LENGUAJES DEFINIDOS POR SISTEMAS DE LINDENMAYER

Por Chistian Choffrut. Profesor de la Universidad de Paris VII.

INTRODUCCION

Presentamos cuatro temas ligados a la teoría de los lenguajes formales y más generalmente a la combinatoria del monoide libre y que utilizaran de manera central la noción de morfismo de monoides libres. Hemos procurado destacar los resultados más importantes y la problemática de cada tema. En vez de una presentación muy formal, hemos preferido ilustrar nuestro discurso con numérosos ejemplos. El lector puede referirse útilmente al panorama dado por Culik en (4) de los mismos temas.

Dado un conjunto X, denotaremos por X^* el monoide libre generado por X i.e. el conjunto de todas las secuencias finitas de elementos de X, munido del producto de concatenación, y por 1 el elemento neutro (o secuencia vacía). Un morfismo h de un monoide libre X^* en otro A^* es definido por una regla que a cada elemento (o letra) de X, asocía una palabra de A^*. La imagen de una palabra cualquiera w X^* por h es obtenida por sustitución de cada ocurrencia de las letras $x \in X$ en w, por la palabra correspondiente de A^*.

Por ejemplo, si $X= \{x,y,z\}$, $A=\{a,b\}$ y $h(a)= ab$ $h(y)= aab$ $h(z)= ba$, tendremos $h(xxyzy)=ab.ab.aab.ba.aab= ababa^2b^2a^3b$

El número de ocurrencias de una letra $x \in X$ en una palabra $w \in X^*$ será denotado por $|w|_x$ y la longitud de w por

$|w| = \Sigma_{x\in X} |w|_x$.

En regla general, los monoides libres que consideraremos a continuación son finitamente generados (i.e. X y A son finitos).

I.- ECUACIONES EN EL MONOIDE-LIBRE (véase (3))

Empezamos por tratar algunos ejemplos que nos darán la intuición de lo que entendemos por "ecuación".

Consideremos el problema siguiente ¿es posible caracterizar las palabras x,y,z de X^* que satisfagan la igualdad: xy=yz?

La respuesta a esa cuestión es dada por la:

Proposición 1

Sean $x,y,z \in X^*$ tres palabras y supongamos $y \neq 1$. Las dos condiciones siguientes son equivalentes:

1) $xy=yz$
2) existen $u,v \in X^*$ y $r \geqslant o$ tales que:
$x=uv$, $z=vu$ & $y=(uv)^r u$

(la implicación 2) $\Longrightarrow$ 1) es obvia. En el sentido 1) $\Longrightarrow$ 2) basta averiguar que $xy=yz$ implica $x^n y = y z^n$ para todo entero $n>o$, y en particular para el menor n tal que $|x^{n-1}| \leqslant y < |x^n|$
Luego se identifican los factores izquierdos de longitud $|y|$ en los dos miembros de la igualdad).

Observación: Dos elementos x,z de un grupo G son conjugados si existe un tercer elemento y G tal que $y^{-1}xy=z$. Esa condición equivale a: $x=(xy).y^{-1}$ & $z=y^{-1}.(xy)$ parecida a la condición 2) de la Proposición donde xy es u & y^{-1} es v. Por extensión las palabras x,y que satisfacen una de las dos condiciones 1) ó 2) son llamadas conjugadas.

Otro ejemplo:

Proposición 2

Sean $x,y \in X^*$ dos palabras que satisfacen una igualdad de la forma:

$$xz_1 \dots z_p = yt_1 \dots t_q$$

donde $p,q \geqslant o$ y $z_1, \dots, z_p, t_1, \dots t_p$ son la palabra x ó y. Entonces existen una palabra $w \in X^*$ y enteros n,m tales que: $x=w^n$ & $y=w^m$

(ese resultado se comprueba por inducion sobre el entero $|x|+|y|$).

Eso nos lleva a la definición siguiente:

Sea X un alfabeto (de incognitas). Una ecuación sobre X es un par (e,e') de palabras. Una solución de (e,e') es un morfismo h de X^* en un monoide libre arbitrario A^* tal que:

$$h(e) = h(e')$$

En la primera Proposición hemos considerado la ecuación $xy=yz$ con tres incognitas, y hemos visto que toda solución $h: X^* \longrightarrow A^*$ es definida por

$$h(x) = uv \qquad h(y) = (uv)^r u \qquad h(z) = vu$$

por algunos u,v A^* $r \geqslant o$.

En la segunda Proposición hemos considerado ecuaciones arbitrarias en dos incognitas x&y. Hemos visto que una solución $h : X^* \longrightarrow A^*$ es definida por: $h(z)=w^n$ & $h(y)=w^m$ por algún w en A^* y $n,m\geqslant o$.

