Parsing of Context-Free Languages

Ota Jirák

December 19, 2007

Parsings systems

A parsing system $\mathbb{P}$ for some grammar G and string $a_1 \dots a_n$ is a tripple $\mathbb{P} = \langle \mathcal{I}, H, D \rangle$

- $\mathcal{I}$ is a set of items, called the *domain* or the *item set* of $\mathbb{P}$,
- H is a finite set of items called the *hypotheses* of $\mathbb{P}$,
- $D \subseteq_{\wp fin} (H \cup \mathcal{I}) \times \mathcal{I}$ is a set of deduction steps.
  We write $\eta_1, \dots \eta_k \vdash \xi$ or $(\eta_1, \dots, \eta_k, \xi)$

inference relation $\vdash$, deduction sequence

inference relation $\vdash$
Let $\mathbb{P} =< \mathcal{I}, H, D >$ be a parsing system. The relation $\vdash \subseteq_{\wp fin} (H \cup \mathcal{I}) \times \mathcal{I}$ is defined by $Y \vdash \xi$ if $(Y', \xi) \in D$ for some $Y' \subseteq Y$.

deduction sequence
Let $\mathbb{P} = \langle \mathcal{I}, H, D \rangle$ by a parsing system. We write $\mathcal{I}^+$ for the set of non-empty, finite sequences $\xi_1, \ldots, \xi_j$, with $j \geq 1$ and $\xi_i \in \mathcal{I}(1 \leq i \leq j)$.
A deduction sequence in $\mathcal{P}$ is a pair $(Y; \xi_1, \ldots, \xi_j) \in_{\wp} (H \cup \mathcal{I}) \times \mathcal{I}^+$, such that $Y \cup \xi_1, \ldots, \xi_{i-1} \dashv \xi_i$ for $1 \leq i \leq j$.
Informal notation $Y \vdash \xi_1 \vdash \cdots \vdash \xi_j$ instead of $(Y; \xi_1, \ldots, \xi_j)$.

set Δ
The set of deduction sequences $\Delta \subseteq_{\wp fin} (H\ cup \mathcal{I}) \times \mathcal{I}^+$ for a parsing system $\mathbb{P} = \langle \mathcal{I}, H, D \rangle$ is defined

$$\Delta = (Y; \xi_1, \ldots, \xi_j) \in_{\wp fin} (H \cup \mathcal{I}) \times \mathcal{I}^+ | Y \vdash \xi_1 \vdash \cdots \vdash \xi_j.$$

relation $\vdash^*$
For a parsing system $\mathbb{P} = \langle \mathcal{I}, H, D \rangle$ we define the relation $\vdash^*$ on $_{\wp fin}(H \cup \mathcal{I}) \times \mathcal{I}$ by

$$Y \vdash^* \xi \text{ if } \xi \in Y \text{ or } Y \vdash \cdots \vdash \xi.$$

Valid items

For a parsing system $\mathbb{P} = \langle \mathcal{I}, H, D \rangle$ the set of valid items is defined by

$\mathcal{V}(\mathbb{P}) = \{\xi \in \mathcal{I} | H \vdash^* \xi\}$.

Example - parsing system CYK 1/2

- Cocke, Younger, and Kasami
- restricted to $\mathcal{CNF}$
- used a triangular matrix T with cell $T_{i,j}$ for all applicable value pairs of i and j.
- output of the algorithm is a *set of items*. $\{[A,i,j] | A \Rightarrow^* a_{i+1} \ldots a_j\}$

Example - parsing system CYK 2/2

- domain of items
  $\mathcal{I}_{CYK} = \{[A, i, j] | A \in N \wedge 0 \leq i < j\}$
- hypotheses representing the string
  $H = \{[a, i-1, i] | a = a_i^1 \leq 1 \leq n\}\}$
- inference rules (set of deduction steps)
  $D^1 = \{[a, i-1, i] \vdash [A, i-1, i] | A \rightarrow a \in P\}$
  $D^2 = \{[B, i, j], [C, j, k] \vdash [A, i, k] | A \rightarrow BC \in P\}$
  $D_{CYK} = D^1 \cup D^2$

Filtering

Purposes of filtering:

- generalization increases the number of steps in parsing process
- filtering decreacing the number of items and deduction steps

Three kinds of filtering:

- static filtering - redundant parts of a parsing schema are discarded,
- dynamic filtering - the validity of some items can be made dependent on the validity of other items,
- step contraction - sequences of deduction steps are replaced by single deduction steps.

Static filtering

A nonterminal $A \in N$ is called *reduced* if:

(i) there are $v, w \in \Sigma^*$ such that $S \Rightarrow^* vAw$,

(ii) there is some $w \in \Sigma^*$ such that $A \Rightarrow^* w$.

A grammar is called reduced if all its nonterminals are reduced.

Dynamic filtering

The relation $\mathbb{P}_1 \longrightarrow_{df} \mathbb{P}_2$ holds if

(i) $\mathcal{I}_1 \supseteq \mathcal{I}_2$

(ii) $\vdash_1 \supseteq \vdash_2$

Reduce the number of valid items, but reduces the possibilities for parallel processing.

Contraction filtering

most powerful

The relation $\mathbb{P}_1 \rightarrow_{sc} \mathbb{P}_2$ holds if

(i) $\mathcal{I}_1 \supseteq \mathcal{I}_2$

(ii) $\vdash_1^* \supseteq \vdash_2^*$

- skipped nullable symbols
- chain of derivations reducing

Conclusion

Parsing schemata provide a general framework for description, analysis and comparison of parsing algorithms.

The End