A representation of context-free grammars with the help of finite digraphs

Krasimir Yordzhev

Faculty of Mathematics and Natural Sciences, South-West University, Blagoevgrad, Bulgaria

Email address: yordzhev@swu.bg (K. Yordzhev)

To cite this article:
Krasimir Yordzhev. A Representation of Context-free Grammars with the Help of Finite Digraphs, American Journal of Applied Mathematics. Vol. 1, No. 1, 2013, pp. 8-11. doi: 10.11648/j.ajam.20130101.12

Abstract: For any context-free grammar, we build a transition diagram, that is, a finite directed graph with labeled arcs, which describes the work of the grammar. This approach is new, and it is different from previously known graph models. We define the concept of proper walk in this transition diagram and we prove that a word belongs to a given context-free language if and only if this word can be obtained with the help of a proper walk.

Keywords: Context-Free Grammar, Finite Digraph, Transition Diagram, Proper Walk, Chomsky Normal Form

1. Introduction

Context-free grammars are widely used to describe the syntax of programming languages and natural languages [1,2]. On the other hand, graph theory is a good apparatus for the modeling of computing devices and computational processes. So a lot of graph algorithms have been developed [4,5]. The purpose of this article is to show how for an arbitrary context-free grammar, we can construct a finite digraph, which describes the work of the grammar. As is well known, any finite state machine (deterministic or nondeterministic finite automaton) and any pushdown automaton can be described by a transition diagram, which is essentially a finite directed graph with labeled arcs [2]. The finite digraph, which we will construct in this work and who will simulate a given context-free grammar will be different from previously known graph models [1,2].

2. The Main Definitions and Notations

If

\[ \Sigma = \{x_1, x_2, \ldots, x_n\} \]

is a finite set, then with \( \Sigma^* \) we will denote the free monoid over \( \Sigma \), i.e., the set of all sequences \( x_{i_1}x_{i_2}\ldots x_{i_n} \), including the empty one which we denote with \( \varepsilon \). Elements of \( \Sigma^* \) are called words, and any subset of \( \Sigma^* \) (formal) language.

Formal (or generative) grammar \( \Gamma \) is a triple

\[ \Gamma = (N, \Sigma, \Pi), \]

where \( N \) and \( \Sigma \) are finite sets, \( N \cap \Sigma = \emptyset \), which are called, respectively, alphabet of nonterminals (or variables) and alphabet of terminals, where \( \Pi \) is a finite subset of the Cartesian product \( (N \cup \Sigma)^* \times (N \cup \Sigma)^* \), and we will write \( u \rightarrow v \) if \( (u, v) \in \Pi \) (see [3]). We will call \( \Pi \) the set of productions or rules.

Let \( \Gamma = (N, \Sigma, \Pi) \) be a formal grammar and let \( w, y \in (N \cup \Sigma)^* \). We will write \( w \rightarrow y \), if there are \( z_1, z_2, u, v \in (N \cup \Sigma)^* \), such that \( w = z_1uz_2 \) and \( y = z_1vz_2 \) and \( u \rightarrow v \) is a production of \( \Pi \). We will write \( w \Rightarrow y \), if \( w = y \), or there are \( w_0, w_1, \ldots, w_r \) such that \( w_0 = w \) and \( w_i \rightarrow w_{i+1} \) for all \( i = 0, 1, \ldots, r - 1 \). The sequence \( w_0 \rightarrow w_1 \rightarrow \cdots \rightarrow w_r \)

is called a derivation of length \( r \).

For every \( A \in N \), a subset \( L(\Gamma, A) \subseteq \Sigma^* \) defined by

\[ L(\Gamma, A) = \{ \alpha \in \Sigma^* | A \Rightarrow \alpha \} \]

is called the language generated by the grammar \( \Gamma \) with the start symbol \( A \) (see [3]).

The formal grammar \( \Gamma' = (N, \Sigma, \Pi) \) is called context-free if all the productions of \( \Pi \) are of the form \( A \rightarrow \omega \), where \( A \in N \), \( \omega \in (N \cup \Sigma)^* \). The language \( L \) is called context-free if there is a context-free grammar \( \Gamma' = (N', \Sigma, \Pi) \) and a nonterminal \( A \in N \), such that \( L = L(\Gamma', A) \).

Transition diagram is called a finite directed graph, all of whose arcs are labelled by an element of a semigroup (see also [2]).

If \( \pi \) is a walk in a transition diagram, then the label of this walk \( l(\pi) \) is the product of the labels of the arcs that
make up this walk, taken in passing the arcs.

