计算科学的新领域:DNA 计算(Ⅱ)

李 燕
(浙江师范大学信息科学与工程学院 金华 321004)

摘 要 DNA 计算是应用分子生物技术进行计算的新方法。从理论上研究 DNA 计算方法,有利于推动理论计算科学的发展。本系列文章应用形式语言及自动机理论技术,系统地探讨了 DNA 分子的可计算性及其计算能力。本文主要介绍 DNA 分子粘接计算模型的文法结构和计算方法,探讨了不同粘接计算模型的计算能力,并证明了 DNA 有穷自动机与正规文法的等价性。

关键词 DNA 粘接计算模型,计算能力,DNA 有穷自动机

A New Field of Compute Science: DNA Computing (Ⅱ)

LI Yan
(College of Information Science and Engineering, Zhejiang Normal University, Jinhua 321004)

Abstract DNA computing is a new method for computation using the technology in molecular biology. The study of DNA computing theory will be of benefit to computing science theory. The series papers systematically discuss the computability and the computational capacity of DNA molecular using the formal language and automata theory. In this paper, we mainly introduce the grammar structures and the computation methods of DNA sticker model, discuss the computational capacity of several DNA sticker models, and prove the equipollence of DNA finite automata and regular grammar.

Keywords DNA sticker model, Computational capacity, DNA finite automata

1 引言

计算的本质是递归,是依据一定的法则对有关符号串进行变换的过程。DNA 分子可以看作是集合$\Sigma=\{A,T,C,G\}$上的字符串[1,2],自然就是应用专门处理字符和字符串的形式语言理论为 DNA 计算建模。建模的关键是将实际的生物分子操作抽象为数学操作方法,计算过程中所采用的操作方法是在理想状态下完成的(不存在误差)。具有互补粘性末端的两个 DNA 分子,能够按照 Watson-Crick 碱基互补原则进行粘接[11]。两个平端的 DNA 分子在连接酶的作用下,可以连接成为一个 DNA 分子。DNA 分子短链在一定的规则控制下进行粘接计算,通过一系列的粘接计算可以达到完全计算。达到完全计算时所生成的完备 DNA 分子就是计算结果。

根据 DNA 分子的基本结构和粘接特性,可以抽象地构造出 DNA 分子粘接计算模型[3~5]。本文给出了粘接模型的文法结构,探讨了不同 DNA 粘接模型的计算能力,并介绍了 DNA 有限自动机[7~13]。

2 DNA 粘接计算模型的文法结构

定义 1 DNA 粘接计算模型的文法结构为 $\gamma(\Sigma,\rho,\Gamma,D)$,其中:

(1)Σ是字母表;(2)$\rho\subseteq\Sigma\times\Sigma$是终结符的互补对应关系;(3)$\Gamma$是 $LR_\rho(\Sigma)$的有限子集,是开始符号的集合。开始符号中的 DNA 双链序列,要求上下链中对应的字母至少包含一个形式为$\begin{bmatrix}a\\b\end{bmatrix}(a\neq\varepsilon,b\neq\varepsilon)$的对应列;(4)$D$ 是 $Z_\rho(\Sigma)\times Z_\rho(\Sigma)$的有限子集,$D$ 中准分子的组合对(u,v)是非终结符的集合。

DNA 粘接计算模型 γ 的计算从 $LR_\rho(\Sigma)$中的"开始符号"开始。开始符号的粘性末端与 D 中非终结符的粘性末端,按照 Watson-Crick 碱基互补方式在相应位置上进行粘接。通过粘接计算,使得开始符号向左或向右延伸,这个过程要求互补的双链序列中没有空位置出现。整个计算过程可表示为一组序列 $x_1\Rightarrow x_2\Rightarrow\cdots\Rightarrow x_k$,其中 $x_1\in\Gamma$。当 $x_k\in FZ_\rho(\Sigma)$时,粘接计算 $\sigma:x_1\Rightarrow^* x_k$ 达到了完全计算。通过完全计算,可以获得具有完备形式的 DNA 分子,这就是计算结果。

