

Conflation of Multisource Geospatial Vector Data

多源地理空间 矢量数据融合

唐文静 著

清华大学出版社



Conflation of Multisource Geospatial Vector Data

多源地理空间 矢量数据融合

唐文静 著



清华大学出版社
北京

内 容 简 介

本书从多源空间数据的概念、研究内容和应用出发,就多源地理空间矢量数据融合的预处理、几何位置融合、要素分类编码融合以及融合过程的不确定性分析等问题进行了探讨。并以海陆地理空间矢量数据的融合为例,讨论了不同来源的1:500 000海图、陆图矢量数据融合的坐标系统一、投影方式统一、几何位置融合、属性编码融合等问题,通过多源矢量空间数据不确定性联合模型来评定海陆地理空间矢量数据融合的质量。

全书共分6章。第1章介绍多源空间数据融合的概念、发展及与其他技术的区别;第2章介绍多源空间数据融合的研究内容、处理过程及进行融合之前需要做的预处理;第3章和第4章介绍多源空间数据融合中的几何位置融合;第5章介绍要素编码融合;第6章对多源空间数据融合结果质量进行评定。

本书可作为地理、测绘、环境、空间信息等专业的本科生和研究生的教学参考用书,还可作为各领域广大地理工作者的参考用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

多源地理空间矢量数据融合/唐文静著. —北京:清华大学出版社,2014

ISBN 978-7-302-38138-9

I. ①多… II. ①唐… III. ①地理信息系统—空间信息系统—数据融合 IV. ①P208

中国版本图书馆CIP数据核字(2014)第227853号

责任编辑:白立军

封面设计:付瑞学

责任校对:焦丽丽

责任印制:李红英

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:北京教图数字印刷有限公司

经 销:全国新华书店

开 本:185mm×230mm

印 张:8.25

字 数:171千字

版 次:2014年10月第1版

印 次:2014年10月第1次印刷

印 数:1~400

定 价:35.00元

产品编号:061795-01

前 言

根据空间数据的表达方式,空间数据可以分为矢量数据和栅格数据,其中矢量数据是GIS和数字制图中最重要数据来源。随着信息技术的进步,各种测绘技术的实用化和获取空间信息途径的多样化,导致地图生产的资料越来越多,如纸质地图、数字地图、卫星遥感影像、GPS测量数据等,空间数据的多源性引发了数据属性的差异,导致空间数据间多层次、多方面的差异,如空间特征的差异、分类分级的差异等,给GIS部门之间的数据共享和数据集成带来极大困难。解决这一问题最有效的方法是多源地理空间矢量数据融合技术。这样研究不同数据源、不同数据精度和不同数据模型的融合问题就有非常重要的现实意义。

(1) 在地图制图过程中合理有效地综合应用各种现势性资料(数据),可以起到保证地图质量、提高地图生产效率的作用。

(2) 通过数据融合可以重新组合专题数据,补充物体的分类分级和属性,进一步改善物体的几何精度。

(3) 可以减少空间数据的多语义性、多时空性、多尺度性、存储格式的不同以及数据模型与存储结构的差异,消除其不一致性,使新生成的地图数据在逻辑一致性及完整性方面得到提高。

(4) 通过融合,可以在最大程度上实现不同空间数据的完全转换,新生成的空间数据在某些方面(如点位精度、详细程度和现势性等方面)都得到极大的提高,这对进一步扩大空间数据的应用范围,提高数据重复使用率,降低数据获取费用,有着非常重要的意义。

(5) 可以促进空间数据的共享和集成:目前空间数据集成主要停留在数据格式统一或一致化的层次上,要实现集成仅依靠数据标准或数据规范是远远不够的,还需要依赖数据内容甚至数据知识等更高层次的集成,空间数据融合技术的研究能为其提供强有力的技术手段。

(6) 可以促进不同的空间数据部门的合作,避免“数据孤岛”的形成。

(7) 可以节约大量的人力、财力,从而促进国民经济的发展。

21世纪是海洋的世纪,为满足我国海洋开发及沿海城市信息化建设的需求,本书以海陆地理空间矢量数据的融合为例来探讨多源数据融合过程中需要解决的关键问题,并以几何数据融合、要素编码融合及其融合结果的不确定性为重点进行介绍。几何位置融合主要是为了解决不同来源数据间同名地物的不一致性问题,包括同名实体的识别匹配