El primer problema planteado por las ecuaciones es determinar expresiones o formulas describiendo el conjunto de todas las soluciones (es posible en los dos ejemplos que hemos tratado pero no es en el caso general, mucho por el contrario ...).

El segundo problema es determinar el número de "parámetros útiles", vale decir, formalizar la intuición que tenemos que nos hace pensar que las soluciones de la ecuación $xy=yz$ dependen de los dos parametros u,v y que las de las ecuaciones de la segunda Proposición, dependen del único parametro w. Eso nos lleva a la noción de "rango" de una solución.

Definición

El rango de una solución dada $h:X^* \longrightarrow A^*$ de la ecuación (e,e') es el menor entero $n>o$ tal que existe una solución $g:X^*\to B^*$ y un morfismo $\Theta:B^* \longrightarrow A^*$ satisfaciendo: 1) $h=\Theta g$, 2) $\Theta(b)\neq 1$ para todo $b\in B$ y 3) cardinalidad $B=n$.

El resultado principal es que el rango de una solución de una ecuación es menor que el número de incognitas. Una demostración elegante de ello puede ser encontrada en (2).

Proposición 3

Sea $(e,e')\in X^*xX^*$ una ecuación con $e\neq e'$, $h:X^* \longrightarrow A^*$ una solución y $r(h)$ su rango.
Se tiene: $r(h) \leqslant |X| - 1$

Como consecuencia, conseguimos una demostración alternativa de la Proposición 2.

II.- CONJUNTOS DE IGUALDAD

El conjunto de igualdad de dos morfismos $h,g:X^* \longrightarrow A^*$ es el conjunto de las palabras sobre las cuales ellos coinciden. Formalmente es definido por:

$$E(h,g) = \{ w\in X^* \mid h(w)=g(w) \}$$

Ejemplo 1: $X= \{x,y\}$ $\quad A= \{t\}$
$h(x)=t^3 \quad h(y)=t^2$
$g(x)=t \quad g(y)=t^2$

Entonces: $E(h,g)= \{w\in X^* \mid |w|_x = |w|_y\}$ es el lenguaje Dyck.

Ejemplo 2: $X= \{x,y,z\}$ $\quad A= \{a,b\}$
$h(x)=abaab \quad h(g)=aabaab \quad h(z)=a$

$g(x)=g(z)=aba \quad g(y)=(aba)^2$

Entonces: $E(h,g)=(xy^*z)^*$

Ejemplo 3: $X=A=\{x,y,z\}$

$h(x)=h(z)=1 \quad h(y)=xyxz$

$g(x)=x \quad g(y)=y \quad g(z)=z$

Entonces: $E(h,g)=(xyxz)^*$

Se observará que el famoso Problema de Correspondencia de Post (P.C.P.) consiste en preguntar si el conjunto de igualdad de dos morfismos dados es vacío.

En efecto sea $n>o$ un entero, A^* un monoide libre y $(u_1,v_1),\ldots,(u_n,v_n)$ n pares de palabras de A^*. Se pregunta si existe un algoritmo que resuelve el problema de saber si existe una secuencia $i_1,\ldots,i_r$ $(1\leqslant i_k\leqslant n$ para $1\leqslant k\leqslant r)$ tal que:

$$(1) \quad u_{i_1}\ldots u_{i_r} = v_{i_1}\ldots v_{i_r}$$

Pongamos $X=\{i \mid 1\leqslant i\leqslant n\}$ y definamos $h,g\colon X^*\longrightarrow A^*$ por:

$$h(i)=u_i \qquad g(i)=v_i \qquad 1<i<n$$

Entonces (1) equivale a:

$$h(i_1\ldots i_r) = g(i_1\ldots i_r)$$

Es conocido que el PCP es indecidible. Recientemente se mostró que sigue siendo indecidible para todo $n\geqslant 10$ fijo. Ese límite ha sido bajado a 9 por Pansiot (14) que da la referencia del resultado anterior. Para los valores menores, no se sabe, excepto en el caso n=2 (Ehrenfeucht & Rozenberg mostraron que es decidible (9)) que sorprendentemente surgió como un problema muy dificil.

La clase de lenguajes obtenidos como conjunto de igualdad es muy amplia. Culik ha mostrado lo siguiente en (5):

Teorema 1

Para todo lenguaje $R\subseteq X^*$ recursivamente enumerable, existen morfismos $h,g\colon A^*\longrightarrow B^*$ y $f\colon A^*\longrightarrow X^*$ tales que $R=f(e(k,g))$ donde $e(h,g)$ es el menor $Y\subseteq A^*$ satisfaciendo $Y^*=E(h,g)$.

