

地层界线, 地层岩性, 序列结构, 对比分析

42-45 对比确定地层界线的一个新方法

p53

刘吉平 张光前

(中国地质大学武汉地遥研究所·武汉·430074)

对比确定地层界线是地质研究中的一项基础工作。本文基于已知标准剖面与待对比剖面的地层岩性序列结构, 运用句法模式识别技术, 提出了一个确定标准剖面中某地层界线在待对比剖面中的位置的新方法。该方法亦可用于两剖面岩性段之间的相似性的对比。

关键词 地层界线对比 地层岩性序列结构 句法模式识别 相似性度量



工作示意图

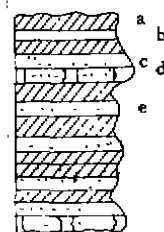
在基础地质和矿床勘探中, 地层界线对比是一项基本工作。在研究矿区钻孔剖面时, 地层的划分及对比对研究层控矿床的含矿层位更具有普遍意义。然而往往由于断层、褶皱等构造作用造成地层的重复或缺失, 使得地层界线的对比存在较大的人为性、不确定性及分歧。因此, 如何尽可能客观地对比地层剖面, 确定各地层界线的位置, 是一个很重要的问题。已有一些对比划分地层界线的定量方法, 如最优分割法, 地层对比动态规划法等, 但其对剖面的地层岩性序列结构考虑不多。而在地层界线对比中, 岩性序列结构是基本依据之一(尽管有时存在相变)。例如地质填图时的界线对比, 经常使用界线标志层及其附近的岩性组合特征。本文即基于地层岩性剖面序列间的相似性度量, 而提出一种地层界线对比划分的新方法。

1 地层剖面结构及其相似性度量

1.1 地层剖面结构的形式化描述

所谓地层剖面结构或地层岩性序列结

构, 是指地层岩性的一维排列关系, 即岩性及其相互顺序关系。如从上到下为细砂岩、粉砂岩、板岩、千枚岩、……、砾岩等的岩性序列即确定了一个岩性序列结构。对这种序列结构的形式化描述, 可引用句法模式识别的方法, 即句子(或称字符串)。图 1 是这种描述方法的一个示例。岩性划分的粗细依具体问题的要求及已有资料的研究精度而定。我们称图 1 中每个唯一地代表一种岩性的字符为一个基本模式元, 称描述该剖面结构的字符串为一个句子。



abcdacaeacaeed

图 1 地层剖面序列结构的句法描述

在对比地层界线时, 仅仅依据岩性及其相互顺序关系, 往往还不能充分客观、准确地确定出界线位置。常常需要同时考虑岩性本身具有的多种定量属性, 如厚度、物质成分、

本文 1996 年 3 月收到, 于纯烈编辑。

本文得到青年基金项目《地学深层次信息提取和高精度预测模型建立方法技术研究》(项目号: 49502042) 资助。

物理性质等等。这时,对地层剖面结构的描述,可用一个附带参数——数值属性向量的句子。设未附带数值属性向量的第 i 个句子为

$$V_i = u_{i1}u_{i2}\cdots u_{im_i}, u_{ij} \in V_T \quad (1)$$

$$i = 1, 2, \cdots, m; j = 1, 2, \cdots, n_i$$

其中 V_T 为基本模式元集。若每一基本模式元 u_{ij} 具有参数向量

$$\bar{q}_{ij} = (q_{ij1}, q_{ij2}, \cdots, q_{ijl_{ij}})$$

其中下标 l_{ij} 表示对不同模式元其参数个数可以不同。那么,第 i 个附带参数向量的句子表示为

$$V_i = u_{i1}(\bar{q}_{i1})u_{i2}(\bar{q}_{i2}), \cdots, u_{im_i}(\bar{q}_{im_i}) \quad (2)$$

