

Neural Precedence Recommender

Filip Bártek Martin Suda

Czech Technical University in Prague

August 30, 2021



Outline

① Introduction

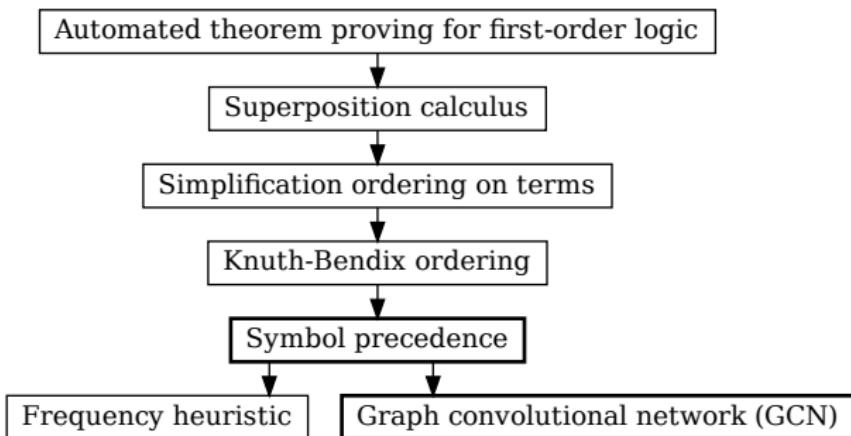
② Architecture

③ Training

④ Evaluation

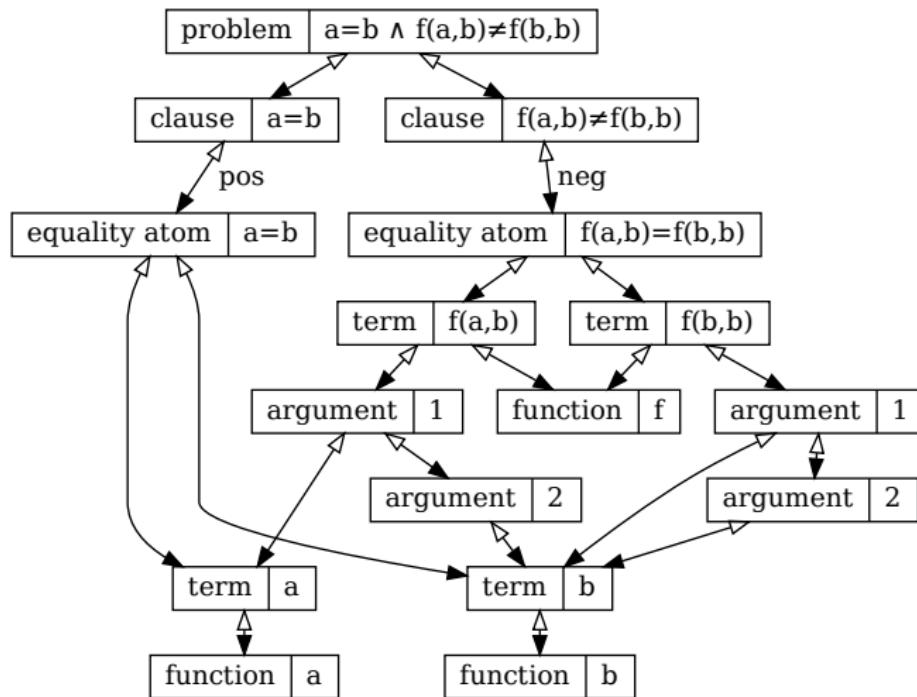
⑤ Summary

Context



Graph representation of a CNF problem

Input problem: $a = b \wedge f(a, b) \neq f(b, b)$



Graph convolutional network (GCN)

