

# 地理信息系统导论

## 一、名词解释

**GIS（地理信息系统）：**是在计算机软硬件支持下，以空间数据库为基础，运用系统工程和信息科学的理论，对空间数据进行科学管理和综合分析，为规划决定，管理和研究提供信息，是介于地理科学，空间科学和管理的新兴边缘学科。

**元数据：**

①“中介数据”，“中继数据”，为描述数据的数据，主要描述数据属性的信息，用来支持如指示存储位置、历史数据、资源查找、文件记录等功能。

②关于数据的数据，对数据库内容的全面描述，其目的是促进数据集的高效利用和充分共享。

**坡度：**坡面的垂直高度与水平方向的距离之比。是地表位置上高度变化率的量度，坡度可以表达为百分数或者度数。

**坡向：**是斜坡方向的量度。坡面的法线在水平面投影的方向。

**栅格：**将空间分割成有规则的网格，在各个网格上给出相应的属性值来表示地理实体的一种数据组织形式。

**栅格数据模型：**使用行、列、像元来构建空间要素。

**矢量数据模型：**离散对象模型。采用点及其  $x$ ,  $y$  坐标来构建点线面空间要素，然后使用点、线、多边形来表示具有清晰空间位置和边界的空间要素。

**拓扑：**研究几何对象在弯曲或拉伸等变换下仍保持不变的性质。它只考虑物体间的位置关系而不考虑它们的形状和大小。

**专题地图：**用来强调一种主题的空间分布的地图。

**DTM：**数字地面模型，利用一个任意坐标系中大量选择的已知  $x$ ,  $y$ ,  $z$  的坐标点对连续地面的一个简单的统计表示。

**DEM：**数字高程模型，一定范围内规则格网点的平面坐标  $(x, y)$  及其高程  $(z)$  的数据集、主要描述区域地貌形态的空间分布。

**DSM：**数字表面模型，包含了地表建筑物，桥梁和树木等高度的地面高程模型。

**DOM:** 数字正射影像, 利用 DEM 对航空航天影像进行正射纠正、接边、色彩调整、镶嵌, 并按照一定范围裁切生成的数字正射影像数据集。

**地理编码:** 地址区配, 在含地址的表格数据与相关图层之间建立联系, 并为表格数据创建一个相应的点要素图层。

**重采样:** 在对采样后形成的由离散数据组成的数字图像中按所需的像元位置或像元间距离重新采样, 以构成几何变换后的新图像。

**空间插值:** 根据已知点数值来估算其他点数值。常用于将离散点的测量数据转换为连续的曲面数据。

**游程编码:** 逐行将相同值的栅格合并, 记录合并后栅格的值及合并栅格的数量, 其目的是压缩数据量, 消除数据间的冗余。

**四叉数:** 按四个象限进行递归分割  $n$  次, 每次分割形成  $2n \times 2n$  个子象限中的属性数值都相同为止, 该子象限就不再分割。一种压缩数据结构, 它把地理空间定量划分为可变大小的网格, 每个网格具有相同性质的属性。(将空间区域按照四个象限进行递归分割 ( $2n \times 2n$ , 且  $n \geq 1$ ), 直到子象限的数值单调为止。凡数值 (特征码或类型值) 呈单调的单元, 不论单元大小, 均作为最后的存储单元。这样, 对同一种空间要素, 其区域网格的大小, 随该要素分布特征而不同。是一种压缩数据结构。)

**数据:** 通过数字化或记录下来可以被鉴别的符号, 是客观对象的表示, 是信息的表达, 只有当数据对实体行为产生影响时才成为信息。

**信息系统:** 具有数据采集、管理、分析和表达数据能力的系统, 它能够为单一的或有组织的决策过程提供有用的信息。包括计算机硬件、软件、数据和用户四大要素。

**不规则三角网模型:** 简称 TIN, 它根据区域有限个点集将区域划分为相连的三角面网络, 区域中任意点落在三角面的顶点、边上或三角形内。如果点不在顶点上, 该点的高程值通常通过线性插值的方法得到 (在边上用边的两个顶点的高程, 在三角形内则用三个顶点的高程)。(TIN, 一种复合矢量数据模型, 它采用一套互不重叠的三角形来近似表示地形。)

