

一种面向对象的时空数据模型

曹志月, 刘 岳
(中国科学院 地理科学与资源研究所, 北京 100101)

An Object-oriented Spatio-temporal Data Model

CAO Zhi-yue, LIU Yue
(Institute of Geographic Sciences and Natural Resources, CAS, Beijing 100101, China)

Abstract: Geographical information has three basic components: space, attribute and time. Future Geographic Information System (GIS) should support the temporal dimension, model and manage all three components efficiently. It is difficult for current GIS software to deal with temporal phenomenon, so spatio-temporal data model has become a hotspot in the GIS area. This paper proposes an object-oriented spatio-temporal data model and describes its data structure, after analyzing the shortcomings existing in the traditional data models.

Key words GIS; object-oriented; spatio-temporal data model

摘 要: 地理信息系统中, 时间、空间和属性是信息的 3 种基本成分, 理想的地理信息系统应支持信息的时态性, 对时空数据进行统一的模拟和管理。由于当前的地理信息系统软件难以处理时态现象, 时空数据模型已成为 GIS 领域的一个研究热点。首先对当前时空数据模型研究现状进行了分析, 在此基础上, 提出了一种面向对象时空数据模型, 并从 3 个方面描述了该模型的数据结构, 包括时空对象结构、时空数据组织方式和时空数据存贮方式。
关键词: 地理信息系统; 面向对象; 时空数据模型

1 时空数据模型研究现状

时空数据模型是一种有效组织和管理时态地理数据、属性、空间和时间语义更完整的地理数据模型。地理信息系统作为一种采集、存储、管理、分析并显示与地球表面及其附近空间数据的计算机系统, 随着应用领域的不断扩大, 对数据的处理提出了更高的要求, 要能够保存并有效地管理历史变化数据, 以方便将来重建历史状态、跟踪变化、预测未来, 这样的 GIS 被称为时态地理信息系统(Temporal GIS, 简称 TGIS)。时态 GIS 的组织核

心是时空数据库, 时空数据模型则是时空数据库的基础。在 TGIS 的研究中, 最基本最迫切的就是时空数据模型的研究, 时空数据模型的优劣, 不仅决定了 TGIS 系统操作的灵活性及功效, 而且影响和制约着 TGIS 其他方面的研究和发展。为此各国学者纷纷进行了大量的努力, 提出了各种时空数据的表示处理方法, 下面对当前各种方法模型进行了分析, 是本文模型提出的基础, 也为对该领域的研究感兴趣的学者们提供一个参考。

1.1 时间作为附加的属性数据

这种方法以关系数据模型为实现基础。由于传统关系模型丰富的语义、较完善的理论和许多高效灵活的实现机制,使人们最初开始尝试在传统空间关系模型中加入时间维,扩充关系模型、关系代数及查询语言,模拟处理时态数据,增强存取管理功能,实现时空数据存取的高效索引技术,从而直接或间接地基于关系模型支持时空数据的存取、表示和处理。

在关系型方法中,根据其扩充时态性所基于的不同级别,可分为关系级方法、元组级方法和属性级方法,它们分别是在关系级、元组级和属性级上处理时间。关系级方法是通过一系列沿时间维的关系快照,模拟整个关系中各目标的历史,其特点是:概念简单、易于理解,但数据高度冗余,单个目标对象的历史状况表达模糊。元组级方法是时间标记在元组级,一旦对象的性质发生任何变化,一个新的元组就被加进关系表中,而所有没发生的性质被重复,这将使存储空间迅速增长,影响查询的反应时间。此外,由于对象的时间与空间关系彼此分离,给空间数据和属性数据的联系带来困难,致使要求知道空间、时间及其他属性的应用时,需建立更多的附加关系。这又引起对内存需求的增加,并且在人机对话环境中潜伏着许多操作上的问题。属性级方法认为时间应当是属性的而不是元组的一部分,意味着模型中的关系不再是以前的正规形式,而是属性随时间变化的,相应提出的 NINF 形式虽能较好地体现时间的结构和特性,减少数据冗余,但不再能利用现成 RDBMS,技术上还有一定难度。

