151. *Sur une classification des langages d'états finis*

Par Masami ITO
Université Kyoto-Sangyo

(Comm. by Kinjirô KUNUGI, M. J. A., Nov. 13, 1972)

1. Introduction. Nous nous rappelons que le nombre des classes de E-équivalence d'un langage d'états finis est fini. Il serait intéressant de voir comment des classes de E-équivalence sont distribuées dans un langage d'états finis. Dans ce mémoire, nous essayons de classifier des langages d'états finis de ce point de vue et d'appliquer cette classification à la théorie de l'espace contextuel [3].

Quant aux notions et aux symboles que nous employons dans ce mémoire, nous voudrions demander aux lecteurs de voir M. Ito [1] et [2].

2. Langage d'états finis de θ-type. Pour un langage d'états finis $\mathcal{L}=(A, L)$, nous introduisons une paire des nombres $\langle p, q\rangle$ et nous l'appelons *la paire d'indices* du langage $\mathcal{L}$. Les nombres p, q s'y présentent respectivement comme les nombres des classes de E-équivalence $\{S(x)\,;\, x\in L\}$, $\{S(y)\,;\, y\in A^+-L\}$.

De plus, nous appelons $\mathcal{L}$ *le langage de θ-type*, si $p/(p+q)\geq\theta$ $(0\leq\theta\leq 1)$.

3. Distribution des classes de E-équivalence. Nous étudions dans cette section un état de la distribution des classes de E-équivalence d'un langage d'états finis, ayant le diamètre n. Dans ce but, nous introduisons les symboles $\mathfrak{S}_n(\theta)$ et $\mathfrak{T}_n(\theta_1, \theta_2)$. $\mathfrak{S}_n(\theta)$ se présente comme l'ensemble de tous les langages de θ-type, qui ont le diamètre n. Nous posons $\mathfrak{T}_n(\theta_1,\theta_2)=\mathfrak{S}_n(\theta_1)-\mathfrak{S}_n(\theta_2)$ $(0\leq\theta_1<\theta_2\leq 1)$.

Nous avons le théorème suivant:

Théorème 1. *Soient n un nombre naturel et θ_1, θ_2 deux nombres quelconques tels que $0\leq\theta_1<\theta_2\leq 1$. Nous avons alors $\mathfrak{T}_n(\theta_1,\theta_2)\neq\emptyset$.*

Démonstration. *En cas de $n=1$.* Posons $A=\{a\}$ et $L=\{a, a^2, a^3, \cdots, a^{r-2}, a^{r-1}, a^{r+s}, a^{r+s+1}, a^{r+s+2}, \cdots\}$; r, s étant deux nombres naturels quelconques. Nous pouvons alors constater aisément que le langage $\mathcal{L}=(A, L)$ a la paire d'indices $\langle r, s\rangle$ et $d(\mathcal{L})=1$. La démonstration sera perfectionnée par le fait que nous pouvons donner n'importe quelle valeur pour les nombres r et s.

En cas de $n>1$. Posons $A=\{a, a_1, a_2, \cdots, a_{n-1}\}$ et $L=\{a, a^2, a^3, \cdots, a^{r-2}, a^{r-1}, a^{r+s}, a^{r+s+1}, \cdots\}\cup\{aa_1^p\,;\, p=1, 2, 3, \cdots\}\cup\{aa_1^{p_1}a_2^{p_2}\,;\, p_1, p_2=1, 2, 3, \cdots\}\cup\cdots\cup\{aa_1^{p_1}a_2^{p_2}a_3^{p_3}\cdots a_{n-1}^{p_{n-1}}\,;\, p_1, p_2, p_3, \cdots, p_{n-1}=1, 2, 3, \cdots\}$; r, s étant deux nombres naturels quelconques. Nous pouvons montrer que le langage $\mathcal{L}=(A, L)$ a la paire d'indices $\langle r+n-1, s+n(n-1)/2+1\rangle$

et $d(\mathcal{L})=n$ par induction sur n. La démonstration sera aussi perfectionnée par le fait que nous pouvons donner n'importe quelle valeur pour les nombres r et s.

4. Nombre des restrictions maximales d'un langage. Dans le mémoire [2], nous avons démontré qu'une restriction maximale d'un langage d'états finis est aussi un langage d'états finis. Dans cette section, nous considérons le problème suivant:

Combien de restrictions maximales (θ-type et leur diamètre est égal à n) d'un langage d'états finis existe-t-il?

Nous donnerons un majorant pour le nombre de telles restrictions maximales. Pour nous préparer à répondre à ce problème, nous introduisons la fonction $I_p(r)$:

Définition de $I_p(r)$. Soit $A=\{a_1, a_2, a_3, \cdots, a_p\}$ un ensemble. Considérons $\mathfrak{A}_r(1\leq r\leq p)$ une famille de sous-ensembles de A, qui comporte les deux propriétés suivantes:

(1) Pour $B\in\mathfrak{A}_r$, B se compose d'éléments dont le nombre est supérieur ou égal à r,

(2) Pour $B_i\in\mathfrak{A}_r(i=1,2,3,\cdots,q,\ B_i\neq B_j(i\neq j),\ q>1)$, nous posons $B=\bigcup_{i=1}^{q} B_i$. Pour n'importe quel B qui est construit à la manière ci-dessus, il existe deux éléments $a, a'\in B$ tels que a, a' ne soient pas contenus en même temps dans un même ensemble $B_i(i=1,2,3,\cdots,q)$.