**拓扑关系:** 指网结构元素结点、弧段、面域之间的空间关系, 主要表现为拓扑邻接、拓扑关联、拓扑包含。根据拓扑关系, 不需要利用坐标或距离, 可以确定一

种地理实体相对于另一种地理实体的位置关系，拓扑数据也有利于空间要素的查询。可根据拓扑关系重建地理实体。

拓扑结构：在点、线和多边形之间建立关联，以及彻底解决邻域和岛状信息处理问题而必须建立的数据结构。这种结构应包括以下内容：唯一标识，多边形标识，外包多边形指针，邻接多边形指针，边界链接，范围（最大和最小  $x$ 、 $y$  坐标值）。

矢量数据结构：空间数据结构：指适合于计算机系统存储、管理和处理的地学图形的逻辑结构，是地理实体的空间排列方式和相互关系的抽象描述。

优：数据精度高、数据冗余低、易于进行网络分析、图像质量高  
缺：数据结构、处理算法复杂、叠合分析困难、技术要求和成本高。

矢量数据结构：利用欧几里得几何学中的点、线、面及其组合体来表示地理实体空间分布的一种数据组织方式。这种数据组织方式能最好地逼近地理实体的空间分布特征，数据精度高，数据存储的冗余度低，便于进行地理实体的网络分析，但对于多层空间数据的叠合分析比较困难。

栅格数据结构：基于栅格模型的数据结构简称为栅格数据结构，指将空间分割成有规则的网格，在各个网格上给出相应的属性值来表示地理实体的一种数据组织形式。

空间索引：指依据空间对象的位置和形状或空间对象之间的某种空间关系按一定的顺序排列的一种数据结构，其中包含空间对象的概要信息。作为一种辅助性的空间数据结构，空间索引介于空间操作算法和空间对象之间，它通过筛选作用，大量与特定空间操作无关的空间对象被排除，从而提高空间操作的速度和效率。

空间数据编码：指将数据分类的结果，用一种易于被计算机和人识别的符号系统表示出来的过程。编码的目的是用来提供空间数据的地理分类和特征描述，同时为了便于地理要素的输入、存储、管理，以及系统之间数据交换和共享的需要。

数据库管理系统：操作和管理数据库的软件系统，提供可被多个应用程序和用户调用的软件系统，支持可被多个应用程序和用户调用的数据库的建立、更新、查询和维护功能。

空间数据库：地理信息系统在计算机物理存储介质上存储的与应用相关的地理空间数据的总和，一般是以一系列特定结构的文件的形式组织在存储介质之上的。

缓冲区分析：根据分析对象的点、线、面实体，自动建立他们周围一定距离的带状区，用以识别这些实体或主体对邻近对象的辐射范围或影响度，以便为某项分析或决策提供依据。

栅格数据结构  
·  
优：数据结构简单、易于算法实现、空间数据叠置和组合容易、输出快速成本低  
缺：数据量大、数据精度低、难以建立空间实体间的拓扑关系、图像质量较低

叠合分析：指在统一空间参照系统条件下，每次将同一地区两个地理对象的图层进行叠合，以产生空间区域的多重属性特征，或建立地理对象之间的空间对应关系。

空间分析：是基于空间数据的分析技术，它以地学原理为依托，通过分析算法，从空间数据中获取有关地理对象的空间位置、空间分布、空间形态、空间形成、空间演变等信息。

网络分析：运筹学模型中的一个基本模型，即对地理网络和城市基础设施网络进行地理分析和模型化。它的根本目的是研究、筹划一项网络工程如何安排，并使其运行效果最好。

空间信息可视化：是地理信息处理的窗口与处理结果的直观表达形式，因而是决策的直观依据。只有把空间数据库中的海量数据转换为直观的图形信息，地理信息处理结果才能为规划、管理与决策提供有力的支撑。

