Generation of Hypergraphs from the N-Best Parsing of 2D-Probabilistic Context-Free Grammars for Mathematical Expression Recognition

Ernesto Noya, Joan Andreu Sánchez, José Miguel Benedí
Pattern Recognition and Human Language Technology Research Center,
Universitat Politècnica de València, Camino Vera s/n, 46022, Valencia, Spain

Introduction

- A new approach for searching mathematical expressions (ME) in large collections of printed document images has been introduced.
- This approach does not require any kind of segmentation, and
- It does not need to have a complete and error-free transcription of the images.
- To reduce the search time, a two-phase solution is proposed.

  a) Offline phase, the posterior probabilities of MEs are calculated from *h*ypergraphs (H) derived from the ME recognition process.

  b) Online phase, these posteriors are used for indexing and searching for MEs in the collection.

- In this paper, we focus on the first offline phase, and the main contributions include:

  1. the computation of the n-best parse trees from a 2D-PCFGs,

  2. the generation of hypergraphs from the n-best parse trees,

N-best parsing of 2D-Probabilistic Context-Free Grammars

Given a *Two-Dimensional Probabilistic Context-Free Grammar*(2D-PCFG), and a set of input connected components, $\boldsymbol{x}$,

- $\mathcal{T}(A, a)$ is the set of trees whose root is (A, a), where A is a non-terminal symbol and a is a input span.
- $p : \mathcal{T} \rightarrow [0, 1]$ is a probabilistic function defined as follow:
  - For terminal rules, $A \rightarrow s$, and $\{x_i\} : 1 \leq i \leq |\boldsymbol{x}|$,

$$p(\langle A, \{x_i\}\rangle) \approx \frac{1}{|\boldsymbol{x}|} \max_s \left\{ \frac{p(s \mid A)\, p(s \mid \{x_i\})}{p(s)} \right\}$$

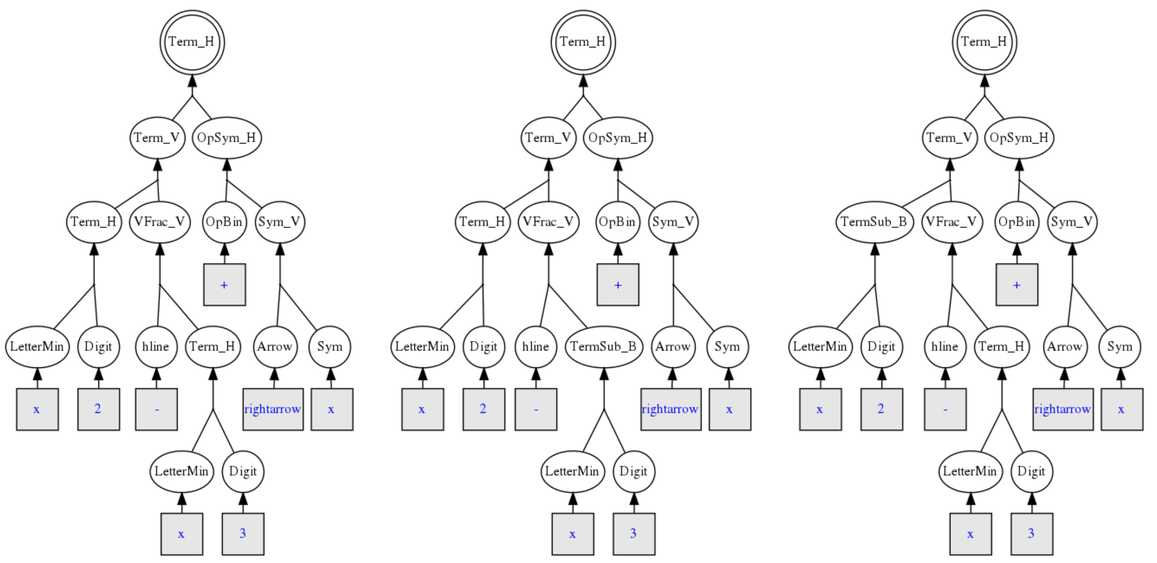
    $p(s|A)$ is the terminal rule probability, $p(s|\{x_i\})$ is provided by a symbol classifier and $p(s)$ is the prior probability of the symbols.
  - For binary rules, $A \rightarrow BC$,

$$p(\langle (A, a), T_1, T_2\rangle) \approx \max_r p(r \mid BC)\, p(BC \mid A)\, p(T_1)\, p(T_2)$$

