

从数据处理到知识处理

——论数学地质的第二次浪潮

丁泰木

(地矿部石油地质中心实验室)

本文从信息的观点出发,紧密结合地质科学的自身特点与数学地质发展的现状,通过国内外的研究动向,论述了将人工智能引入到数学地质的理由,强调开拓地质学家智能活动量化的新领域,把计算机的应用从当前的数据处理水平提高到科学知识处理水平,并提出了当前数学地质发展的第二次浪潮的新见解。文中还介绍了专家系统的面貌和PROLOG语言的基本特点。

国际数学地质协会前主席E·H·T·Whitten在展望未来数学地质的任务时曾经这样说过:为了使数学地质在地球科学中成为一门真正具有决定性意义的重要学科,而不是像现在很多情况下仅起锦上添花的附属作用,就必须把主要的努力放在对地质过程用数学方法来进行思维和认识,建立有效的、可靠的数学模型,对地质体进行定量化,并付诸实践。

自牛顿以来,一门学科如不能充分地运用数学,便不成其为真正科学的信念已经深深地渗入了人们的意识。地球科学也不例外,1840年地质学家Lyell首先把数学的方法运用到地质领域,通过古生物化石的统计分析对第三系地层进行划分。但是自那以后,地质科学量化的工作并没有取得多大的进展,地质学家依然主要依靠记录和定性描述的方法收集实际地质资料,通过分析归纳认识规律并建立相应的理论。因此,长期以来人们普遍认为地质科学是一门主要依靠归纳推理的科学。从本世纪四十年代开始,由于物理的、化学的方法和手段在地球科学领域中得到了广泛的应用,其技术也有了急剧的发展,由此产生了大量的数据,因而对数学的要求已经到了必不可少的地步。电子计算机——数学地质。

的应用,起了催化剂的作用,自此之后才真正地形成了现代地球科学定量化的研究领域

诚然,在地球科学领域里应用计算机开展定量化工作以来,在技术上已经获得了很大的进步,并以此取得了社会经济效益。然而,也不能忽略这样的现实:时至今日,由于以数值数据为信息处理基本单元的传统软件的局限性(指对地质科学研究对象的特性及地质事件与地质过程的描述等),由这种方式构成的软件模式已经满足不了地质科学定量化范围与深度日益发展的要求,于是人们根据现在的技术水平(软件发展等)探讨一种更新的、更为灵活的、更能反映地质科学研究特点的软件模式是很必然的。

一、地球科学研究对象的特性与地质学家的认识模式

地球科学研究最重要的特点来自于研究对象本身——地球。无论是对一个地质事件（如沉积、成矿、岩浆侵入、造山运动等）发生机制的探讨，或是对一个地质过程（如成矿演化、古环境变迁、古生物演化等）发展的追踪，这些都是立足于对现今“遗迹”的观测或实验而对遥远过去发生的事件的推测。对这种宏伟的时空实体的研究决定了本学科研究独特的性质——在众多未知和不确定性的因素所构成的可能性空间中艰苦地搜索。法国数学家、动态系统的理论的创立者H·Poincare认为，从原则上说，虽然后来的情况完全可以由过去的情况所决定，但实际上即使原先几个本身丝毫不带随机因素的简单确定性系统，也会因为初始条件中微小的差异逐渐加强，由于系统综合的作用而导致最终出现根本不同的结果。因此，即使短时间内的行为可以预测，时间一长便不能准确预测，或者根本不能预测了。这种对于“混沌”现象研究的科学原理正揭示了地球科学研究对象的特性：成因机制和演化过程的不确定性和不可逆性、实体表现的多变性和复杂性、人们认识的主观性以及认识效果检验的困难性。

