SOME RESULTS OF REGULAR HYPER - LANGUAGE

Dang Huy Ruan
Faculty of Mathematics - College of Natural Sciences - VNU
Phung Van On
Vietnam Maritime University

I. INTRODUCTION

The class of regular languages on finite words has been studied in the theory of formal languages. In [3,4], the lower and supper limits for the complexity of finite automaton recognizing regular expressions and generated schemata is shown. The approach to language on finite words is infinite in [2]. In this paper, we shall deal with the regular hyper- language, regular language on infinite words.

We introduce the definition of hyper-words, their limits, operation on language and hyper - language, hyper - iterations of a language, regular hyper-language, hyper- sources, hyper-automaton recognizing languages by limits of hyper-words of states. We obtain some basic results on recognization, closeness with some operations.

II. ELEMENTARY CONCEPTS

1. Hyper-word: Let $\sum = \{a_1, a_2...a_n\}$ be an alphabet. An infinite sequence $\alpha = a_{i_1}a_{i_2}...$ of characters in $\sum$ is called an infinite word or a hyper-word on $\sum$. The set of all hyper-words on $\sum$ is denoted by $\sum^\infty$ (we recall that $\sum^*$ is the set of all finite words on $\sum$).

The hyper-word $\alpha = a_{i_1}a_{i_2}...$ is called cyclic with τ period, beginning at the τ_1 place (τ, τ_1 - are positive integers) if for any integer $i, i \geq \tau_1 + 1$, we have $a_i = a_{i+\tau}$.

We write $\alpha.\beta.$ for the compound product of the finite word α with the hyper-word β.

2. Limit of hyper-word: Let $\alpha = a_{i_1}a_{i_2}...$ be a hyper- word on $\sum$. The set of all $a \in \sum$ for which there is an infinite sequence of indexes $j_1, j_2, j_3, ...$ such that $a_{j_k} = a$ with $k = 1, 2, 3, ...$ is called the limit of hyper-word α and is denoted by lim (α).

Comment: If $\sum$ is finite then all hyper-words $\alpha \in \sum^\infty$ have the limit, that means lim $(\alpha) \neq \emptyset$.

3. Hyper-language: A subset of $\sum^\infty$ is called a hyper- language on the alphabet $\sum$. We consider the following operations on the family of hyper-languages:

+ The compound product of a language M_1 with a hyper-language M_2 on the alphabet $\sum$ is a hyper-language on $\sum$ (denoted by $M_1.M_2$) which is defined by

$$M_1.M_2 = \{\alpha.\beta \mid \alpha \in M_1, \;\; \beta \in M_2\}.$$

+The union of two hyper-language M_1, M_2 on $\sum$ is a hyper- language $M_1 \cup M_2$ on $\sum$

$$M_1 \cup M_2 = \{\alpha \in \sum{}^{\infty} \mid \alpha \in M_1 \text{or} \alpha \in M_2\}.$$

+ The intersection of two hyper-language M_1, M_2 on $\sum$ is a hyper-language $M_1 \cap M_2)$ on $\sum$

$$M_1 \cap M_2 = \{\alpha \in \sum{}^{\infty} \mid \alpha \in M_1 \text{ and } \alpha \in M_2\}.$$

+ The subtraction of two hyper-language M_1, M_2 on $\sum$ is a hyper- language on $\sum$ (denoted by $M_1 \setminus M_2$)

$$M_1 \setminus M_2 = \{\alpha \in M^{\infty} \mid \alpha \in M_1 \text{and} \alpha \notin M_2\}.$$

+ The complement of a hyper-language M on $\sum$ is a hyper-language on $\sum$, denoted by $C(M)$

$$C(M) = \{\alpha \in \sum{}^{\infty} \mid \alpha \notin M\}.$$

+ The hyper-iteration of a language M on $\sum$ is M^{∞}, that is a hyper-language on $\sum$, defined by

$$M^{\infty} = \{x_1 x_2 ... \in \sum{}^{\infty} \mid x_i \in M, i \geq 1\}.$$

Note that the symbol M^{∞} is accordant with the symbol of hyper-language $\sum^{\infty}$ We have $\emptyset^{\infty} = \emptyset$ and for any hyper- language M, $\emptyset.M = \emptyset$.