Initial embedding of node d :

$$h_d^{(0)} = (\text{feature vector}) \oplus (\text{trainable vector})$$

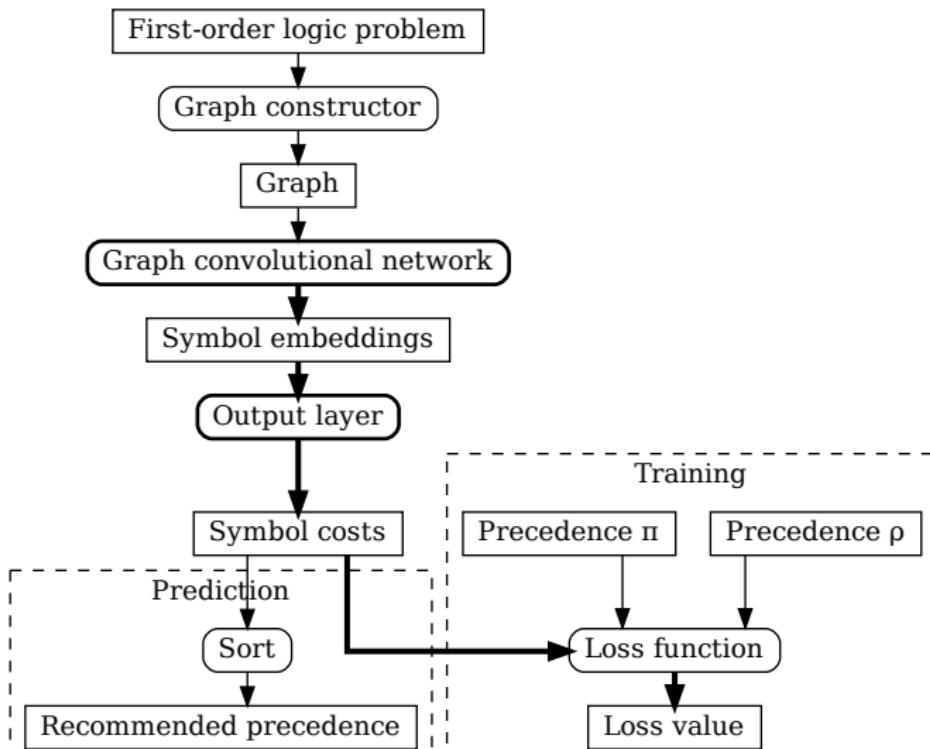
Feature vector:

- Clause: role (axiom, assumption, negated conjecture)
- Symbol: introduced in preprocessing, in conjecture

Propagation rule for layer l :

$$h_d^{(l+1)} = \sum_{r \in \mathcal{R}} \sigma \left(\sum_{s \in \mathcal{N}_d^r} \frac{1}{\sqrt{|\mathcal{N}_s^r|} \sqrt{|\mathcal{N}_d^r|}} (W_r^{(l)} h_s^{(l)} + b_r^{(l)}) \right)$$

Precedence recommender overview



Training: Precedence cost

Reminder: c_i is the cost of the i -th symbol.

Cost of symbol precedence π over signature of length n

$$C(\pi) = \frac{2}{n(n+1)} \sum_{i=1}^n i \cdot c_{\pi(i)}$$

Lemma (Precedence cost minimization)

The precedence cost C is minimized by any precedence that sorts the symbols by their costs in non-increasing order:

$$\operatorname*{argmin}_{\rho \in \text{Perm}(n)} C(\rho) = \text{argsort}^-(c_1, \dots, c_n)$$

Training: Loss function

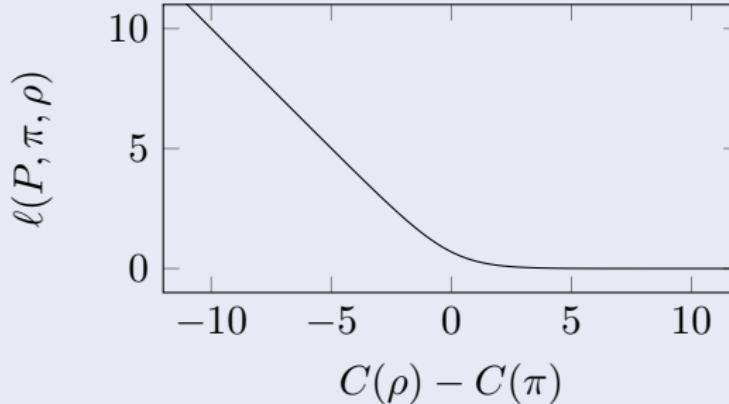
Training example $\pi \prec_P \rho$

Precedence π is better than precedence ρ for problem P .