研究对象的特性导致了研究方法鲜明的特点。地球科学研究不同物理学、化学以及其他科学，由于完成一个地质事件或地质过程的时间非常漫长而又不可重复。因此，其观测仅限于对过去事件局部的记录，这就使得地球科学成为一门历史性的、描述性的科学。我们不妨以粗略的信息的观点对一个很一般的地质研究过程进行观察便可见一斑：地质学家通过学习和对大量的事实进行观察，从中获得了信息→通过不断地归纳-演绎，使各种信息得以过滤并在其间建立有机的联系→形成概念、经验、知识，并把它储存在头脑里→对新的情况产生联想，即新情况与存储在头脑中的信息模式（事实与概念）的比较→再通过推理与判断→形成新的知识。如此往返不断循环，运用知识又产生新知识，形成理论，表述于字里行间之中。由此可见，一个经过了地质学家深思熟虑而产生的地质描述，便是对地质体质与量的综合表达，其中蕴含着丰富的信息。因此，有人称地球科学“既是一门科学，又是一门艺术”不无道理。

二、科学定量化的基本条件及数学地质的困难

电子计算机以其高速准确的演算能力及严格确定的逻辑证明能力对当代所有的学科都产生了极大的诱惑力。一个研究领域能够借助计算机运用数学的形式进行研究的条件如何？从原则上说，对客体任何量、任何质以及质与量统一的关系，数学都可以提供应用的辅助方法。可是，这只有当这个研究领域的问题业已相当成熟，概念业已表述得十分清楚，其中各种量的关系比较确定时，数学才能应用于这个领域。这就是说，下列的条件是基本的：

1. 与人们对研究对象的认识水平有关，即研究领域是否已经达到了可以建立理论的水平。
2. 已经建立的理论是否具有与作为形式化条件的数学表达相适应的逻辑简明性——

概念模型。

3. 反映这些理论内在联系规律的语言及推理规律的形式化程度。

4. 选择合适的数学表达式。

如前所述，地球科学研究是在一个非常宏伟的时空体系中进行的，因此它不像物理体系那样具有精确清晰的逻辑规律和确定的因果关系。尽管如此，力图说明其形成机制和存在事实而建立在各种观点基础上的各种地质理论，无可否认地经过了千百年来人们不断地探索和实践，已经在相当宽的范围里揭示了一些大家公认的，实践业已证实的规律。这些理论与规律，便构成了地球科学定量化的基础。因此，计算机的出现，也大大地推动了地球科学定量化的发展。然而，众所周知，现有计算机是严格依赖人们为之编制的程序来运行的，而这种程序是以高精度的数学刻划为前提的。数学化要求的确定性与简明性与地质研究的不确定性和复杂性，时常使数学地质技术人员陷入困境。

目前，在地质研究中在定量测定基础上产生的数据，对于一个地质体的认识而言，是非常有限的。但是地质学家在其工作中却获得了更多难以用数值表达的信息，这些信息一方面来自于对研究对象观察所作的定性描述，另一方面则来自于对研究过程感知、经验、推理、判断的总结，这就是基本的地质科学知识。这些知识一般不具有通常意义下的数学表达形式，但它却包含着对地质体认识的大量质与量的信息，是地质研究最基本的手段和表达形式。如何开拓这方面的量化研究并运用计算机实现之，不仅对地质科学研究具有很重要的意义，而且也将是使数学地质在地球科学中成为重要学科关键的一步。

三、从数据处理到知识处理

纵观数学地质近二十多年来发展的大致情况，无论是最初开展的地质多元统计分析，或是随后发展起来的地质过程的数学模拟，也无论旨在矿物储量与品位计算的地质统计学，或是属于信息管理的地质数据储存、检索、自动处理和显示等等，无不都是以数值数据为基本单元的代表和操纵为特征。如前所述，人们实际上在一个地质体中能够搜集到的定量数据是非常有限的，而且通常的地质研究的方法是非数学性的，数据仅是孤立的信息数据，之间以及与现实的地质背景之间的关系只有当地质科学人员运用学科知识给予解释之后才有意义。因此，基于这种方法或模式的信息处理对于地球科学量化是不充分的，它存在着明显的局限性。

由国际斯坦福人工智能研究所的知识工程专家同美国地质调查所的地质专家共同研制的PROSPECTOR（1976年—1981年），则是完全不同的另外一类信息的处理系统，它以学科领域知识为信息处理的基本单元。

