有必要。如果仅给出一系列模拟图形上的点的位置坐标，很难保证两个不同软件绘出相同的曲线或面来。

1.3.2 模拟图形方向缺乏灵活性

与其他线和面元素一样，模拟图形在图形表现属性上也有类似的要求，如方向、线型等。《地球空间数据交换格式》未能给出足够的例子和说明，但从目前已有的描述来看，其方向是不可逆的（仅可顺时针方向），因此，可能会满足不了有一些有专门要求的应用需求。

1.4 注记

CNSDTF 注记段的格式非常复杂，这是由于 CNSDTF 将图形表现属性与非图形属性合二为一的缘故。如果将图形表现属性提取出来，将大大缩短 CNSDTF 文件的长度。

对于目前 CNSDTF 注记段的格式说明尚不完备，有些甚至无法真正地操作。

2 对标准功能扩充的建议

作为一个数据交换标准，不可能定义一个与每一个应用平台所使用的数据结构完全一致的数据结构，但应尽可能多地包容各种不同数据结构所包含的数据，且数据结构还要简单。除此之外，这个数据结构还应该能够扩充，以适应特殊数据的需求。当然，用户扩充的结构本身必须符合规范给出的数据结构要求。为此，建议对现有规范作一些必要的扩充，使之具有更强的兼容性、易用性和可扩充性。

2.1 对《地球空间数据交换格式》文件头部分的扩充

2.1.1 增加注释行符号

既然目前的《地球空间数据交换格式》还是

ASCII 纯文本的格式, 就可以考虑增加注释行, 即读者可以用来作一些标记或说明等, 以增加 CNSDTF 数据的可读性。如像 BASIC 语言里的单引号“'”、C 语言里的“/ * */”注释块或“//”注释行和批文件 BAT 中的 REM 标记等。为此, 建议增加

[Comment: <字符><CR>]

Comment: 任意非空白字符, 在此后的内容为说明, 系统不必处理。附加部分、缺省可以是 (#) 号。

2.1.2 增加一系列是否使用 ID 号的选项

使用 ID 号可以满足不同需求用户的使用, 建议添加的选项有:

[UseRGBColor: <T|F|Y|N>]

[UseFontName: <T|F|Y|N>]

[UseLevelName: <T|F|Y|N>]

...

说明:

<UseRGBColor>: := T|F|Y|N

T 和 Y 表示使用 RGB 颜色, F 和 N 表示使用颜色号, 并提供颜色号与 RGB 颜色的对照表。附加部分、缺省为 Y。

<UseFontName>: := T|F|Y|N

T 和 Y 表示使用字体名, F 和 N 则表示使用字体号, 并提供字体号与字体名的对照表。附加部分、缺省为 Y。

<UseLevelName>: := T|F|Y|N

T 和 Y 表示使用图层名, F 和 N 则表示使用图层号, 并提供图层号与图层名的对照表。附加部分、缺省为 Y。

事实上, 此类对照表还可以有很多, 如图形表现属性中的线宽, 在有的软件平台上采用了线宽号(线宽 ID 号), 而在有的软件平台上则指定输出线宽的具体数值。

因为, 此类说明是一种带有可由用户扩充意义的定义, 可在一般规定里指定, 以 Use×××××××格式标识的附加部分内容, 可能会有相应的对照表出现。

2.1.3 增加文件坐标系统的偏移设置

我国大部分的地图采用高斯-克吕格横轴投影的平面直角坐标系, 该方式使赤道的纵坐标为 0, 因此, 我国的大部分疆土位于纵坐标 3 000 多 km 的位置上; 另一方面, 由于采用了分带投影, 不同中央子午线的分带具有了完全相同的坐标, 且中央子午线以西的坐标值为负。为区分两个不同位置, 传统的做法是在横坐标上加上 500km 的

常数再冠以带号。因此, 我国疆土的任一位置的坐标小数点前都具有 7 位纵坐标和 8 位横坐标(以 m 为单位)。

有时, 由于作业范围并不一定很大, 特别是对于大比例尺测图而言, 为了方便, 常常舍去了纵、横坐标的前几位数据。在进行数据的交换或共享时, 有必要将省略掉的几位数据完整地还原回来。还原的最简单方法就是由数据的提供者告诉用户, 这些数据已经舍去了几位什么样的数值。甚至数据的使用者可以根据需要人为地加上某个常数, 平移整个作业区域。

建议增加设置坐标系统偏移量的可选项字段:

[OffsetX: <浮点>]

[OffsetY: <浮点>]