The regular hyper language: We define the set of regular hyper-languages on $\sum$ as following:

Definition. *The set of all regular hyper-language on $\sum$ consists of*

a) *The elements R^{∞} where R is a regular language on $\sum$.*

b) *All compound products $R_1.R_2$ where R_1 is a regular language and R_2 is a regular hyper-language on $\sum$.*

c) *All unions $R_1 \cup R_2$ where R_1, R_2 are regular hyper-language on $\sum$.*

4. Hyper-source: A hyper- source on $\sum$ is a finite directed graph G on $\sum$ with the set of vertices S, the beginning v and the set of the ends $\{v_1, v_2, ... \; v_n\}$ such that for each bow, it is assigned to a character $a \in \sum$ (called main bow) or an empty word (denoted by ϵ, called empty bow).

A hyper-line in a hyper-source G is an infinite sequence $\pi : w_1, \rho_1, w_2, \rho_2, ...$ where w_i, $i = 1, 2, ...$ is a vertex of G, and ρ_i is a bow of G from w_i to w_{i+1}.

The hyper-line π generates a hyper-word $[\pi] = a_{i_1} a_{i_2} ...$ where, $a_{i_j} \in \sum$ assigned to bow $\rho_{i_j}, j = 1, 2, ...$ The vertex w_1 is called the beginning of π if there is an infinite set of indexes i such that $w = w_i$. The set of all limit points of π is denoted by lim (π).

Each hyper-source G generates a hyper-language (denoted by $\|G\|$) which contains all hyper-words $\alpha \in \sum^{\infty}$ such that each hyper-line having the beginning ν (which is the beginning of G) satisfies $[\pi] = \alpha$ and there is at least one end of G belonging to lim (π).

Comment: $\|G\| \neq \emptyset$ if and only if there is in G a right close line which contains at least one end and one main bow.

A hyper-source G is deterministic if it has no empty bow, and on two bows having the different beginning from one vertex must be assigned to two different characters .

5. Hyper-automaton: A quintuple $V = (\sum, Q, q_1, \varphi, F)$ is called a hyper-automaton, where:

$\sum$ is a finite set of symbols, called input alphabet;

$Q \neq \emptyset$, a finite set of states, $Q = \{q_1, ..., q_n\}$.

$q_1 \in Q$, an initial state;

$F \subseteq Q$, the set of terminal states;

$\varphi : Q \times \sum^{*} \to Q$ is a transitional function of V :

$$\varphi(q_i, a) = q_j,$$

$$\varphi(q, a_1 a_2 ... a_n) = \varphi(...\varphi(\varphi(q, a_1), a_2)...a_n)$$

The extension of φ is $\overline{\varphi} : Q \times \sum^{\infty} \to Q^{\infty}$ which assigns each hyper-word $\alpha = a_1 a_2 ...$ to the hyper-word of states $q_{i_1} a_{i_2}$:

$$\overline{\varphi}(q_1, \alpha) = \overline{\varphi}(q_1, a_1 a_2 ...) = \varphi(q_1, a_1)\varphi(q_1, a_1 a_2)... = q_{i_1} q_{i_2} ...$$

A hyper-word $\alpha \in \sum^{\infty}$ is recognized by the hyper- automaton V if lim $\overline{\varphi}(q_1, \alpha) \cap F \neq \emptyset$.

The set of hyper-words recognized by the hyper-automaton V, denoted by R, consists of all languages recognized by the hyper-automaton V : $R = \{\alpha \in \sum^{\infty} \mid \lim \overline{\varphi}(q_1, \alpha) \cap F \neq \emptyset\}$.

III. MAIN RESULTS

Theorem 1. *Any hyper-language R recognized by the hyper- automaton V is a regular hyper-language.*

To prove this theorem, we consider the following lemmas:

Lemma 1. *Let $R_{ij}; i = 0, 1, ..., n; j = 1, 2, ..., n$ be regular languages; $X_1, X_2, ..., X_n$ be the set of words satisfying*

$$X_1 = X_1 R_{11} \cup ... \cup X_n R_{n1} \cup R_{01},$$

....

$$X_n = X_1 R_{1n} \cup ... \cup X_n R_{nn} \cup R_{0n}$$

Then, $X_1, X_2, ..., X_n$ are regular languages. (see[2], p.92).