弧段：有序的坐标集合，用于表示在给定的比例尺上窄到无法表示为面的地理要素。

SQL 查询：SQL 结构化查询语言，是一种数据库查询和程序设计语言，用于存取数据以及查询、更新和管理关系数据库系统。同时也是数据库脚本文件的扩展名。

视域：从一个或多个观察点可以看见的地表范围。

提取视域的过程称为视域分析。

流量栅格：显示已填建热网上每个像元水流流向的栅格。

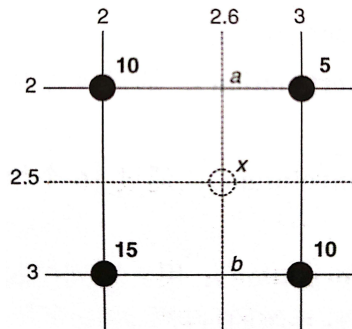
流域：将水和其中物质排向共同出水口的区域范围。

流域分析：包括水流方向计算，流域边界提取和河网分析。

## 二、计算题

### 1、双线性插值的计算

双线性插值法使用原始图像 4 个最邻近像元值计算出新图像的像元值，图中的像元  $x$  代表新图像上的像元，其像元值需要由原始图像推导得出。像元  $x$  在原始图像的相应位置是  $(2.6, 2.5)$ ，其 4 个邻近像元的图像坐标为  $(2, 2)$ 、 $(3, 2)$ 、 $(2, 3)$  和  $(3, 3)$ ，对应的像元值分别是 10、5、15 和 10。



进行双线性插值，首先，沿着扫描线 2 和 3 进行二次插值，从而推导出  $a$  和  $b$  插值：

$$a = 0.6(5) + 0.4(10) = 7$$

$$b = 0.6(10) + 0.4(15) = 12$$

接下来，进行  $a$  和  $b$  的第三次插值，推导出  $x$  的插值：

$$x = 0.5(7) + 0.5(12) = 9.5$$

### 2、描述性统计量

描述性统计量概括数据集的数值，假设数据集按升序排列：

- 值域：最大值与最小值之差。
- 中值：中间值，或者第 50 个百分位数值。
- 第一个四分位数：第 25 个百分位数值。 *第 25% 的数字*
- 第三个四分位数：第 75 个百分位数值。
- 平均值：数据值的平均值。
- 方差：用来衡量数据相对于均值的离散程度。
- 标准差：方差的平方根。
- Z 得分：由  $(x - \text{mean}) / s$  计算的标准得分，式中  $s$  代表标准差。

注释栏中包含了 1990—2000 年间美国各州人口变化率的描述性统计量数值。

DC	-5.7	VT	8.2	AL	10.1	NM	20.1
ND	0.5	NE	8.4	MS	10.5	OR	20.4
WV	0.8	KS	8.5	MD	10.8	WA	21.1
PA	3.4	SD	8.5	NH	11.4	NC	21.4
CT	3.6	IL	8.6	MN	12.4	TX	22.8
ME	3.8	NJ	8.6	MT	12.9	FL	23.5
RI	4.5	WY	8.9	AR	13.7	GA	26.4
OH	4.7	HI	9.3	CA	13.8	ID	28.5
IA	5.4	MO	9.3	AK	14.0	UT	29.6
MA	5.5	KY	9.6	VA	14.4	CO	30.6
NY	5.5	WI	9.6	SC	15.1	AZ	40.0
LA	5.9	IN	9.7	TN	16.7	NV	66.3
MI	6.9	OK	9.7	DE	17.6		



该数据集的描述性统计量如下所示：

- 平均值:13.45
- 中位数:9.7
- 值域:72.0
- 第一个四分位数:7.55, 位于 MI (6.9) 和 VT (8.2) 之间
- 第三个四分位数:17.15, 位于 TN (16.7) 和 DE (17.6) 之间
- 标准差:11.38
- 内华达 (66.3) 的 Z 得分:4.64

$T = -0.4744 \times 57.296 = -27.181$   
 因为  $S < 0$  且  $n_x < 0$ , 则  
 坡向 =  $270 - (-27.181) = 297.181$   
 作为比较, Fleming 和 Bofter 算法得到  $S_f$  的值为 17.16; Sharpack 和 Akin 算法得到  $S_f$  的值为 17.39, 而坡向值分别为 299.06 和 296.56。