¹OffsetX: X 方向偏移量, 在输入时, 将 X 坐标加偏移量。附加部分、缺省为 0。

¹OffsetY: Y 方向偏移量, 在输入时, 将 Y 坐标加偏移量。附加部分、缺省为 0。

2.2 对《地球空间数据交换格式》结构的扩充

2.2.1 图形表现属性表

对于非图形属性, 《地球空间数据交换格式》已经提供了比较完整的属性数据结构和属性数据段, 通过几何图形数据和“注记”段中的“目标标识码”可以方便地实现图形对象与非图形属性的连接。但现有的《地球空间数据交换格式》中, “要素类型参数”段对图形表现属性的描述结构则很难满足电子地图转换的需求。

通常, 图形表现属性包括图层、颜色、显示用的线型、线宽、显示比例、旋转方向和是否显示、填充与否、填充方式、线型方向等。其中图层、颜色、线型和线宽等应该用 ID 号来表示, 而图层名、颜色 RGB、线型名和线宽尺寸等具体值可以通过专门的对照表来说明; 显示比例、旋转方向(由于仅支持二维数据, 因此不考虑用旋转矩阵)可以直接用实数来表示; 是否显示、填充与否、可否捕捉等则可以用整型变量或布尔变量来确定。

另一方面, 图形表现属性的应用除了图形元素点、线和面外, “注记”段也有非常复杂的描述。目前的《地球空间数据交换格式》定义了注记中常用的一部分图形表现属性, 有些属性如“行距”、“字间距”、“竖排”、“沿元素”等属性还无法传递。

在《地球空间数据交换格式》里, 要素类型参数定义为:

<要素类型参数>: := FeatureCodeBegin <CR>

{ <要素类型编码>, <要素类型名称>, <几何类型>, <缺省颜色>, <属性表名> } { <用户项> } <CR>

FeatureCodeEnd<CR>

考虑与现有格式兼容,建议改<缺省颜色>为<图形表现属性表名>。另外,<用户项>也可以是其他不同的<图形表现属性表名>。事实上,图形表现属性的内容很多,其内容也可能随着图形对象的不同而不同。因此,定义的“图形表现属性表”的结构也可以是不同的,如“填充”这一图形表现属性仅针对“面”对象;“竖排”这一图形表现属性仅针对“注记”对象。同时,还应该支持一个特征连接几个“图形表现属性表”,如此可以有效地控制 VCT 文件的冗余度。

“图形表现属性表”的结构可以在“属性数据结构”段内定义,其数值同样可以放在“属性数据”段内。不同的是,“属性数据”中的“目标标识码”字段应设为“-1”,以表示不针对某一图形对象,而是针对某类特征的全部图形对象。

2.2.2 增加对照表

对照表可以是一个比较通用的结构,基本字段有两个,编码 ID 和对应的名称或数值,也可以增加别称或其他说明连接。对照表的结构说明可以放在“属性数据结构”中,其内容同样放在“属性数据”段内。与“图形表现属性表”一样,其“属性数据”内容中的“目标标识码”字段应设为“-1”,以表示不针对某一图形对象。

对照表的项数根据文件头中 UseXXXXXX 确定,当该项值为 N 或 F 时,就有相应对照表存在。使用时,仅需将对照表的内容取代编码 ID 即可。

2.2.3 增加函数类图形的结构

在早期 GIS 中,定义的空间图形基本上只有点、线(串)和由封闭的线(串)组成的面。这样的定义大大简化了 GIS 中图形处理和相应空间实体运算的复杂程度。单从以往的 GIS 角度来说,这样的定义通常是可行的,但往往不能满足一些特殊的需求,尤其是在电子地图数据的交换中,现有的《地球空间数据交换格式》无法妥善处理模拟图形即用函数来表示的图形的交换。

模拟图形在电子地图和工程设计中用得十分广泛,因为它的线条具有平滑的表现能力,这是一般线串所做不到的。

一个模拟图形上的坐标对应于确定的一个函数,该函数中作为自变量的坐标值并不一定出现在该图形上。圆和椭圆可以用圆心、长半径、短半径和方向来惟一确定,作为自变量的圆心坐标就不出现在作为显示图形的圆周上。当长半径和短半径相等时,为圆;不一致时,为椭圆,这时可以定

义旋转方向为长半轴的方向。

弧是圆的一部分,在圆的基础上增加起始角和扫角两个参数,扫角值可以带符号,即方向的含义,它将确定弧的顺或逆时针方向。

其他的曲线,如 B 样条曲线等,可以通过线型描述表中数学表达式的定义来惟一确定。因为从理论上讲,这一类的图形有无穷多种,因此,定义的线型描述表的结构必须能满足任意曲线的描述。