Lemma 2. *Let $V = (\sum, \{q_1, ..., q_n\}, \varphi, q_1, F)$ be a finite automaton. Let R^k_{ij} be the set of finite words leading V from the state q_i to the state q_j such that it does not pass any state q_l with $l > k$ (note that i, j may be greater than k.) In other words, $R^k_{ij} = \{w \in \sum^* | \varphi(q_i, w) = q_j$ and for all prefixes v of $w, v \neq w, v \neq \epsilon, \varphi(q_i, v) = q_l$ with $l \leq k\}$.*

Then R^k_{ij} is a regular language (see [1], p.29).

Proof of the Theorem 1:

Let $V = (\sum, Q, q_1, \varphi, F)$ with $Q = \{q_1, ..., q_n\}$ be an automaton recognized by the hyper-language $R = \{\alpha \in \sum^\infty \mid \lim \overline{\varphi}(q_1, \alpha) \cap F \neq \emptyset\}$. We shall construct a regular hyper-language M such that $M = R$.

Let $M_i = \{w \in \sum^* \mid w \neq \epsilon, \varphi(q_1, w) = q_i\}, i = 1, 2, ..., n$.

From the lemma 2, $R^n_{ij} = \{w \in \sum^* \mid \varphi(q_i, w) = q_j\}; i, j = 1, ..., n$ is regular and satisfies the following conditions:

$$M_1 = M_1 R^n_{11} \cup ... \cup M_n R^n_{n1} \cup R^n_{11}$$

....

$$M_n = M_1 R^n_{1n} \cup ... \cup M_n R^n_{nn} \cup R^n_{1n}$$

The lemma 1 says that $M_1, ..., M_n$ is regular.

The set $M = \bigvee_{qj \in F} M_j (R^n_{jj})^\infty$ is a hyper- language.

We must prove $M = R$.

1) $M \subseteq R$: Let $\alpha \in M, \alpha = a_1 a_2 a_3 ...$, we show that $\alpha \in R$.

For $\alpha \in M$, there are $r \in \{1, 2, ..., n\}$ such that $q_r \in F$; a natural number s that $\alpha_0 = a_1 ... a_s \in M_r$ and the hyper- word $a_{s+1} a_{s+2} ... \in (R^n_{rr})^\infty$. Therefore, there exists an infinite sequence of increasing indexes $i_1 = s + 1, i_2, i_3, ...$ such that $\alpha_j = a_{i_j} a_{i_j+1} ... a_{i_{j+1}-1} \in R^n_{rr}$ with $j = 1, 2, ...,$ i.e. , $\varphi(q_{i_j}, \alpha_j) = q_r$ and $q_r \in F$. That means $q_r \in \lim \overline{\varphi}(q_1, \alpha)$, i.e., $\lim \overline{\varphi}(q_1, \alpha) \cap F \neq \emptyset$, which implies $\alpha \in R$.

2) $R \subseteq M$. Let $\alpha \in R, \alpha = a_1 a_2 a_3 ...$, we must prove $\alpha \in M$.

Because $\lim \overline{\varphi}(q_1, \alpha) \cap F \neq \emptyset$, there exists $r \in \{1, 2, ..., n\}, q_r \in F$ such that $\varphi(q_1, a_1 ... a_j) = q_r$, for a set of increasing indexes j. Therefore, there exist s such that $\varphi(q_1, a_1 ... a_s) = q_r$, i.e.; $\alpha_1 ... \alpha_s \in M_r$ and a sequence of increasing indexes $i_1 = s + 1, i_2, i_3, ...$ that $\varphi(q_r, \alpha_j) = q_r$ with $\alpha_j = a_{i_j} a_{i_j+1} ... a_{i_{j+1}-1}$, i.e., $\alpha_j \in R^n_{rr}, j = 1, 2, ...$ That means $a_{s+1} a_{s+2} ... \in (R^n_{rr})^\infty$. Hence, $\alpha \in M_r (R^n_{rr})^\infty$, i.e., $\alpha \in M$.