和调整变换两个过程,同名实体的识别匹配是解决几何数据融合的基础,调整变换是在同名实体匹配的基础上,建立两个地图之间的局部坐标转换关系,它是解决几何数据融合的关键;要素分类编码融合是为了解决不同来源数据间属性数据的差异,从要素分类编码融合的原则和步骤进行探讨;数据融合是一个复杂的不确定信息处理过程,融合过程的不确定性分析是为了科学的度量和表达数据质量,从不确定信息的角度对融合结果的质量进行评定。

多源地理空间矢量数据融合是地理信息领域致力解决的问题。本书是作者在密切跟踪该领域技术研究成果的基础上总结而成的,是一本全面论述多源地理空间矢量数据融合技术的著作,全书图文并茂,深入浅出,可读性强,并将理论与实践有机结合,以期为进一步学习、研究和应用打下基础。但多源地理空间矢量数据融合技术是一个复杂的问题,在本书撰写过程中参考了国内外有关数据融合方面的大量文献,直接引用的有数百篇。在此,向相关作者表示衷心的感谢。

本书的出版得到国家自然科学基金(No. 61170161)、山东省自然科学基金(No. ZR2012FQ029)、鲁东大学校基金(No. LY2010014)的资助。另外,本书的编写还得到鲁东大学信电学院院长邹海林教授、哈尔滨工程大学自动化学院许兆新教授的大力支持,在此对这些同志的鼓励和帮助表示衷心的感谢。

特别感谢清华大学出版社,感谢责任编辑及其他参与此书编辑工作的各位老师为本书顺利出版而付出的辛勤劳动。

由于作者水平有限,书中不足之处在所难免,恳请广大读者和同行批评指正。

唐文静

2014年7月于烟台

目 录

第 1 章 绪论	1
1.1 引言	1
1.2 多源空间数据融合	1
1.2.1 空间数据定义	1
1.2.2 空间数据特点	2
1.2.3 多源空间数据	3
1.2.4 多源空间数据融合	5
1.3 多源空间矢量数据融合发展	9
1.3.1 矢量数据融合研究	9
1.3.2 数据匹配技术研究	11
1.3.3 数据合并变换技术研究	13
1.4 与其他技术的区别	14
1.4.1 与空间数据集成的区别	14
1.4.2 与制图综合的区别	16
1.4.3 与地图叠置操作的区别	17
1.5 本书的结构安排	18
第 2 章 多源地理空间矢量数据融合的理论基础及预处理	19
2.1 多源空间矢量数据融合的研究内容及应用	19
2.1.1 多源空间矢量数据融合的研究内容	19
2.1.2 多源空间矢量数据融合的应用	21
2.2 多源地理空间矢量数据融合的预处理	23
2.2.1 空间数据模型	23
2.2.2 坐标系转换原理	26
2.2.3 地图投影的基本原理	32
2.2.4 多源地理空间矢量数据融合的预处理	36
2.3 海陆空间矢量数据融合的预处理	37
2.4 本章小结	39

第 3 章	多源地理空间矢量数据的几何位置融合之匹配	40
3.1	多源空间同名地物的匹配策略	40
3.2	基于空间关系的同名地物的匹配	41
3.2.1	点实体的匹配	41
3.2.2	线实体匹配	43
3.2.3	面实体匹配	45
3.2.4	同名实体匹配算法分析	46
3.3	基于空间相似性的实体匹配算法	47
3.3.1	算法提出	47
3.3.2	相似度的计算	49
3.3.3	算法步骤	51
3.4	基于本体的同名地物的匹配	53
3.4.1	本体匹配的基本理论	53
3.4.2	语义相似度的计算	54
3.4.3	实例分析	57
3.5	海陆图几何位置融合之实体匹配	57
3.5.1	仿真实验	57
3.5.2	结果分析	60
3.6	本章小结	62
第 4 章	多源地理空间矢量数据的几何位置融合之调整	63
4.1	几种调整变换算法及分析	63
4.1.1	基于同名点三角剖分的地物调整方案	63
4.1.2	基于拓扑关系的合并变换算法	64
4.1.3	基于平差原理的地物调整方案	65
4.1.4	调整变换算法分析	66
4.2	评价因素分析	66
4.2.1	同名要素对周围地物的影响度	67
4.2.2	同名要素位置准确度的影响	68
4.2.3	同名要素在不同图上的重要性	68
4.3	基于多评价因素的调整变换算法	68
4.3.1	点状要素调整变换算法	69
4.3.2	线状要素调整变换算法	71
4.3.3	面状要素调整变换算法	73
4.4	海陆图几何位置融合之调整仿真	75