2.2.4 增加部分常量的定义

定义常量本身就是制定规范或标准的行为。将一些计算机操作系统或软件平台默认的设置量加以常量(可以是 ID 号)来定义,这会使处于不同应用的数据规范到对数据相同的认识上来。如定义“BLUE”为“0x0000FF”、“RED”为“0xFF0000”,相应地,对字体表现属性中“字型”、“字号”等也可以进行定义。

2.2.5 增加记录的结束符

《地球空间数据交换格式》提供的是 ASCII 的 VCT 文件。在整个文件结构的 6 个数据段中,除了最后的“属性数据”段里使用了“TableEnd”作为每一属性字段的结束标志外,内容相当大的“几何图形数据”和“注记”段内未给每一图形记录或文字记录设置任何结束标志。当一个 VCT 文件在“几何图形数据”和“注记”段内出现任何多行或少行的错误时,该段的内容将产生极大的混乱,甚至导致程序出错和崩溃。因此,建议在“几何图形数据”和“注记”段内对每一图形对象和注记增加起始或结束标志,就像 DXF 格式里使用了“0”作为记录的起始标志一样。增加起始或结束标志将极大地方便数据的阅读和调试,也大大增强了系统的冗余性,一个很大的数据文件不会再因为很小的一点错误而导致该错误后面的数据全部丢失。

3 结 语

《地球空间数据交换格式》给出了二维矢量数据的通用交换格式,该标准的推广对于在我国地理空间信息的广泛共享,其作用是明显的,对于数字中国的建设无疑也起到了积极的作用。但标准中对于电子地图数据交换层面上的考虑有待于进一步的补充和完善,本文建议扩充的部分结构和功能将不会破坏现有标准的结构和原则,但却可以大大增强现有标准的适用范围,特别是对于纯电子地图数据的转换。

参 考 文 献

- 1 GB/T 17798-1999. 中华人民共和国国家标准《地球空间数据交换格式》. 北京: 中国标准出版社, 1999
- 2 龚健雅. 关于我国空间数据标准若干问题的思考. 测绘标准化, 1997, 13(1): 2~3
- 3 龚健雅, 薛明. 我国空间数据交换格式. 中国地理信息系统协会、中国海外地理信息系统协会'98 学术年会, 北京, 1998. 207~217
- 4 王艳东, 龚健雅, 黄俊韬, 等. 基于中国地球空间数据交换格式的数据转换方法. 测绘学报, 2000, 29(2): 142~148
- 5 徐志斌, 谭海樵, 史永锋, 等. 用 XML 实现 CNSDTF 的矢量数据交换格式. 中国矿业大学学报, 2002, 31(5): 344~348

作者简介: 陈泽民, 高级工程师, 博士生. 主要从事地图学与地理信息系统研究.

E-mail: chenzemin@hotmail.com

Application Research on CNSDTF-VCT

CHEN Zemin¹

(1 Department of Urban and Resources, Nanjing University, 22 Hankou Road, Nanjing 210093, China)

Abstract: CNSDTF is a national commend standard for the transformation of geo-spatial vector data. There are some difficulties when transferring digital map with this CNSDTF. This paper expounds a new concept that there are two kinds of attribute classes in geo-spatial vector data after researching this standard earnestly, and discuss the extension of functions for CNSDTF briefly.

Key words: vector data; position data; non-graph attribute

About the author: CHEN Zemin, senior engineer, Ph. D candidate, majors in geography and GIS.

E-mail: chen-zemin@hotmail.com

(责任编辑: 洪远)

(上接第 446 页)

Automatic Extraction of Skeleton and Center of Area Feature

CHEN Tao¹ AI Tinghua¹

(1 School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: Reviewing the traditional vector parallel line based and raster based method for extracting the skeleton and center of polygon, this paper presents an approach to the automatic extraction of skeleton and center based on the constraint Delaunay triangulation. The triangle in the local joint location acts as the root of the extraction tree. This paper calculates the area size split by the above triangle and makes the extension continue in the two directions whose areas are bigger. The operation repeats until all extensions ends at the vertex of polygon. The connection of all the midpoints of triangle's edges in sequence forms the skeleton of polygon. The algorithm is developed in detail and its properties are analyzed by experiment.

Key words: Delaunay triangulation; polygon shape center; skeleton searching

About the first author: CHEN Tao, postgraduate, majors in map generalization, electronic map.

E-mail: qwe123asd@etang.com

(责任编辑: 晓平)