地质科学知识，是地质学家通过思维活动所得到的有意义的、具有一定结构的信息集合，它与那种简单的地质事实数据的集合不同，是信息之间以及信息与地质背景之间相关联，能作为智能行动（如理解、推理、判断、归纳、演绎、类比、评价等）基础的事实。以这种形式为基本单元的信息处理系统，除了运用通常的算法，更重要的是使用了推理机制。这不仅使信息处理的方式更类似于地质学家的工作活动过程，便于自然而

正确地描述地质现象和地质过程，演绎出正确的结论，而且一般具有能向用户复示这个结论的推理过程，列举推论路径上所使用的规则，帮助解释如何得到它们的结论或纠正其错误的功能。

PROSPECTOR是一个基于计算机的矿藏勘探咨询专家系统，它存储有一般地质学与矿物学的知识，同时也存储了某些特定地区如密西西比河流域和美国各大山脉特定的地质知识，可供地质学家在矿藏勘探和资源估价时作智能助手。1982年，美国一家勘探公司把它应用于华盛顿州的小瀑布勘探和铅矿开发，结果在一个原先为探矿专家们所未料到的地点发现了矿藏，据估计此处矿藏将为这家公司带来几百万—1亿美元的利益。

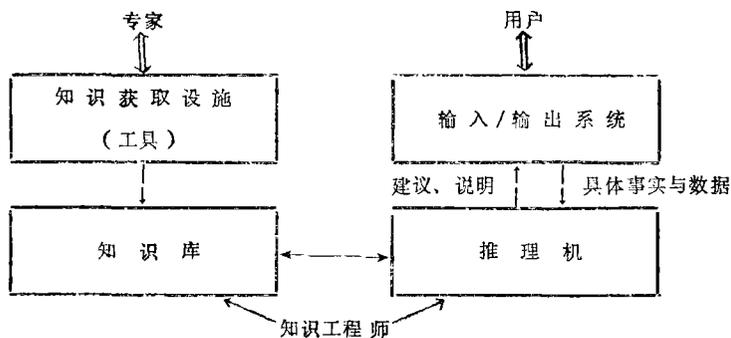
地质专家系统的出现和成功，不仅在国外吸引了有关专家的广泛兴趣，应用的范围迅速地铺开，同时也唤起了国内有关科技人员极大的热情。八十年代初以来，各种应用的地质专家系统逐渐研制出来，例如吉林大学与地质科学院矿床所合作的“勘探专家系统”，与航空物探总队合作的“航空物探专家系统”，与武汉地院合作的“金矿专家系统”，还有胜利油田与自动化所合作的“石油测井解释系统”，海洋石油勘探开发研究中心研制的“油气资源评价系统(PRES)”等等。

作为一种新型结构的软件系统，地质专家系统正在自身的研制与应用之中 迈开步伐，崭露头角。

四、专家系统的基本结构与PROLOG语言的特点

专家系统是六十年代中期发展起来的人工智能的新兴分支，而且是目前已经取得实际应用效益的一门技术。简单说，它是一个计算机程序系统，里面存储有使它能够专家水平上工作的知识和解决问题的能力。它把知识的描述与专家解决问题的控制策略有机地结合起来，模拟专家认识问题，解决问题的过程，通过推理与判断达到目标。人们借助于它，只需要告诉计算机“做什么”(What to do)而将“怎样做”(How to do)的任务交由计算机自行安排实现。

一个专家系统的基本结构可由下图表达，它包含知识库(Knowledge Base)，推理机(Inference Engine)和灵活的人机接口(Man-Machine Interface)三部分。



摘自E. A. Eigenbaum

知识库,即学科专家解决问题的领域知识的集合。它来自于两个方面:一方面是范畴的事实,即广泛共有的知识,如教科书上的知识和常识;另一方面则是经验性的知识,也称为试探式(heuristic)的知识,这是一种根据特定领域的经验进行试探和推理的知识,往往需经多年的工作实践总结而成。

一个好的知识库的设计是一个专家系统研制最为重要的一环。也就是说,要把地质学家毕生勘探的思考方法和主观判断从分散的、有时甚至连专家自己也不能自觉地认识到其实质的细节中,由不断变化的信息或数据点提炼为简单明了的法则,并转化为计算机内代表知识的具有一定结构形式的符号,这种从非形式化的领域向形式化的领域的转化是一项非常艰苦的劳动,犹如砂中淘金一般。