经过完全计算所产生的具有完备形式 DNA 分子的集合表示为 $LFZ_n(\gamma)$。LFZ 表示分子语言,下标 n 表示"没有限制",也就是说,除了达到完全计算之外没有其它任何限制条件;

$LFZ_n(\gamma)=\{w|x\Rightarrow^* w,x\in\Gamma,w\in FZ_\rho(\Sigma)\}$。

由 DNA 粘接计算模型 γ 产生的语言表示为:

$$L_n(\gamma)=\left\{w\left|\begin{bmatrix}w\\w'\end{bmatrix}\in LFZ_n(\gamma),w\in\Sigma^*,w'\in\Sigma^*\right.\right\}。$$

由于 DNA 分子的上下链中的碱基是互补的,因此 $L_n(\gamma)$相当于是 $LFZ_n(\gamma)$的编码序列。

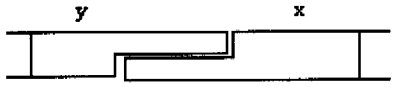
3 粘接操作方法

设开始符号为 x,$x=x_1x_2x_3$,其中 $x_2\in FZ_\rho(\Sigma)-\left\{\begin{bmatrix}\varepsilon\\\varepsilon\end{bmatrix}\right\}$,$x_1,x_3\in\begin{pmatrix}\Sigma^*\\\varepsilon\end{pmatrix}\cup\begin{pmatrix}\varepsilon\\\Sigma^*\end{pmatrix}$。

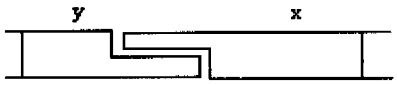
$\varphi(y,x)$表示 $x\in Z_\rho(\Sigma)$和 $y\in Z_\rho(\Sigma)$向左延伸的粘接操

李 燕 博士,副教授,主要研究领域为自动机理论,DNA 计算,人工智能。

作结果。粘接结果的形式可分为以下几种：

(1) 

当 $y=y'\begin{pmatrix}w_1\\ \varepsilon\end{pmatrix}$，$x_1=\begin{pmatrix}\varepsilon\\ w_2\end{pmatrix}$ 时，其中 $w_1,w_2\in\Sigma^*$，$\begin{bmatrix}w_1\\ w_2\end{bmatrix}\in FZ_\rho(\Sigma)$，$y'\in L_\rho(\Sigma)$；那么 $\varphi(y,x)=y'\begin{bmatrix}w_1\\ w_2\end{bmatrix}x_2x_3$。

(2) 

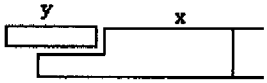
当 $y=y'\begin{pmatrix}\varepsilon\\ w_2\end{pmatrix}$，$x_1=\begin{pmatrix}w_1\\ \varepsilon\end{pmatrix}$ 时，其中 $w_1,w_2\in\Sigma^*$，$\begin{bmatrix}w_1\\ w_2\end{bmatrix}\in FZ_\rho(\Sigma)$，$y'\in L_\rho(\Sigma)$；那么 $\varphi(y,x)=y'\begin{bmatrix}w_1\\ w_2\end{bmatrix}x_2x_3$。

(3) 

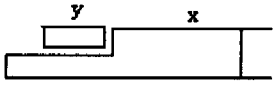
当 $y=y\begin{pmatrix}w_1\\ \varepsilon\end{pmatrix}$，$x_1=\begin{pmatrix}w_2\\ \varepsilon\end{pmatrix}$ 时，其中 $w_1,w_2\in\Sigma^*$ 时，那么，$\varphi(y,x)=\begin{bmatrix}w_1w_2\\ \varepsilon\end{bmatrix}x_2x_3$。

(4) 

