Tạp chí Tin học và Điều khiển học, T.28, S.1 (2012), 52–63

XÁC ĐỊNH ĐỘ KHÔNG NHẬP NHẰNG CỦA NGÔN NGỮ CHÍNH QUY THEO ÔTÔMAT

ĐẶNG QUYẾT THẮNG[1], NGUYỄN ĐÌNH HÂN[2], PHAN TRUNG HUY[3]

[1] *Trường Đại học Sư phạm Kỹ thuật Nam Định*

[2] *Trường Đại học Sư phạm Kỹ thuật Hưng Yên*

[3] *Trường Đại học Bách khoa Hà Nội*

Tóm tắt. Trong bài báo này, chúng tôi giới thiệu và nghiên cứu về độ không nhập nhằng của ngôn ngữ, một khái niệm mới làm mịn khoảng trống giữa khái niệm mã và tích không nhập nhằng. Với những ngôn ngữ không phải là mã nhưng có độ không nhập nhằng hữu hạn đủ lớn, ta có thể sử dụng để mã hóa thông tin bí mật ... Chúng tôi đề xuất giải thuật tính độ không nhập nhằng của ngôn ngữ chính quy, với độ phức tạp thời gian $\mathcal{O}(n^4)$ cho trường hợp tổng quát xét trên các ôtômat đa định, $\mathcal{O}(n^2 logn)$ với trường hợp riêng cho các ôtômat đơn định, ở đó n là số trạng thái của ôtômat hữu hạn đoán nhận những ngôn ngữ này.

Abstract. In this paper, we introduce and study on unambiguous degree of languages, a new concept which fills the gap between code and unambiguity product notions. We can use some languages which are not codes but having finite unambiguous degree large enough for encrypting secret information... We propose algorithms to determine unambiguous degree of regular languages with time complexity $\mathcal{O}(n^4)$ for the general case of non-deterministic automata, $\mathcal{O}(n^2 logn)$ for the case of deterministic automata, where n is the number of states of finite automata recognizing these languages.

1. MỞ ĐẦU

Khái niệm tích không nhập nhằng của hai ngôn ngữ được đề xuất bởi Schützenberger (1955) có liên quan chặt chẽ với khái niệm mã. Các tính chất đại số của mã dựa trên tích không nhập nhằng được nghiên cứu sâu sắc bởi Schützenberger (1955), Gilbert and Moore (1959) và các tác giả khác (xem [1, 3, 5, 6, 7, 8]). Gần đây, nghiên cứu lý thuyết mã có xu hướng sử dụng các yếu tố điều khiển, đa trị, nhập nhằng để mở rộng khái niệm tích, từ đó xây dựng những lớp mã mới, chẳng hạn như Z-mã dựa trên tích zigzag [9, 10], T-mã dựa trên tích trộn có điều khiển [11], C-mã dựa trên tích nhập nhằng sử dụng yếu tố ngữ cảnh [13, 14] và $\diamond$-mã [15]. Giữa khái niệm mã và tích không nhập nhằng có một khoảng trống, dựa vào khái niệm độ không nhập nhằng của ngôn ngữ do chúng tôi đề xuất thì mã có độ không nhập nhằng $k = \infty$, các ngôn ngữ có độ không nhập nhằng k được trải rộng từ 0 đến ∞. Từ đó, nhận được một phân bậc chặt trên lớp toàn bộ các ngôn ngữ. Với $k = 0$ là lớp tất cả các ngôn ngữ, $k = \infty$ là lớp mã, $k = 2$ liên quan đến tích không nhập nhằng. Đã có những nghiên cứu về tính không nhập nhằng của đối tượng biểu diễn ngôn ngữ như: văn phạm nhập nhằng, không nhập nhằng; ôtômat nhập nhằng, không nhập nhằng,... Ở đây chúng tôi nghiên cứu về độ không nhập nhằng của ngôn ngữ, một khái niệm mới có liên quan chặt chẽ với mã. Về mặt

ứng dụng, ta có thể sử dụng những ngôn ngữ có độ không nhập nhằng đủ lớn, không nhất thiết là mã để mã hóa thông tin mật. Đối phương tấn công vào các hệ mã không là mã sẽ phức tạp hơn, chi phí cao hơn tấn công vào các hệ mã là mã. Do đó, xây dựng giải thuật xác định độ không nhập nhằng của ngôn ngữ chính quy được đoán nhận bởi ôtômat có ý nghĩa quan trọng trong lý thuyết và ứng dụng.

Bài báo đưa ra khái niệm độ không nhập nhằng của ngôn ngữ, giải thuật xác định độ không nhập nhằng của ngôn ngữ chính quy được đoán nhận bởi ôtômat hữu hạn. Giải thuật này có độ phức tạp thời gian $\mathcal{O}(n^4)$ cho trường hợp xét trên các ôtômat đa định tùy ý, $\mathcal{O}(n^2 logn)$ với trường hợp các ôtômat đơn định, ở đó n là số trạng thái của ôtômat hữu hạn. Có thể dùng kỹ thuật tương tự [16] cho phép xây dựng giải thuật cỡ đa thức theo chỉ số tương đẳng cú pháp của ngôn ngữ chính quy. Nếu áp dụng kỹ thuật này với ngôn ngữ chính quy được cho bởi ôtômat (nói chung là đa định) thì cần tính tương đẳng từ ôtômat đa định với độ phức tạp thời gian là hàm mũ.

Tiếp theo, Mục 2 sẽ trình bày một số khái niệm ngôn ngữ, mã, độ không nhập nhằng của ngôn ngữ, ôtômat và đồ thị. Mục 3 trình bày các giải thuật mở rộng ôtômat. Mục 4 đề xuất phương pháp xác định độ không nhập nhằng của ngôn ngữ chính quy và cuối cùng là phần kết luận của bài báo.

2. MỘT SỐ KHÁI NIỆM

Cho bảng chữ cái hữu hạn Σ, dãy $w = a_1a_2...a_n$, $a_i \in \Sigma$, $i = 1,..,n$ gọi là một từ hay một xâu trên Σ, số ký tự có trong xâu w gọi là độ dài xâu và ký hiệu là $|w|$, xâu có độ dài bằng 0 gọi là xâu rỗng và ký hiệu là ε, tập các xâu trên Σ ký hiệu là Σ^*. Tập $X \subseteq \Sigma^*$ gọi là ngôn ngữ trên Σ. Cho hai ngôn ngữ $X, Y \subseteq \Sigma^*$, tích ghép hai ngôn ngữ X và Y ký hiệu là XY được xác định như sau: $XY = \{w \in \Sigma^* \mid w = uv, u \in X, v \in Y\}$. Ký hiệu $X^0 = \{\varepsilon\}$, $X^1 = X$, $X^n = X^{n-1}X$, với $n \geq 1$ và $X^+ = \bigcup_{i=1}^{\infty} X^i$, $X^* = X^0 \cup X^+$. Về cơ sở xin xem thêm tài liệu [1].