1.2 时间作为新的一维

由于高维的拓扑矢量 GIS 允许时空联结操作和拓扑查询,使得这种模型克服了地理关系模型的限制,减少了数据冗余。但这种方法也存在缺点:第 1,要获得高维对象构建的有效算法还有根本的困难;第 2,空间数据和非空间数据的变化还必须分别处理;第 3,时间和空间在性质上和潜在查询的参考基础上有重大不同。这有可能造成在 GIS 应用中高维表达的语义不正确;第 4,不可能利用现有的 GIS 和 DBMS 支持高维对象。

1.3 面向对象方法

Michael F. Worboys^[1], Donna J. Peuquet^[2] 等首先关注面向对象 (Object-oriented) 技术 (简称 OO 技术) 在时空数据建模中的运用。尽管面向对

象技术在建模概念、理论基础和实现技术上还没有达成共识,不够成熟,但它以更自然的方式对复杂的时空信息模型化,是支持时空复杂对象建模的最有效手段,它的最基本优点是打破关系模型范式的限制,直接支持对象的嵌套和变长记录。近年来,面向对象技术已越来越引起时空数据研究者的兴趣。从 OO 技术逐渐引入地理信息模淹的历程来看,面向对象地理信息模型的开发可分为 3 个阶段:①OO 方法模拟纯空间信息^[3~5],②OO 方法模拟时空信息^[6],③OO 方法模拟时空及属性信息^[7]。本文中时空模型的研究是基于第 3 个阶段的。

2 面向对象时空数据模型

基于对多种时空数据模型的总结分析,本文提出了一个基于 OO 设计思想的时空数据模型。其基本框架如图 1 所示。

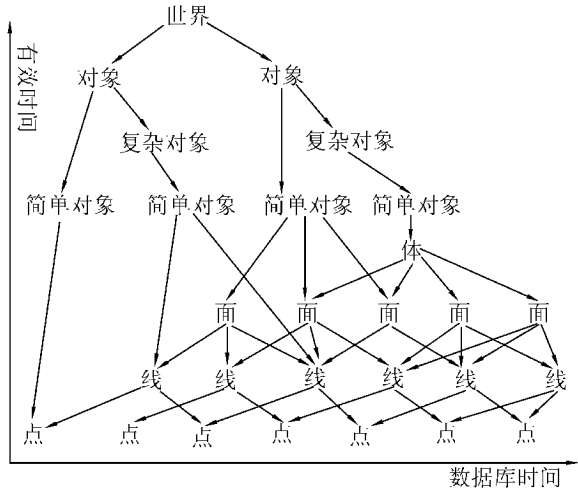


图 1 面向对象时空数据模型的基本框架

Fig. 1 Fundamental frame for object-oriented spatio-temporal data model

该模型的核心,是以面向对象的基本思想组织地理时空。其中对象是独立封装的具有惟一标识的概念实体。每个地理时空对象中封装了对象的时态性、空间特性、属性特性和相关的行为操作及与其他对象的关系。时间、空间及属性在每个时空对象中具有同等重要的地位,不同的应用中可根据具体重点关心的方面,分别采用基于时间 (基于事件)、基于对象 (基于矢量) 或基于位置 (基于栅格) 的系统构建方式。本文重点讨论基于时间的数据组织方式和数据存贮方式。

2.1 时空对象结构

模型中时空对象的一般结构为： $\langle \text{OBJ} : \{ \text{O-ID}, \text{Attr}(t), \text{Spatial}(t), \text{Temporal}(T_v, T_d), \text{Actions} \} \rangle$ 。对象结构中，对地理信息的 3 个基本成分：时间、空间和属性分别进行了描述。