    $p(BC|A)$ is the binary rule probability and $p(r|B, C)$ is the probability that regions B and C are arranged according to spatial relationship r.

The N-Best parsing allows to obtain $T^n(S, \boldsymbol{x})$ for $1 \leq n \leq N$ from:

$$\mathcal{T}^n(A, a) = (\mathcal{T}^{n-1}(A, a) - \{T^{n-1}(A, a)\}) \cup \{\langle T(A, a), T^{p+1}(B, b), T^q(C, c)\rangle\} \cup \{\langle T(A, a), T^p(B, b), T^{q+1}(C, c)\rangle\}$$



$\frac{x2}{x3} + \vec{x}$ $\quad$ $\frac{x2}{x_3} + \vec{x}$ $\quad$ $\frac{x_2}{x3} + \vec{x}$

Generation of Hypergraphs

- A **hypergraph** is definied as $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of nodes and $\mathcal{E}$ is a set of hyperarcs.
- A **node** is defined as $v = (n(v), s(v))$, where $n(v)$ is the node tag and $s(v)$ is the span (or set of connected components) associated with this node.
- A **hyperarc** is defined as $\mathcal{H} = (H(e), T(e),\ t(e), p(e))$, where the tail $T(e)$ and head $H(e)$ are subsets of $\mathcal{V}$, $t(e)$ is a transcription associated with the hyperarc and $p(e)$ is a score defined as:

$\forall e \in \mathcal{E} :\ n(H(e)) = A;\ n(T(e)) = (s);\ s(H(e)) = \{x_i\};$

$$p(e) = \frac{p(s \mid A)\, p(s \mid \{x_i\})\, p(\{x_i\})}{p(s)}$$

$\forall e \in \mathcal{E} :\ n(H(e) = A;\ n(T(e)) = (B, C);\ s(H(e)) = a$

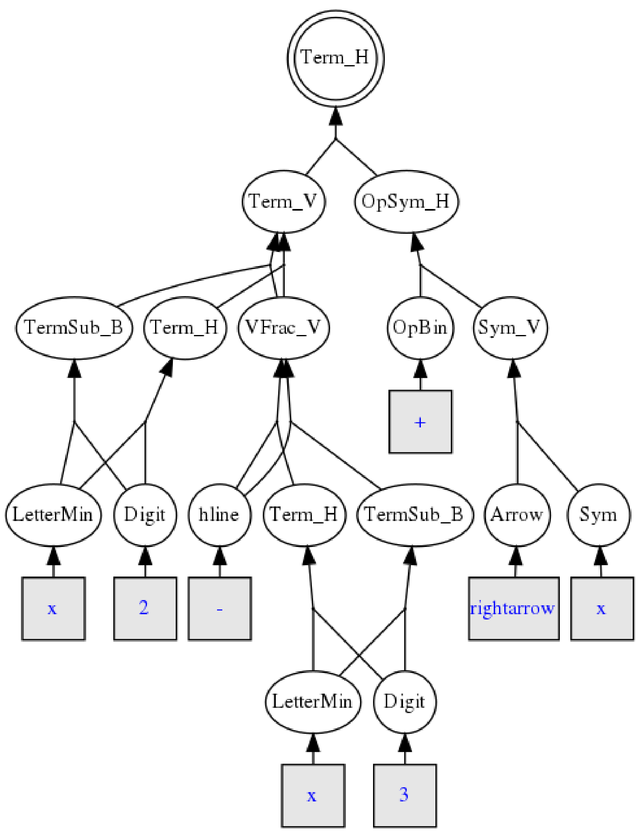
$$p(e) = p(r \mid BC)\, p(BC \mid A)$$

- A **complete tree** $\boldsymbol{t}$ is a sequence of hyperarcs that completely covers the input ME represented by $\boldsymbol{x}$.