4.4.1	仿真实验	75
4.4.2	结果分析	82
4.5	本章小结	84
第 5 章	多源地理空间矢量数据的编码融合	85
5.1	地理要素分类编码模型	85
5.2	地理要素分类分级与编码方法	86
5.2.1	地理要素分类分级原则与方法	86
5.2.2	地理要素属性编码原则与方法	89
5.3	多源地理空间数据分类编码融合的原则及步骤	92
5.3.1	要素分类编码融合的原则	93
5.3.2	要素分类编码融合的步骤	94
5.4	海陆图要素分类编码融合的实现	95
5.4.1	海陆图要素属性编码差异	95
5.4.2	海陆图要素属性编码融合的实现	96
5.5	本章小结	100
第 6 章	多源空间矢量数据融合的不确定性分析	101
6.1	不确定性概述	101
6.1.1	数据不确定性研究意义	101
6.1.2	数据不确定性研究现状	103
6.1.3	数据不确定性研究方法工具	104
6.2	多源空间矢量数据融合的不确定性产生原因	107
6.2.1	空间对象自身的不确定性	107
6.2.2	空间对象抽象的不确定性	108
6.2.3	空间数据处理过程中产生的不确定性	109
6.2.4	空间数据使用中的不确定性	109
6.3	多源空间矢量数据不确定性模型	110
6.3.1	单源空间矢量数据不确定性模型	110
6.3.2	空间数据不确定性传播定律	111
6.3.3	多源空间矢量数据融合不确定性模型	112
6.4	海陆空间矢量数据不确定分析	112
6.5	本章小结	114
	参考文献	115

第1章 绪 论

1.1 引 言

空间数据是地理信息系统(Geographic Information System, GIS)的操作对象,它们的获取在 GIS 工程里占有很重要的地位。不同部门根据特定的应用目的,同一地区、同一比例尺的空间数据往往采用不同的空间数据标准、特定的数据模型和特定的空间物体分类分级体系进行重复采集。这不仅造成了人力、财力的巨大浪费,还引发了空间数据的多语义性、多时空性、多尺度性、存储格式不同以及数据模型与存储结构的差异等,给 GIS 部门之间的数据共享和数据集成带来极大困难^[1,2]。为了充分利用已有信息、减少数据采集的高额开销, GIS 应用部门经常需要将这些不同来源的地图数据合并在一起使用,这就产生了多源地理空间数据融合技术。目前的空间数据的种类和数量很多,并且空间数据之间存在冗余性、互补性和合作性^[3~5]。冗余性是指它们对同一空间实体的表示、描述和解译相同;互补性指空间信息来自不同的角度来测量空间实体;合作性是指不同的空间信息在处理和分析过程中对其他空间信息具有依赖关系。将现有的、多个“单一表现”的数据库联系在一起,建立表示同一实体的不同空间对象联系,来维持一致性,这个过程也依赖于空间数据融合技术^[6,7,8]。

矢量数据是 GIS 和数字制图中最重要数据源,矢量数据之间的融合是应用最广泛的空间数据融合形式,也是空间数据融合研究的重点。从需求分析上说,需要进行矢量数据融合的情况一般为对数据信息进行更改、更新、增加以适应某种特定的需要。随着科技的发展,传统的海洋经济增长方式不能满足国民经济建设的需要,对海洋的开发必然要向沿岸转移。所以,就海陆地理空间中的矢量数据融合技术进行研究是必要的,有重要意义。它既可为建立应用领域信息共享统一地理信息系统平台打下基础,又可为沿岸城市的海洋开发、城市发展、规划决策提供支撑环境。

1.2 多源空间数据融合

1.2.1 空间数据定义

空间数据(Spatial Data)是指以地球表面空间位置为参照的自然、社会和人文经济景

观数据,可以是图形、图像、文字、表格和数字等。它是由系统的建立者通过数字化仪、扫描仪、键盘、磁带机或其他系统通信输入 GIS,是系统程序作用的对象,是 GIS 所表达的现实世界经过模型抽象的实质性内容^[9]。它可以表示空间实体的位置、形状、大小及其分布特征等诸多方面信息,可以描述来自现实世界的目标,具有定位、定性、时间和空间关系等特性。定位是指在一个已知的坐标系里空间目标都具有唯一的空间位置;定性是指有关空间目标的自然属性,它伴随着目标的地理位置;时间是指空间目标是随着时间变化的;空间关系通常又称为拓扑关系。

