

人工智能原理在古生物化石归类中的应用

王益锋 张逸昆

(南京大学地球科学系)

一、引 言

当前定性的形态描述仍作为古生物研究的基本手段,因此,对古生物学来说,定性资料的自动化处理显得尤其迫切。实际上,在微机普及的今天,这完全可以借助于计算机的自动推理功能得以实现。自动推理方法免去了古生物研究中的一些机械的和重复性的劳动。本文着重讨论人工智能原理在古生物化石归类中的应用以及在微机上实现的途径。

二、基本原理

设所考虑的古生物化石具有 n 个特征要素,这里将它们看作 n 个变量

$$V_i, i = 1, 2, \dots, n,$$

V_i 既可以取定性值,也可以取定量值。这样,每一类型的化石与以变量 V_i 为坐标的 n 维空间中的某一点群相对应。也许我们会想到,这种对应关系可用表格的形式表示。表的列对应于变量 V_i ;表的行与化石类型相对应;第 i 行第 j 列的值就是第 i 类化石 V_i 的取值。然而,在实际应用中,直接用表格形式表示古生物分类系统是不适宜的。在通常的古生物形态描述中,对于每一类型的化石,并不就全部特征进行逐一描述,而是其中几个重要的特征。这样势必造成表中许多位置是多余的。因此,表格的表示方式很不经济。另外,还有一些描述句型难以转变成表格的形式。

上述这种对应关系的表示应该考虑到古生物分类系统的自身特点。事实上,用谓词公式表示古生物分类系统是很方便的。首先,古生

物的归属关系可表示成一棵树(如插图 1)。树

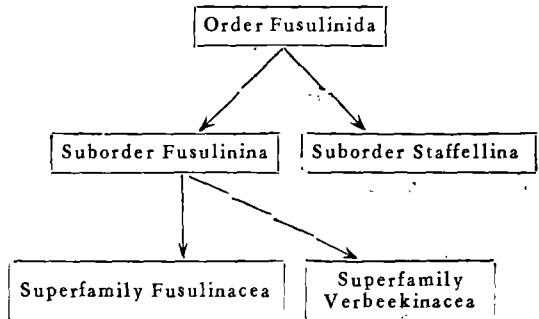


插图 1 一个归属关系树形图的实例

的结点代表某类型的化石;树枝代表归属关系。然后将树形图用谓词公式表示。例如对插图 1 来说,相应的谓词公式如下:

BELONG (Suborder Fusulinina, Order Fusulinida)

BELONG (Suborder Staffellina, Order Fusulinida)

BELONG (Superfamily Fusulinacea, Suborder Fusulinina)

BELONG (Superfamily Verbeekinacea, Suborder Fusulinina)

此处 BELONG (X_1, X_2)—— X_1 属于 X_2 。

化石的特征也可用谓词公式写出:如

TEST(x , large)—— X 化石的壳体是大的。

另外,化石的特征与分类的关系采用蕴含形式的谓词公式表示,如

COMPOSITION (x , Calcic) \Rightarrow BELONG (X , Suborder Fusulinida)

这里 COMPOSITION(X_1, X_2)—— X_1 化石的成分是 X_2 。

根据人工智能原理, 这些谓词公式分为两大类。第一类公式是以蕴含形式出现, 如上述的化石特征与归属关系的谓词公式, 称为规则; 第二类公式是以非蕴含形式出现的, 如待鉴定化石的特征描述, 称为事实。这样古生物化石自动鉴定问题归结为: 从某些特征事实和规则出发, 借助于谓词演算, 解决古生物化石的自动归属。

为了处理方便, 我们将规则化成或分解成以下形式

$$L_1 \wedge L_2 \wedge \cdots \wedge L_m \Rightarrow L$$

L_1, L_2, \cdots, L_m, L 均为文字 (literal)。古生物分类中经常遇到两类型古生物化石的对比问题。它可表示成如下形式

$$[\text{BELONG}(X, S_1) \vee \text{BELONG}(X, S_2)] \wedge P \Rightarrow \text{BELONG}(X, S_1) \wedge \sim \text{BELONG}(X, S_2)$$