### 3、栅格计算坡度和坡向

下表表示高程栅格的一个 3\*3 窗口。高程以 m 为单位, 像元大小为 30m。

1006	1012	1017
1010	1015	1019
1012	1017	1020

用 Horn 算法计算中央像元的坡度和坡向:

$$n_x = (1006 + 2 \times 1010 + 1012) - (1017 + 2 \times 1019 + 1020) = -37$$

$$n_y = (1012 + 2 \times 1017 + 1020) - (1006 + 2 \times 1012 + 1017) = 19$$

$$S = [(-37)^2 + (19)^2]^{1/2} / (8 \times 30) = 0.1733$$

$$S_p = 100 \times 0.1733 = 17.33$$

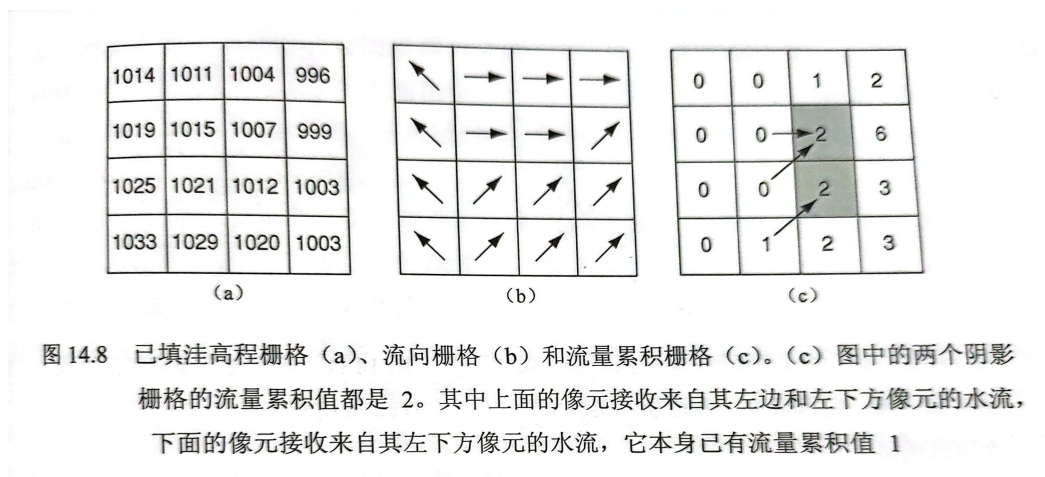
$$D = \arctan(n_y/n_x) = \arctan(19/-37) = -0.4744$$

$$T = -0.4744 \times 57.296 = -27.181$$

因为  $S < 0$  且  $n_x < 0$ , 则

$$\text{坡向} = 270 - (-27.181) = 297.181$$

### 4、流量栅格计算



### 5、游程编码

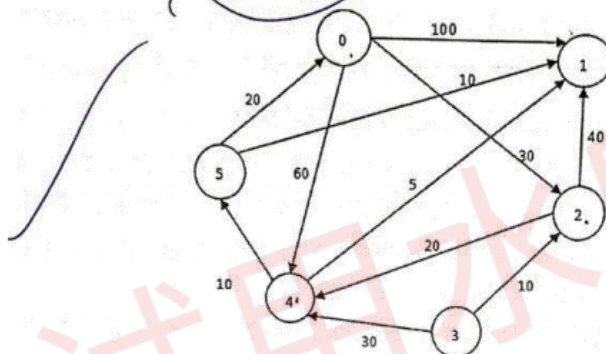
0	2	2	5	5	5	5	5
2	2	2	2	2	5	5	5
2	2	2	2	3	3	5	5
0	0	2	3	3	3	5	5
0	0	3	3	3	3	5	3
0	0	0	3	3	3	3	3
0	0	0	0	3	3	3	3
0	0	0	0	3	3	3	3

沿行方向进行编码:  $(0, 1)$ ,  $(2, 2)$ ,  $(5, 5)$ ;  $(2, 5)$ ,  $(5, 3)$ ;  $(2, 4)$ ,  $(3, 2)$ ,  $(5, 2)$ ;  $(0, 2)$ ,  $(2, 1)$ ,  $(3, 3)$ ,  $(5, 2)$ ;  $(0, 2)$ ,  $(3, 4)$ ,  $(5, 1)$ ,  $(3, 1)$ ;  $(0, 3)$ ,  $(3, 5)$ ;  $(0, 4)$ ,  $(3, 4)$ ;  $(0, 5)$ ,  $(3, 3)$ .