Chaque $\mathfrak{A}_r$, qui a les propriétés (1) et (2), contient ainsi comme ses éléments quelques sous-ensembles de A. Lorsque l'on fixe r, il existe donc au moins un $\mathfrak{A}_r$ qui contient des éléments dont le nombre est maximum. Nous posons $I_p(r)$ le nombre des éléments de cette famille $\mathfrak{A}_r$.

Nous ne calculons pas $I_p(r)$ dans ce mémoire. Nous nous bornons à remarquer que $I_p(r)\leq\min(3^{p/2}, 2^{p-r})$ (pour $r\geq p/2$, nous pouvons montrer que $I_p(r)=2^{p-r}$).

Rappelons nous le fait suivant [2]:

Soit $\mathcal{M}=(B, M)$ une restriction maximale d'un langage $\mathcal{L}=(A, L)$. Nous avons alors $M=\bigcup_{x\in M} S(x)$.

Nous utilisons ce théorème. Supposons désormais que toutes les restrictions maximales d'un langage $\mathcal{L}=(A, L)$ aient des formes $(A, \cdot)$.

Proposition. *Soit p le premier terme de la paire d'indices d'un langage d'états finis $\mathcal{L}=(A, L)$. Alors, le nombre des restrictions maximales (leur diamètre est égal à n et leur premier terme de la paire d'indices est supérieur ou égal à r) du langage $\mathcal{L}$ n'est pas supérieur à $I_p(r)$.*

Démonstration. Posons $\{S(x)\,;\,x\in L\}=\{S_1, S_2, S_3, \cdots, S_p\}$. $\{S(x)\,;\,x\in A^+\}$ se présente comme l'ensemble des classes de E-équivalence de $\mathcal{L}$. Soit $\mathcal{N}=(A, N)$ une restriction maximale de $\mathcal{L}$, dont le diamètre est égal à n et dont le premier terme de la paire d'indices est supérieur ou égal à r. Remarquons que l'on peut représenter $N=\bigcup_{i=1}^{l} S_{k(i)}(S_{k(i)} \neq S_{k(j)}$, si $i\neq j$), où $1\leq k(i)\leq p(i=1,2,3,\cdots,l)$.

Posons $\mathfrak{N}=\{S_{k(1)},S_{k(2)},\cdots,S_{k(l)}\}$. A savoir $\mathfrak{N}$ est une expression de $\mathcal{N}$. Remarquons que des restrictions maximales différentes ont respectivement des expressions différentes. Nous pouvons ainsi donner une expression pour une restriction maximale.

Soit $\mathfrak{B}_r$ l'ensemble des expressions pour les restrictions maximales de $\mathcal{L}$, où leur diamètre est égal à n et leur premier terme de la paire d'indices est supérieur ou égal à r. $\mathfrak{B}_r$ est alors une famille de sous-ensembles de $\mathfrak{L}=\{S_1,S_2,S_3,\cdots,S_p\}$. Si nous pouvons montrer que $\mathfrak{B}_r$ a les propriétés (1), (2) exposées à l'occasion de la définition de $I_p(r)$, nous pouvons arriver à la conclusion de la proposition.

Propriété (1). Soit $\mathcal{N}=(A,N)$ une restriction maximale de $\mathcal{L}$, dont le diamètre est égal à n et dont le premier terme de la paire d'indices est supérieur ou égal à r. Posons $\{T(x)\,;\,x\in A^+\}$ l'ensemble des classes de E-équivalence de $\mathcal{N}$ et $\mathfrak{N}=\{S_{k(1)},S_{k(2)},S_{k(3)},\cdots,S_{k(l)}\}$ l'expression de $\mathcal{N}$.

Du fait que $S(x)\subseteq T(x)\ (x\in A^+)$ ([2], p. 100), nous concluons que le nombre cardinal de l'ensemble $\{S(x)\,;\,x\in N\}$ n'est pas inférieur à celui de $\{T(x)\,;\,x\in N\}$. En conséquence, $\mathfrak{N}$ contient des éléments dont le nombre est supérieur ou égal à r.

Propriété (2). Procédons par réduction à l'absurde.

Supposons que $\mathfrak{B}_r$ n'ait pas la propriété (2). Dans ce cas, il y a $q(q>1)$ éléments $\mathfrak{N}_1,\mathfrak{N}_2,\mathfrak{N}_3,\cdots,\mathfrak{N}_q$ de $\mathfrak{B}_r$ tels que, pour chaque paire $S,S'\in\mathfrak{N}$ (où $\mathfrak{N}=\bigcup_{j=1}^{q}\mathfrak{N}_j$), il existe quelque $\mathfrak{N}_j(1\leq j\leq q)$ contenant S,S'. Considérons un langage $\mathcal{N}=(A,N)$, où $N=\bigcup_{S\in\mathfrak{N}}S$. Voyons que $\mathcal{N}$ est une restriction de $\mathcal{L}$, ayant le diamètre n.