推理机则是用来解决问题,运用知识的推理策略的集合,旨在根据已经存储的知识,推导出解决问题的方法并得出结论。如同专家通常思考问题,解决问题一样,推理的形式既可有演绎型的,也有递归型的,或是确定性的,或是不确定性的。例如PROSPECTOR采用的就是主观Bayes法的不确定性推理算法。在PROSPECTOR中,领域知识用规则表示,形式是

$$\boxed{E} \frac{LS, LN}{\text{(规则名)}} \boxed{H} P(H)$$

即证据E以某种程度与假设H相关。其中LS为充分性度量($0 \leq LS \leq +\infty$),表示证据E出现时对假设H的支持程度;LN称为必要性度量($0 \leq LN \leq +\infty$),表示证据E不出现时对假设H的支持程度。 $P(H)$ 则为专家给出的一个先验概率,以表示假设H在对E一无所知时以概率 $P(H)$ 成立。PROSPECTOR的推理过程就是根据证据E的概率 $P(E)$,利用规则(LS, LN)把假设H的先验概率 $P(H)$ 更新为后验概率 $P(H/E)$ 的过程,它通过这种概率转化在推理网络中逐层传播最终达到选择假设的目的。

一个软件系统如不能与用户“友好相处”,最后结果将无人过问。因此,一个灵活而又令人自然适从的人机接口设计无疑是重要的。专家系统不但要善于解决问题,而且首先要善于提出问题或引导用户提出合适的问题。专家系统的目标在于按照专家的思维方式工作,因此如能采用这种方式的语言来描述是再适合不过的了。逻辑程序设计提供了这种基本方式,而PROLOG(Programming in Logic)就是逻辑程序设计语言之一。

PROLOG语言是以谓词逻辑关系为基础的一种逻辑型语言,它以一种近似于自然语言语义的形式化方式来描述。所谓逻辑关系 $R(X_1, \dots, X_n)$ 就是对于某一组值 (X_1, \dots, X_n) 给出一个真或假的逻辑判断的函数 $R: (X_1, \dots, X_n) \mapsto \text{真或假}$ 。PROLOG与过程式语言如BASIC, FORTRAN, PASCAL, COBOL等不同,无需把求解的算法一步一步地按顺序规定在程序中,而只需把一些必要的事实和规则,即认识解决问题的知识交给它,PROLOG就可以通过这些知识使用与之分开的控制策略,即推理机一边试探着进行判断,一边寻找答案。所以说PROLOG语言是陈述式的或启发式的。它由描述对象之间关系的事实和证明对象之间关系的规则所组成。

五、数学地质的第二次浪潮

以地质数据为信息处理基本单元,运用与此相适应的过程式程序设计语言描述,以揭示地球科学研究的某些规律和事实的定量化研究,在推动地球科学研究发展的同时,也为数学地质自身的发展作出过很大的贡献,并将继续发展作出贡献。如果说以此特征标志数学地质发展的一个阶段称之为第一次浪潮,那么,以地质知识为信息处理基本单元,运用更接近于人类思维方式的逻辑程序设计描述地质学家观察研究对象,解决地球科学问题为特征的新一代的软件系统的发展,已经开创了数学地质发展的另一个阶段,这便是数学地质的第二次浪潮。

如前所述,知识是具有一定结构的信息集合,是人们在社会实践中通过思维获得的东西。在现实世界中,一个地质学家通常是运用地质知识观察问题解决问题,而不仅限于在数据中解决问题(数据求解问题仍然必须运用学科知识)。地质知识包括了地质数据并比地质数据更接近于地质背景,它更完整地反映地质体质与量的统一关系。数学地质第二次浪潮的基本特点就在于把数值定量的精度与思维机制的活性有机地结合起来,在更广阔的领域里,更深刻的层次上,以更灵活的方式为刻划地质事实和地质过程,解决复杂的地质问题提供合适的手段。

