

Another Approach to The Finiteness Problem in Context-free Grammars.

Lee, Myung-Joon

Dept. of Computer Science

(Received April 30, 1985)

〈Abstract〉

In this paper the general finiteness problem in context-free grammars is defined and related definitions and theorems are developed. The solvability of this problem is showed by the property of φ relation which is the key concept of the developed formalism. Moreover a testing algorithm for this problem can be easily derived by this formalism.

Context-free grammar에서의 유한성문제에 대한 접근방식을 달리한 연구

이 명 준

전자계산학과

(1985. 4. 30 접수)

〈요 약〉

본 논문에서는 Context-free grammar에서의 일반적 유한성 문제가 정의되었으며 그와 관련있는 새로운 정의 및 정리가 연구되었다. 이 문제의 해결성은 주로 ω relation을 이용하여 증명되었는데 이 φ relation의 개념이 본 연구의 전개에 있어서 가장 중요한 역할을 차지하게 된다. 또한 관련 문제에 대한 testing algorithm은 연구전개과정에 나타나게 되는 여러가지 결과에 의해 간단하면서도 효율적으로 형성될 수 있다.

which will be defined later.

I. Introduction

The finiteness problem in a grammar G is "Can it be determined that $L(G)$ is finite or not?" The solvability of this problem in context-free grammars is a known property.^(2,4) These traditional approaches are based on "uvwxy theorem". But possible testing algorithms developed from these approaches are very inefficient, and this inefficiency will be enlarged in the environments of the general finiteness problem in context-free grammars

We will introduce some new definitions and developed theorems, where φ relation is the key concept of our formalism and will play an important role in proving the solvability of the general finiteness problem in context-free grammars and in developing efficient testing algorithms.

II. Notations and Preliminaries

The basic definitions and theorems in this paper are consistent with those of Aho and

Ullman.⁽¹⁾ Notational conventions are also stated.

A context-free grammar (CFG) is a quadruple (N, T, P, S) , where N, T, P , and S stand, respectively, for a set of nonterminal symbols, a set of terminal symbols, a set of productions (each of which is of the form $A \rightarrow \alpha$), and a start symbol in N . Given a grammar G , V (the vocabulary) stands for $N \cup T$ and V^* for the reflexive-transitive closure of V and the transitive closure is denoted by V^+ . The language generated by G is denoted by $L(G)$.

Lower-case Greek letters such as α, β, γ and ω denote strings in V^* , lower-case Roman letters toward the beginning of the alphabet (a, b , and c) are terminals, whereas those near the end (u, v , and w) are strings in T^* : upper-case letters toward the beginning of the alphabet (A, B , and C) are nonterminals, whereas those near the end (X, Y , and Z) are symbols in V . An empty string is denoted by Λ .

[Definition 2.1] We say that a symbol $X \in V$ is useless in a CFG $G = (N, T, P, S)$ if there does not exist a derivation of the form $S \xrightarrow{*} wXy \xrightarrow{*} wxy$ where w, x , and y in T^* .

[Definition 2.2] We say that a CFG $G = (N, T, P, S)$ is Λ -free if either

- (1) P has no Λ -productions, or
- (2) There is exactly one Λ -production $S \rightarrow \Lambda$ and S does not appear on the right side of any production in P .

[Definition 2.3] We say that a CFG $G = (N, T, P, S)$ is cycle-free if there is no derivation of the form $A \xrightarrow{+} A$ for any A in N .

[Definition 2.4] We say that G is proper if it is cycle-free, is Λ -free, and has no useless symbols.

[Theorem 2.1] Given a CFG G , we can find a proper CFG G' which is $L(G) = L(G')$ and equivalently $L(G, S) = L(G', S)$ if we use the notation of Definition 3.1.

[Definition 2.5] A labeled ordered tree D is a

derivation tree with root A for a CFG if

- (1) The root of D is labeled A .
- (2) If D_1, \dots, D_k are the subtrees of the direct descendants of the root of D , D_i is labeled X_i , then $A \rightarrow X_1 \dots X_k$ is a production in P . D_i must be a derivation tree with root X_i for G if X_i is a nonterminal, and D_i is a single node labeled X_i if X_i is a terminal.
- (3) alternatively, if D_1 is the only subtree of the root D and the root of D_1 is labeled A , then $A \rightarrow A$ is a production in P .

III. General Finiteness Problem in Context-free Grammars.

In this section, we will introduce the extended language concept in G and the general finiteness problem in CFG and related new definitions and developed theorems.

[Definition 3.1]

Let $\alpha = X_1 X_2 \dots X_n \in V^+$. The language generated by α in G , denoted by $L(G, \alpha)$, is defined to be

$$L(G, \alpha) = \{ \omega \mid \omega \in T^* \text{ and } \omega = \omega_1 \omega_2 \dots \omega_n, \}$$

where $X_i \xrightarrow{*} \omega_i$ for each $i, 1 \leq i \leq n$.

In addition, the language generated by Λ is $L(G, \Lambda) = \{ \Lambda \}$. This definition comes from (3). And the traditional language notation $L(G)$ with $G = (N, T, P, S)$ can be denoted by $L(G, S)$. The following Lemma is immediate from Definition 3.1 and the well-known language concatenation.

[Lemma 3.1] $L(G, \alpha_1 \alpha_2 \dots \alpha_n) = L(G, \alpha_1) \cdot L(G, \alpha_2) \dots L(G, \alpha_n)$

[Definition 3.2] The general finiteness problem in CFG is "Can it be determined for a given CFG G that $L(G, \alpha)$ is finite or infinite?"