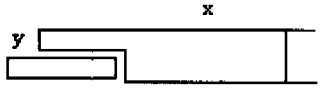
当 $y=y\begin{pmatrix}\varepsilon\\ w_1\end{pmatrix}$，$x_1=\begin{pmatrix}\varepsilon\\ w_2\end{pmatrix}$ 时，其中 $w_1,w_2\in\Sigma^*$，那么 $\varphi(y,x)=\begin{bmatrix}\varepsilon\\ w_1w_2\end{bmatrix}x_2x_3$。

(5) 

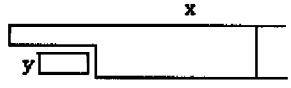
当 $y=\begin{pmatrix}w_1w_2\\ \varepsilon\end{pmatrix}$，$x_1=\begin{pmatrix}\varepsilon\\ w_3\end{pmatrix}$ 时，其中 $w_1,w_2,w_3\in\Sigma^*$，$\begin{bmatrix}w_2\\ w_3\end{bmatrix}\in FZ_\rho(\Sigma)$；那么，$\varphi(y,x)=\begin{pmatrix}w_1\\ \varepsilon\end{pmatrix}\begin{bmatrix}w_2\\ w_3\end{bmatrix}x_2x_3$。

(6) 

当 $y=\begin{pmatrix}w_1\\ \varepsilon\end{pmatrix}$，$x_1=\begin{pmatrix}\varepsilon\\ w_2w_3\end{pmatrix}$ 时，其中 $w_1,w_2,w_3\in\Sigma^*$，$\begin{bmatrix}w_1\\ w_3\end{bmatrix}\in FZ_\rho(\Sigma)$；那么，$\varphi(y,x)=\begin{pmatrix}\varepsilon\\ w_2\end{pmatrix}\begin{bmatrix}w_1\\ w_3\end{bmatrix}x_2x_3$。

(7) 

当 $y=\begin{pmatrix}\varepsilon\\ w_2w_3\end{pmatrix}$，$x_1=\begin{pmatrix}w_1\\ \varepsilon\end{pmatrix}$ 时，其中 $w_1,w_2,w_3\in\Sigma^*$，$\begin{bmatrix}w_1\\ w_3\end{bmatrix}\in FZ_\rho(\Sigma)$；那么，$\varphi(y,x)=\begin{pmatrix}\varepsilon\\ w_2\end{pmatrix}\begin{bmatrix}w_1\\ w_3\end{bmatrix}x_2x_3$。

(8) 

当 $y=\begin{pmatrix}\varepsilon\\ w_3\end{pmatrix}$，$x_1=\begin{pmatrix}w_1w_2\\ \varepsilon\end{pmatrix}$ 时，其中 $w_1,w_2,w_3\in\Sigma^*$，$\begin{bmatrix}w_2\\ w_3\end{bmatrix}\in FZ_\rho(\Sigma)$；那么，$\varphi(y,x)=\begin{pmatrix}w_1\\ \varepsilon\end{pmatrix}\begin{bmatrix}w_2\\ w_3\end{bmatrix}x_2x_3$。

按照 Watson-Crick 碱基互补配对原则，上述的粘接计算(1)和(2)的粘性末端必须互补，并且长度一致。(3)和(4)不能退火，它们是通过连接酶绑结到一起的。按照上述的方法，同样可以列出 $\varphi(x,y)$ 的操作方法。至于双链序列是向左还是向右延伸，主要取决于分子本身的排列形式和它们的粘性末端。

4 DNA 粘接计算模型的计算能力

定义 2 DNA 粘接计算模型 $\gamma=(\Sigma,\rho,\Gamma,D)$ 被称为是：

(1)正规的，如果所有的 $(u,v)\in D$，有 $u=\varepsilon$。

(2)单方向的，如果对于每一对 $(u,v)\in D$，有 $u=\varepsilon$ 或者 $v=\varepsilon$。

(3)简单的，如果所有的 $(u,v)\in D$，有 $u,v\in\begin{pmatrix}\Sigma^*\\ \varepsilon\end{pmatrix}$ 或者 $u,v\in\begin{pmatrix}\varepsilon\\ \Sigma^*\end{pmatrix}$。