Let \( \Gamma = (N, \Sigma, \Pi) \) be a context-free grammar and let

\[ N' = \{ A' \mid A \in N \} \times \{ x' \mid x \in \Sigma \}. \]

We consider the monoid \( T \) with the set of generators \( N \cup \Sigma \cup N' \cup \Sigma', \) and with the defining relations

\[ A'A = \varepsilon, \quad x'x = \varepsilon, \quad (1) \]

where \( A' \in N', \quad A \in N, \quad x' \in \Sigma', \quad x \in \Sigma \) and \( \varepsilon \) is the empty word (the identity of the monoid \( T \)). We set by definition \((x')^i = x, (A')^i = A,\) if \( \omega = z_1 \ldots z_k \in T,\) then \( \omega' \) will mean \( z'_1 \ldots z'_{i-1} \ldots z'_i \) and \( \varepsilon' = \varepsilon.\)

Let

\[ R = \Sigma' \times T = \{(\omega, z) \mid \omega \in \Sigma^*, z \in T\}. \]

In \( R \) we introduce the operation \( * \) as follows: If \((\omega_1, z_1), (\omega_2, z_2) \in R,\) then

\[ (\omega_1, z_1) * (\omega_2, z_2) = (\omega_1 \omega_2, z_2 z_1). \quad (2) \]

It is easy to see that \( R \) with this operation is a monoid with identity \((\varepsilon, \varepsilon).\)

3. Finite Digraphs and Context-free Grammars

**Definition 1.** For the context-free grammar \( \Gamma, \) we construct a transition diagram \( H(\Gamma) \) with the set of vertices \( V = \{ u_A, v_A \mid A \in N \} \) and the set of arcs \( E \subseteq V \times V \) which is made as follows:

1. For each production \( A \rightarrow \alpha, \) where \( A \in N, \alpha \in \Sigma^* \) there is an arc from \( u_A \) to \( v_A \) with a label \((\alpha, \varepsilon)\) (see Fig. 1);
2. For each production \( A \rightarrow \alpha B \) where \( A, B \in N, \alpha \in \Sigma^*, z \in (N \cup \Sigma)^* \) there is an arc from \( u_A \) to \( u_B \) with a label \((\alpha, z); \)
3. For each production \( A \rightarrow z_1 B_1 \alpha_1 B_2 z_2 \) where \( A, B_1, B_2 \in N, \alpha \in \Sigma^*, z_1, z_2 \in (N \cup \Sigma)^* \) there is an arc from \( v_B_1 \) to \( v_B_2 \) with a label \((\alpha, B'_2 \alpha'); \)
4. For each production \( A \rightarrow z B \alpha, \) where \( A, B \in N, \alpha \in \Sigma^*, z \in (N \cup \Sigma)^* \) there is an arc from \( v_B \) to \( v_A \) with a label \((\alpha, \varepsilon).\)
5. There are no other arcs in \( H(\Gamma) \) except described in 1 + 4.

![Figure 1. Production A \rightarrow \alpha, A \in N, \alpha \in \Sigma^* and the corresponding part of the transition diagram H(\Gamma).](image1)

If the production is of the form

\[ A \rightarrow \alpha_0 B_1 \alpha_1 B_2 \alpha_2 \ldots \alpha_{k-1} B_k \alpha_k, \]

where \( \alpha_i \in \Sigma^* \) ( \( i = 0,1, \ldots k), A, B_j \in N \) ( \( j = 1,2, \ldots, k), \) then the corresponding part of the transition diagram is shown in Figure 2.

![Figure 2. Production A \rightarrow \alpha_0 B_1 \alpha_1 B_2 \alpha_2 \ldots \alpha_{k-1} B_k \alpha_k and the corresponding part of the transition diagram H(\Gamma).](image2)

3. There are no other proper walks in \( H(\Gamma) \) except described in 1 and 2.

The following theorem is true.

**Theorem 1.** Let \( \Gamma = (N, \Sigma, \Pi) \) be a context-free grammar and let \( A \in N. \) Then \( \alpha \in L(\Gamma, A) \) if and only if there is a proper walk in \( H(\Gamma) \) with the start vertex \( u_A, \) the final vertex \( v_A \) and with a label \((\alpha, \varepsilon).\)

Proof. Necessity. Let \( \alpha \in L(\Gamma, A). \) Then there is a derivation \( A \Rightarrow \alpha. \) The proof we will make by induction on the length of the derivation.

Let a production \( A \rightarrow \alpha \) exists. Then according to Definition 1, item 1 there exists an arc \( u_A v_A \) with a label \((\alpha, \varepsilon).\)

Suppose that for all \( A \in N \) and for all \( \alpha \in L(\Gamma, A) \) for
which there is a derivation $A \Rightarrow \alpha$ no longer than $t$, there is a proper walk with the start vertex $u_A$, the final vertex $v_A$ and labeled $(\alpha, \varepsilon)$.

Let $\alpha \in L(\Gamma, A)$, and there is a derivation $A \Rightarrow \alpha$ of length $t + 1$. Let $A \Rightarrow \alpha_0B_1\alpha_1 \ldots B_k\alpha_k$ be the first production of this derivation, where $\alpha_i \in \Sigma^*$. B $\in$ N, $i = 0, 1, \ldots, k$, $j = 1, 2, \ldots, k$. Then there is a derivation $A \Rightarrow \alpha_0B_1\alpha_1 \ldots B_k\alpha_k \Rightarrow \alpha_0\beta_1\alpha_1 \ldots \beta_k\alpha_k = \alpha$,

where $\beta_1 \in L(\Gamma, B_1)$ and for all $\beta_i$ there exist derivations $B_i \Rightarrow \beta_i$ with the length less than or equal to $t$, $i = 1, 2, \ldots, k$.