Định nghĩa 2.1. Cho bảng chữ cái Σ, tập $X \subseteq \Sigma^*$ được gọi là mã nếu với mọi $m, n \geq 1$ và với mọi $x_1, \ldots, x_n$, $y_1, ..., y_m \in X$, nếu có

$$x_1 \ldots x_n \;=\; y_1 \ldots y_m \quad \text{thì suy ra} \quad m = n \quad \text{và} \quad x_i = y_i, \text{ với } i = 1,..,n.$$

Nói cách khác, tập X là mã nếu mọi xâu trong X^+ chỉ có một phân tích duy nhất thành các xâu trong X. Vì $\varepsilon.\varepsilon = \varepsilon$ nên mọi tập mã đều không chứa xâu rỗng.

Định nghĩa 2.2. Cho $X \subseteq \Sigma^*$ và số tự nhiên $k \geq 0$, khi đó:

i) Tập X được gọi là *k-không nhập nhằng* nếu với mọi số nguyên $m \geq 1$ và với mọi $x_1, ..., x_k$, $y_1, ..., y_m \in X$, nếu có

$$x_1 \ldots x_k \;=\; y_1 \ldots y_m \quad \text{thì suy ra} \quad k = m \quad \text{và} \quad x_i = y_i, \text{ với } i = 1,..,k$$

.

Ngược lại (tồn tại từ $w \in X^*$, $w = x_1...x_k = y_1...y_m$ mà $k \neq m$ hoặc $x_1 \neq y_1$) thì tập X được gọi là *k-nhập nhằng*. Quy ước mọi tập X đều là 0-không nhập nhằng.

ii) Nếu có số k hữu hạn lớn nhất sao cho X là k-không nhập nhằng thì k được gọi là *độ không nhập nhằng* của X, khi đó ta gọi $k + 1$ là *độ nhập nhằng* của X.

iii) Nếu X là k-không nhập nhằng với mọi k thì ta nói rằng X có độ không nhập nhằng ∞ và X không có độ nhập nhằng.

Ví dụ 2.1. Cho $\Sigma = \{a, b, c, d, e, f\}$, $X = \{\varepsilon, a, ab, be, ec, cf, fd, d\}$, ta thấy X có độ không nhập nhằng là 0, X là 1- nhập nhằng vì tồn tại $w = \varepsilon = \varepsilon\varepsilon$.

Ví dụ 2.2. Cho $\Sigma = \{a, b\}$, $X = \{a, ab, b\}$, ta thấy X có độ không nhập nhằng là 0, X là 1-nhập nhằng vì tồn tại $w = (ab) = (a)(b)$ mà $ab \neq a$.

Ví dụ 2.3. Cho $\Sigma = \{a, b, c, d, e, f\}$, $X = \{a, ab, be, ec, cf, fd, d\}$, ta có thể dễ dàng kiểm tra bằng định nghĩa X có độ không nhập nhằng là 2, nhưng X là 3-nhập nhằng vì tồn tại từ $w = (ab)(ec)(fd) = (a)(be)(cf)(d)$ mà $ab \neq a$.

Ví dụ 2.4. Cho $k \geq 1$ là một số tự nhiên tùy ý, ta xét bảng chữ $\Sigma = \{c, a_1, b_1, \ldots, a_k, b_k\}$ và $X = \{c, ca_1, a_1b_1, b_1a_2, \ldots, b_{k-1}a_k, a_kb_k, b_k\}$. Dễ thấy, X là k-không nhập nhằng và X là $(k+1)$-nhập nhằng.

Nhận xét 2.1. Cho $X \subseteq \Sigma^*$, khi đó ta có các tính chất hiển nhiên:

(*i*) Nếu X là k-không nhập nhằng thì X là $(k-1)$- không nhập nhằng với $k \geq 1$.

(*ii*) Nếu X là k-nhập nhằng thì X là $(k+1)$- nhập nhằng với $k \geq 1$.

(*iii*) Nếu $\varepsilon \in X$ hoặc X là 1-nhập nhằng thì X luôn có độ không nhập nhằng 0, độ nhập nhằng 1.

(*iv*) Nếu X có độ nhập nhằng k thì k là số nhỏ nhất sao cho X là k-nhập nhằng và ngược lại.

(*v*) Nếu X có độ không nhập nhằng vô hạn thì X là mã.

Phân bậc ngôn ngữ theo khái niệm không nhập nhằng: Ta ký hiệu $\mathcal{L}_k$ lớp ngôn ngữ là k-không nhập nhằng, $\mathcal{L}_0$ là lớp tất cả các ngôn ngữ, $\mathcal{L}_\infty = \bigcap_{i=0}^{\infty} \mathcal{L}_i$ là lớp mã. Từ Ví dụ 2.2, 2.4 và Nhận xét 2.1, ta nhận được một phân bậc chặt:

$$\mathcal{L}_\infty \subsetneq \ \ldots \ \subsetneq \mathcal{L}_2 \subsetneq \mathcal{L}_1 \subsetneq \mathcal{L}_0.$$

Dưới đây ta nhắc lại một số khái niệm về ôtômat và đồ thị có hướng được dùng trong các phần sau:

Ôtômat hữu hạn là một bộ 5 $\mathcal{A} = (Q, \Sigma, E, I, F)$, với Q là tập hữu hạn các trạng thái, Σ là bảng chữ cái hữu hạn, $E \subseteq Q \times \Sigma \times Q$ là tập hữu hạn các cung (không chứa cung rỗng), $I \subseteq Q$ là tập các trạng thái đầu, $F \subseteq Q$ là tập các trạng thái kết thúc.

Cho cung $e = (q_1, a, q_2) \in E$, ta nói rằng e rời q_1 đến q_2, q_1 là trạng thái trước của e ký hiệu $p[e]$, q_2 là trạng thái sau của e ký hiệu $n[e]$, a là nhãn của e ký hiệu $l[e]$. Cho $q \in Q$, ta ký hiệu $E[q]$ là tập các cung rời q.

Ôtômat hữu hạn $\mathcal{A}$ được gọi là đơn định nếu $\mathcal{A}$ có duy nhất một trạng thái đầu và với mỗi trạng thái $q \in Q$, với mỗi $a \in \Sigma$ có nhiều nhất một cung rời q với nhãn a.

Cho $\pi = e_1 ... e_k \in E^*$, với $e_1 = (p_0, a_1, p_1)$, $e_2 = (p_1, a_2, p_2)$, $\ldots$, $e_k = (p_{k-1}, a_k, p_k)$ được gọi là đường đi từ p_0 đến p_k. Đường đi π gọi là đi qua trạng thái q nếu q là trạng thái trước

hoặc trạng thái sau của một cung thuộc π. Một đường đi thành công trong ôtômat $\mathcal{A}$ là một đường đi từ trạng thái đầu đến trạng thái kết thúc. Từ $w = a_1a_2...a_k$ gọi là nhãn của đường đi π. Tập hợp nhãn của các đường đi thành công trong ôtômat $\mathcal{A}$ ký hiệu là $\mathcal{L}(\mathcal{A})$ và gọi là ngôn ngữ đoán nhận bởi $\mathcal{A}$.