Theorem 2. *For each regular hyper-language M on $\sum$, there always exists a deterministic automation V recognized by M.*

To prove this theorem, we need two following lemmas:

Lemma 3. *For each regular hyper-language M on $\sum$, there always exists a hyper-source G such that $M = \|G\|$ (see [2], p.102).*

Lemma 4. *For each regular hyper-source G on $\sum$, there always exists a deterministic hyper-automaton recognized by $\|G\|$.*

Prove of lemma 4

Let G be a hyper-source on $\sum = \{a_1, a_2, ..., a_m\}$ with the set of vertices $S = \{\nu_1, \nu_2, ..., \nu_n\}$ whose initial vertex is ν_1 and whose set of ends $F_G = \{\nu_{i1}, \nu_{i2}, ..., \nu_{ik}\}$. Then $\|G\| = \{\alpha \in \sum^\infty \mid \exists$ hyper-line $\pi = \nu_1 \rho_1 \nu_{j2} \rho_2 \nu_{j3} \rho_3 ...$ so that $[\pi] = \alpha$ and $\lim(\pi) \cap F_G \neq \emptyset\}$.

Let G be a deterministic hyper-source. We construct a deterministic hyper-automaton $V = (\sum, Q, \varphi, v_1, F_v)$, where Q is the set of all vertices of G (i.e., $Q \equiv S$). The function φ is defined as $\varphi(q, a) = p$ if there is a bow assigning to a in G from q to p. The set of ends $F_v \equiv F_G$. Denote by $R = \{\alpha \in \sum^\infty$ lim $\overline{\varphi}(v_1, \alpha) \cap F_v \neq \emptyset\}$, the set of words recognized by the hyper-automaton V. We have to prove $\|G\| = R$.

1) $\|G\| \subseteq R$: Let $\alpha \in \|G\|$. Then exists a hyper-line $\pi = \nu_1 \rho_1 \nu_{j2} \rho_2 \nu_{j3} \rho_3 ...$ with $[\pi] = \alpha, \alpha = a_{j_1} a_{j_2} ... \in \sum^\infty$, where a_{jk} is the character assigning to the bow ρ_k and lim $(\pi) \cap F_G \neq \emptyset$. From $\lim(\pi) \cap F_G \neq \emptyset$, there is an infinite sequence of increasing indexes j such that the vertices $\nu_{ji} = \nu$ and $\nu \in F_v$, i.e., $\lim \overline{\varphi}(v_1, \alpha) \cap F_v \neq \emptyset$. Therefore $\|G\| \subseteq R$.

2) $R \subseteq \|G\|$: Let $\alpha \in \sum^\infty \alpha = a_{j_1} a_{j_2} ... \in \sum^\infty$, such that lim $\overline{\varphi}(v_1, \alpha) \cap F_v \emptyset$.

We denote $\overline{\varphi}(v_1, \alpha) = \nu_{j_1} \nu_{j_2} ... \in Q^\infty$. Since $\lim \overline{\varphi}(v_1, \alpha) \cap F_v \neq \emptyset$, there is a sequence of infinite increasing indexes $i_1, i_2, ...$ such that $\nu_{i_1} = \nu_{i_2} = ... = \nu \in F_v$ and $\varphi(v_1, a_1 ... a_{i_k - 1}) = \nu_{ik}; k = 1, 2, ...$ From the way of constructing the automaton, according to the hyper word $\alpha = a_{j_1} a_{j_2} ... \in R$ we have a hyper-line in the hyper-source G : $\pi = \nu_1 \rho_1 \nu_{j2} \rho_2 \nu_{j3} \rho_3 ...$ so that $[\pi] = \alpha$ and $\varphi(\nu_{i_k}, a_{i_k} = \nu_{i_k + 1}$, where, from the about result, there is an infinite sequence of increasing indexes $i_1, i_2, ...$ such that $\nu_{i_1} = \nu_{i_2} = ... \nu \in F_G$, i.e., $\lim \pi \cap F_G \neq \emptyset$. Therefore, $R \subseteq \|G\|$.