其中 O-ID 是指对象的标识，可惟一地确定一个对象； $\text{Attr}(t)$ 是对象中随时间变化的非空间属性的描述； $\text{Spatial}(t)$ 是对象中随时间变化的空间特性的描述； $\text{Temporal}(T_v, T_d)$ 是对象的时态性的描述，反映对象的产生、状态的改变及消亡的生命历程，这里认为有效时间 T_v 和数据库时间 T_d 是正交的。

对象中另一个重要的部分 Actions 是描述对象的行为操作部分。其中定义对象的时间、空间及属性的各种运算操作，实现同类对象或不同类对象间的互相联系，使对象的数据和操作部分紧密联系起来，通过 OO 的多态性和继承性，以自然简洁的方式实现运算符重载，各种拓扑关系集合关系的重载。特别对时态关系及操作符的实现，对空间、非空间属性随时间变化的描述，都具有很好的灵活性、可扩充性和可维护性。

2.2 时空数据组织方式

在上述时空对象的结构中，对象的时间、空间和属性是对象中重要的 3 部分信息，不同的时空过程或不同的应用目的，可分别以基于属性、基于位置、或基于时间的角度来理解对象中的信息，相应的实现方式有基于向量、基于栅格和基于时间的方式。这是由信息变化的主导性决定的。下面来具体讨论基于时间的时空数据组织方式。由于模型中采用了 T_v, T_d 的双时序结构，我们分别从这 2 维时序角度来组织整个系统中信息变化的过去、现在和将来。

2.2.1 对象的 T_v 组织结构

对任意对象，标识为 O-ID-X，可以属于点类、线类、面类、体类、简单对象类或复杂对象类。如图 2 建立对象 O-ID-X 的有效时间历史的动态双向链表，其中对象 O-ID-XTv_i 是对象 O-ID-Z 的一个实例。

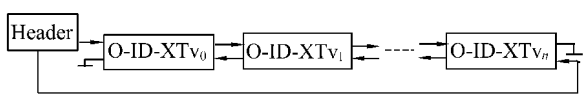


图 2 对象 O-ID-X 的有效时间历史的动态双向链表
Fig. 2 Dynamic double linked list for valid time history of the object O-ID-X

系统中每个对象对映惟一个有效时间链表，描述该对象在现实世界中的实际变化历史。链表按对象 O-ID-X 的 T_v 顺序排列，采用基于基态的动态多级索引方式存贮。 T_v 的采样是基于变化的，意味着：①对象 T_v 链表中任意相邻 2 项的空间、非空间属性不能完全相同，至少有一项不同；②对象 T_v 链表中相邻 2 项的 T_v 之间，即在有效时间段 $V_i: (T_{v_i} \sim T_{v_{i+1}})$ 内，对象属性没发生变化，或发生的变化在相应的系统中可忽略不考虑，还可能是发生的变化是可以由特定的规律确定的。

2.2.2 对象的 T_d 组织结构

对任意对象，标识为 O-ID-X，是点类、线类、面类、体类、简单对象类或复杂对象类的一个实例，如图 3 建立对象 O-ID-X 的数据库时间历史的动态双向链表，其中对象 O-ID-XTd_i 是对象 O-ID-X 在数据库时间段的一个状态实例。

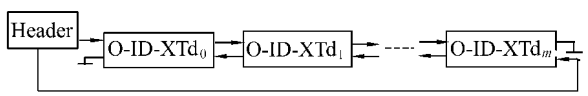


图 3 对象 O-ID-X 的数据库时间历史的动态双向链表
Fig. 3 Dynamic double linked list for database time history of the object O-ID-X