另外，还有简单单方向的、简单正规的粘接模型。在单方向的粘接模型中，向左的延伸与向右的延伸是相互独立的。在正规的粘接模型中只有向右延伸的序列，因此开始符号必须是 x_1x_2，其中 $x_1\in FZ_\rho(\Sigma)$，$x_2\in\begin{pmatrix}\Sigma^*\\ \varepsilon\end{pmatrix}\cup\begin{pmatrix}\varepsilon\\ \Sigma^*\end{pmatrix}$。

定理 1 正规 DNA 粘接计算模型能够产生属于正规语言的语言。

证明：设一个正规粘接模型为 $\gamma=(\Sigma,\rho,\Gamma,D)$。

构造一个上下文无关文法 $G=(N,T,S,P)$，$N=\{S\}\cup\{\begin{pmatrix}u\\ \varepsilon\end{pmatrix},\begin{pmatrix}\varepsilon\\ \bar{u}\end{pmatrix}\}$，$u,\bar{u}\in\Sigma^*$。使其能够模仿正规粘接计算模型的计算过程。

P 包含下面的规则：

(1) $S\to\begin{bmatrix}x_1\\ x_2\end{bmatrix}\begin{pmatrix}u_1\\ u_2\end{pmatrix}$，其中 $\begin{bmatrix}x_1\\ x_2\end{bmatrix}\begin{pmatrix}u_1\\ u_2\end{pmatrix}\in\Gamma$，$\begin{pmatrix}u_1\\ u_2\end{pmatrix}\in\begin{pmatrix}\varepsilon\\ \Sigma^*\end{pmatrix}\cup\begin{pmatrix}\Sigma^*\\ \varepsilon\end{pmatrix}$，$\begin{bmatrix}x_1\\ x_2\end{bmatrix}\in FZ_\rho(V)$。

(2) $\begin{pmatrix}u_1\\ u_2\end{pmatrix}\to\begin{bmatrix}w_1\\ w_2\end{bmatrix}\begin{pmatrix}u_1'\\ u_2'\end{pmatrix}$，其中 $\begin{pmatrix}u_1\\ u_2\end{pmatrix},\begin{pmatrix}u_1'\\ u_2'\end{pmatrix}\in\begin{pmatrix}\varepsilon\\ \Sigma^*\end{pmatrix}\cup\begin{pmatrix}\Sigma^*\\ \varepsilon\end{pmatrix}$，$\begin{bmatrix}w_1\\ w_2\end{bmatrix}\in FZ_\rho(V)$。

粘接末端 $\begin{pmatrix}u_1\\ u_2\end{pmatrix}$ 相当于非终结符。按照粘接末端的形式，在 D 的粘接对中应该有一对形式为 $(\begin{pmatrix}\varepsilon\\ \varepsilon\end{pmatrix},\begin{pmatrix}x_1\\ x_2\end{pmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix}\begin{pmatrix}u_1'\\ u_2'\end{pmatrix})$ 的粘接对。其中，$\begin{pmatrix}x_1\\ x_2\end{pmatrix}\in\begin{pmatrix}\Sigma^*\\ \varepsilon\end{pmatrix}\cup\begin{pmatrix}\varepsilon\\ \Sigma^*\end{pmatrix}$，$\begin{bmatrix}y_1\\ y_2\end{bmatrix}\in FZ_\rho(V)$，使得 $\begin{bmatrix}u_1x_1y_1\\ u_2x_2y_2\end{bmatrix}=\begin{bmatrix}w_1\\ w_2\end{bmatrix}$。

(3) $\begin{pmatrix}\varepsilon\\ \varepsilon\end{pmatrix}\to\varepsilon$。当没有粘接末端提供时，推导完成。

