

束流发射度形状的句法模式识别分析

何卫宁

(中国科学院高能物理研究所,北京)

摘 要

本文利用句法模式识别的理论和技術,处理束流发射度形状的畸变问题.这种方法首先把发射度形状转换成符号串描述,然后对该符号串进行句法分析.在句法分析过程中,利用句法引导的翻译技术消除发射度形状畸变.

一、引 言

北京 35MeV 质子直线加速器予注入器 (750keV) 的实测发射度形状如图 1 所示,发射度形状发生了畸变.这种畸变是由于离子源发射的离子除质子 H^+ 外,可能还有 H_2^+ 和 H_3^+ . 图 1 中, AA' 代表所需要的 H^+ 离子的发射度形状,而 BB' 为 H_2^+ 离子的发射度形状.为了正确地分析质子束流的发射度,以便研究和调试加速器的性能,应该把 BB' 从发射度形状中删去.

从原理上讲,可以采用最小二乘拟合法求出一个最小二乘拟合椭圆.但这种办法有一个弱点,在拟合时 AA' 和 BB' 被同等地对待,这就必然造成偏差.

让我们从另外一个角度看待这个畸变问题.一个具有束流发射度概念的人,能容易地从测量结果的形状上识别出畸形部分和正常部分.通常,实际的发射度形状近似为椭圆,如果认为 H^+ 的椭圆总应是最大的,那么把测量结果中最大的椭圆识别出来,问题就解决了.在这里,人主要是利用其形状识别能力而不是数学运算能力.因此,如果让计算机取代人完成这项工作,就必须使计算机具有形状识别能力和有关发射度的知识.

为使计算机具有形状识别能力,我们引入句法模式识别方法.在这种方法中,一个复杂的对象总是被分解成若干基元,即基本元素.给每个基元相应的符号名称,待识别的对象就完全地被基元符号组成的符号串所描述.对象的结构就反映在该符号串的结构上.用一组句法规则来描述串的结构,句法规则指定了如何才能由基元构造出合法的串,而这些串实际上代表一些特定的对象.这样,识别某一未知对象的过程就转变成检查描述该未知对象的基元串是否与给定的句法相容.在下面一节中,将讨论如何建立一个实际句

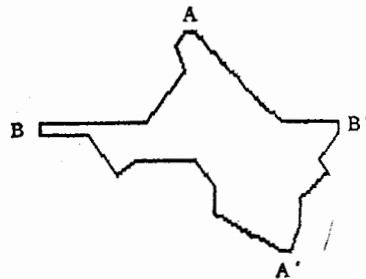


图 1 实测发射度形状

法模式识别系统来分析发射度形状。

二、句法模式识别系统

从结构上看,一个句法模式识别系统分成两个主要部分:基元提取器和句法分析器。基元提取器把待识别的形状转换成基元串,然后送给句法分析器。句法分析器根据设计好的句法分析基元串的结构。

1. 基元提取器

设计一个实际的基元提取器的主要问题是选择什么样的基元来描述形状。

发射度测量的数据存放于数组 $A(M, N)$ 中。其中:

$$A(i, j) = \begin{cases} 1 & \text{表示点 } (i, j) \text{ 在形状上。} \\ 0 & \text{表示点 } (i, j) \text{ 不在形状上。} \end{cases}$$

因此,发射度形状可视为离散平面上由 1 组成的一块区域。注意到物体的形状信息主要体现在边界上,因此需要用符号化的方法描述区域的边界。离散平面上的曲线可以用链码描述^[2]。图 2(a) 给出了链码的定义。链码实际上刻画了曲线上相邻两点间的关系。由于边界曲线连续,故一串链码就能描述边界曲线。图 2(b) 给出了一个例子。