这里 P 为描述特征的谓词公式, S_1, S_2 表示两个类型的古生物化石。经等价变换上式可简化为:

$$P \Rightarrow \sim \text{BELONG}(X, S_2)$$

一个古生物化石分类系统中, 规则集合是很庞大的。显然, 直接从中选取合适的规则是很费事的。因此, 有必要对规则集合作划分, 对规则的选取加以控制。正如前面所述, 古生物分类系统具典型的层次结构, 可用树形图表示 (如插图 1), 通常古生物化石鉴定顺序由高层次到低层次。如瓣类化石, 从目→亚目→超科→科→亚科→属。根据这种关系, 将规则集合 D 划分为若干个子集 d_1, d_2, \cdots, d_n , 使得其与树形图的节点 (除树叶外) 一一对应。对应于某一结点的子集是由这样的规则组成: 这些规则用来确定化石标本的下一层次的归属, 即确定化石属于那一个子结点。这种划分大大地提高了规则选取的效率, 同时也便于规则的获取, 因为在建立规则时只要分别考虑各层次各类型古生物化石的逻辑关系即可, 而不必同时考虑整个分类系统。以插图 1 为例, 对应于结点 Suborder Fusulinina 的规则子集为

$$r_1: \text{SPIR}(x_1, \text{fine developed}) \wedge \text{PRESENT}(x_1,$$

$$\text{tunnel}) \Rightarrow \text{BELONG}(x_1, \text{Superfamily Fusulinacea})$$

$$r_2: \text{PRESENT}(x_2, \text{parachomata}) \wedge \text{PRESENT}(x_2, \text{foramina}) \Rightarrow \text{BELONG}(x_2, \text{Superfamily Verbeekinacea})$$

$$r_3: \sim \text{SPIR}(x_3, \text{fine developed}) \Rightarrow \sim \text{BELONG}(x_3, \text{Superfamily Fusulinacea})$$

$$r_4: \sim \text{PRESENT}(x_4, \text{tunnel}) \Rightarrow \sim \text{BELONG}(x_4, \text{Superfamily Fusulinacea})$$

$$r_5: \sim \text{PRESENT}(x_5, \text{parachomata}) \Rightarrow \sim \text{BELONG}(x_5, \text{Superfamily Verbeekinacea})$$

$$r_6: \sim \text{PRESENT}(x_6, \text{foramina}) \Rightarrow \sim \text{BELONG}(x_6, \text{Superfamily Verbeekinacea})$$

这里 $\text{PRESENT}(x_1, x_2)$ ——In x_1 x_2 is (are) present; $\text{SPIR}(x_1, x_2)$ ——In x_1 spirotheca is x_2

下面我们讨论具体的推理过程。在鉴定化石时, 我们常不清楚哪些特征是最关键的。倘若把全部特征描述输入计算机, 那么显然太麻烦了。因此, 我们只能采用计算机向用户询问的方式进行。这样, 应用后向推理方式是最合适的。推理的过程分为两步:

1, AND/OR 树的生成。设经过推理, 待鉴定化石已归属于归属关系树形图中的某一结点 (即类型)。通过考察归属关系图就可知道该结点的子结点。这样就可以产生 AND/OR 树的根结点, 也即目标结点 (goal node)。还是以插图 1 为例, 设化石已归到结点 Suborder Fusulinina, 该结点有两个子结点, 相应地可以写出 4 个 AND/OR 树根结点:

$$(1) \text{BELONG}(\text{the Sample}, \text{Superfamily Fusulinacea})$$

$$(2) \sim \text{BELONG}(\text{the sample}, \text{Superfamily Fusulinacea})$$

$$(3) \text{BELONG}(\text{the sample}, \text{Superfamily Verbeekinacea})$$

$$(4) \sim \text{BELONG}(\text{the Sample}, \text{Superfamily Verbeekinacea})$$

对于一目标结点, 通过规则的匹配和合一就可

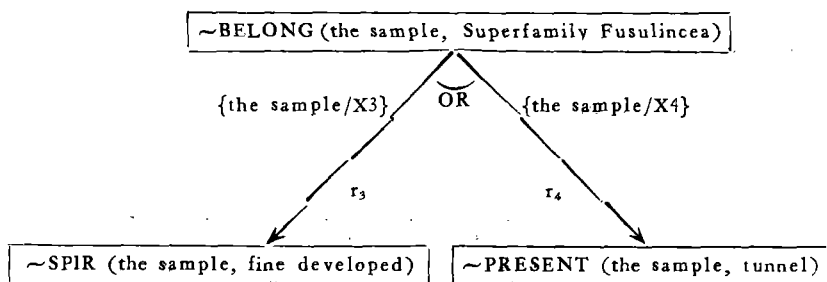


插图 2 AND/OR 树的一个实例

以产生一棵 AND/OR 树。应用前面的规则集合,对于第(2)个目标结点,我们有如插图 2 的 AND/OR 树。在这里遇到两次合一匹配。所用到的变元置换分别为 $\{the\ sample/X3\}$, $\{the\ sample/X4\}$ 。