空间数据适用于描述所有呈二维、三维甚至多维分布的关于区域的现象,空间数据不仅能够表示实体本身的空间位置及形态信息,而且还有表示实体属性和空间关系(如拓扑关系)的信息。在空间数据中不可再分的最小单元称为空间实体,空间实体是对存在于这个自然世界中地理实体的抽象,主要包括点、线、面以及体等基本类型:如把一根电线杆抽象成为一个点,该点可以包含电线杆所处的位置信息、电线杆的高度信息和其他一些相关信息;可以把一条道路抽象为一条线,该线可以包含这条道路的长度、宽度、起点以及道路等级等相关信息;可以把一个湖泊抽象为一个面,该面可以包含湖泊的周长、面积和湖水的质量信息等;在空间对象建立后,还可以进一步定义其相互之间的关系,这种相互关系称为“空间关系”,又称为“拓扑关系”,如可以定义点-线关系、线-线关系、点-面关系等。因此可以说空间数据是一种可以用点、线、面以及体等基本空间数据结构来表示人们赖以生存的自然世界的数据库。

根据空间数据对空间实体不同特征的描述,可将空间数据分为以下三类^[3]。

1. 属性数据

属性数据主要用于描述空间实体的属性特征,如实体的名称、类型、等级等。

2. 几何数据

几何数据主要用于描述空间实体的空间特征,有时也称为位置数据或者定位数据,如点的平面坐标、地理坐标等。

3. 关系数据

关系数据主要用于描述空间实体之间的关系,多数情况下描述的是拓扑关系,如邻接、包含和关联等。

1.2.2 空间数据特点^[1,10]

空间数据具有数据的一般性质,又具备自身的一些特性,这些自身特性是引发地理空间数据多源性的原因。

1. 抽样性

为了能以数字的方式对空间物体进行描述,必须以有限的抽样数据表述无限的连续物体。空间物体的抽样不是对空间物体的随机选取,而是对物体形态特征点的有目的的

选取,其抽样的基本准则就是力求准确描述物体的全局和局部的形态特征。抽样方法根据物体的形态特征不同而不同,直接影响空间分析结果的有效性。空间分析中的拟合和插值,就是根据抽样数据重建连续的空间物体。

2. 概括性

空间物体的概括性指对物体形态的化简综合及对物体的取舍,它区别于数据的详细性,是空间数据处理的一种手段。在空间数据库中,由于主题不同,人们可能会舍去较为次要的地物,或者对一些地物的形态在抽样的基础上进行进一步化简,这种化简不是因为比例尺的限制,而是空间数据应用环境和任务的需要。

3. 多态性

空间数据的多态性具有两层含义,一是同样地物在不同情况下的形态差异。例如,就形态而言,城市在地理空间占据一定范围的地域,可认为它是面状地物,但在比例尺较小的空间数据库中,城市是作为点状地物处理的;河流在现实世界中是具有一定宽度的条带状的面状地物,但在空间数据表示中,可能表示为单线河流或双线河流,即在大多数空间分析中河流是作为线状物体处理的。二是不同地物占据同样的空间位置。大多表现为社会经济人文数据与自然环境数据在空间位置上的重叠。例如,长江是水系要素,但同时在不同的地段上与省界、县界相重叠。

4. 空间性

空间性是指空间物体的位置、形态以及由此产生的系列特性。它是空间数据的最主要特性,如果不考虑空间物体的空间性,空间分析就失去意义。作为非空间数据,两个城市之间的关系可以用一般的数值和逻辑关系来描述,如人口的多少、经济的发展程度等;但是作为空间数据,两个城市之间的关系就会增加,如距离、方位、空间相互作用等。空间性不但导致空间物体的位置和形态的分析处理,同时导致空间相互关系的分析处理。

1.2.3 多源空间数据

由于地理实体的不确定性、人类认识表达能力的局限性、测量误差、数字化采集误差及地理空间数据在计算机中表达的局限性等因素的影响,不同的数据获取手段(遥感、GPS、实地勘测、地图)、不同的专业领域数据、不同的比例尺、不同的获取时间和使用不同软件系统所获取的同一地区的地理空间数据存在差异。这种同一地区多次获取的地理空间数据称为多源地理空间数据。