Soit x,y deux éléments de N. Nous montrons que $\operatorname{dis}(x,y)_{\mathcal{N}} =\operatorname{dis}(x,y)_{\mathcal{L}}$. D'après $x,y\in N$, il existe deux éléments S,S' de $\mathfrak{N}$ tels que $x\in S, y\in S'$. De la remarque donnée ci-dessus, nous pouvons trouver quelque $\mathfrak{N}_j(1\leq j\leq q)$ tel que $S,S'\in\mathfrak{N}_j$. Posons $\mathcal{N}_j=(A,N_j)$, où $N_j=\bigcup_{S\in\mathfrak{N}_j}S$. Remarquons que $\mathcal{N}_j$ est une restriction maximale de $\mathcal{L}$, ayant le diamètre n. De ce fait, nous avons $\operatorname{dis}(x,y)_{\mathcal{N}_j}=\operatorname{dis}(x,y)_{\mathcal{L}}$. D'après $N_j\subset N\subseteq L$, nous avons $\operatorname{dis}(x,y)_{\mathcal{L}}\leq\operatorname{dis}(x,y)_{\mathcal{N}}\leq\operatorname{dis}(x,y)_{\mathcal{N}_j}$.

Par conséquent, nous avons $\operatorname{dis}(x,y)_{\mathcal{N}}=\operatorname{dis}(x,y)_{\mathcal{L}}$.

Montrons que $d(\mathcal{N})=n$. Soient x,y deux éléments de N. Nous pouvons trouver alors quelque $\mathcal{N}_j(1\leq j\leq q)$ d'après le raisonnement exposé ci-dessus. Il s'ensuit que $\operatorname{dis}(x,y)_{\mathcal{N}}=\operatorname{dis}(x,y)_{\mathcal{N}_j}\leq d(\mathcal{N}_j)$. En conséquence, nous déduisons que $d(\mathcal{N})\leq n$.

Posons maintenant $N_j=\bigcup_{S\in\mathfrak{N}_j}S$ et $\mathcal{N}_j=(A,N_j)$ pour tout $j(1\leq j\leq q)$. Remarquons d'abord que $\mathcal{N}_j$ est une restriction maximale de $\mathcal{L}$, ayant le diamètre n. Il est évident que nous pouvons choisir deux éléments x, y parmi des éléments de N_j tels que dis $(x, y)_{\mathcal{N}_j}=$ dis $(x, y)_{\mathcal{L}}=n$, pour chaque N_j. D'après $N_j \subset N$ et dis $(w, z)_{\mathcal{N}}=$ dis $(w, z)_{\mathcal{L}}$ $(w, z \in N)$, nous concluons que dis $(x, y)_{\mathcal{N}}=$ dis $(x, y)_{\mathcal{L}}=n$. Par conséquent, nous déduisons que $d(\mathcal{N}) \geq n$. Il en résulte enfin que $d(\mathcal{N})=n$.

Nous voyons ainsi que le langage $\mathcal{N}$ est une restriction de $\mathcal{L}$, ayant le diamètre n. Mais, cette affirmation est en contradiction avec le fait que $\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_q$ sont des restrictions maximales de $\mathcal{L}$, qui ont le diamètre n. Pour cette raison, $\mathfrak{B}_r$ doit avoir la propriété (2).

Théorème 2. *Soit p le premier terme de la paire d'indices d'un langage d'états finis $\mathcal{L}=(A, L)$. Alors, le nombre des restrictions maximales (θ-type et leur diamètre est égal à n) de $\mathcal{L}$ n'est pas supérieur à $I_p([\theta(n+1)])$. [ ] s'y présente comme le symbole de Gauss.*

Démonstration. Soit $\mathcal{M}=(A, M)$ une restriction maximale de θ-type de $\mathcal{L}$, dont le diamètre est égal à n. Posons $\{T(x) ; x \in A^+\}$ l'ensemble des classes de E-équivalence de $\mathcal{M}$. Il n'est pas difficile de constater que le nombre cardinal de cet ensemble est supérieur ou égal à $n+1$. Nous concluons ainsi que le premier terme de la paire d'indices de $\mathcal{M}$ est supérieur ou égal à $[\theta(n+1)]$. En vertu de la proposition, nous pouvons arriver à la conclusion du théorème.

Références

[1] M. Ito: Sur l'extension et la restriction d'un langage associé à l'espace contextuel. Proc. Japan Acad., **48**, 94–97 (1972).

[2] ——: Sur la restriction maximale d'un langage. Proc. Japan Acad., **48**, 98–100 (1972).

[3] S. Marcus: Introduction mathématique à la linguistique structurale. Dunod (1967).