Đồ thị có hướng G là một cặp $G = (V, E)$, V là tập các đỉnh, E là tập các cặp có thứ tự gồm hai phần tử của V gọi là các cung. Nếu $e = (u, v)$ là cung của đồ thị có hướng G thì ta nói đỉnh v kề u, kí hiệu $Next(u) = \{v \in V | (u, v) \in E\}$. Nếu V, E là hữu hạn thì ta gọi G là đồ thị hữu hạn có hướng.

Đường đi từ đỉnh u đến đỉnh v trên đồ thị có hướng G là dãy đỉnh $x_0, \ldots, x_n$, với $u = x_0$, $v = x_n$, $(x_i, x_{i+1}) \in E$, $i = 0, .., n - 1$.

3. MỘT SỐ GIẢI THUẬT MỞ RỘNG ÔTÔMAT

Mục này ta sẽ xét một số kỹ thuật mở rộng ôtômat để xây dựng *ôtômat lưỡng cực*, *ôtômat tích* từ ôtômat hữu hạn cho trước. Các ôtômat mở rộng này được dùng để thiết kế giải thuật xác định độ không nhập nhằng của ngôn ngữ chính quy ở phần sau.

3.1. Ôtômat lưỡng cực

Ôtômat hữu hạn $\mathcal{A}$ (đa định) được gọi là ôtômat *lưỡng cực* nếu $\mathcal{A}$ có một trạng thái đầu, một trạng thái kết thúc, không có cung đến trạng thái đầu và không có cung rời trạng thái kết thúc.

Cho ôtômat hữu hạn $\mathcal{A} = (Q, \Sigma, E, I, F)$ đoán nhận ngôn ngữ $X = \mathcal{L}(\mathcal{A}) \subseteq \Sigma^+$, ta xây dựng ôtômat lưỡng cực $\mathcal{A}' = (Q', \Sigma, E', I', F')$ cũng đoán nhận X như sau:

(*i*) $Q' = Q \cup \{s, f\}, s, f \notin Q$ và $s \neq f, I' = \{s\}, F' = \{f\}$.

(*ii*) $E' = E_1 \cup \{(s, a, q) | (p, a, q) \in E_1, p \in I\}$, với $E_1 = E \cup \{(p, a, f) | (p, a, q) \in E, q \in F\}$.

Để đơn giản, ký hiệu $\mathcal{A}' = (Q', \Sigma, E', s, f)$ và gọi s là *cực vào* (trạng thái khởi đầu), f là *cực ra* (trạng thái kết thúc). Hàm biểu diễn giải thuật xây dựng ôtômat lưỡng cực $\mathcal{A}'$ từ ôtômat hữu hạn $\mathcal{A}$ được ký hiệu là $D(\mathcal{A})$, giải thuật có độ phức tạp thời gian $\mathcal{O}(|Q| + |E|)$.

Cho ôtômat lưỡng cực $\mathcal{A}$ đoán nhận ngôn ngữ $X = \mathcal{L}(\mathcal{A}) \subseteq \Sigma^+$, ta xây dựng ôtômat *mở rộng* $\mathcal{A}'$ đoán nhận X^+ (tức là $X^+ = \mathcal{L}(\mathcal{A}')$) bằng cách bổ sung một cung rỗng đi từ cực ra tới cực vào của $\mathcal{A}$. Giải thuật xây dựng ôtômat mở rộng được thực hiện bởi một hàm ký hiệu là $Ex(\mathcal{A})$.

Nhận xét 3.1. Cho ôtômat $\mathcal{A} = (Q, \Sigma, E, I, F)$, $c = |\Sigma|$ coi là hằng số, $n = |Q|$, $m = |E|$.

(*i*) Nếu $\mathcal{A}' = D(\mathcal{A})$ thì $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$.

(*ii*) Nếu $\mathcal{A}$ là đơn định thì với mỗi đỉnh $p \in Q$ có tối đa c cung ra, suy ra số cung tối đa của $\mathcal{A}$ là $m = nc$. Vậy, giải thuật xây dựng ôtômat lưỡng cực có độ phức tạp thời gian là $\mathcal{O}(|Q|)$.

(*iii*) Nếu $\mathcal{A}$ là đơn định thì: số trạng thái của $D(\mathcal{A})$ và $Ex(D(\mathcal{A}))$ không quá $n + 2$, do đó có cỡ $\mathcal{O}(n)$; số cung của $D(\mathcal{A})$ không quá $2m + 2c = 2nc + 2c$, $Ex(D(\mathcal{A}))$ không quá $2nc + 2c + 1$, do đó cùng cỡ $\mathcal{O}(n)$.

3.2. Ôtômat tích

Phép lấy tích ôtômat (xem [2, 4]) được sử dụng trong nhiều ứng dụng để tạo ra ôtômat phức hợp từ những ôtômat đơn giản. Cho hai ôtômat lưỡng cực hoặc mở rộng như đã xét ở trên: $\mathcal{A}_1 = (Q_1, \Sigma, E_1, s_1, f_1)$ và $\mathcal{A}_2 = (Q_2, \Sigma, E_2, s_2, f_2)$. Ôtômat tích của $\mathcal{A}_1$ và $\mathcal{A}_2$ được ký hiệu là $Prod(\mathcal{A}_1, \mathcal{A}_2) = (Q, \Sigma, E, (s_1, s_2), (f_1, f_2))$, ở đó $Q \subseteq Q_1 \times Q_2$, E được xác định theo quy tắc sau:

(i) $\forall (q_1, a, p_1) \in E_1, \forall (q_2, a, p_2) \in E_2, a \in \Sigma \quad \Rightarrow \quad ((q_1, q_2), a, (p_1, p_2)) \in E.$

(ii) $\forall (q_1, \varepsilon, p_1) \in E_1, \forall (q_2, \varepsilon, p_2) \in E_2 \quad \Rightarrow \quad ((q_1, q_2), \varepsilon, (p_1, p_2)) \in E.$

(iii) $\forall (q_1, \varepsilon, p_1) \in E_1, \forall (q_2, a, p_2) \in E_2, a \in \Sigma \quad \Rightarrow \quad ((q_1, q_2), \varepsilon, (p_1, q_2)) \in E.$

(iv) $\forall (q_1, a, p_1) \in E_1, \forall (q_2, \varepsilon, p_2) \in E_2, a \in \Sigma \quad \Rightarrow \quad ((q_1, q_2), \varepsilon, (q_1, p_2)) \in E.$

(v) E chỉ chứa các cung đã xét ở bốn trường hợp trên.

Giải thuật xây dựng ôtômat tích thực hiện bắt đầu từ trạng thái (s_1, s_2), rồi theo quy tắc ở trên để xác định các trạng thái và cung của ôtômat tích như sau.