2. 将 AND/OR 树分解为 AND 树。插图 2 可分解成插图 3。

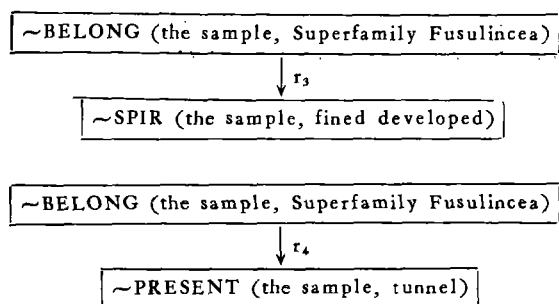


插图 3 AND 树形图

接下去对每棵 AND 树进行求证。在求证中,通过人机对话,根据 AND 树计算机随时向用户索取待鉴定化石的特征。然后再判定 AND 树是否成立。只要 AND/OR 树中的一棵 AND 树成立,则 AND/OR 的目标结点得证。

这样,通过对每个目标结点进行求证,最后可以确定待鉴定化石归于哪一类型。当然也有可能不属于任何类型,这就说明是一新类型,这时计算机应该及时向用户提出。当鉴定完一层次后,计算机就转入下一层次的鉴定,一直进行到所需要的归属为止。

另外,在鉴定过程中,计算机将用户提供的事实都储存下来。在提问前,计算机自动地在已储存的事实中寻找,若找不到,再向用户提问。此外,我们还应对树采取一些修剪措施。

这样就避免向用户重复索取同一事实的问题。

三、程序的基本结构

以上只是讨论了人工智能原理在古生物分类中的应用的一些基本问题。关于人工智能的细节问题可参考文献 (Nils J. Nilsson, 1982)

根据上述原理,我们用 BASIC 语言编制了古生物化石的计算机自动分类的程序。程序主要包括以下模块:

1, 控制模块。根据归属关系从规则集中选择合适的子集,并将其调入内存。同时根据归属关系产生 AND/OR 树的根结点;

2, 推理模块。该模块根据根结点形成 AND/OR 树,并将其分解为 AND 树。然后逐一求证;

3, 显示模块,其基本功能是将谓词公式转变成通常的英语句子,然后在屏幕上显示。该模块在必要时还可以显示推理的过程;

4, 修改规则库模块。用户通过该模块输入或删除规则等。这增加了程序的灵活性。

我们用䗴类化石作为实例,进行了试验。结果是令人满意的。分类的效率比人工提高了十几倍。该程序相对于知识库是相对独立的,因此也适用于其它的古生物分类系统。另外,程序是用 BASIC 语言编写的,因此它可以在大多数微机上实现。当然,本文讨论的内容是比较粗浅的,只是抛砖引玉。可以预料,人工智能对于古生物资料的智能检索的应用是有广阔前景的。

参 考 文 献

Nils J. Nilsson, 1982: Principles of artificial inte-

lligence, -Springer-Verlag Berlin Heidelberg
New York.

[1986 年 11 月 10 日收到]

APPLICATION OF ARTIFICIAL INTELLIGENCE PRINCIPLE TO PALEONTOLOGIC TAXONOMY

Wang Yi-feng and Zhang Yi-kun

(Department of Earth Sciences, Nanjing University)

Summary

This paper discusses the automatic classification of paleontologic fossils by micro-computer. It is pointed out that the system of paleontologic taxonomy can be represented by predicate formulas and the fossils can be automatically classified by means of predicate calculus. Taking into consideration the features of paleontologic taxonomy,

the authors have put forward a control strategy for the rule selection, which greatly raises the operating efficiency. The paper also discusses the computer program for paleontologic taxonomy. Through practice it has been proved that the approach given here is very effective.