系统中每个对象对应惟一的一个数据库时间链表，描述该对象在数据库中被表示、更新、处理的历史过程，链表按对象的 T_d 顺序排列， T_d 的序列是由系统中操作处理的时间确定的，这意味着：①对象的 T_d 链表中，任意相邻的 2 项的空间、非空间属性不能完全相同，即至少有 1 项不同；②链表中相邻 2 项的 T_d 之间，即数据库时间段： $(T_{d_i} \sim T_{d_{i+1}})$ 内，对象属性没发生变化，也就是说在 T_d 时刻，库中记录的对象特征，直到在下一个数据库时刻 $T_{d_{i+1}}$ ，系统对该对象特性进行更新，在这期间 $(T_{d_i} \sim T_{d_{i+1}})$ 库中记录的对对象的特性保持不变。

2.2.3 系统的 T_v 组织结构

首先我们来看，在点类 CLASS POINT 中，假设有对象 P_1, P_2, \dots, P_n 。各对象分别有自己的 T_v 链表： P_1 的 T_v 链表： $\text{ID-}P_1\text{-}T_{v1} \leftrightarrow \text{ID-}P_1\text{-}T_{v2} \leftrightarrow \dots \leftrightarrow \text{ID-}P_1\text{-}T_{vk}$ ； P_2 的 T_v 链表： $\text{ID-}P_2\text{-}T_{v1} \leftrightarrow \text{ID-}P_2\text{-}T_{v2} \leftrightarrow \dots \leftrightarrow \text{ID-}P_2\text{-}T_{v1}$ ； \dots ； P_n 的 T_v 链表： $\text{ID-}P_n\text{-}T_{v1} \leftrightarrow \text{ID-}P_n\text{-}T_{v2} \leftrightarrow \dots \leftrightarrow \text{ID-}P_n\text{-}T_{vm}$ 。

若我们需要检索 T_{vi} 时刻的所有点类对象，

或满足某一条件的部分点类对象时,就需要对每个点类对象的 T_v 链表进行检索,当这种需求较多时,就会带来严重的时间开销,为此,可建立如图 4 的点类对象的 T_v 组织结构。

CLASS POINT;

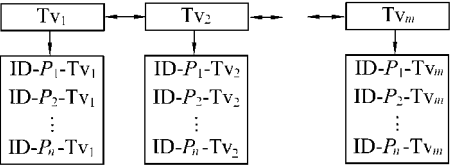


图 4
Fig. 4

链表是基于 T_v 时序的动态双向链表。由于在链表中只是存贮了指向对象某一时刻的实例的指针,而不是拷贝存贮整个对象实例的内容,所以采用这一组织方式,既可提高系统的时间效率,也节省了存储空间开销。

同理我们可以建立线类对象、面类对象、体类对象、简单对象类及复杂对象类的 T_v 组织结构。这时,从所建的各类的 T_v 组织结构中,我们可以很方便分别检索到某 T_v 时刻的某一类中的所有或部分对象。但当需要对整个系统检索某 T_{vi} 时刻的所有类或部分满足条件的类中的对象时,就需要去遍历点类、线类、面类、体类及简单对象、复杂对象各类的 T_v 组织结构,这样系统的执行速度会随着这种类型的检索需求的增多、对象类的增多以及各类中对象历史的增长而严重降低。为此,本文中进一步探讨了如何在上述建立的各类对象的 T_v 组织结构上,建立系统的 T_v 组织结构和系统的 T_d 组织结构。

系统的 T_v 组织结构如图 5 所示。

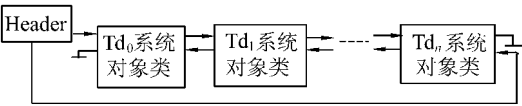


图 5
Fig. 5

其中 T_{vi} 系统对象类结构如下:

CLASS POINT; ID- P_1 - T_{vi} ID- P_2 - T_{vi} ...ID- P_n - T_{vi} ;

CLASS LINE; ID- L_1 - T_{vi} ID- L_2 - T_{vi} ...ID- L_n - T_{vi} ;

CLASS ZONE; ID- Z_1 - T_{vi} ID- Z_2 - T_{vi} ...ID- Z_n - T_{vi} ;