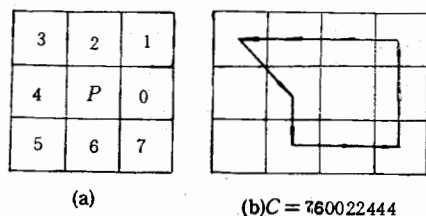


图 2

(a) 链码定义 (b) 一个例子: 曲线链码串

利用边界跟踪算法^[3]可求得发射度边界的链码描述:

$$E = c_1 c_2 \cdots c_m, \quad c_i \in \{0, 1, 2, 3, 4, 5, 6, 7\},$$

很难构造一个句法产生上述链码串,主要原因是链码描述对噪声敏感。另一方面,单个链码很难作为一个句法成份出现。因此,需要消除噪声和提取更高一层的边界描述。可以按照下述算法^[3]消除噪声并提取线段基元:在边界曲线上移动一个定长的链码片断,两相邻片断间的夹角反映了曲线的起伏情况。把低于某一个阈值的点视为噪声起伏,而高于此阈值的点可视为有意义的转折点。在有意义的转折点之间可认为是直线段。这样就获得了形状的线段基元描述:

$$E = l_1 l_2 \cdots l_n,$$

其中 l_i 表示直线段。由于线段在构成形状时有很大的自由度,通常会导致上下文敏感句法^[4],这种句法很难分析,所以提取诸如山峰、平台等形状基元是有益的。图 3 中的 (a)、(b)、(c) 和 (d) 表示山峰形状基元 HILL 的形成过程,用基元产生语言来描述。基元产生

语言的每一条语句是一条基元产生规则。基元产生规则说明了“ \rightarrow ”号右端的变量以何种方式才能产生左端的变量。花括号中的表达式表示右端的变量应满足的条件,称为基元产生条件表达式。例如,图 3(b)中的规则说明一个 UP-SIDE 变量加上一个 LINE 变量,当这个 LINE 变量的倾角在 270° 到 360° 之间,就形成一个 HILL 变量。

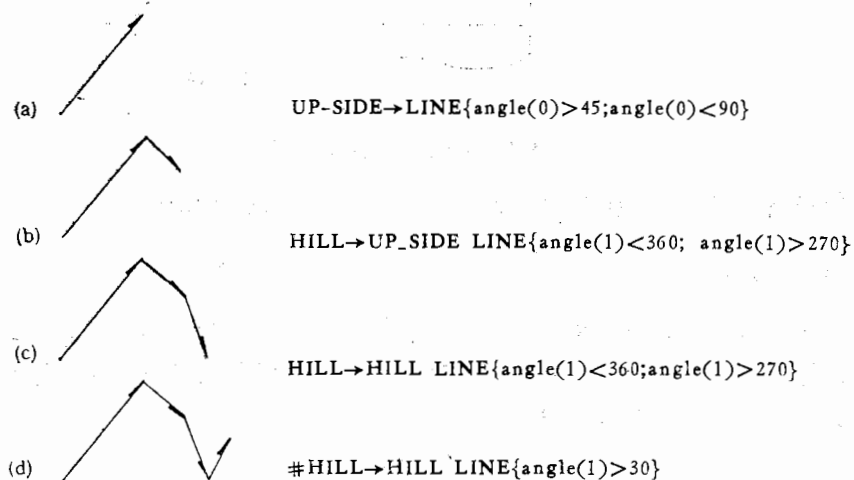


图 3 基元 HILL 的形成过程

基元提取器的结构如图 4 示。该基元提取器实际上是一个改型的有限状态自动机^[5]。它不断地从发射度线段基元描述串中读入基元,直到线段基元串被读空为止。每读入一个基元,它在基元产生规则中找出一条可用的规则,根据这条规则把自己的状态切换到新状态上。当前读入基元和当前提取器状态确定了一组可用规则的候选者。在这一组候选者中,产生条件表达式演算为真的规则是可用规则。当可用的规则是一条带“#”的规则时,一个形状基元就被提取出来了,并送入输出串。就这样,基元提取器把发射度形状的线段基元描述转换成形状基元描述。