这一发展阶段的另一个特点在于更进一步密切了数学地质与地质科学领域其他学科的关系。要从地质书上或资料中吸取知识,特别是从地质学家的头脑中逐个挖掘知识,仔细而又准确地了解专家是怎样运用他的知识,并把这些归纳起来最终转化成计算机能够非常准确地模仿其思想方式工作的计算机代码,这是一个相当费劲而又需要十分细致的工作过程,它是软件专家与地质专家紧密配合,双方知识相互渗透,反复琢磨、不断完善的结果。从事这种转化的数学地质专家,他不仅要熟悉通常软件人员开展的数据处理型的工作,而且还要完成这类工作与地质领域知识有机统一的系统,倘若他不具备“双栖”的知识和才干是不可能的。

地质应用软件的研制从以往数据的水平上深入到仿照地质学家思维,运用知识(也包括了大量的数据)求解地质问题,这种进步不仅反映了地球科学定量化发展的要求,同时也反映了计算机科学发展的影响。旨在模拟人类思维活动,以电子计算机为物理背景,应用形式化手段研制能模拟人类的智能行为进行问题求解的人工智能研究的兴起,不仅大大地推动了计算机科学自身的发展,由此而产生的硬件,特别是各种智能软件的成功,给包括地球科学在内的所有学科都带来了良好的机会。近几年来,无论是逻辑型程序设计语言(如PROLOG),或是函数型程序设计语言(如LISP),或是在此基础上发展起来的兼有两者之优的LOGLISP语言以及各种智能软件开发工具,都有很大的发展和广泛的应用,而且把应用的范围逐步地扩展到微型计算机系统。这种良好的软件支持环境,显然也为数学地质第二次浪潮的形成和发展提供了良好的物质条件和精神条件。

控制论的创立人N.Wiener说过,数学的伟大使命在于从混沌中发现秩序,混沌之中也存在有序。数学地质向智能化发展,这不是对地质科学传统的直感式工作方法的迁

就,而是在认识现实世界——地质体的过程中,以更高的层次从渊源千万年至上亿年的实体中发现规律。与此同时,这种发展也将促使地质科学在日益发展起来的信息社会中,逐步地改变传统的工作方式,使地质学家运用思维的活性与信息环境的要求有机地统一起来,推动地质科学的现代化。

总之,以地质知识作为信息处理基本单元为基本特征的数学地质第二次浪潮正在高涨,这是数学地质在当代科学技术发展的协同作用下深化的必然,随着它在实践应用中的不断发展,它在地球科学研究中将发挥越来越大的作用,最终成为一个具有决定性意义的重要学科。

参 考 文 献

- [1] 赵瑞清, 1986, 专家系统初步, 气象出版社。
- [2] 徐道一, 1985, 数学地质与地质信息的新进展, 全国数学地质在煤田地质中应用学术讨论会资料。
- [3] E.A. 费根鲍姆, 1984, 人工智能对世界的挑战, 新华社参考消息编辑部。
- [4] E. 丽奇, 1986, 人工智能引论, 广东科技出版社。
- [5] F.P. 阿格特伯格等, 1980, 地质数学, 科学出版社。

DATA PROCESSING TOWARD KNOWLEDGE PROCESSING

——A Discussion on 'the Second Wave' of Mathematical Geology

Ding Taimu

(Central Laboratory of Petroleum Geology,
Ministry of Geology and Mineral Resources)

Abstract

Can computers do the job as geologists do by a mode of thinking, judging and reasoning? This is an important goal which mathematical geologists are seeking for. As viewed with new informations combined with the characters of geological science and the recent developments in mathematical geology at home and abroad, the author of this paper describes the reasons why 'the artificial intelligence' is introduced into the researches of mathematical geology. It emphasizes to open up a frontier for the quantification of geologist intelligence and weed through recent data processings to bring forth the knowledge processings. It is proposed that we are coming into the time of 'Second Wave' of mathematical geology.

Based on the developments of geologist intelligence in mathematical geology, i.e. the combination of precision in value quantification with the activated thinkings, it will provide a flexible solution for the complicated researches of geology in wider and profound degree. Finally, the introduction is briefly made on the expert system and PROLOG language, so that it will help to keep abreast current developments in geology at home and abroad.