CLASS BODY; ID- B_1 - T_{vi} ID- B_2 - T_{vi} ...ID- B_n - T_{vi} ;

CLASS SMPOBJ; ID- S_1 - T_{vi} ID- S_2 - T_{vi} ...ID- S_n - T_{vi} ;

CLASS CMPOBJ; ID- C_1 - T_{vi} ID- C_2 - T_{vi} ...ID- C_n - T_{vi} 。

在如上建立的系统的 T_v 组织结构中,是以各类对象的 T_v 组织结构为基础的,因此不仅保持了各类对象的 T_v 组织结构的特点及功能,便于实现各类对象某 T_v 时刻的检索,而且方便了对整个系统在某 T_v 时刻(或时区段)的检索。同样只存贮指向对象实例指针,而不是实例内容的拷贝存贮,所以是一种降低存贮开销,提高系统执行效率的机制。

2.2.4 系统的 T_d 组织结构

系统的 T_d 组织结构如图 6 所示。

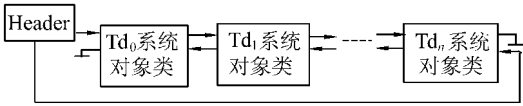


图 6
Fig. 6

其中 T_{di} 系统对象类结构如下:

CLASS POINT; ID- P_1 - T_{di} ID- P_2 - T_{di} ...ID- P_n - T_{di}

CLASS LINE; ID- L_1 - T_{di} ID- L_2 - T_{di} ...ID- L_n - T_{di}

CLASS ZONE; ID- Z_1 - T_{di} ID- Z_2 - T_{di} ...ID- Z_n - T_{di}

CLASS BODY; ID- B_1 - T_{di} ID- B_2 - T_{di} ...ID- B_n - T_{di}

CLASS SIMPLE; ID- S_1 - T_{di} ID- S_2 - T_{di} ...ID- S_n - T_{di}

CLASS COMPLEX; ID- C_1 - T_{di} ID- C_2 - T_{di} ...ID- C_n - T_{di}

2.2.5 建 议

上述分别讨论并建立了对象的 T_v 组织结构,对象的 T_d 组织结构和系统的 T_v 组织结构,系统的 T_d 组织结构。这 4 种结构都是基于对象的 T_v 历史链表和 T_d 历史链表,以不同方式为实现不同

的检索目的而建立的。其中, 对象的 T_v 组织结构, 描述了某一具体地理对象在物理世界中真实的从产生、发展变化到消亡, 以及再生等整个生命期的历史过程, 所以对象的 T_v 组织结构应该是这 4 种组织结构的核心, 而且与其他 3 种组织结构不同的是, 它存储了对象实例的内容而不是指针。

对后 3 种组织结构, 作者建议根据具体的时空过程或应用目的, 采用动态、静态相结合的建立方法, 即某一类组织结构是便于应用中某些方面的检索、操作分析等功能的, 若具体应用中, 对这些方面的要求较频繁, 则采用静态建立的方式; 反之, 若在具体应用中, 对这些方面的要求较少, 则可采用动态建立(使用时再建立)的方式。

2.3 时空数据存贮方式

在时空数据结构的研究中, 人们已经探讨了时空数据的多种存贮方式, 但都不是很完善。本文提出了基于“基态修正方法”(base state with amendments)的动态多级索引方式的基态修正存贮法。现已提出的几种对基态的修正方式如图 7 所示。

