Solutions

October 17, 2013

1 Problem 3

We will say that a grammar has a cycle if there is a reachable non-terminal A such that $A \stackrel{+}{\Rightarrow} A$, i.e. it is possible to derive the nonterminal A from A by a nonempty sequence of production rules.

Show that if a grammar has a cycle, then it is not LL(1).

1.1 Solution 1

A grammar is not LL(1) if a symbol A is not nullable and there are two derivations rules for which the first sets are not disjoint. We will exhibit such rules in our grammar.

Let the rules of the cycle be the following, where L_i and R_i are sequences of non-terminals:

```
A1 -> L2 A2 R2
A2 -> L3 A3 R3
...
An -> L1 A1 R1
```

Because A1 rewrites to A1, it requires all Li and Ri to be nullable.

Then, first(A2) $\subset$ first(A1), and because it is true for all Ai, it follows that all first(Ai) are equal.

Because we suppose that A is productive, there exists at least one of the Ai which is directly productive, which means we can derive a word from it without using again one of the rules.

```
Ai -> FF
Ai -> Li+1 Ai+1 Ri+1
```

$\emptyset \neq$ first(FF) $\subset$ first(Ai). Besides,
first(Ai) $\subset$ first(Ai+1) $\subset$ first(Li+1 Ai+1 Ri+1).
Therefore first(FF) $\cap$ first(Li+1 Ai+1 Ri+1) $\neq \emptyset$
which means that the grammar is not LL(1). QED.

1.2 Solution 2

There is a theorem, that asserts that ambiguous grammars cannot be LL(1). Therefore, if A is productive, then given a word where A is in the derivation tree, we could rewrite another derivation tree by replacing A by the chain $A \stackrel{+}{\Rightarrow} A$. So the grammar is ambiguous and it cannot be LL(1).

2 Problem 6

Assume a grammar in Chomsky normal form has n non-terminals. Show that if the grammar can generate a word with a derivation having at least 2^n steps, then the recognized language should be infinite.

2.1 Solution

Let G be such a grammar in Chomsky normal form with n non-terminals N_1... N_n. Let t the derivation tree of a word $w = w_1 w_2 \dots$ containing 2^n derivation steps, without counting the ultimate steps $N_i \rightarrow a$ where N_i is a non-terminal and a is a terminal.

Because of the Chomsky Normal Form, each non-terminating rule has the form $N_i \rightarrow N_j N_k$. Therefore, each step adds exactly one new non-terminal to the derivation. Because we start with one non-terminal N_1, after the derivation of w we end up with $2^n + 1$ non-terminals, which all yield terminals. So the size of w is $2^n + 1$.

We now show that in the derivation tree of w, there is a path from N_1 to one the terminals w_k that contains at least $n + 1$ non-terminals. If the maximum height was n, then the tree could produce at most 2^n terminals. Indeed, with a full binary tree, there would be 1 node at height 0, 2 nodes at height 1, 4 at height 2.... and 2^n at height n. Because we have $2^n + 1$ terminals at the end, it means that the height is at least $n + 1$. Therefore there is a path $N_1 \rightarrow \dots \rightarrow N_{i_{n+1}} \rightarrow w_k$.

Because there are only n different non-terminals, two of these non-terminals are the same, let us name it M. So $N_1 \rightarrow \dots \rightarrow M \rightarrow \dots \rightarrow M \rightarrow \dots N_{i_{n+1}} \rightarrow w_k$.

Therefore we can split the word w into: $w_I w_L w_M w_R w_F$

where w_M is the word derived by the second M and $w_L w_M w_R$ the word derived by the first M.

Because the first M does not immediately rewrites to the second M, at least one of the two w_L and w_R is not empty.

Therefore, by replacing the second derivation tree starting from the second M by the one starting by the first M, we can generate the word:

$w_I w_L^2 w_M w_R^2 w_F$ wich is in the recognized language.

By reccurrence, we show that $w_I w_L^i w_M w_R^i w_F$ is in the recognized language and is of size each time at least $2^n + i$.

Therefore the recognized language is infinite. QED.