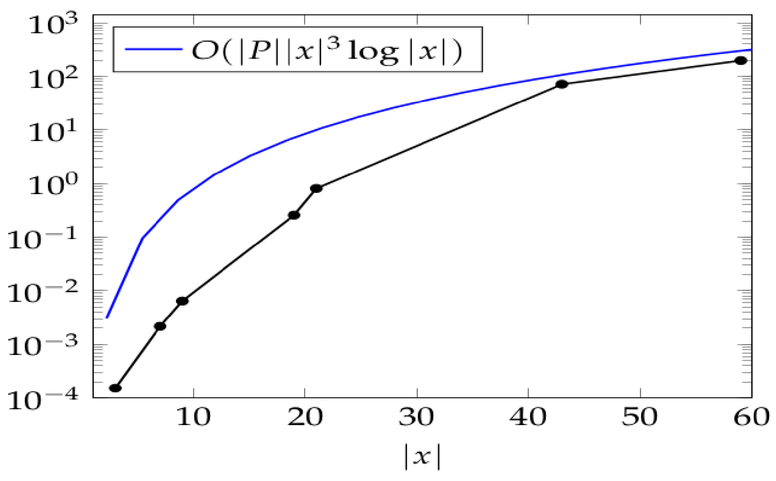
$$p(\boldsymbol{t}, \boldsymbol{x}) \approx \prod_{e \in \psi(\boldsymbol{t})} p(e) \stackrel{\text{def}}{=} p_{\mathcal{H}}(\boldsymbol{t}, \boldsymbol{x})$$

where $\psi(\boldsymbol{t})$ is the set of all hyperarcs that make up $\boldsymbol{t}$.

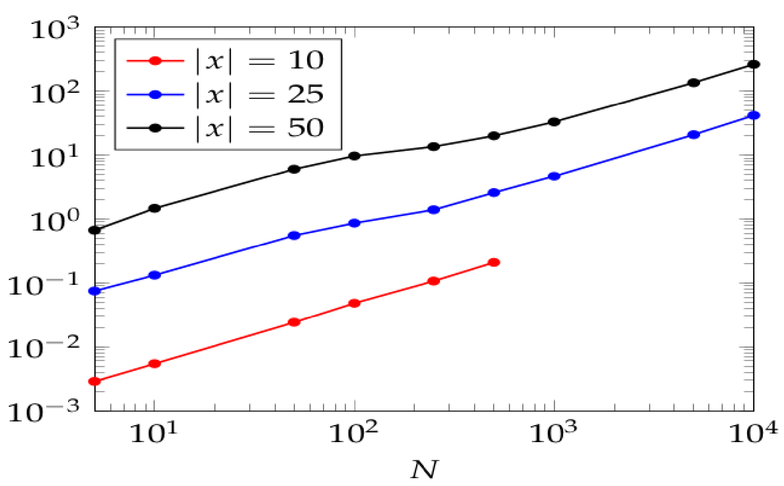
$$p(\boldsymbol{x}) \approx \sum_{\boldsymbol{t}} p_{\mathcal{H}}(\boldsymbol{t}, \boldsymbol{x}) \stackrel{\text{def}}{=} p_{\mathcal{H}}(\boldsymbol{x})$$



Experimental evaluation of the algorithms



1-best with horizontal and vertical relationships.



The red curve, $|x| = 10$, has 500 hypothesis.

Conclusions

- A proposal for generating hypergraphs from the N-Best parsing of 2D-PCFGs for ME recognition has been presented.
- A formal framework for the development of inference algorithms (*inside* and *outside*) and normalization strategies of hypergraphs has been also presented.
- Preliminary experiments have been reported to check the behavior of the proposed algorithms.