随着信息技术的进步,空间数据的类型更加复杂多样,导致了空间数据的多源性,其差异表现主要可以概括为以下几个层次^[3~5,11,12]。

(1) 多语义性。GIS研究对象的多种类型特点决定了地理信息的多语义性。不同GIS解决问题的侧重点有所不同,因而会存在语义分异。

(2) 多时空性和多尺度。GIS数据具有很强的时空特性,GIS会根据系统需要采用

不同尺度对地理空间进行表达,不同的观察尺度具有不同的比例尺和不同的精度。

(3) 获取手段多源性。获取地理空间数据的方法多种多样,包括来自现有系统、图表、遥感手段、全球定位系统(Global Positioning System, GPS)手段等。不同手段获得的数据其存储格式和处理手段都各不相同。

(4) 存储格式多源性。

(5) 空间基准不一致。不同来源的空间数据有不同的坐标参考体系和不同的投影方式。

空间数据的多源性引发了数据属性(数学基础、比例尺、用途、时间、精度等)的不同,导致了空间数据的差异,这些差异是多层次和多方面的,给空间数据融合带来困难。多源地理空间矢量数据间可能存在的差异包括^[2,14~19]如下。

1. 数据模型的差异

空间数据模型是在实体概念的基础上发展起来的,它包含两个基本内容,即实体组和实体之间的相互关系^[20]。不同的数据模型有其不同的表示客观事物的方法和不同的处理空间数据联系的方式,即使是同种模型,数据组织方式也会不同,如线实体可能用面条状结构,也可能用结点弧段拓扑结构。

2. 分类分级的差异

地图应用目的不同是导致分类分级差异的根本原因。对分类分级的描述一般体现在属性编码上,不同系统的编码方法一般而言也各有特点。

3. 空间特征的差异

空间特征,表现为地理实体的几何特征(地理实体的位置、形状、大小及其分布特征)和实体间的空间关系。造成地物空间特征差异的原因主要是同一地区的空间数据被多次获取,在数据采集和制图的过程中都会存在各种各样的误差,其中较为重要的误差包括如下。

(1) 解释差异。由于应用目的不同,不同来源的地图中对同一地物的解释存在差异,表现在点采集的密度、小特征的取舍(见图 1.1)、对象的分割等方面。如同一建筑物实体,地形图从作为地形要素的角度来决定其与周围地物的取舍;而房产图则主要从产权产籍管理的角度来决定其分合。

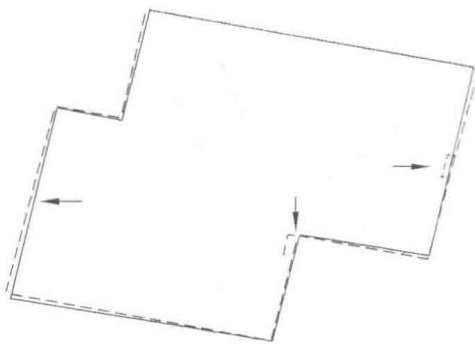


图 1.1 1:500 房产图和地籍图建筑物实体小特征提取差异^[16]

(2) 表达方式差异。主要表现在两幅图所采用的数据结构不同,和对同一种地物要素使用不同的实体类型来表达。

(3) 数据精度差异。例如,相同比例尺的地形图和地籍图,前者对地物的空间坐标位置的

精度要求较高,后者只要求权属界线有较高的精度。

(4) 数据详细程度差异。主要由于地图比例尺不同,受地图综合的影响,对于同一实体信息量大小不同。例如,一条河流在大比例尺下比在小比例尺下具有更多点信息。另外,由于数据获得时间不同,也会因为现实世界的变化而导致两幅图中实体数据不一致。

(5) 同名实体的几何位置差异。不同来源的地图,其几何数据一般都存在差异,引起几何位置差异的原因有很多,如不同的应用目的,导致了不同的采集规范(见图 1.2);多次数字化时,即使采用相同的规范,但不同的作业人员具有不同的素质;不同的更新时间等。

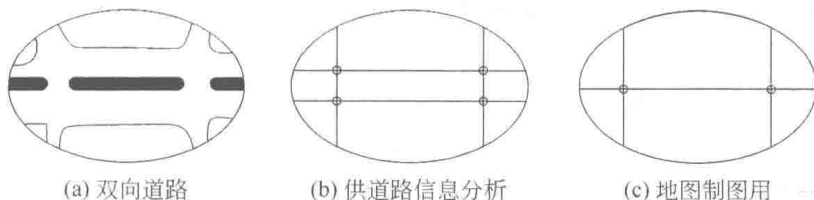


图 1.2 不同应用目的的数据采集差异^[2]