例如,一个考虑岩性层厚度的岩性序列,其描述即如图 2 所示。

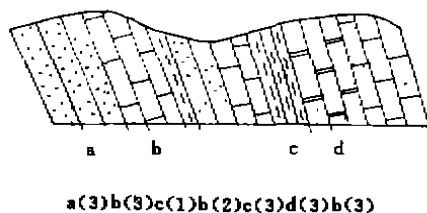


图 2 附带厚度值属性的地层剖面序列结构的描述

1.2 地层剖面结构之间的相似性计算

两个地层剖面结构之间的相似性,即是两个句子之间的相似性。其计算的方法不止一种。对于不附带数值属性向量的句子(1),N. V. Findler (1979)提出了一个方法,称之为 Findler 相似性度量 $F(x, y)$ 。其定义为

$$F(x, y) = \frac{\sum \min \{P(x : \alpha), P(y : \alpha)\}}{\sum \max \{P(x : \alpha), P(y : \alpha)\} |\alpha|} \quad (3)$$

其中 $P(S : \alpha)$ 表示字符串(句子) S 的子串 α 在 S 中发生的次数(包括重叠部分); $|\alpha|$ 表示子串 α 的长度(字符个数); $\sum$ 表示对 x, y 中的所有子串求和。

Findler 相似性度量满足如下两个原则:

①两个串间固定个数的不同元素(字符)当串无限增长时越来越不重要;②若两个串间只

有固定个数的元素不同,则当串无限增长时其相似性度量值的极限为 1。

对于附带数值属性向量的句子(2),笔者将(3)式的定义加以扩展,其原理与过程简述如下。

为要确定字符串 $V_i, V_j (i \neq j)$ 之间的相似性 $F'(V_i, V_j)$,我们不仅要比较两个字符串中基本模式元的异同,还应比较它们各自参数向量 $\bar{q}_{ik}$ 与 $\bar{q}_{jk}$ 之间异同。也就是在比较两个字符串的子串时,对于“相同”的模式元,我们要依据它们各自参数向量之间的差异来确定它们的何种程度上“相同”。为此,首先对两个字符串中同一模式元定义一个参数向量之间的相似性度量 $S(\bar{q}_{ik}, \bar{q}_{jk})$ (对应于 $u_{ik} = u_{jk}$)。考虑用距离的倒数来定义

$$S(\bar{q}_{ik}, \bar{q}_{jk}) = \frac{1}{d(\bar{q}_{ik}, \bar{q}_{jk}) + 1} \quad (4)$$

其中距离 $d(\dots)$ 可选择任一合适的距离度量,如欧氏距离。式中分母加 1 是为了使该式总有意义并归一化。当 $S=1$ 时,认为两模式元完全相同;当 $S=0$ 时,则认为二者实际上是不同的(数值属性完全不同)。

推广的 Findler 相似性的计算步骤如下:

①将各基本模式元的参数向量正规化

$$q'_{ij} = \frac{q_{ij} - \bar{Q}_j}{S_j}, S_j^2 = \frac{\sum_{i=1}^{m'} (q_{ij} - \bar{Q}_j)^2}{m' - 1}$$

$$i = 1, 2, \cdots, m'; j = 1, 2, \cdots, l_{u_{ij}}$$

其中 m' 为含有模式元 u_{ij} 的所有字符串中 u_{ij} 的总个数; $\bar{Q}_j$ 为 m' 个 u_{ij} 的参数向量中第 j 个参数的平均值; S_j 为 m' 个 u_{ij} 的参数向量中第 j 个参数的标准差; q_{ij} 为第 i 个模式元 u_{ij} 参数向量中第 j 个原始值; q'_{ij} 为对应于 q_{ij} 的规格化值。当然,用极差变换代替上述变换亦可。