Function Prod$(\mathcal{A}_1, \mathcal{A}_2)$
Input: $\mathcal{A}_1$, $\mathcal{A}_2$ là ôtômat lưỡng cực hoặc mở rộng.
Output: $\mathcal{A} = (Q, \Sigma, E, s, f)$ là ôtômat tích của $\mathcal{A}_1$ và $\mathcal{A}_2$.
// Giải thuật dùng một hàng đợi S

1. $Q = \{(s_1, s_2)\}$; $CQpush(S, (s_1, s_2))$;
   $s = (s_1, s_2)$; $f = (f_1, f_2)$; $E = \phi$;
2. While $S \neq \phi$ do
3. $(q_1, q_2) \leftarrow CQPop(S)$;
4. for each $(e_1, e_2) in E[q_1] \times E[q_2]$ do
5. t = true; $label = \varepsilon$;
6. case
   $l[e_1] = l[e_2]$ and $l[e_1] \neq \varepsilon$: $(p_1, p_2) = (n[e_1], n[e_2])$; $label = l[e_1]$;
   $l[e_1] = l[e_2] = \varepsilon$: $(p_1, p_2) = (n[e_1], n[e_2])$;
   $l[e_1] = \varepsilon$ and $l[e_2] \neq \varepsilon$: $(p_1, p_2) = (n[e_1], q_2)$;
   $l[e_1] \neq \varepsilon$ and $l[e_2] = \varepsilon$: $(p_1, p_2) = (q_1, n[e_2])$;
   else t = false;
   end case;
7. if t = true then
   if $(p_1, p_2) \notin Q$ then
   $Q = Q \cup \{(p_1, p_2)\}$; $CQPush(S, (p_1, p_2))$;
   $E = E \cup \{((q_1, q_2)$, label, $(p_1, p_2))\}$;
8. Return Prod$(\mathcal{A}_1, \mathcal{A}_2)$.

Nhận xét 3.2

(i) Tương tự như phân tích của Mohri trong [2, 4], giải thuật xây dựng ôtômat tích có độ phức tạp thời gian là $\mathcal{O}((|Q_1| + |E_1|)(|Q_2| + |E_2|))$. Theo Nhận xét 3.1, nếu $\mathcal{A}_1$, $\mathcal{A}_2$ là đơn định thì giải thuật có độ phức tạp thời gian là $\mathcal{O}(|Q_1||Q_2|)$.

(*ii*) Cho ôtômat lưỡng cực $\mathcal{A}_1 = (Q_1, \Sigma, E_1, s_1, f_1)$ đoán nhận ngôn ngữ L, ôtômat mở rộng $\mathcal{A}_2 = (Q_2, \Sigma, E_2, s_2, f_2)$ đoán nhận L^+, ta có:

+ Trên ôtômat tích $Prod(\mathcal{A}_1, \mathcal{A}_2)$, nhãn của đường đi giữa hai trạng thái kế tiếp (f_1, q_i) và (f_1, q_j) (hoặc (p_i, f_2) và (p_j, f_2), hoặc (s_1, q_i) và (f_1, q_j), hoặc (p_i, s_2) và (p_j, f_2)) là từ thuộc L.

+ Trên ôtômat tích $Prod(\mathcal{A}_2, \mathcal{A}_2)$, nhãn của đường đi giữa hai trạng thái kế tiếp (f_2, q_i) và (f_2, q_j) (hoặc (p_i, f_2) và (p_j, f_2), hoặc (s_2, q_i) và (f_2, q_j), hoặc (p_i, s_2) và (p_j, f_2)) là từ thuộc L.

4. XÁC ĐỊNH ĐỘ KHÔNG NHẬP NHẰNG CỦA NGÔN NGỮ

Để xác định độ không nhập nhằng của $X \subseteq \Sigma^*$ được đoán nhận bởi ôtômat hữu hạn $\mathcal{A}$, trước hết ta giải quyết bài toán trên đồ thị như dưới đây.

4.1. Bài toán về đường đi hợp lệ trên đồ thị

Ta xét đồ thị hữu hạn có hướng (có thể có khuyên) $G = (V, E)$, có hai đỉnh đặc biệt là *đỉnh khởi đầu* s và *đỉnh kết thúc* f, với $s \neq f$, các đỉnh còn lại là *đỉnh kiểm soát* hoặc *không kiểm soát*. Đường đi π từ đỉnh s đến đỉnh f gọi là *đường đi hợp lệ* nếu π đi qua ít nhất một đỉnh kiểm soát. Đường đi hợp lệ π gọi là có *giá* $k \geq 1$ nếu π đi qua k đỉnh kiểm soát.

Bài toán 1. Cho đồ thị hữu hạn có hướng G như ở trên, tìm giá nhỏ nhất của đường đi hợp lệ (nếu có) trên G.

Để giải bài toán trên, đầu tiên ta xây dựng *đồ thị sao chép* $G' = (V', E')$ từ đồ thị $G = (V, E)$ bằng cách sử dụng kỹ thuật sao chép đồ thị như sau:

(*i*) Với $v \in V$ sao chép thành hai đỉnh $(v, 1)$ và $(v, 2)$ của V'.

(*ii*) Với $(u, v) \in E$:

+ Sao chép thành hai cung $((u, 1), (v, 1))$, $((u, 2), (v, 2))$ của E'.

+ Nếu u là đỉnh kiểm soát thì bổ sung vào E' cung $((u, 1), (v, 2))$.

(*iii*) Trên G' ta gán trọng số là 1 cho các cung đi đến đỉnh (v, i), $i = 1, 2$ mà v là đỉnh kiểm soát, các cung còn lại được gán trọng số 0.

Dưới đây là giải thuật xây dựng đồ thị sao chép G' từ đồ thị G:

Function XCopy(G)
Input: Đồ thị có hướng $G = (V, E)$.
Output: Đồ thị $G' = (V', E')$ là đồ thị sao chép từ G.
// Giải thuật dùng mảng: contr$[q] = 1$ *khi và chỉ khi đỉnh q là đỉnh kiểm soát*

1. $V' = \phi$; $E' = \phi$;
2. For each u in V do $V' = V' \cup \{(u, 1), (u, 2)\}$;
3. For each u in V do
4. For each v in $Next(u)$ do
$E' = E' \cup \{((u, 1), (v, 1)), ((u, 2), (v, 2))\}$;
if $contr[v] = 1$ then // gán trọng số cho các cung
$w[((u, 1), (v, 1))] = 1$; $w[((u, 2), (v, 2))] = 1$;

else
$w[((u,1),(v,1))] = 0;\quad w[((u,2),(v,2))] = 0;$
if $contr[u] = 1$ then
$E' = E' \cup \{((u,1),(v,2))\};$
if $contr[v] = 1$ then // gán trọng số cho các cung
$w[((u,1),(v,2))] = 1;$
else
$w[((u,1),(v,2))] = 0;$

5. Return G'.

Nhận xét 4.1.

(*i*) Đồ thị sao chép G' có $|V'| = 2n$, $|E'| \leq 3m$, với $|V| = n$, $|E| = m$.