## 6、迪杰斯特拉算法（最短路径计算）

Dijkstra 算法亦称为标号算法，是一个按路径长度递增的次序产生最短路径的算法。

下图所示为一个网络的带权有向图，图中有 6 个结点，11 条线路，现用 Dijkstra 算法求取从节点  $V_0$  到其它节点的最佳路径。



图：一个网络的带权有向图

计算步骤如下：

(1) 对 0 做标记：

$$D(1)=100; D(2)=30; D(3)=\infty; D(4)=60; D(5)=\infty. \text{ 即 } 2 \text{ 最短. } 0 \rightarrow 2$$

(2) 对 2 做标记：

$$D(1)=\{100, 30+40\}=70; D(3)=\{\infty, 30+\infty\}=\infty; D(4)=\{60, 30+20\}=50; D(5)=\{\infty, 30+\infty\}=\infty; \text{ 即 } 4 \text{ 最短. } 0 \rightarrow 2 \rightarrow 4$$

(3) 对 4 做标记：

$$D(1)=\{70, 50+5\}=55; D(3)=\{\infty, 50+\infty\}=\infty; D(5)=\{\infty, 50+10\}=60; \text{ 即 } 1 \text{ 最短. } 0 \rightarrow 2 \rightarrow 4 \rightarrow 1$$

(4) 对 1 做标记：

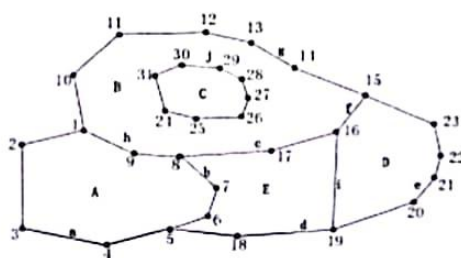
$$D(3)=\{\infty, 55+\infty\}=\infty; D(5)=\{60, 55+\infty\}=60; \text{ 即 } 5 \text{ 最短. } 0 \rightarrow 2 \rightarrow 4 \rightarrow 1 \rightarrow 5$$

(5) 对 5 做标记：

$$D(3)=\{\infty, 60+\infty\}=\infty;$$

综上，用 Dijkstra 算法求取的从节点  $V_0$  到节点  $V_1$  的最佳路径为： $V_0 \rightarrow V_2 \rightarrow V_4 \rightarrow V_1$ ；节点  $V_0$  到节点  $V_2$  的最佳路径为： $V_0 \rightarrow V_2$ ；节点  $V_0$  到节点  $V_3$  的最佳路径为： $V_0$  无法到达  $V_3$ ；节点  $V_0$  到节点  $V_4$  的最佳路径为： $V_0 \rightarrow V_2 \rightarrow V_4$ ；节点  $V_0$  到节点  $V_5$  的最佳路径为： $V_0 \rightarrow V_2 \rightarrow V_4 \rightarrow V_5$ 。

## 7、点、线、面的拓扑关系



多边形号	弧段号	周长	面积	中心点坐标
A	h,b,a			
B	g,f,c,h,j			
C	j			
D	e,i,f			
E	e,i,d,b			

多边形文件

弧段号	起始点	终止点	左多边形	右多边形	弧段号	点号
a	5	1	O	A	a	5,4,3,2,1
b	8	5	E	A	b	8,7,6,5
c	16	8	E	B	c	16,17,8
d	19	5	O	E	d	19,18,5
e	15	19	O	D	e	15,23,22,21,20,19
f	15	16	D	B	f	15,16,
g	1	15	O	B	g	1,10,11,12,13,14,15
h	8	1	A	B	h	8,9,1
i	16	19	D	E	i	16,19
j	31	31	B	C	j	31,30,29,28,27,26,25,24,31

弧段文件

弧段点文件

## 8、MD 码计算

一种按位操作的方法：

如行号为2、列号为3的栅格的MD（注意：其始行列号从0计）

步骤：

(1) 行、列号为二进制  $I_b = 10$   $J_b = 11$

(2) I行J列交叉

$$\begin{array}{c} \diagup \quad \diagdown \\ \diagdown \quad \diagup \\ 1 \quad 1 \quad 0 \quad 1 = 13 \end{array}$$

(3) 再化为十进制。

实质上是按左上、右上、左下、右下的顺序，从零开始对每个栅格进行自然编码。前后两个MD码之差即代表了叶结点的多少。

A 0	A 1	A 4	A 5
A 2	B 3	B 6	B 7
A 8	A 9	B 12	B 13
A 10	A 11	B 14	B 15

红果研上还有一些除了名词解释和简答的题可以了解，变成测绘综合之后是没考过的，而且矿大的本科重点也不是这个，但是以防万一还是要了解。