②计算相应于 $|\alpha|$ 的 $|\alpha'|$ 值

$$|\alpha'| = \sum_{k=1}^{|\alpha|} S(\bar{q}_{ik}, \bar{q}_{jk}) \quad (5)$$

式中 $|\alpha|$ 即为(3)式中子串 α 的长度, $S(\bar{q}_{ik}, \bar{q}_{jk})$ 由(4)式计算。 $0 < |\alpha'| \leq |\alpha|$, 当两个

子串字符相同且对应参数向量也完全相同时, $|\alpha'| = |\alpha|$; 当两个子串字符相同而对应参数向量的差异趋于无穷大时, $|\alpha'| \rightarrow 0$ 。

③将(5)式代入(3)式, 可得推广的 Findler 相似性计算公式

$$F'(x, y) = \frac{\sum_{\alpha \in P(x, y)} \min \left\{ \sum_{P(x, \alpha)} |\alpha'|, \sum_{P(y, \alpha)} |\alpha'| \right\}}{\sum_{\alpha \in P(x, y)} \max \left\{ \sum_{P(x, \alpha)} |\alpha'|, \sum_{P(y, \alpha)} |\alpha'| \right\}} \quad (6)$$

式中 $\sum_{P(S, \alpha)} |\alpha'|$ 表示对 α 在 S 中出现的所有位置分别计算 $|\alpha'|$ 值并求和, 其他符号意义同(3)式。

2 基于剖面结构的地层界线对比方法

方法的原理乃基于两个剖面在界线附近的岩性结构及岩性属性之间的相似性比较。其基本假设是在一个地质条件具有可比性或相似性的有限范围内, 同一地层界线在不同剖面中的该界线附近的有限区段内, 其岩性序列结构及岩性属性具有最大相似性。由此出发, 设计如下对比方法: 已经充分研究而确立的地层界线所在剖面为标准剖面, 截取该剖面中以该界线为中心(也可以该界线为顶界或底界)、长度为 l 的一段岩性为标准岩性段(简称标准段)。 l 的确定以能充分识别该界线为准则。考虑有实用性的地层界线标准往往不止一个, 因此标准段亦可有多个(这样更适合存在相变的情形)。将各标准段按“标准化程度”由高到低排序。然后求所有标准段与待对比剖面中的所有子段的相似性。其相似性最大的子段即为待对比剖面中所求界线的位置。具体过程如下:

设 $B = \{b_i | i = 1, 2, \dots, m\}$, B 为标准段集, b_i 为第 i 个标准段; $A = \{a_i | i = 1, 2, \dots, n\}$, A 为待对比剖面集, a_i 为第 i 个待对比剖面; $\beta_j \supset a_j$ 为长度为 $|b_i|$ 的 a_j 的子段; $F(b_i, \beta_j)$ 为 b_i 与 β_j 的相似性,

$$\text{则 } F'(b_i, a_j) = \max_{\substack{\beta_j \in B \\ \beta_j \supset a_j}} F(b_i, \beta_j)$$

所对应的位置即为待对比剖面 a_j 中所

求界线所在岩性段位置。 $F(b_i, \beta_j)$ 用(6)式计算。当考虑 b_i 的标准性程度时, 可对 $F(b_i, \beta_j)$ 乘一按 b_i 给定的权值。

算法:

输入: 标准段集 B , 待对比剖面集 A ;

输出: b_i , 对应于 F' 的 β_j 及其在 a_j 中的位置, F', β_j 。

(1) $j = 1$;

(2) $i = 1, \max = 0$;

(3) 求 $F'(b_i, a_j)$;

(4) 若 $\max < F'(b_i, a_j)$, 则 $\max = F'$;

(5) $i = i + 1$, 若 $i \leq m$, 返回(3); 否则, 输出 $i, j, \max, \beta_j$ 及位置;