(*ii*) Tập các đỉnh dạng (v,k) cảm sinh trong G' đồ thị con G_k, $k = 1,2$. Mỗi đồ thị con G_k đều đẳng cấu với G. G' thực chất là sự kết nối có chọn lọc của hai đồ thị con G_1, G_2, chỉ có các cung đi từ G_1 đến G_2 mà không có chiều ngược lại.

(*iii*) Giải thuật *XCopy* có độ phức tạp thời gian $\mathcal{O}(|V| + |E|)$.

Vai trò của đồ thị sao chép G' trong việc xác định đường đi hợp lệ có giá nhỏ nhất trên G được cho bởi bổ đề sau:

Bổ đề 4.1. *Cho đồ thị G như ở trên và $G' = XCopy(G)$, ta có: i) Trên G' có đường đi ngắn nhất từ $(s,1)$ đến $(f,2)$ độ dài k khi và chỉ khi trên G có đường đi hợp lệ với giá nhỏ nhất k. ii) Trên G' không tồn tại đường đi từ $(s,1)$ đến $(f,2)$ khi và chỉ khi trên G không có đường đi hợp lệ.*

Chứng minh. i) ($\Rightarrow$) Trên G' có đường đi ngắn nhất π' từ $(s,1)$ đến $(f,2)$ độ dài k. Theo cách xây dựng đồ thị sao chép G' thì π' có dạng:

$$(u_0,1),\ldots,(u_m,1),(u_{m+1},2),\ldots,(u_n,2),$$

với $(s,1) = (u_0,1), (f,2) = (u_n,2)$ và trên π' có ít nhất một đỉnh $(u_m,1) \in V'$ mà $u_m \in V$ là đỉnh kiểm soát. Tương ứng với π', ta có đường đi hợp lệ π có giá bằng k trên G như sau:

$$u_0,\ldots,u_m,u_{m+1},\ldots,u_n,$$

với $s = u_0$, $f = u_n$. π cũng là đường đi hợp lệ có giá nhỏ nhất bằng k trên G, thật vậy: giả sử có đường đi hợp lệ θ trên G có giá $l < k$, khi đó θ có dạng:

$$v_0,\ldots,v_p,v_{p+1},\ldots,v_q,$$

với $s = v_0$, $f = v_q$. Vì θ là đường đi hợp lệ nên ta có thể giả sử v_p là đỉnh kiểm soát, theo cách xây dựng G' thì ta có đường đi θ' từ $(s,1)$ đến $(f,2)$ độ dài $l < k$ như sau:

$$(v_0,1),\ldots,(v_p,1),(v_{p+1},2),\ldots,(v_q,2),$$

với $(s,1) = (v_0,1)$, $(f,2) = (v_q,2)$. Điều này mâu thuẫn với π' là đường đi ngắn nhất từ $(s,1)$ đến $(f,2)$ độ dài k trên G'.

($\Leftarrow$) Trên G ta có đường hợp lệ π với giá nhỏ nhất k, khi đó π có dạng:

$$u_0,\ldots,u_m,u_{m+1},\ldots,u_n,$$

với $s = u_0$, $f = u_n$ và trên π có ít nhất một đỉnh $u_m \in V$ là đỉnh kiểm soát. Tương ứng với π ta có đường đi π' từ $(s,1)$ đến $(f,2)$ độ dài k trên G' như sau:

$$(u_0, 1), \ldots, (u_m, 1), (u_{m+1}, 2), \ldots, (u_n, 2),$$

với $(s, 1) = (u_0, 1)$, $(f, 2) = (u_n, 2)$. π' cũng là đường đi ngắn nhất từ $(s, 1)$ đến $(f, 2)$ độ dài k trên G', thật vậy: giả sử có đường đi θ' từ $(s, 1)$ đến $(f, 2)$ độ dài l trên G' mà $l < k$, khi đó θ' có dạng:

$$(v_0, 1), \ldots, (v_p, 1), (v_{p+1}, 2), ...(v_q, 2),$$

với $(s, 1) = (v_0, 1)$, $(f, 2) = (v_q, 2)$ và v_p là đỉnh kiểm soát. Theo cách xây dựng G' thì ta có đường đi hợp lệ θ trên G có giá $l < k$ như sau:

$$v_0, \ldots, v_p, v_{p+1}, \ldots, v_q,$$

với $s = v_0$, $f = v_q$. Điều này mâu thuẫn với π là đường đi hợp lệ có giá nhỏ nhất k trên G.

ii) Dễ dàng suy ra từ các điều kiện của kỹ thuật sao chép đồ thị. ■

4.2. Giải thuật xác định độ không nhập nhằng của ngôn ngữ chính quy

Bài toán 2. Cho ngôn ngữ chính quy $X \subseteq \Sigma^*$ được đoán nhận bởi ôtômat hữu hạn $\mathcal{A}$. Hãy xác định độ không nhập nhằng của Xdựa trên cấu trúc của $\mathcal{A}$.

Giả sử cho $\mathcal{A} = (Q, \Sigma, E, I, F)$ và $X = \mathcal{L}(\mathcal{A})$, ta xét các trường hợp sau:

1) Nếu $\varepsilon \in X$ thì X có độ không nhập nhằng 0: kiểm tra ε thuộc X hay không tương đương với việc kiểm tra $I \cap F$ có khác rỗng không. Bước này có thể thực hiện bằng giải thuật ký hiệu là $Epsilon(\mathcal{A})$ có độ phức tạp thời gian là $O(n)$, ở đó $n = |Q|$, với sự biểu diễn I, F bằng hai mảng: $InI(q) = 1 \Leftrightarrow q \in I$ và $InF(q) = 1 \Leftrightarrow q \in F$.

2) Nếu $\varepsilon \notin X$ thì $X \subseteq \Sigma^+$: ta xây dựng các ôtômat $\mathcal{A}_1 = D(\mathcal{A}) = (Q_1, \Sigma, E_1, s_1, f_1)$, $\mathcal{A}_2 = Ex(\mathcal{A}_1) = (Q_2, \Sigma, E_2, s_2, f_2)$, $\mathcal{A}_3 = Prod(\mathcal{A}_1, \mathcal{A}_2)$ và $\mathcal{A}_4 = Prod(\mathcal{A}_2, \mathcal{A}_2)$.

