Notes on Parsing as Logical Deduction

Chris Dyer

October 21, 2014

1 CKY Parser

Given a CFG in Chomsky Normal Form \((N, \Sigma, S, R)\), and a sentence \(x = x_1, x_2, \ldots, x_n\) the logic program for the CKY algorithm is:

- **items:** \([N, N, N]\)
- **goals:** \([S, 0, n]\)
- **axioms:** \([X, i, i + 1] (X \to x_i) \in R\)
- **inference rules:**
  - \([X, i, k] [Y, k, j] [Z, i, k]\)

An item \([X, i, j]\) says “starting with nonterminal \(X\), it is possible to derive the terminal string \(x_i, x_{i+1}, \ldots, x_j\).”

2 Earley Parser

Given an arbitrary CFG \((N, \Sigma, S, R)\), and a sentence \(x = x_1, x_2, \ldots, x_n\) the logic program for the Earley algorithm is:

- **items:** \([N \to (N \cup \Sigma)^* \bullet (N \cup \Sigma)^*, N, N]\)
- **goals:** \([S' \to S \bullet, 0, n]\)
- **axioms:** \([S' \to \bullet S, 0, 0]\)
- **inference rules:**
  - \([X \to \alpha \bullet x_{j+1} \beta, i, j]\)
  - \([X \to \alpha x_{j+1} \bullet \beta, i, j + 1]\) (SCAN)
  - \([X \to \alpha \bullet Y \beta, i, j]\) \((Y \to \gamma) \in R\) (PREDICT)
  - \([Y \to \bullet \gamma, j, j]\)
  - \([X \to \alpha \bullet Y \beta, i, j]\) \([Y \to \gamma \bullet, j, k]\)
  - \([X \to \alpha Y \bullet \beta, i, k]\) (COMPLETE)

Note: \(\alpha, \beta, \gamma\) may be arbitrary strings of terminals and nonterminals, including the empty string. An item \([X \to \alpha \bullet \beta, i, j]\) says, in English, “starting with nonterminal \(X\), it is possible to generate derive a sequence of terminal symbols \(x_i, x_{i+1}, \ldots, x_j\) using the \(\alpha\) part of the grammar rule \(X \to \alpha \beta\).”