Reminder: $C(\pi)$ is the predicted cost of precedence π .

Loss on training example $\pi \prec_P \rho$

$$\ell(P, \pi, \rho) = -\log \text{sigmoid}(C(\rho) - C(\pi))$$



Evaluation

Symbol cost model	Success count ¹		Improvement	
	Mean	Std	Absolute	Relative
GCN (pred. and func.)	3951.6	1.62	+182.0	1.048
GCN (predicate only)	3923.6	2.24	+154.0	1.041
GCN (function only)	3874.2	1.83	+104.6	1.028
Frequency (baseline)	3769.6	3.07	0.0	1.000

¹Total number of validation problems: 7648. Number of repetitions: 5.

Summary

- GCN over a graph representation of a FOL problem predicts symbol costs
- Sorting symbols by symbol costs yields a precedence
- Training is performed on oriented precedence pairs " $\pi \prec_P \rho$ " ("Precedence π is better than precedence ρ in problem P .)
- Combination of two GCNs outperforms the "frequency" heuristic by 4.8 %

Thank you for your attention!

Filip Bártek

[filip.bartek@cvut.cz](mailto:filiip.bartek@cvut.cz)

Acknowledgement: This work was supported by the Czech Science Foundation project no. 20-06390Y (JUNIOR grant), the project RICAIP no. 857306 under the EU-H2020 programme, and the Grant Agency of the Czech Technical University in Prague, grant no. SGS20/215/OHK3/3T/37.

Notation overview

$\pi(i)$	index of the i -th symbol in precedence π
c	vector of symbol costs (output of the GCN)
c_i	cost of the i -th symbol
$C(\pi)$	cost of symbol precedence π
$\ell(P, \pi, \rho)$	loss for training example $\pi \prec_P \rho$

Prediction of symbol precedence π

Require: Problem P

Ensure: Symbol precedence π

$c \leftarrow \text{GCN}(P)$ {Forward pass through the GCN to obtain vector of symbol costs c }

$\pi \leftarrow \text{argsort}^-(c_1, \dots, c_n)$ {Sort symbols by c in non-increasing order to obtain precedence π }

return π

Proof sketch: Precedence cost minimization

$$C(\pi) = \frac{2}{n(n+1)} \sum_{i=1}^n i \cdot c_{\pi(i)}$$

Lemma (Precedence cost minimization)

$$\operatorname*{argmin}_{\rho \in \text{Perm}(n)} C(\rho) = \text{argsort}^-(c_1, \dots, c_n)$$

Example

$$C\left(\begin{array}{|c|c|c|}\hline & & \\ \hline \end{array}\right) < C\left(\begin{array}{|c|c|c|}\hline & & \\ \hline \end{array}\right)$$

Proof:

$$\begin{aligned} C\left(\begin{array}{|c|c|c|}\hline & & \\ \hline \end{array}\right) - C\left(\begin{array}{|c|c|c|}\hline & & \\ \hline \end{array}\right) &\propto ((2 \cdot 3 + 3 \cdot 4) - (2 \cdot 4 + 3 \cdot 3)) \\ &= 18 - 17 = 1 > 0 \end{aligned}$$

Concise representation of training examples

Cost of symbol precedence π over signature of length n

$$C(\pi) = \frac{2}{n(n+1)} \sum_{i=1}^n i \cdot c_{\pi(i)} = \frac{2}{n(n+1)} \sum_{i=1}^n c_i \cdot \pi^{-1}(i)$$

Loss of training example $\pi \prec_P \rho$

$$\begin{aligned}\ell(P, \pi, \rho) &= -\log \text{sigmoid}(C(\rho) - C(\pi)) \\ &= -\log \text{sigmoid} \frac{2}{n(n+1)} \sum_{i=1}^n i(c_{\rho(i)} - c_{\pi(i)}) \\ &= -\log \text{sigmoid} \frac{2}{n(n+1)} \sum_{i=1}^n c_i(\rho^{-1}(i) - \pi^{-1}(i))\end{aligned}$$

Concise representation of $\pi \prec_P \rho$: $\rho^{-1} - \pi^{-1}$