Ta có thể coi $\mathcal{A}_4$ như một đồ thị có hướng G_4 (hay $\mathcal{A}_4$ xác định một đồ thị) có đỉnh khởi đầu (s_2, s_2), đỉnh kết thúc (f_2, f_2), các trạng thái (f_2, q) với $q \neq f_2, s_2$ là đỉnh kiểm soát, các trạng thái còn lại là đỉnh không kiểm soát, một cung bất kỳ của $\mathcal{A}_4$ xác định một cung của đồ thị G_4. Trong trường hợp $\varepsilon \notin X$, ta thiết lập kết quả dưới đây:

Định lý 4.1. *Cho $X \subseteq \Sigma^+$ được đoán nhận bởi ôtômat hữu hạn $\mathcal{A}$, cho các ôtômat $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ và đồ thị G_4 được xác định như ở trên, khi đó:*

(*i*) *X có độ không nhập nhằng $k = 0$ khi và chỉ khi trên $\mathcal{A}_3$ có đường đi thành công đi qua ít nhất một trạng thái (p, f_2), với $p \neq f_1, s_1$.*

(*ii*) *X có độ không nhập nhằng hữu hạn $k > 0$ khi và chỉ khi trên G_4 có đường đi hợp lệ với giá nhỏ nhất k.*

(*iii*) *X có độ không nhập nhằng vô hạn $k = \infty$ khi và chỉ khi trên G_4 không có đường đi hợp lệ.*

Chứng minh. i) ($\Rightarrow$) X có độ không nhập nhằng $k = 0$, suy ra X có độ nhập nhằng $l = 1$. Vậy, theo Định nghĩa 2.2 thì tồn tại $x_1 \in X$, $y_1, ..., y_m \in X$, $m > 1$, $x_1 = y_1...y_m$ sao cho $x_1 \neq y_1$. Với $x_1 \in X$, trên $\mathcal{A}_1$ có đường đi thành công π với nhãn x_1:

$$s_1 \xrightarrow{x_1} f_1,$$

tương tự với xâu $y_1...y_m$, trên $\mathcal{A}_2$ có đường đi thành công θ với nhãn $y_1...y_m$:

$$s_2 \xrightarrow{y_1} f_2 \xrightarrow{\varepsilon} s_2 \xrightarrow{y_2} f_2 \xrightarrow{\varepsilon} s_2 \ \ldots \ s_2 \xrightarrow{y_m} f_2,$$

theo tích ôtômat, trên $\mathcal{A}_3$ có đường đi thành công ρ tạo nên từ π và θ như sau:

$$(s_1,s_2) \xrightarrow{y_1} (p_1,f_2) \xrightarrow{\varepsilon} (p_1,s_2) \xrightarrow{y_2} (p_2,f_2) \xrightarrow{\varepsilon} (p_2,s_2) \dots (p_{m-1},s_2) \xrightarrow{y_m} (f_1,f_2),$$

với $p_i \neq s_1,\ f_1, i = 1,...,m-1$. Vậy với $m > 1$, trên $\mathcal{A}_3$ có đường đi thành công đi qua ít nhất một trạng thái (p,f_2), với $p \neq f_1, s_1$.

($\Leftarrow$) Trên $\mathcal{A}_3$ có đường đi thành công ρ đi qua ít nhất một trạng thái (p,f_2), với $p \neq f_1, s_1$. Vậy, theo tích ôtômat và Nhận xét 3.2 thì trên $\mathcal{A}_1$ có đường đi thành công với nhãn $x_1 \in X$, trên $\mathcal{A}_2$ có đường đi thành công với nhãn $y_1...y_m$, ở đó $y_1,...,y_m \in X$, mà $x_1 = y_1...y_m$. Vì ρ đi qua ít nhất một trạng thái (p,f_2), với $p \neq f_1, s_1$ nên $m > 1$. Vậy X có độ nhập nhằng 1, hay X có độ không nhập nhằng 0.

ii) ($\Rightarrow$) X có độ không nhập nhằng hữu hạn $k > 0$, suy ra X có độ nhập nhằng $l = k+1 > 1$. Vậy, theo Định nghĩa 2.2 thì l là số nhỏ nhất:

$$\exists w \in X^*, w = x_1...x_l = y_1...y_m, \text{ với } x_1,..,x_l,y_1,..,y_m \in X, \text{ thỏa mãn } l \neq m \text{ hoặc } x_1 \neq y_1. \quad (1)$$

Vì l là nhỏ nhất thỏa (1) nên ta có $x_1 \neq y_1$. Từ $w = x_1...x_l$, trên $\mathcal{A}_2$ có đường đi thành công π với nhãn $w = x_1...x_l$:

$$s_2 \xrightarrow{x_1} f_2 \xrightarrow{\varepsilon} s_2 \xrightarrow{x_2} f_2 \xrightarrow{\varepsilon} s_2 \dots s_2 \xrightarrow{x_l} f_2,$$

với $w = y_1...y_m$, trên $\mathcal{A}_2$ có đường đi thành công θ với nhãn $y_1...y_m$:

$$s_2 \xrightarrow{y_1} f_2 \xrightarrow{\varepsilon} s_2 \xrightarrow{y_2} f_2 \xrightarrow{\varepsilon} s_2 \dots s_2 \xrightarrow{y_m} f_2,$$

theo tích ôtômat, trên G_4 có đường đi ρ từ (s_2,s_2) đến (f_2,f_2) tạo nên từ π và θ như sau:

$$(s_2,s_2) \xrightarrow{x_1} (f_2,q_1) \xrightarrow{\varepsilon} (s_2,q_1) \xrightarrow{x_2} (f_2,q_2) \xrightarrow{\varepsilon} (s_2,q_2) \dots (s_2,q_{l-1}) \xrightarrow{x_l} (f_2,f_2),$$

do l nhỏ nhất thỏa (1) nên $q_i \neq f_2,\ s_2,\ i = 1,..,l-1$. Vậy, ρ là đường đi hợp lệ với giá $k = l-1$. Đường đi ρ cũng là đường đi hợp lệ với giá nhỏ nhất k trên G_4, thật vậy: giả sử có đường đi hợp lệ ρ' trên G_4 có giá nhỏ nhất $h < k$, suy ra ρ' có h đỉnh kiểm soát như sau:

$$(s_2,s_2) \xrightarrow{u_1} (f_2,p_1) \xrightarrow{\varepsilon} (s_2,p_1) \xrightarrow{u_2} (f_2,p_2) \xrightarrow{\varepsilon} (s_2,p_2) \dots (s_2,q_h) \xrightarrow{u_{h+1}} (f_2,f_2),$$

ở đó $p_i \neq f_2,\ s_2,\ i = 1,..,h$. Theo Nhận xét 3.2 thì tồn tại $w = u_1...u_{h+1} = v_1...v_m$, với $u_1,...,u_{h+1},\ v_1,...,v_m \in X$ thỏa $u_1 \neq v_1$. Vậy X là $h+1$ nhập nhằng, hay độ không nhập nhằng của X là nhỏ hơn hoặc bằng $h < k$. Điều này là mâu thuẫn với X có độ không nhập nhằng hữu hạn k.