粘接模型是应用 D 中的粘接对，通过碱基互补使得粘接末端的 *DNA* 链向右延长。通过上述产生式的构造方法，上下文无关文法 G 可以利用产生式规则中非终结符的转换，来

模仿粘接计算模型的计算过程，使得上下文无关文法所产生的语言与正规 DNA 粘接计算模型 γ 所产生的语言是相同。

由于在上下文无关文法 G 中产生式的形式为 $X\to uX$，$X\to u$，因此该上下文无关文法 G 所产生的语言 $L(G)$ 是正规的。所以正规 DNA 粘接计算机模型能够产生属于正规语言的语言。 □

定理 2 单方向 DNA 粘接计算模型能够产生属于正规语言的语言。

证明：设单方向粘接计算模型为 $\gamma=(\Sigma,\rho,\Gamma,D)$。$d$ 为 Γ 或 D 中最长的粘接末端或最长的单链序列。

构造上下文无关文法 $G=(N,T,S,P)$，其中，$N=\{\binom{u}{\varepsilon}_l,\binom{u}{\varepsilon}_r,\binom{\varepsilon}{u}_l,\binom{\varepsilon}{u}_r\mid u\in\Sigma^*,0\leqslant|u|\leqslant d\}\cup\{S\}$，下标 l 表示左边，下标 r 表示右边。

P 包含下面的规则：

(1) $S\to\binom{u_1}{u_2}_l\begin{bmatrix}x_1\\x_2\end{bmatrix}\binom{v_1}{v_2}_r$，其中：$\binom{u_1}{u_2}\begin{bmatrix}x_1\\x_2\end{bmatrix}\binom{v_1}{v_2}\in\Gamma$，$\binom{u_1}{u_2},\binom{v_1}{v_2}\in\binom{\varepsilon}{\Sigma^*}\cup\binom{\Sigma^*}{\varepsilon},\begin{bmatrix}x_1\\x_2\end{bmatrix}\in FZ_\rho(V)$。

(2) $\binom{u_1}{u_2}_l\to\binom{u_1'}{u_2'}\begin{bmatrix}w_1\\w_2\end{bmatrix}$，其中，$\binom{u_1}{u_2},\binom{u_1'}{u_2'}\in\binom{\varepsilon}{\Sigma^*}\cup\binom{\Sigma^*}{\varepsilon},\begin{bmatrix}w_1\\w_2\end{bmatrix}\in FZ_\rho(V)$。

粘接末端 $\binom{u_1}{u_2}_l$ 相当于非终结符。按照粘接末端的形式，应用 D 中右边为空的粘接对，使得 DNA 链向左延长。所以在 D 的粘接对中有一对的形式应该为，$(\binom{u_1'}{u_2'}_l\begin{bmatrix}x_1\\x_2\end{bmatrix}\binom{y_1}{y_2},\binom{\varepsilon}{\varepsilon})$，$\binom{y_1}{y_2}\in\binom{\Sigma^*}{\varepsilon}\cup\binom{\varepsilon}{\Sigma^*},\begin{bmatrix}x_1\\x_2\end{bmatrix}\in FZ_\rho(V)$。使得 $\begin{bmatrix}x_1y_1u_1\\x_2y_2u_2\end{bmatrix}=\begin{bmatrix}w_1\\w_2\end{bmatrix}$。这里 $\begin{bmatrix}w_1\\w_2\end{bmatrix}$ 可以为 $\begin{bmatrix}\varepsilon\\\varepsilon\end{bmatrix}$。

(3) $\binom{u_1}{u_2}_r\to\begin{bmatrix}w_1\\w_2\end{bmatrix}\binom{u_1'}{u_2'}_r$。其中：$\binom{u_1}{u_2},\binom{u_1'}{u_2'}\in\binom{\varepsilon}{\Sigma^*}\cup\binom{\Sigma^*}{\varepsilon},\begin{bmatrix}w_1\\w_2\end{bmatrix}\in FZ_\rho(V)$。