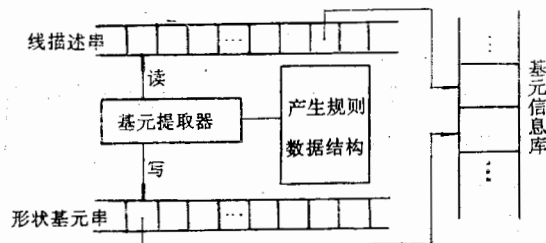


图 4 基元提取器结构

由于在书写基元产生规则时需要选定很多参数(角度参数),这给系统的调试和运行带来不便。我们采用动态数据结构技术解决这个问题。基元产生规则在系统初始化时调入,编译后形成图 5 所示的数据结构。图中 HILL 基元用指针 A 和 B 引导了一串规则,这是因为 HILL 基元可能对应许多条产生规则。每一产生规则都可以有条件表达式,指针

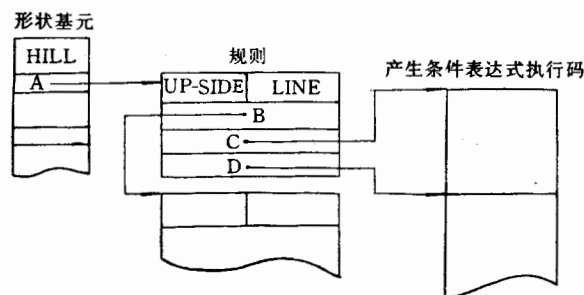


图5 基元产生规则数据结构

C和D用来指向相应的产生条件表达式可执行代码区的起点和终点。

2. 句法分析器

句法分析器的结构如图6所示。句法分析器读入形状基元串，进行句法分析后写入输出串。输出串描述了消除畸变后的发射度形状。我们利用句法引导的翻译技术^[4]来消除畸变。一般地说，句法规则具有形式：

$$A \rightarrow X_1 X_2 \cdots X_n,$$

它表示句法变量A由句法变量 X_1, X_2, \cdots, X_n 组成。句法引导的翻译规则具有形式：

$$A \rightarrow X_1 X_2 \cdots X_i \{T_i\} X_{i+1} \cdots X_n,$$

它表示在正确地识别出 $X_1 X_2 \cdots X_i$ 后，可执行翻译动作 T_i 。

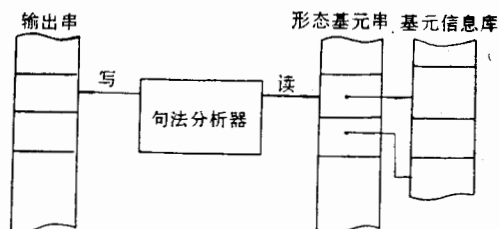


图6 句法分析器结构

由于发射度形状已由高层次基元描述，句法引导翻译规则可以简单地写为：

$$\langle \text{EMITTANCE} \rangle \rightarrow \langle \text{H} \rangle \langle \text{HLIST} \rangle$$

$$\langle \text{HLIST} \rangle \rightarrow \langle \text{H} \rangle \langle \text{HLIST} \rangle$$

$$\langle \text{HLIST} \rangle \rightarrow \varepsilon \quad (\varepsilon \text{ 意为空})$$

$$\langle \text{H} \rangle \rightarrow \langle \text{MHILL} \rangle \{ \text{输出 } \langle \text{MHILL} \rangle \text{ 到输出串} \}$$

$$\langle \text{H} \rangle \rightarrow \langle \text{DHILL} \rangle \{ \text{修剪 } \langle \text{DHILL} \rangle \text{ 后输出到输出串} \}$$

$\langle \text{MHILL} \rangle$ 和 $\langle \text{DHILL} \rangle$ 无法继续用句法定义，只能利用基元的几何属性确定一个基元应属于 $\langle \text{MHILL} \rangle$ 或是 $\langle \text{DHILL} \rangle$ 。基元提取器已把基元的几何属性存放于基元信息库中(参见图4)供句法分析器使用。