(6) 拓扑结构差异。对特定地物的关心程度以及制图综合等因素都将导致同一地物在不同图上的拓扑关系上的差异,表现在结点的度的不同,弧段的左右多边形号不同等。

1.2.4 多源空间数据融合

多源空间数据融合是一个新兴的研究领域,涉及信号处理、概率统计、信息论、模式识别、人工智能、模糊数学、不确定性理论、计算机科学等交叉学科。由于空间数据融合是一个非常广泛的领域,因此并没有形成统一的定义。

根据空间数据的表达方式,可以把空间数据分为矢量数据和栅格数据,因此,空间数据的融合就分为矢量数据的融合和栅格数据的融合。

1. 栅格数据融合

对于地理空间栅格数据的融合已有系统而深入的研究,通常可分为三级:数据级融合、特征级融合和决策级融合^[21,22],如图 1.3 所示。

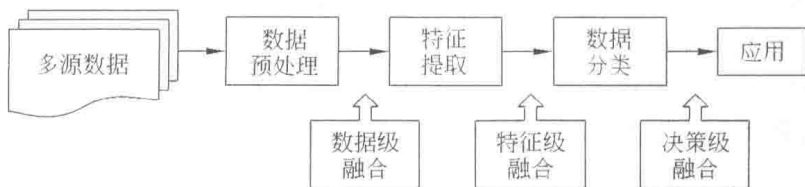


图 1.3 数据融合的 3 个等级

数据级融合是直接在采集到的原始数据层上进行的融合,在多源数据未经预处理之前就进行数据的综合和分析,这是最低层次的融合。如在遥感图像处理过程中对来自两个不同特性的图像的加权融合过程。这种融合的主要优点是尽可能多地保持信息,提供其他融合层次所不能提供的细微信息,具有最高精度和保真度。但局限性也是很明显的,它所处理的数据量太大,故处理代价高、处理时间长、实时性差。

特征级融合属于中间层次融合,它首先对来自传感器的原始信息进行特征提取(可以是目标的边缘、方向、速度等),产生特征矢量,然后对特征矢量进行融合处理。一般来说,提取的特征信息应是像素信息的充分表示量或充分统计量,然后按特征信息对多传感器数据进行分类、汇集和综合。特征级融合的优点在于实现了可观的信息压缩,有利于实时处理,并且由于所提供的特征直接与决策分析有关,因而融合结果能最大限度地给出决策分析所需要的特征信息。其缺点是比数据级融合精度差。

决策级融合是一种高层次的融合,它首先对每一数据进行属性说明,然后对其结果加以融合,得到目标或环境的融合属性说明。决策级融合结果可为指挥控制决策提供依据。因此,决策级融合必须从具体决策问题的需求出发,充分利用特征级融合所提取的测量对象各类特征信息,采用适当融合技术来实现。决策级融合是三级融合的最终结果,是直接针对具体决策目标的,融合结果直接影响决策水平。决策级融合的主要优点:具有很高的灵活性;系统对信息传输带宽要求较低;能有效地反映环境或目标各个侧面的不同类型信息;当一个或几个传感器出现错误时,通过适当的融合,系统还能获得正确的结果,所以具有很强的容错性;通信量小,抗干扰能力强;具有很好的开放性,对传感器的依赖性小;融合中心处理代价低,处理时间短。但是,决策级融合对原传感器信息进行预处理以获得各自的判定结果,所以预处理代价高。

2. 矢量数据融合

矢量数据主要是利用欧几里得(Euclid)几何学中的点、线、面及其组合体来表示地理实体的空间分布,是GIS和数字制图中最重要数据源。矢量数据之间的融合是应用最广泛的空间数据融合形式,也是空间数据融合研究的重点,本书主要是围绕多源地理空间矢量数据的融合来展开的。由于它的应用广泛性,因此并没有统一的定义,已有文献主要有以下几种提法。

1993年,Saalfeld定义数据融合是地图编绘过程中的一个操作,用来“编辑以消除同一地区不同地图之间的差异”^[23]。

1998年,Cobb从数据集成的角度出发,认为“数据融合是将同一地区两个或两个以上不同的地图数据库合并成一个数据库,新生成的地图数据库在某些方面如点位精度、详细程度或现势性等比源数据库好”^[24],即处理的目的是为了获得较高质量的信息。

2003年,郭黎结合空间数据的特点,认为多源地理空间矢量数据融合是指按某种特定的应用目的,将同一地区不同来源的空间数据,采用一定的方法和原则,重新组合专题属性数据,统一物体的分类分级和属性,改善地理空间实体的几何精度,提高地理数据生

产效率和质量,消除以下差异:

- ① 空间物体在不同的空间数据模型中多次采集所产生的数据描述上的差异;
- ② 相同或不同的数据模型采用不同的分类分级方法采集所产生的要素属性差异;
- ③ 空间数据的应用目的不同表现在要素综合详细程度上的差异以及多次数字化所产生的几何位置差异^[2]。

这也是目前比较公认的一种定义。

多源地理空间矢量数据融合是指按某种特定的应用目的,将同一地区不同来源的空间数据,采用一定的方法进行匹配,按照一定的原则对数据进行融合,包括重新组合专题属性数据,改善地理空间实体的几何精度,提高地理数据生产效率和质量,最终产生新的质量更高的数据^[25],空间矢量数据融合的框架流程如图 1.4 所示。这个定义指出了以下几点。

(1) 矢量数据融合研究的对象是同一地区的不同矢量数据,这些数据存在数据的不一致性,包括同一地物在不同图上几何位置、几何形状、拓扑关系、属性数据等方面的不一致性。

(2) 根据不同的研究对象采用不同的匹配方法与融合原则。正确而有效的匹配方法是数据融合的关键技术,融合的原则包括一些经验型的原则和一些有效的算法。

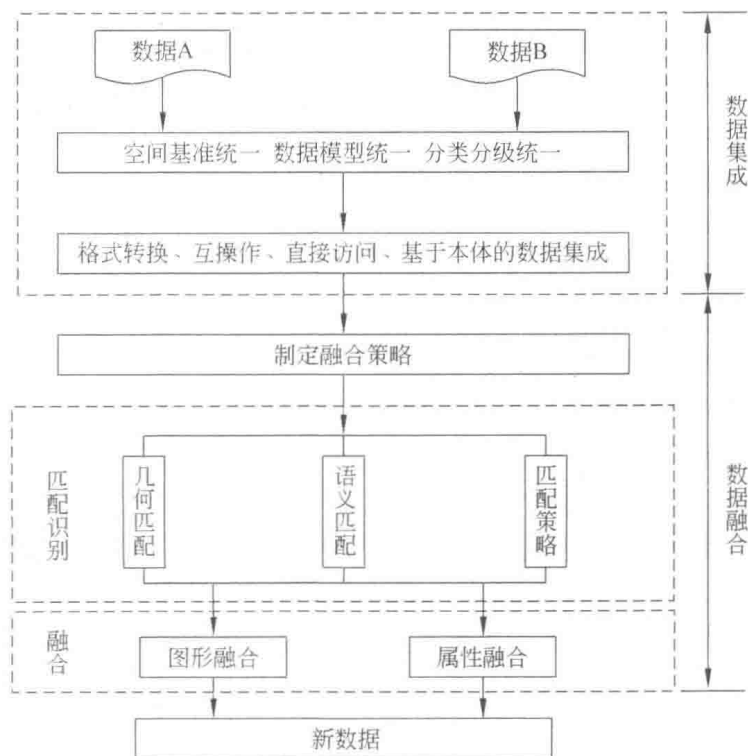


图 1.4 地理空间矢量数据融合的框架流程^[1]

(3) 融合的目的是提高数据质量,包括改善几何精度和丰富属性信息。

(4) 融合的结果是产生新的数据,新数据部分或者全部集成了两种源数据的优点,如高的点位精度、好的现势性、丰富的属性信息等。

数据融合技术的本质特点一方面体现在输入信息的多源性,另一方面体现在多种信息处理技术的综合应用。目前尚未有系统成型的数据融合理论,其中,地图数据库合并(Conflation of Geographic Database)或者地图合并(Map Conflation)技术在解决空间数据的垂直集成以及基于特征的空间信息融合问题方面显示了它独特的优势,故被认为是空间数据采集和集成技术中的核心技术之一^[26~28]。地图合并是在同名实体匹配的基础上,调整相关地图实体的几何位置,实现同一地区不同来源地图数据库的集成和信息融合。从定义中看出,地图合并主要有两个步骤:实体的匹配(Entity Matching)和实体的调整变换(Entity Adjusting)^[29,30]。同名实体的识别或匹配就是通过分析空间实体的差异和相似性识别出不同来源图中表达现实世界同一地物或地物集的过程。相关实体的调整变换包含两方面,一方面是图形上的(几何位置的配准和协调),另一方面是语义上的(属性数据的转换或合并)。

各种测绘技术的实用化和获取空间信息途径的多样化,导致地图生产的资料越来越多,如纸质地图、数字地图、卫星遥感影像、GPS 测量数据等,研究不同数据源、不同数据精度和不同数据模型的融合问题有着非常重要的现实意义^[16,31~35]。