($\Leftarrow$) Giả sử trên G_4 có đường đi hợp lệ ρ với giá nhỏ nhất $k > 0$, suy ra ρ đi qua k đỉnh kiểm soát và có dạng như sau:

$$(s_2,s_2) \xrightarrow{x_1} (f_2,q_1) \xrightarrow{\varepsilon} (s_2,q_1) \xrightarrow{x_2} (f_2,q_2) \xrightarrow{\varepsilon} (s_2,q_2) \dots (s_2,q_k) \xrightarrow{x_{k+1}} (f_2,f_2),$$

vì ρ có giá nhỏ nhất $k > 0$ nên trên ρ không có đỉnh (f_2,f_2) trừ đỉnh cuối. Vậy ta có k là nhỏ nhất sao cho: $\exists w \in X^*,\ w = x_1...x_{k+1} = y_1...y_m$, với $x_1,...,x_{k+1},\ y_1,...,y_m \in X$ thỏa mãn $k+1 \neq m$ hoặc $x_1 \neq y_1$. Hay X có độ nhập nhằng $k+1$, nghĩa là X có độ không nhập nhằng k.

iii) ($\Rightarrow$) X có độ không nhập nhằng vô hạn $k = \infty$, suy ra với mọi $l > 0$ hữu hạn bất kỳ và với mọi $x_1,...,x_l,\ y_1,...,y_m \in X$, nếu có

$$w = x_1...x_l = y_1...y_m \quad \text{thì suy ra} \quad l = m \text{ và } x_i = y_i, \text{ với } i = 1,..,l. \quad (2)$$

từ $w = x_1...x_l$, trên $\mathcal{A}_2$ có đường đi thành công π với nhãn $w = x_1...x_l$:

$$s_2 \xrightarrow{x_1} f_2 \xrightarrow{\varepsilon} s_2 \xrightarrow{x_2} f_2 \xrightarrow{\varepsilon} s_2 \ldots s_2 \xrightarrow{x_l} f_2,$$

với $w = y_1...y_m$, trên $\mathcal{A}_2$ có đường đi thành công θ với nhãn $y_1...y_m$:

$$s_2 \xrightarrow{y_1} f_2 \xrightarrow{\varepsilon} s_2 \xrightarrow{y_2} f_2 \xrightarrow{\varepsilon} s_2 \ldots s_2 \xrightarrow{y_m} f_2,$$

theo tích ôtômat và theo (2) thì trên G_4 có đường đi ρ từ (s_2, s_2) đến (f_2, f_2) tạo nên từ π và θ như sau:

$$(s_2, s_2) \xrightarrow{x_1} (f_2, f_2) \xrightarrow{\varepsilon} (s_2, s_2) \xrightarrow{x_2} (f_2, f_2) \xrightarrow{\varepsilon} (s_2, s_2) \ldots (s_2, s_2) \xrightarrow{x_l} (f_2, f_2),$$

ở đó, trên ρ không có đỉnh kiểm soát. Hay nó cách khác, trên G_4 không có đường đi hợp lệ.

($\Leftarrow$) Trên G_4 không có đường đi hợp lệ, suy ra mọi đường đi ρ từ (s_2, s_2) đến (f_2, f_2) không có đỉnh kiểm soát (f_2, q), với $q \neq f_2, s_2$. Tương ứng với mỗi đường đi ρ, trên $\mathcal{A}_2$ có đường đi thành công π với nhãn $w = x_1...x_l$:

$$s_2 \xrightarrow{x_1} f_2 \xrightarrow{\varepsilon} s_2 \xrightarrow{x_2} f_2 \xrightarrow{\varepsilon} s_2 \ldots s_2 \xrightarrow{x_l} f_2,$$

và trên $\mathcal{A}_2$ có đường đi thành công θ với nhãn $y_1...y_m$:

$$s_2 \xrightarrow{y_1} f_2 \xrightarrow{\varepsilon} s_2 \xrightarrow{y_2} f_2 \xrightarrow{\varepsilon} s_2 \ldots s_2 \xrightarrow{y_m} f_2,$$

vì ρ không có đỉnh kiểm soát nên với mọi l hữu hạn và $x_1, ..., x_l, y_1, ..., y_m \in X$, nếu có

$$w = x_1...x_l = y_1...y_m \quad \text{thì suy ra} \quad l = m \text{ và } x_i = y_i, \text{ với } i = 1, .., l.$$

Vậy X có độ không nhập nhằng vô hạn $k = \infty$. ■

Từ Định lý 4.1 và Bổ đề 4.1 ở trên, cho phép ta xây dựng giải thuật xác định độ không nhập nhằng của ngôn ngữ chính quy $X \subseteq \Sigma^*$ dưới đây:

Giải thuật UnambDe($\mathcal{A}$)

Input: Ôtômat hữu hạn $\mathcal{A}$ (n đỉnh, m cung) và $X = \mathcal{L}(\mathcal{A}) \subseteq \Sigma^*$.

Output: Độ không nhập nhằng của X.

1. If $Epsilon(\mathcal{A})$ then Return 0;
2. $\mathcal{A}_1 = D(\mathcal{A})$; $\mathcal{A}_2 = Ex(\mathcal{A}_1)$;
3. $\mathcal{A}_3 = Prod(\mathcal{A}_1, \mathcal{A}_2)$; //cỡ n^2 trạng thái và m^2 cung
4. If có đường đi thành công trên $\mathcal{A}_3$ đi qua (p, f_2), với $p \neq f_1, s_1$ then
   Return 0; // Độ không nhập nhằng của X là 0
5. $\mathcal{A}_4 = Prod(\mathcal{A}_2, \mathcal{A}_2)$; //cỡ n^2 trạng thái và m^2 cung
6. $G = XCopy(\mathcal{A}_4)$; // cỡ $2n^2$ đỉnh và $3m^2$ cung
7. If có đường đi ngắn nhất từ $((s_2, s_2), 1)$ đến $((f_2, f_2), 2)$ trên G độ dài k then
   Return k;
   else Return ∞. //Không có đường đi từ $((s_2, s_2), 1)$ đến $((f_2, f_2), 2)$ trên G.

Đánh giá độ phức tạp thời gian của giải thuật:

Độ phức tạp thời gian của bước 1 là $\mathcal{O}(n)$, bước 2 là $\mathcal{O}(n+m)$, bước 3 và 5 là $\mathcal{O}((n+m)^2)$, bước 6 là $\mathcal{O}(n^2 + m^2)$.

Để thực hiện bước 4, ta áp dụng giải thuật $DFS(DepthFirstSearch)$ trong [12] như sau: trên $\mathcal{A}_3$, áp dụng giải thuật DFS từ (s_1, s_2) tìm đến các trạng thái (p, f_2), với $p \neq f_1, s_1$ và đánh dấu; xây dựng $\mathcal{A}_3^R$ (nhận được từ $\mathcal{A}_3$ nhờ đảo ngược các cung), áp dụng giải thuật DFS từ (f_1, f_2) tìm đến các trạng thái đã đánh dấu. Nếu tìm được ít nhất một trạng thái đã đánh dấu thì ta có đường đi thành công trên $\mathcal{A}_3$ đi qua ít nhất một trạng thái (p, f_2), với $p \neq f_1, s_1$. Toàn bộ bước 4 có độ phức tạp thời gian là $\mathcal{O}(n^2 + m^2)$.

Bước 7, ta dùng giải thuật $Dijkstra$ trong [12] để tìm đường đi ngắn nhất từ $((s_2, s_2), 1)$ đến $((f_2, f_2), 2)$ trên G. Bước này có độ phức tạp thời gian $\mathcal{O}(n^2 logn^2 + m^2)$.