Let G be a non-deterministic hyper-source. We construct the deterministic hyper-source G' from G and prove the equivalence of their recognizance of hyper-languages. Summing up the above result we have the proof of lemma □

Let G' be the hyper-source to be received when G is determined . The set of vertexes of G' in the set of all subsets of S, containing $\emptyset$ and S, too. The initial of G' is $\{v_1\}$, the set of ends $F_{G'}$ of G' containing subsets of S which containing at least one of ends of G.

The vertexes and the bows of G' are defined as follows: From the initial $\{v_1\}$ of G', for each $a \in \sum$, the succeeding of $\{v_1\}$ is the set of vertexes $C = \{s | s \in G$: there is a bow from v_1 to s which assigns the character $a\}$. In that case, on G' we assign the character a to the bow from $\{v_1\}$ to C.

Suppose that we have determined a vertex C of G'. For each $a \in \sum$, the succeeding D of C is the set of vertexes of G such that $D = \bigvee_{\nu \in C} \theta(\nu, a)$, where $\theta(\nu, a)$ is the set of vertexes of G connecting to ν by a bow assigning to the character a.

In the case of finite languages, it is known that G and G' are recognized the same regular language. We shall prove that so are the hyper-source G and G'.

If the hyper-line $\pi = w_1, \rho_1, w_2, \rho_2 ...$ in G having $[\pi] \in \|G\|$ then there is an infinite set of indexes i such that $w_1 = w \in F_G$, that means from π we can take an infinite set of complete finite lines $\pi_i = w_1, \rho_1, w_2, \rho_2, ..., w_i$ (i.e., the beginning part of π ending at an end w_i) such that the word generating is contained in the finite language recognized by G. Because G and G' are recognized by the same finite languages, for each π_i of G, there exists a respective $\pi_i' \in G'$ such that they generate the same word. Let i tend to the infinite, we obtain the hyper-line π and π', respectively in G and G' such that they generate the same word belonging to $\|G\|$ and $\|G'\|$. Therefore, $\|G\| \subseteq \|G'\|$. The same argument can be applied to the reciprocal one. That means $\|G\| = \|G'\|$.

The theorem 2 is followed from the lemmas 3 and 4.

Theorem 3. *A hyper-language is closed with the operation of intersection and complement.*

Proof: Let M_1, M_2 be regular hyper-languages. Theorem 2 shows that there exist two hyper-automaton $V = (\sum, Q, \varphi, F, q_1)$ and $V' = (\sum, Q', \varphi', F'; q_1')$ recognized by M_1, M_2, i.e., $M_1 = \{\alpha \in \sum^\infty \mid \lim \overline{\varphi}(q_1, \alpha) \cap F \neq \emptyset\}$ and $M_2 = \{\alpha \in \sum^\infty \mid \lim \overline{\varphi}'(q_1', \alpha) \cap F' \neq \emptyset\}$.

1) $M_1 \cap M_2$ is hyper-language. We consider a hyper-automaton

$$V'' = (\sum, Q \times Q', \varphi'', F \times F', (q_1, q_1')) \text{ with } \varphi''(q_1, q_1'), a) = (\varphi(q_1, a), \varphi'(q_1', a)).$$

Let $R = \{\alpha \in \sum^\infty \mid \lim \overline{\varphi}''((q_1, q_1'), \alpha) \cap (F \times F') \neq \emptyset\}$. We prove $M_1 \cap M_2 = R$.

Let $\alpha \in M_1 \cap M_2$, we have $\alpha \in M_1$ and $\alpha \in M_2$. Therefore, $\lim \overline{\varphi}(q_1, \alpha) \cap F \neq \emptyset$ and $\lim \overline{\varphi}'(q_1', \alpha) \cap F' \neq \emptyset$. It follows $\lim \overline{\varphi}(q_1, \alpha), \lim \overline{\varphi}'(q_1', \alpha)) \cap (F \times F') \neq \emptyset$ or $\lim \overline{\varphi}"((q_1, q_1'), \alpha) \cap (F \times F') \neq \emptyset$, i.e., $\alpha \in R$.