(6) $j = j + 1$, 若 $j \leq n$, 返回(2); 否则, 结束。

3 应用实例

在某矿区收集 57 个钻孔岩性剖面资料, 从中挑选 4 个地层划分相对完整可靠的钻孔剖面为标准剖面。再从这 4 个剖面中截取从 $\epsilon_2 l / \epsilon_2 t$ 界线(中寒武龙哈组与田蓬组的分界线)之上的 9 个岩性分层作为标准段。(根据地质资料, 更合适的取法应以界线为中心对称截取, 但因为钻孔中该界线之下一般见矿, 而钻孔一般又终止于矿层下的岩性层, 所以无法对称地在该界线两侧截取分层)。将此 4 个标准段和待对比的 57 个标本的岩性序列(包括岩性厚度值属性)作为输入, 用上节所述算法计算, 得到 57 个钻孔剖面中 $\epsilon_2 l / \epsilon_2 t$ 界线的计算位置。比较钻孔中该界线的原划分位置与计算位置, 结果令人满意。除 4 个包含标准段的剖面完成吻合外, 其余的吻合率达 81.2%。并且, 结果还表明, 对某些钻孔计算结果比人工划分更为合理。如某钻孔原来 $\epsilon_2 l / \epsilon_2 t$ 界线在第 17 个分层处, 计算后划分在第 11 个分层处(图 3)。考虑区域上第 16 分层的大破碎带一般在该界线之下, 故其计算位置比原划分位置更合理些。

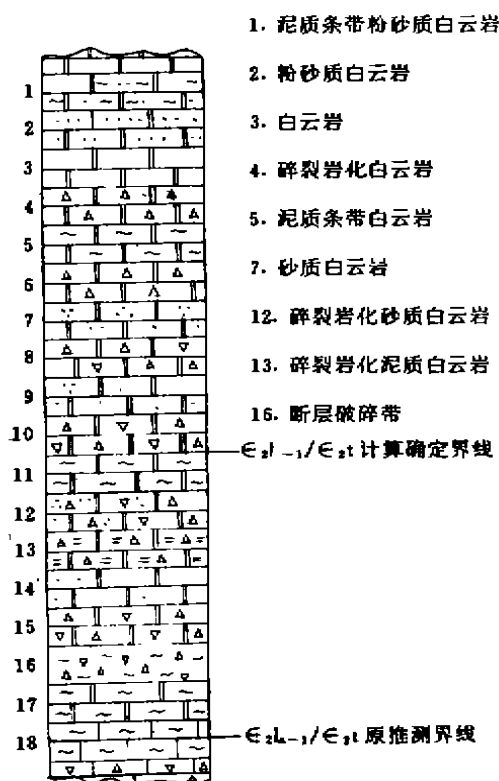


图3 ZK51-35孔原地层界线与
计算确定界线对比

4 结论

基于地层岩性序列结构的地层界线对比方法比已有方法更加符合地层界线对比的实际准则,并且它将岩性序列结构与岩性的其他属性统一考虑,使划分对比更加客观合理。应用表明,这一方法是有效的。最后还有必要补充三点:一是对标准剖面要有真正细致、客观的研究,识别界线的可靠性只是对标准剖面而言;二是最好要有地层或岩性的其他数值属性特征,如物质成份、物理性质等,这样,对比的精确度会提高;三是本方法亦可用于两剖面岩性段之间相似性的对比。

感谢陈永清副教授对作者完成此文给予的帮助。

参考文献

- 1 周冠雄. 计算机模式识别—结构方法(第一版). 武汉:华中理工大学出版, 1987, 130~148.
- 2 赵鹏大等. 矿床统计预测(第二版). 北京:地质出版社, 1994, 243~252.

A NEW METHOD TO DETERMINE STRATIGRAPHIC BOUNDARY

Liu Jiping Zhang Guangqian

Based on the basic principle which is used to determining a unknown stratigraphic boundary according to the structure of the stratigraphic sequence in which boundary has been identified, a new method is introduced here by syntactic approach in pattern recognition. It's application is also discussed by this paper.

Key words: stratigraphic correlation, the structure of stratigraphic sequence, pattern recognition, similarity