For the first production $A \Rightarrow \alpha_0B_1\alpha_1 \ldots B_k\alpha_k$ of the derivation $A \Rightarrow \alpha$ we have:

1. According to Definition 1, item 2 there exists an arc $\rho$ from $u_A$ to $u_{B_1}$, with a label $((\alpha_0, \alpha_1B_2 \ldots B_k)\alpha)$. Let $\alpha \in L(\Gamma, A)$, and there are

$$l(\pi_1) = \ldots \alpha_k \beta \alpha \ldots \beta_k \alpha,$$

where $\alpha \in L(\Gamma, A)$.

Sufficiency. Let $\pi$ be a proper walk in $H(\Gamma)$ with the start vertex $u_A$, the final vertex $v_A$ and labelled $l(\pi) = (\alpha, \varepsilon)$. The necessity is proved.

Suppose that for all $\alpha \in N_A$ and for all proper walks with the start vertex $u_A$, the final vertex $v_A$, length less than or equal to $t$ and with a label $(\alpha, \varepsilon)$ is satisfied $\alpha \in L(\Gamma, A)$. Let $\pi$ be a proper walk with the start vertex $u_A$, the final vertex $v_A$, labelled $(\alpha, \varepsilon)$ and length $t + 1$. Since $\pi$ is a proper walk, then there is a vertex $B_i \in N$ such that the first arc $\rho$ in the walk $\pi$ starts at $u_A$ and finishes at $u_{B_i}$. Hence there is a production $A \Rightarrow \alpha_0B_1\alpha_1 \ldots B_k\alpha_k$ in $\Gamma$ such that $l(u_{B_i}) = (\alpha_0, \alpha_1B_2 \ldots B_k)\alpha$ for some $B_i \in N$, $\alpha_j \in \Sigma^*$, $j = 2, 3, \ldots, k$. According to Definition 2, $\pi$ is given by $\pi = \rho\pi_1\pi_2 \ldots \pi_k\tau$, where $\pi_i$ are proper walks with the start vertices $u_{C_i}$ and the final vertices $v_{C_i}$ for some $C_i \in N$ (obviously $C_1 = B_1$).

Such a grammar is said to be in Chomsky Normal Form. As is well known [2], every context-free language can be generated by a grammar in Chomsky normal form.

An algorithm that recognizes whether a word $\alpha \in \Sigma^*$ belongs to the language $L$, that is generated by a grammar in Chomsky normal form. This algorithm works in time $O(n^2)$, where $n = |\alpha|$. Described in section 3 graph model will help to understand in detail this algorithm for more accurate description.

If the production is of the form $A \Rightarrow BC$, i.e. it is one of the productions of grammar in Chomsky normal form, then the appropriate fragment of the transition diagram is shown in Figure 4.

Obviously, if we consider a context-free grammar in Chomsky normal form, then the corresponding transition diagram is in quite simple and convenient form.
5. Conclusions and Future Work

Described in this article a graph model of an arbitrary context-free grammar is new and original. This is different from the familiar until now graph models.

Some polynomial algorithms for testing the inclusion of any regular or linear language in a group language are described in [7]. These algorithms are based on the search for a cycle in the corresponding transition diagram. This idea can be used for any context-free language \( L \) by the transition diagram described in Section 3. The difficulty here is to find, in general, the proper walk, which corresponds to any given word \( \alpha \in L \). We hope this open problem to be solved in the near future.

The research is partly supported by the project SRP-B3/13, funded by the Ministry of Education, Youth and Science (Bulgaria) and South-West University “Neofit Rilsky”, Blagoevgrad.

References

[1] A.V. Aho, J.D. Ullman, The theory of parsing, translation and computing. Vol. 1, 2, Prentice-Hall, 1972.
[2] J.E. Hopcroft, R. Motwani, J.D. Ullman, Introduction to automata theory, languages, and computation. Addison-Wesley, 2001.
[3] G. Lallemant, Semigroups and combinatorial applications. John Wiley & Sons, 1979.
[4] I. Mirchev, Graphs. Optimization algorithms in networks. SWU “N. Rilsky”, Blagoevgrad, 2001.
[5] M. Swami, K. Thulasirman, Graphs, networks and algorithms. John Wiley & Sons, 1981.
[6] K. Yordzhev, An n^2 Algorithm for Recognition of Context-free Languages. Cybern. Syst. Anal. 29, No.6, (1993), 922-927.
[7] K. Yordzhev, Inclusion of Regular and Linear Languages in Group Languages. International J. of Math. Sci. & Engg. Appls. (IJMSEA) ISSN 0973-9424, Vol. 7 No. 1 (2013), 323-336.