Tổng hợp lại, giải thuật xác định độ không nhập nhằng của ngôn ngữ chính quy có độ phức tạp thời gian $\mathcal{O}(n^4)$, ở đây ta xem $\mathcal{O}(m) = \mathcal{O}(n^2)$ khi đồ thị dày cạnh.

Trường hợp $\mathcal{A}$ là ôtômat đơn định có n trạng thái: vì theo Nhận xét 3.1 ta có cỡ trạng thái và cung của ôtômat $\mathcal{A}_2$ là $\mathcal{O}(n)$, khi đó ôtômat tích $\mathcal{A}_4$ có số trạng thái và cung cỡ $\mathcal{O}(n^2)$ nên bước 7 có độ phức tạp thời gian $\mathcal{O}(n^2 logn^2 + m^2)$, hay $\mathcal{O}(n^2 logn)$. Cũng theo Nhận xét 3.1, từ bước 1 đến bước 6 có độ phức tạp thời gian $\mathcal{O}(n^2)$. Tổng hợp lại trong trường hợp $\mathcal{A}$ đơn định, giải thuật có độ phức tạp thời gian $\mathcal{O}(n^2 logn)$ nếu coi lực lượng của bảng chữ cái Σ là hằng số.

Hệ quả 4.1. *Giải thuật UnambDe xác định chính xác độ không nhập nhằng của ngôn ngữ chính quy được đoán nhận bởi ôtômat $\mathcal{A}$ có n trạng thái, với độ phức tạp thời gian là $\mathcal{O}(n^4)$ nếu $\mathcal{A}$ là đa định, là $\mathcal{O}(n^2 logn)$ nếu $\mathcal{A}$ đơn định.*

5. KẾT LUẬN

Nghiên cứu các mô hình ôtômat nâng cao và ứng dụng của nó là một trong các xu hướng nghiên cứu hiện đại được nhiều nhà khoa học - công nghệ quan tâm. Trong bài báo này, các bài toán có ý nghĩa về lý thuyết cũng như ứng dụng thực tiễn được nghiên cứu bao gồm:

- Giới thiệu và nghiên cứu về độ không nhập nhằng của ngôn ngữ, đưa ra một phân bậc mịn và chặt trên lớp các ngôn ngữ có độ không nhập nhằng từ 0 đến ∞. Điều đó cho thấy lớp các ngôn ngữ có độ không nhập nhằng cao thật rộng lớn, tuy không là mã, nhưng vẫn có khả năng mã hóa thông tin mật, nâng cao khả năng chống tấn công, tiềm năng ứng dụng lớn. Nghiên cứu đặc trưng của lớp các ngôn ngữ và phép toán trên chúng cũng là các chủ đề lý thú, các vấn đề này sẽ được nghiên cứu trong các công trình tiếp theo.

- Đề xuất giải thuật xác định độ không nhập nhằng của ngôn ngữ chính quy được đoán nhận bởi ôtômat hữu hạn nhờ kỹ thuật sao chép đồ thị, với độ phức tạp thời gian là đa thức. Điểm mạnh của thuật toán này là áp dụng cho cả ôtômat đơn định hoặc đa định, ta không phải chuyển từ ôtômat đa định sang đơn định với giá phải trả cho thuật toán chuyển đổi có độ phức tạp thời gian cỡ lũy thừa. Thuật toán này cũng cho biết ngôn ngữ được kiểm định có là mã hay không, do đó cũng xem như một thuật toán kiểm tra mã.

TÀI LIỆU THAM KHẢO

[1] J. Berstel, D. Perrin, *Theory of Codes*, Academic Press, New York (1985).

[2] M. Mohri, Edit-Distance of Weighted Automata: General Definitions and Algorithms, *International Journal of Foundations of Computer Science* **14** 6 (2003) 957–982.

[3] G. Lallement, *Semigroups and Combinatorial Theory*, Wiley, New York, 1979.

[4] M. Mohri, F. Pereir7a, M. Riley, *Speech Recognition with Weighted Finite-State Transducers*, Springer Handbook of Speech Processing, Springer, 2007.

[5] E. N. Gilbert and E. F. Moore, *Variable length binary encodings*, Bell System Tech, J. (1959) 933—967.

[6] J. Devolder, M. Latteux, I. Litovsky, L. Staiger, Codes and infinite words, *Acta Cybernetica* **11** 4 (1994), Szeged.

[7] M. Anselmo, A non-ambiguous language factorization problem, *In Proceedings of Developments in Language Theory* (1999) 141–152.

[8] M. Anselmo, A non-ambiguous decomposition of regular languages and factorizing codes, *Discrete Applied Mathematics* **126** Issues 2-3 (2003) 129–165.

[9] D.L. Van, B. Lesaëc and I. Litovsky, On coding morphisms for zigzag codes, *Informatique théorique et applications* **26** 6 (1992) 565–580.

[10] M. Anselmo, On zigzag codes and their decidability, *Theoretical Computer Science* **74** (1990) 341—354.

[11] A. Mateescu, G.D. Mateescu, G. Rozenberg and A. Salomaa, *Shuffle–Like Operations on ω–words*, New Trends in Formal Languages, Lecture Notes in Computer Science, Springer–Verlag, Berlin, Heidelberg, **1218** (1990) 395—411.

[12] Cormen, Thomas H.; Leiserson, Charles E., Rivest, Ronald L., Stein, Clifford. *Introduction to Algorithms (3rd ed.)*. MIT Press and McGraw-Hill (1990), ISBN 0-262-03384-4.

[13] Phan Trung Huy, Vũ Thành Nam, Về một hình thức mã mới, *Kỷ yếu Hội thảo Quốc gia lần thứ 6 tại Thái Nguyên, tháng 8/2003 "Một số vấn đề chọn lọc của công nghệ thông tin"*, NXB Khoa học Kỹ thuật, (2003) 164—170.

[14] Phan Trung Huy, Vũ Thành Nam, Mã luân phiên và mã tiền ngữ cảnh, *Kỷ yếu Một số vấn đề chọn lọc của công nghệ thông tin, lần thứ 8 tại Dà Nẵng*, NXB Khoa học Kỹ thuật, (2004) (188-–197).

[15] Hồ Ngọc Vinh, Nguyễn Đình Hân, Phan Trung Huy, Mã với tích biên và độ trễ giải mã, *Tạp chí Công nghệ Thông tin và Truyền thông* **V-1** 4 (**24**) (2010) 46–56.

[16] Nguyễn Đình Hân, Hồ Ngọc Vinh, Phan Trung Huy, Đỗ Long Vân, Thuật toán xác định tính chất mã của ngôn ngữ chính quy, *Tạp chí Tin học và Điều khiển học* **27** 1 (2010) 1—9.

Nhận bài ngày 05 - 12 - 2011
Nhận lại sau sửa ngày 15 - 04 - 2012