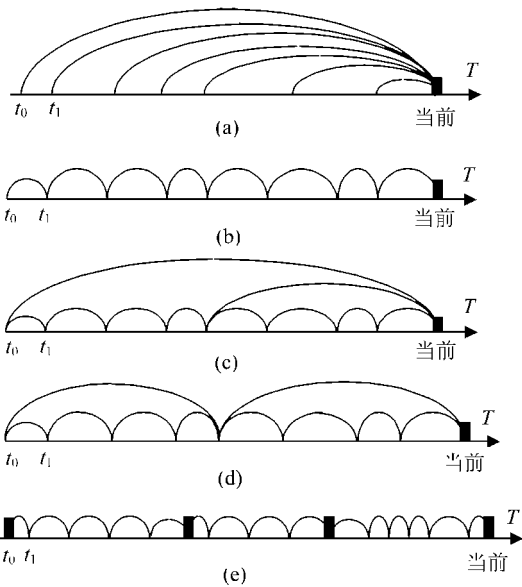


图 7 对基态的修正方式

Fig. 7 Base state amending methods

方式 a 把每个历史状态都作为相对于现在状态的修正。这种方法当库中记录了较久远的历史状况时, 则每次更新基态时, 都要修改所有的差文件, 这是不现实的。

方式 b 避免了方式 a 的缺陷, Langran^[8] 由此选择了 b, 然而对很远的过去状态进行检索时, 方

式 b 要对几乎整个历史状况进行阅读操作, 同样也是不理想的。

方式 c、方式 d 是张祖勋等人提出的 2 种改进方法^[9], 对方式 a、方式 b 的不足给予了一定的弥补, 但当历史很久远时, 同样存在着方式 a、方式 b 中的问题。

本文提出的动态多级索引方法如方式 e。在整个历史状况中动态地设立多个基态, 基态间的差文件数, 称为基态距。可根据具体应用由用户或系统确定基态距阈值, 系统中, 随着对象历史状况的插入、修正, 以及对历史时期的检索频繁程度而动态地创立基态。即当新的变化状况不断加入, 基态间差文件距大于某阈值时, 就要自动创建新的基态; 当用户对某时间段的检索较频繁时, 特别对于连续变化的时空过程, 差文件的恢复中大量的连续变化属性的推导过程会占去较大的时间开销, 通过动态创建差文件和基态的方法, 缓解了这一问题。

与方式 a ~ 方式 d 相比较, 动态多级索引方法在数据存取效率上得到了提高, 而不受历史久远情况的影响。然而其中一个重要的且需要灵活掌握的关键是基态距阈值的确定, 因为这一方法中, 增加了基态的占用空间, 降低了检索的时间开销, 其中基态距阈值的确定, 是对这 2 方面的权衡, 阈值太小会浪费空间, 阈值太大又会占用过多时间, 所以, 基态距阈值的确定要根据具体情况合适选择。

3 结 论

对时间和空间信息处理的实际需要促使了 GIS 中时空数据模型的研究。本文提出了一种面向对象的时空数据模型, 对它的数据结构进行了描述, 其特性可归纳如下:

- 1. 在时空对象中, 对地理信息的时间、空间和属性 3 部分进行了统一的表示。
- 2. 数据结构中支持了有效时间和数据库时间的双时序。
- 3. 对整个系统和具体对象的历史都采用了动态多级索引方式的基态修正存贮法, 存贮效率较高, 数据冗余少。
- 4. 基于双时序的数据组织方式, 使对象及系统的状态变化历史表示直观, 检索方便。
- 5. 对象的空间表示能力和系统其他功能的可扩充性强, 运用范围广, 适用于基于属性(矢

量)、基于空间(栅格)和基于时间(变化事件) 的应用领域,也适用于连续变化和离散变化的时空过程。

为了验证研究内容,在 WIN 98 平台, VC++ 6.0 开发环境下,设计了武汉市市政管理时空实验系统,模拟了武汉市市政管理中不断变化的时空信息,具体实现了面向对象的时空数据模型中提出的时空数据组织结构、存贮结构的设计方法及其主要的时空查询检索功能的实现过程。关于模型的更新和查询功能,拟另文介绍。

参考文献:

[1] WORBOYS M F. Object-oriented Approaches to Geo-referenced Information [J] . Int J Geographical Information Systems, 1994, 8(4): 225-245.

[2] PEUQUET D J, NIU D. An Event-based Spatio-temporal Data Model (ESTDM) for Temporal Analysis of Geographical Information Systems [J] . Int J Geographical Information Systems, 1995, 9(1): 7-24.