三、分析系统的实现与实验事例

发射度形状句法模式识别分析系统已用 C 语言实现。该系统使用自上而下的编译原理^[4]把基元产生规则编译成伪指令。伪指令可由一伪处理器运行。基元提取器在适当的时候启动此伪处理器演算某一产生条件表达式。句法分析器也用自上而下编译原理分析形状基元串。

这个系统已成功地用于发射度形状分析。图 7 和图 8 给出了两组实验事例。从这两组事例中可看出修剪的结果与人的直觉符合。

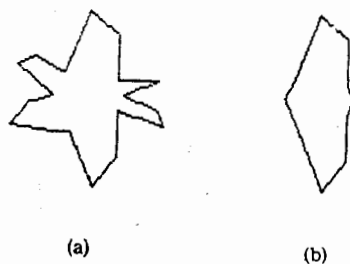


图 7

(a) 原形状 (b) 修剪后形状

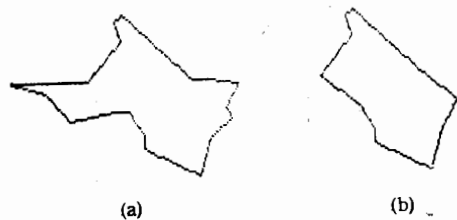


图 8

(a) 原形状 (b) 修剪后形状

下面给出进行发射度形状分析时使用的一组基元产生规则。其中 Line 表示发射度形状描述中的线段基元; Ang 为角度基元, 它描述两相邻线段基元间角度的变化。

基元产生规则

```

Side→Line
Sidel→Side Line
Sidel→Sidel Line
#Side→Side Ang {angle(2)>180}
#Hilli→Sidel Ang {angle(2)>180}
Turn→Ang {angle(1)>180}
#Turn→Turn Line
Hill→Side Ang {angle(2)<180}
Hill→Sidel Ang {angle(2)<180}
Hill→Hill Ang {angle(2)<180}
Hill→Hill Line
#Hill→Hill Ang {angle(2)>180}

```

四、结束语

把句法模式识别方法用于束流发射度形状分析已取得较好的效果。尽管这种方法会

损失一些传统意义上的精确度,但它能解决一些数字方法难以处理的问题。对一些往往需要人介入才能完成的数据分析问题,都可考虑利用句法模式识别方法。

感谢薛景瑄研究员、王书鸿副研究员在本文工作期间所给予的帮助。

参 考 文 献

- [1] T. Pavlidis, Algorithms for Computer Graphics and Image Processing, Computer Science Press, 1982
- [2] T. Pavlidis, A Review of Algorithms for Shape Analysis, Computer Graphics and Image Processing 7, 1978, 243—258
- [3] H. Freeman, and C. S. David, A Corner Finding Algorithm for Chain Coded Curves, IEEE Trans. on Computer C-26, 1977, 297—303
- [4] A. H. Aho, and J. D. Ullman, Principles of Compiler Design, Addison Wesley Pub. Comp., 1977
- [5] P. M. Lewis, and D. J. Rosenkrantz, and R. E. Stearns, Compiler Design, Theory, Addison Wesley Pub. Comp., 1976

SYNTACTIC ANALYSIS OF DEFORMED EMITTANCE SHAPE MEASUREMENTS

HE WEINING

(Institute of High Energy Physics, Academia Sinica, Beijing)

ABSTRACT

The practical emittance measurement of a Proton Linear Accelerator inevitably involves deformation. This paper suggests a syntactic pattern recognition method to handle the deformation problem. This approach consists of two steps. The first step is to use a set of shape primitives to describe the contour of the emittance shape. The second step is to analyze the description syntactically. Syntax-Directed Translation scheme is employed to remove the deformation.