Inversely, let $\alpha \in R$, i.e., $\lim \overline{\varphi}"((q_1, q_1'), \alpha) \cap (F \times F') \neq \emptyset$, then

$$(\lim \overline{\varphi}(q_1, \alpha), \lim \overline{\varphi}'(q_1', \alpha)) \cap (F \times F') \neq \emptyset,$$

i.e.,

$$\lim \overline{\varphi}(q_1, \alpha) \cap F \neq \emptyset,$$

and

$$\lim \overline{\varphi}'(q_1', \alpha) \cap F' \neq \emptyset \text{ or } \alpha \in M_1,$$

and

$$\alpha \in M_2, \text{ i.e., } \alpha \in M_1 \cap M_2.$$

Therefore, $M_1 \cap M_2 = R$, i.e.,$M_1 \cap M_2$ is hyper-regular.

2. $C(M_1) = \sum^\infty \backslash M_1$ is hyper regular. Consider the hyper- automaton $V = (\sum, Q, \varphi, Q \backslash F, q_1)$, we shows that $C(M_1)$ is recognized by the hyper-automaton V. We note that if $\sum$ is finite then each hyper-word $\alpha \in \sum^\infty$ has the limit i.e., $\lim(\alpha) \neq \emptyset$. Therefore if $\alpha \in \sum^\infty \backslash M_1$, i.e., $\alpha \in \sum^\infty$ and $\alpha \notin M_1$ then $\lim \overline{\varphi}(q_1, \alpha) \cap Q \neq \emptyset$ and $\lim \overline{\varphi}(q_1, \alpha) \cap F = \emptyset$, that implies $\lim \overline{\varphi}(q_1, \alpha) \cap Q \backslash F \neq \emptyset$, i.e., α is recognized by V. Similarly, the inverse part is proved.

Corollary. *$M_1 \setminus M_2$ is hyper regular.*

This result is received from two above parts and $M_1 \setminus M_2 = M_1 \cap (\sum^{\infty} \setminus M_2)$.

REFERENCES

[1] Nguyễn Văn Ba: Ngôn ngữ hình thức, 1997.

[2] В.Б Кудрявцев, С.В. Алешен и А.С. Подколин. *Введение в Теорю Автоматов.* Москва "Наука" 1998.

[3] Данг Зуи Руан. *Конечноавтоматная Сложность Порождающих Схем Специального Вида.* Discrete Mathematics Banach Center Publications, Vol **7.** PWN-Polish Scientific Publishers. Warsaw 1982.

[4] Данг Зуи Руан. *Сложность Конечного Автомата, Соотвестующего Обобщенному Регулярному Выражению.* Докл. АН СССР **213** (1973), 1.

TẠP CHÍ KHOA HỌC ĐHQGHN, KHTN, t.XV, n^{0}1 - 1999

MỘT SỐ KẾT QUẢ VỀ LỚP SIÊU NGÔN NGỮ CHÍNH QUY

Đặng Huy Ruận
Khoa Toán - Đại học KH Tựnhiên - ĐHQG HàNội
Phùng Văn Ổn
Trường Đại học Hàng hải Việt Nam

Lớp ngôn ngữ chính qui với từ hữu hạn đã được nghiên cứu khá kỹ lưỡng trong lý thuyết ngôn ngữ hình thức. Trong phạm vi của bài này, chúng tôi nghiên cứu lớp ngôn ngữ chính qui với từ vô hạn, gọi là siêu ngôn ngữ chính qui.

Trên cơ sở các định nghĩa về từ vô hạn, định nghĩa về giới hạn của chúng và các phép toán trên siêu ngôn ngữ, chúng tôi đã thu được một số kết quả ban đầu về vấn đề đoán nhận của các siêu ngôn ngữ chính quy thông qua các siêu otomat với các từ vào vô hạn và các siêu từ trạng thái tương ứng, các siêu nguồn với các siêu đường của nó và một số tính chất của siêu ngôn ngữ chính quy như tính đóng với các phép lấy phần bù, phép nhân siêu ngôn ngữ.