按照粘接末端的形式，应用 D 中左边为空的粘接对，使得 DNA 链向右边延长。在 D 的粘接对中有一对的形式应该为$(\binom{\varepsilon}{\varepsilon},\binom{x_1}{x_2}\begin{bmatrix}y_1\\y_2\end{bmatrix}\binom{u_1'}{u_2'})$，其中$\binom{x_1}{x_2}\in\binom{\Sigma^*}{\varepsilon}\cup\binom{\varepsilon}{\Sigma^*}$，$\begin{bmatrix}y_1\\y_2\end{bmatrix}\in FZ_\rho(V)$，使得$\begin{bmatrix}u_1x_1y_1\\u_2x_2y_2\end{bmatrix}=\begin{bmatrix}w_1\\w_2\end{bmatrix}$。

(4) $\binom{\varepsilon}{\varepsilon}_l\to\varepsilon,\binom{\varepsilon}{\varepsilon}_r\to\varepsilon$。当没有粘接末端提供时，推导完成。

在推导过程中，DNA 链向左的延长操作同 DNA 链向右的延长操作是相互独立的。因此，总是可以使用 D 中形式为$(\binom{z_1}{z_2},\binom{\varepsilon}{\varepsilon})$或$(\binom{\varepsilon}{\varepsilon},\binom{z_1}{z_2})$的粘接对，来完成粘接推导过程。这样文法 G 利用产生式规则中非终结符的转换，来控制推导过程的方式，与粘接计算模型利用 D 中的粘接对来控制粘接操作可视为相同。

在文法 G 中不存在形式为 $X\Rightarrow^* uXv$ 的推导，u、v 是非空字符串。所以，语言 $L(G)$ 是正规的。 □

5 DNA 有穷自动机

自动机是与文法的工作方向相反的语言定义装置。DNA 自动机与普通自动机的数据结构是不同的。DNA 有穷自动机是按照 DNA 分子互补的双链序列定义的，是由 Watson-Crick 带、两个读写头和一个有穷状态控制器组成，根据需要还可以增加辅助存储器。Watson-Crick 带包括上下两条带，上带和下带分别对应分子的上链和下链，对应字母是互补的。两个读写头，一个读上链的符号，另一个读下链的符号。有穷状态控制器中存储了有穷状态集合中的元素，它对状态的控制转换与普通的自动机一样。辅助存储器的存储方式与 Watson-Crick 带相同。

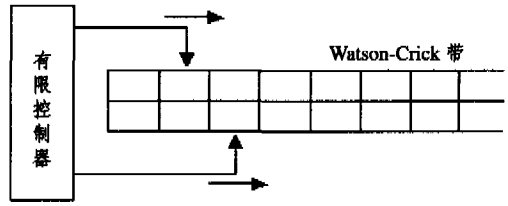


图 1 DNA 有穷自动机

定义 3 DNA 有穷自动机的结构为 $M=(\Sigma,Q,\rho,q_0,F,\delta)$，其中：

(1) Σ 是输入字母表；(2) Q 是状态集合，Σ 和 Q 是不相交的字母表；(3) $\rho\subseteq\Sigma\times\Sigma$ 是在 Σ 上互补的对应关系；(4) q_0 是初始状态；(5) 转换函数 $\delta:Q\times\Sigma\times\Sigma\to\Psi(Q)$。

$q'\in\delta(q,\binom{x_1}{x_2})$ 表示自动机处在状态 q，通过双链序列上链中的 x_1 和 x_2 下链中的 x_2，进入状态 q'。相应地，DNA 有穷自动机的转换过程表示为：$\binom{x_1}{x_2}q\binom{y_1}{y_2}\binom{z_1}{z_2}\vdash\binom{x_1}{x_2}\binom{y_1}{y_2}q'\binom{z_1}{z_2}$。其中，$\binom{x_1}{x_2},\binom{y_1}{y_2},\binom{z_1}{z_2}\in\binom{\Sigma^*}{\Sigma^*}$，$q,q'\in Q$，$\begin{bmatrix}x_1y_1z_1\\x_2y_2z_2\end{bmatrix}\in FZ_\rho(\Sigma)$，并且 $q'\in\delta(q,\binom{y_1}{y_2})$。还可用产生式规则 $q\binom{x_1}{x_2}\to\binom{x_1}{x_2}q'$ 的形式来表示状态的转换。