En ciertos casos es posible precisar a que familia de lenguajes pertenece un conjunto de igualdad.

Teorema 2 (12)

Sea $h\colon X^*\longrightarrow A^*$ un morfismo y $g\colon X^*\longrightarrow X^*$ el morfismo identidad ($g(x)=x$ para todo $x\in X$). Entonces:

$E(h,g)=Y^*$ donde y es un subconjunto de X^* de cardinalidad menor o igual a X.

Diremos que un morfismo $h:X^* \longrightarrow A^*$ es <u>elemental</u> si para toda factorización $h=h_2h_1$ donde $h_1:X^* \longrightarrow B^*$ y $h_2:B^* \longrightarrow A^*$, se tiene $|B|>|A|$ (por ejemplo $X=A=\{x,y\}$ $h(x)=xy$ $h(y)=y$). El teorema siguiente es válido bajo condiciones más generales.

<u>Teorema 3</u> (9)

Sean $h,g:X^* \longrightarrow A^*$ dos morfismos, uno de los cuales es elemental. Entonces $E(h,g)$ es un lenguaje racional.

Diremos que un morfismo $h:X^* \longrightarrow A^*$ es cíclico si existe una palabra u A^* tal que para todo $x \in X$ existe un $p \geqslant o$ con $h(x)=u^p$ (es el caso de h en el ejemplo 1 y de g en el ejemplo 2).

Se conjetura (7) que cuando $X=\{x,y\}$ i.e. X consiste en sólo dos letras, $E(h,g)=Y^*$ donde y es finito, si al menos uno de los morfismos no es cíclico. Este problema aunque probablemente en vía de ser resuelto, está todavía abierto.

III. <u>SISTEMAS DE LINDENMAYER</u>

Un DOL es definido por un alfabeto X, un morfismo $h:X^* \rightarrow X^*$ y una palabra (llamada <u>axioma</u>) w X^*. Un tal DOL se denota por:

$$H=\{(X,h,w)\}$$

El lenguaje generado por H es el subconjunto

$$L(H)=\{h^n(w) \mid n \geqslant o\}$$

y la secuencia asociada por H es la secuencia de las palabras $w,h(w),\ldots,h^n(w),\ldots$

Ejemplo 1: $X=\{x_1,x_2y_1,y_2,z_1,z_2\}$ $w=x_2y_2z_2$

$$\begin{array}{ll} h(x_1)=x_1 & h(x_2)=x_1x_2 \\ h(y_1)=y_1 & h(y_2)=y_1y_2 \\ h(z_1)=z_1 & h(z_2)=z_1z_2 \end{array}$$

Entonces $h^n(w)=x_1^{n-1}x_2y_1^{n-1}y_2z_1^{n-1}z_2$

Ejemplo 2: $X=\{x,y,z,t\}$

$$h(x)=xt \quad h(y)=zy \quad h(z)=t \quad h(t)=z \quad w=xy$$

Entonces $h^n(w)=x(tz)^ny$

Ejemplo 3: $X = \{x,y,z\}$

$h(x)=x \quad h(y)=yxy \quad h(z)=yxz \quad w=xyxz$

Entonces $h^n(w)=(xy)^{p_n}x\ z$ donde p_n satisface una recurrencia lineal.

Ejemplo 4: $X = \{x,y,z\}$

$g(x)=xyx \quad g(y)=y \quad g(z)=z \quad w=xy\ xz$

Entonces $h^n(w)=(xy)^{q_n}x\ z$ donde q_n satisface una recurrencia lineal.

La palabra DOL esta compuesta de iniciales. La "L" se refiere a Lindenmayer, la "D" a determinista en el sentido que la regla de sustitución o de reescritura tal como fue definida en la introducción, es determinista (cada letra se reescribe en una palabra determinada) y la "O" se refiere al entero cero en el sentido que la sustitución de cada letra es independiente de la letras vecinas. Los demás sistemas de Lindenmayer que no trataremos aquí se obtienen aflojando esas condiciones y otras (como por ejemplo, dividir el alfabeto X en dos: terminal y no terminal.

Los problemas planteados por los DOL son de tipo decidibilidad (por un H dado ¿es decidible si L(H) es finito? o por un H y una palabra $u \in X^*$ dados ¿es decidible si u L(H)?). Sin lugar a duda el problema que más ha llamado la atención es el problema si se puede decidir si dos D0l $G=(X(may),g,w)$ y $H=(X,h,w)$ son equivalentes i.e. si ellos definen una misma secuencia: $h^n(w)=g^n(w)$ para todo $n \geq o$ (como en los ejemplos 3 y 4).

Se ha mostrado de dos maneras diferentes que el problema es decidible ((6) & (9)). En esa segunda versión se introduce la noción fundamental de morfismo elemental.