[3] WORBOYS M F. Object-oriented Models of Spatio-temporal Information [A] . Proceedings of GIS/ LIS' 92 [C] . At-

lanta GA: ACSM, 1992. 825-834.

[4] EGENHOFER M J, FRANK A. Object-oriented Modeling in GIS, Inheritance and Propagation [A] . Proceedings of Auto-Carto 9[C] . Baltimore: ACSM, 1989. 588-598.

[5] WORBOYS M F, HEAMSHOW H M, MAGUIRE D J. Object-oriented Data Modeling for Spatial Databases [J] . International Journal of Geographical Information System, 1990, 4(4): 369-383.

[6] WORBOYS M F. A Model for Spatio-temporal Information [A] . Proceedings of 5th Int. Symposium on Spatial Data Handling [C] . Charleston: IGU Commission on GIS, 1992. 602-611.

[7] SHI Wen-zhong. Objected-oriented Approach for Spatial, Temporal and Attribute Data Modeling [A] . Proceedings of GIS/ LIS' 95 [C] . Nashville: Nashville Convention Center, 1995.

[8] LANGRAN G. Temporal GIS Design Trade-offs [A] . Proceedings of GIS/ LIS' 88 [C] . San Antonio: ACSM, 1988. 890-899.

[9] ZHANG Zi-xun, HUANG Ming-zhi. Temporal GIS Data Structure Discussion [J] . Bulletin of Surveying and Mapping, 1996, (1): 19-22. (in Chinese)

第六届全国矿山测量学术讨论会征文通知

为了交流矿山测量科技成果,促进我国矿山测量科学技术的发展,更好地为经济建设和社会发展服务,并为 2003 年国际矿山测量学术大会在中国召开做准备,中国测绘学会、中国煤炭学会、中国金属学会决定于 2002 年下半年联合召开综合性的全国矿山测量学术讨论会。

征集论文内容包括: 1. “数字矿山”及其关键技术; 2. 矿山测量数据采集、计算机管理、机助制图; 3. 开采沉陷与防护、“三下”采矿; 4. GPS、GIS、RS 及其在矿山的应用; 5. 矿区地籍测量、土地信息系统与土地资源管理; 6. 矿区生态环境监测与治理、土地复垦及生态环境重建; 7. 矿产资源信息获取、评价、开发管理、保护、经济分析与矿区可持续发展; 8. 矿山测量的任务、发展方向、组织管理、教育与人才培养; 9. 其他测绘新技术新理论在矿山的应用。

提交的论文首先由本单位审查,论文要求立论正确、有创新、简明扼要、重点突出,图表清晰。全文一般不超过 6 000 字,提要不超过 800 字,论文格式以 2001 年《测绘通报》上论文样式为准,但正文可以不用分栏方式排版。论文请以打印稿一份和 word 格式软盘同时邮寄,也可以用 E-mail 方式传送,但必须同时邮寄打印稿一份,论文无论录用与否,无特殊说明,一般不退稿,请自留底稿。

已在全国性刊物上发表的论文不予征集, 论文截止日期: 2002 年 6 月 20 日。

论文可寄至如下任意一个地址:

- 1. 江苏省徐州市中国矿业大学环测学院 张书毕、汪云甲收, 邮编: 221008
E-mail: zhangshi@sina.com Wj4139@cumt.edu.cn
 - 2. 河北省唐山市新华西道 21 号煤科总院唐山分院矿测所 滕永海收, 邮编: 063012
 - 3. 辽宁省沈阳市和平区文化路 3 号巷 11 号东北大学 马洪滨、高福聚收, 邮编: 110006
- 论文被接受后, 组织者将通知作者, 其他事项另行通知。

中国测绘学会矿山测量专业委员会
中国煤炭学会矿山测量专业委员会
中国金属学会矿山测量专业委员会