DNA 有穷自动机识别的语言为：

$L_u(M)=\{w_1\mid q_0\begin{bmatrix}w_1\\w_2\end{bmatrix}\Rightarrow^*\begin{bmatrix}w_1\\w_2\end{bmatrix}q_f\}$，其中，$\begin{bmatrix}w_1\\w_2\end{bmatrix}\in FZ_\rho(\Sigma)$，$q_f\in F$，$w_1\in\Sigma^*$，$w_2\in\Sigma^*$。

DNA 自动机开始工作时，处于状态 q_0，两个读写头分别位于上、下链的第一个字母前。工作时，读写头向右移动，根据当前的状态和字母，对应转换函数转换状态。当读写头从输入序列开始，通过向右的移动到达了序列的最右端，并进入终止状态，自动机停机，即到达了接受状态，表明这个字符串能被 DNA 自动机接受。能够被 DNA 自动机识别的字符串都属于集合 $FZ_\rho(\Sigma)$，这些串中的字母属于字母表 Σ，并且上链和下链对应位置上的字母符合互补关系 ρ。

定理 3 终止状态的单字母 DNA 有限自动机能够识别正规语言。

证明：设接受正规语言 $L\in REG$ 的有限自动机 $M=(\Sigma,$

(下转第 187 页)

算法的时间复杂度。

4 实验结果及分析

我们选用 UCI 机器学习数据库中的 7 个数据库在 PC 机(Intel-Pentium 2GHz, 256MB RAM, Win2000 Professional, Microsoft Access)上进行实验,分别采用文[5](简称算法 a)和本文中的约简算法 2 进行属性约简。实验结果如表 2 所示。

从表 2 可以看出,算法 2 在效率上相对于算法 a 有一定提高,这同该算法需要计算分明矩阵及其对应的最小简化的析取范式有直接关系。

表 2 约简算法比较

数据库名称	实例数	属性集合 P 中属性个数	属性集合 D 中属性个数	算法 a 执行时间(s)	算法 2 执行时间(s)
Postoperative Patient	90	8	1	0.406	0.032
Monk's Problems (1)	432	6	1	0.156	0.033
Monk's Problems (2)	432	6	1	47.906	0.703
Monk's Problems (3)	432	6	1	0.141	0.046
Hayes-Roth Database	132	5	1	0.984	0.076
Teaching Assistant Evaluation	151	5	1	1.578	0.078
BUPA liver disorders	345	6	1	19.672	3.031
Balance-scale database	625	4	1	29.469	0.546
Car Evaluation Database	1728	6	1	2912.501	8.481

结论 高效的约简算法是 Rough 集应用于知识发现的基础。因此,寻求快速的 Rough 集相对约简算法具有重要的意义。本文介绍了一种相对约简的计算方法,采用已知的约简 $RED_Q(U-\{x_0\},P)$ 计算约简 $RED_Q(U,P)$ 的思想,提出了一种相对约简的计算方法。由于该方法不用计算分明矩阵的中间环节,节省了空间和时间,提高了运行效率。理论分析和实验结果表明,该约简算法在效率上较现有的算法有显著提高。

参 考 文 献

1 Pawlak Z. Rough sets. International Journal of Computer and Information Science, 1982, 11(5): 341~356

2 刘清. Rough 集及 Rough 推理. 北京:科学出版社, 2001

3 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法. 北京:科学出版社, 2001

4 Wong S K M, Ziarko W. On optimal decision rules in decision tables. Bulletin of Polish Academy of Sciences, 1985(33): 693~696