IV. CONJUNTOS DE PRUEBA

Sea $L \subseteq X^*$ un conjunto cualquiera. Quisieramos poder verificar si dos morfismos dados $h,g:X^* \to A^*$ coinciden sobre L (i.e. $L \subseteq E(h,g)$). Lo más natural es esperar que existe un número finito de palabras $w_1,w_2,\ldots,w_r$ en L, sobre las cuales basta averiguar que h y g coinciden. Eso nos lleva a la:

Definición

Sea $L \subseteq X^*$ y L_o un subconjunto finito de L. Diremos que L_o es un conjunto de prueba de L si para todos $h,g:X^* \to A^*$ se tiene:

$h(w)=g(w)$ para todo $w \in L_0$ implica $h(w)=g(w)$ para todo $w \in L$

Ejemplo 1: Sea $X=\{x,y\}$ y $L \in X^*$ tal que existen dos palabras w_1,w_2 L satisfaciendo:

$$|w_1|_x|w_2|_y \neq |w_1|_y|w_2|_x$$

Para todos $h,g:X^* \longrightarrow A^*$ que coinciden sobre L y en particular sobre w_1 y w_2 tenemos $|h(x)|=|g(x)|$ y $|h(y)|=|g(y)|$ y luego $h(x)=g(x)$ y $h(y)=g(y)$. Con esa observación queda claro que w_1,w_2 es un conjunto de prueba de L.

Ejemplo 2: $X=\{x,y\}$ y $L=\{x^ny^n \mid n \geqslant o\}$. Se puede verificar que $\{xy,x^2y^2\}$ es un conjunto de prueba de L.

Ehrenfencht ha conjeturado (13) que todo lenguaje L posee un conjunto de prueba. Este problema ha sido resuelto por dos vías diferentes (8) y (11) en el caso de un alfabeto binario (X= x,y}). El segundo trabajo muestra que siempre existe un conjunto de prueba con tres o menos elementos. El problema general (X tiene más de dos elementos) esta todavía abierto.

Por ciertas familias de lenguajes (los libres de contexto por ejemplo (1)) se ha mostrado que existen conjunto de prueba efectivamente constructibles. Es interesante observar que si fueramos capaces de construir efectivamente un conjunto de prueba para cada DOL, tendriamos una solución nueva al problema de equivalencia de los DOL.

REFERENCIAS

(1) J.ALBERT, K.CULIK & J.KARHUMAKI, Test sets for context free languages and algebraic systems of equations, Research report (5-81-16, Department of Computer Science, University of Waterloo, Canada, 1981.

(2) J.BERSTEL, D.PERRIN, J.F.PERROT & A.RESTIVO, Sur le theoreme du defaut, J. of Algebra, 60, (1979), 169-180

(3) C.CHOFFRUT, Capitulo IX del libro colectivo "Combinatories on words" Addison Wesley, (1982).

(4) K.CULIK, Homomophisms: Decidability, Equality and Test Sets, Comunicacion al "International Symposium on Formal Language Theory, Santa Barbara, California, December 10-14, 1979.

(5) K.CULIK, A purely homomorphic characterization of recursively enumerable sets, JACM vol.6 (1979), 345-350.

(6) K CULIK & J.FRIS, The decidability of the equivalence problem

for DOL-systems, Inf. and Control, 35, (1977), 20-39.

(7) K.CULIK & J.KARHUMAKI, On the equality sets for homomorphisms on free monoids two generators, RAIRO, Informatique Théorique, 14, (1980), 349-369.

(8) K.CULIK & A.SALOMAA, Test sets and checking words for homomorfism equivalence, J.Comput System Scí., 1980.

(9) A.EHRENFEUCHT & G.ROZENBERG, Elementary morphisms and a solution of the DOL sequence equivalence problem, Theoret Comput. Sci., 7, (1978), 169-183.

(10) A.EHRENFERUCHT, J.KARHUMAKI & G.ROZENBERG, The (generalized) Post correspondence problem with lists consisting of two words is decidable, Theoret. Comput. Sci., por salir.

(11) A.EHRENFERUCHT, J.KARHUMAKI & G.ROSENBERG, On binary equality sets and a solution to the Ehren Feucht conjecture in the binary case, sin publicar.

(12) G.T.HERMAN & A.WALKER, Context-free languages in biological systems, Internat. J. Comput. Math, 4, (1975), 369-391.

(13) M.KARPINSKI, ed., New Scottish Book of Problems, en preparación.

(14) PANSIOT, algo asi como "A note on PCP", Information Processing Letters, (1981),

(15) G.ROZENBERG & A.SALOMAA, "The Mathematical Theory of L Systems, Academic Press, 1980.

* * * * * *