[Definition 3.3] Let φ be a relation such that $\varphi \subseteq N \times N$ and $A \varphi B$ if and only if $A \rightarrow \alpha B \beta \in P$. The reflexive transitive closure of φ is denoted by φ^* and the transitive closure of φ is denoted

by φ^+ .

Then the following Lemma is obvious by the concept of \Rightarrow .

[Lemma 3.2] $A\varphi^+B$ if and only if $A\overset{+}{\Rightarrow}\alpha B\beta$.

[Lemma 3.3] Let G be a proper CFG.

Then $L(G, A)$ is infinite if $A\varphi^+A$.

proof. Assume that $A\varphi^+A$. Then there exists a derivation

$A\overset{k}{\Rightarrow}\alpha A\beta$ for some $k \geq 1$ since by Lemma 3.2, $A\overset{+}{\Rightarrow}\alpha A\beta$. Since this derivation step can be applied infinitely, there exists a derivation $A\overset{n}{\Rightarrow}\alpha^n A\beta^n$ for any $n \geq 1$. Hence there exists a derivation $A\overset{n}{\Rightarrow}\alpha^n A\beta^n \overset{+}{\Rightarrow} W_1 W_2 W_3$, where $\alpha^n \overset{*}{\Rightarrow} W_1$, $A \overset{*}{\Rightarrow} W_2$, $\beta^n \overset{*}{\Rightarrow} W_3$, where $W_i \in T^*$, since G has no useless symbols. Notice that either α or β can not be Λ since G is cycle-free. Hence Length of $W_1 W_2 W_3$ is greater than or equal to $(n+1)$ by the fact that G is Λ -free. Therefore the length of $W_1 W_2 W_3$ becomes infinity when n becomes infinity. This completes the proof.

[Theorem 3.1] Let G be a proper CFG. Then

$L(G, B)$ is infinite if and only if there exists a nonterminal A such that $B\varphi^*A$ and $A\varphi^+A$.

proof. Suppose that there exists a nonterminal A such that $B\varphi^*A$ and $A\varphi^+A$. Then there exists a derivation $B\overset{+}{\Rightarrow}\alpha A\beta$, and from this, $L(G, B) \supseteq L(G, \alpha A\beta)$ is immediate from Definition 3.1.

But, $L(G, \alpha A\beta)$ is infinite since

$$L(G, \alpha A\beta) = L(G, \alpha) \cdot L(G, A) \cdot L(G, \beta)$$

(by Lemma 3.1) and

$L(G, A)$ is infinite (by $A\varphi^+A$ and

Lemma 3.3).

Thus $L(G, B)$ is infinite. This proves the if-part of this theorem.

To complete our proof, we will prove the contraposition of the only-if part. Suppose that there does not exist a nonterminal A such that $B\varphi^*A$ and $A\varphi^+A$. Then the depth of any derivation tree with root B for G is at most the number of nonterminals in N since any

path from the root to a leaf can have each nonterminal at most once. Further any node in a derivation tree can have a finite number of child nodes since the length of the righthand side in any production is finite. Thus $L(G, B)$ is finite.

This complete our proof.

[Theorem 3.2] Let G be a proper CFG. Then $L(G, \alpha)$ is infinite if and only if there exist a nonterminal B in N such that

$\alpha = X_1 X_2 \dots X_n$, $B = X_i$ for some i ($1 \leq i \leq n$) and $B\varphi^*A$ and

$A\varphi^+A$ for some nonterminal A in N .

proof. Immediate from Lemma 3.1 and Theorem 3.1

And the following Lemma is trivial from Lemma 3.1

[Lemma 3.4] Let G be a CFG. Then $L(G, \alpha)$ is infinite if and only if there exists a nonterminal B such that

$\alpha = X_1 X_2 \dots X_n$, $B = X_i$ for some i ($1 \leq i \leq n$)

and $L(G, B)$ is infinite.

[Lemma 3.5] Let $G = (N, T, P, S)$ be a CFG.

Then it can be determined that $L(G, B)$ is finite or infinite for any nonterminal B in N .

proof. In virtue of Theorem 2.1 and Definition 3.1 we can find a proper CFG G' such that $L(G, B) = L(G', B)$ for any B in N . And then by Theorem 3.1, we can determine $L(G', B)$ is finite or not.

Consequently Lemma 3.4 and Lemma 3.5 give us the following theorem.

[Theorem 3.3] The general finiteness problem in CFG is solvable.

7. Conclusion

The general finiteness problem in CFG is defined with the extended language concept in G , and the solvability of the general finiteness problem in CFG is showed by the property of φ relation which is the key concept of our formalism. Moreover, a testing algorithm for

the problem can be easily derived by using a transformation algorithm from a *CFG* to the corresponding proper *CFG* and an algorithm for computing the transitive closure of φ .

References

1. Aho, A.V., and Ullman, J.D., "The Theory of parsing, Translation, and Compiling," Volume 1: Parsing, Prentice-Hall, Inc. 1972.
2. Hopcroft, J.E., and Ullman, J.D., "Formal Languages and their relation to automata," Addison-Wesley, Inc. 1969.
3. Back house, L.C., "Syntax of Programming languages," Prentice-Hall, Inc. 1979.
4. Harrison, M.A., "Introduction to Formal Language Theory." Addison-Wesley, Inc. 1978.