5 Skowron A, Rauszer C. The discernibility matrics and function in information system. In: Slowinski R, ed. Intelligent Decision Support Handbook of Application and Advances of the Rough sets Theory. Dordreecht: Kluwer Academic Publishers, 1991. 331~362

(上接第 157 页)

Q, q_0, F, δ),构造终止状态的单字母 DNA 有限自动机 $M=(\Sigma, Q, q_0, F, \delta, P)$,其中:

$\rho=\{(a,a)\mid a\in\Sigma\}$;

$Q'=Q\cup\{q_f\}, q_f\notin Q$;

$$F(q,a)=\begin{cases}\{q_f\}, \delta(q,a)\cap F\neq\phi\\ \phi, \delta(q,a)\cap F=\phi\end{cases}, q\in Q, a\in\Sigma;$$

$$\delta'(q,\binom{a}{\varepsilon})=\delta(q,a)\cup F(q,a), q\in Q, a\in\Sigma;$$

$$\delta'(q_f,\binom{\varepsilon}{a})=\{q_f\}, a\in\Sigma;$$

$\delta'(q,\binom{x}{y})=\phi$,除上面之外的其它任何情况。

识别双链序列 $\begin{bmatrix}w_1\\ w_2\end{bmatrix}$,$M'$ 首先识别上链,这个识别过程同普通的有限自动机 M 一样,只是到了识别上链的最后一步时,有限自动机 M 到了终止状态,而 M' 进入了状态 q_f,接着从左向右依照 $\delta'(q_f,\binom{\varepsilon}{a})=\{q_f\}$ 识别双链序列的下链,这样就可以完成整个识别过程。 □

小结 本文介绍了根据 DNA 分子的粘接特性抽象出的 DNA 分子粘接计算模型的文法结构及其计算方法,证明了不同 DNA 粘接计算模型的计算能力,给出了 DNA 有限自动机的结构,进一步从生物学的角度证明了 DNA 有限自动机与正规文法的等价性。

参 考 文 献

1 Adleman L M. Molecular computation of solutions to combinational problems [J]. Science, 1994, 266(5187): 1021~1023

2 Adleman L M. On Constructing a Molecular Computer [J]. In: DNA based computers, American Mathematical Society, 1996 (27): 1~21

3 Kari L, et al. DNA computing, sticker system, and universality. Acta Informatica, 1998, 35(5): 401~420

4 Kari L, et al. At the Crossroads of DNA Computing and Formal Languages: Characterizing Recursively Enumerable Languages by Insertion-Deletion Systems [A], In: Proc. of 3rd DIMACS Workshop on DNA- Based Computers, Philadel-phia, 1997. 318~333

5 Paun G H, Rozenberg G. sticker systems. Theoretical Computer Sci, 1998, 205

6 邹海明,等. 形式语言、自动机和语法分析[M]. 武昌:华中工学院出版社, 1985

7 Sipser M. 计算理论导引[M]. 北京:机械工业出版社, 2000

8 Roweis S, Erik W, et al. A sticker based model for DNA computation [J]. Journal of Computational Biology, 1998, 5(4): 615~629

9 Cox J C, Cohen D S, Ellington A D. The complexities of DNA computation [J]. Trends in Biotechnology, 1999, 17(4): 151~154

10 Gatterdam R W. Splicing systems and regularity [J]. Intl. Journal of Computer Mathematics, 1998(31): 63~67

11 Garzon M H, et al. Biomolecular Computing and Programming [J]. IEEE Trans. on Evolutionary Computation, 1999, 3(3): 236~250

12 Sakakibara Y. DNA computers: A new computing paradigm [J]. Journal of Photopolymer Science and Technology, 1998, 11(4): 681~686

13 Martin-Vide C, et al. University results for finite H systems and for Watson-Crick finite automata. Computing with Biomolecules, Springer, Berlin, 1998. 200~220