(1) 在地图制图过程中合理有效的综合应用各种现势性资料(数据),可以起到保证地图质量、提高地图生产效率的作用。

(2) 通过数据融合可以重新组合专题数据,补充物体的分类分级和属性,进一步改善物体的几何精度。

(3) 可以减少空间数据的多语义性、多时空性、多尺度性、存储格式的不同以及数据模型与存储结构的差异,消除其不一致性,使新生成的地图数据在逻辑一致性及完整性方面得到提高。

(4) 通过融合,可以在最大程度上实现不同空间数据的完全转换,新生成的空间数据在某些方面(如点位精度、详细程度和现势性等方面)都得到极大提高,这对进一步扩大空间数据的应用范围、提高数据重复使用率、降低数据获取费用有着非常重要的意义。

(5) 可以促进空间数据的共享和集成。目前空间数据集成主要停留在数据格式统一或一致化的层次上,要实现集成仅依靠数据标准或数据规范是远远不够的,还需要依赖数据内容甚至数据知识等更高层次的集成,空间数据融合技术的研究能为其提供强有力的技术手段。

(6) 可以促进不同的空间数据部门的合作,避免“数据孤岛”的形成。

(7) 可以节约大量的人力、财力,从而促进国民经济的发展。

目前,数据融合技术已经被广泛应用到生产生活的各个领域,多源地理空间数据的融

合已经成为迫切需要解决的问题。

1.3 多源空间矢量数据融合发展

1.3.1 矢量数据融合研究

矢量数据融合的相关理论和实践起源于20世纪80年代中期美国人口调查局的一个实验项目。1980年为了完成十年一次的人口普查,美国人口调查局数字化了全国278个大中城市的地图,共5500多幅图。这些地图数据包含有丰富的属性,如名称、街区地址范围、街区类型等。1983年,美国人口调查局获得美国地质测量局(USGS)的电子地图数据,但没有属性数据。人口调查局将已有地图数据与新的USGS的高精度的DLG数据匹配或合并以便能检测并消除两种图的不一致性,将属性信息转到USGS地图数据库中。这就是世界上最早的地图融合研究^[28]。

之后许多学者进行了相应的研究。1987年,Lupien和Moreland指出不同来源地图数据融合主要有两个过程:识别同名实体和通过橡皮筋对准匹配实体,这些步骤是反复迭代进行的^[36]。1996年,Lemarie和Raynal提出应该建立一个通用的地理数据融合工具箱^[37]。1998年,Cobb讨论了美国矢量数据标准格式文件VPF之间的融合技术^[24],由于VPF数据带有丰富的属性信息(地物编码、实体几何类型等),提出了基于知识的属性数据匹配策略,取得了较好的效果。1999年,Walter和Fritsch研究了欧洲地理数据文件GDF和德国地形制图地图数据库ATKIS的融合技术,侧重讨论了两类数据库中道路实体的匹配问题^[28]。2004年,韩国学者Hoseok Kang阐明了地图融合的概念框架和根据几何关系及拓扑关系一致性的两个相关地图融合模型^[8]。2005年,Jan-Henrik Haunert研究了一种基于橡皮筋变换的拓扑一致性转换实体要素的方法来融合地图,实现了对地图数据库的更新^[38]。

英国、日本和美国等沿海发达国家早已完成了不同比例尺系列的地形图、海图、影像等数字产品的出版任务,但由于战略思想与装备技术的变革,不同数字产品的一致性问题、不同数据源的融合问题引起了决策机构的高度重视^[19,39]。英国海道测量局按照先由栅格数据、再到矢量数据的技术思路,现已形成了由数据源的评价、海道测量数据库、图像域属性编辑、栅格海图数据库、栅矢转换技术、电子海图数据库、纸质海图数据库、航海通告的自动改正、海洋测绘产品出版等技术工序组成的海洋测绘产品的生产体系。日本的海洋测绘数字产品是完全按照IHO的S-57数据格式出版的。日本水路部早已于1994年正式出版了符合IHO标准的电子海图(ENC),并于1998年实现了数字海图的自动改正,同时正式出版发行了与ENC配套的潮汐表、航标表等数字产品。美国的NIMA自始至终坚持数字战场环境地理信息的综合性与一致性,无论是数据库的形式、测绘产品的形态、还是技术体系的组成,完全是按照海、陆、空、天一体化的战场综合保障方式实施的。