GRASP: RANDOMISED SEMIRING PARSING

WILKER AZIZ
University of Amsterdam, ILLC
w.aziz@uva.nl

WHAT'S GRASP ABOUT?

Open source

- semiring parsing
- weighted deduction
- approximate forest rescoring
- inference by **sampling**

https://github.com/wilkeraziz/grasp

DEDUCTIVE PARSING

Abstracts away from implementation details.

- **items** represent intermediate results in the deduction process
- $\frac{A_1 \ldots A_k}{B} C_1 \ldots C_j$ is a deduction rule
- **item derivations**
  - $D = \langle r \rangle$ where r is a **grammar rule**
  - $D = \langle b : D_{a_1}, \ldots, D_{a_k} \rangle$ where $\frac{a_1 \ldots a_k}{b} c_1 \ldots c_j$ is an instantiation of $\frac{A_1 \ldots A_k}{B} C_1 \ldots C_j$ and $D_{a_1}, \ldots, D_{a_k}$ are item derivations

Hypergraph representation

- *items* are mapped to **nodes**
- $\frac{a_1 \ldots a_k}{b} c_1 \ldots c_j$ are mapped to **edges**
  - b is mapped to a **head** node
  - $a_1 \ldots a_k$ are mapped to **tail** nodes

SEMIRING PARSING

The same deductive parser, multiple tasks!

FOREST	parse forest
VITERBI	max
INSIDE	sum
BOOLEAN	recognition
COUNTING	counting
1-BEST	argmax
K-BEST	top derivations

Value recursion

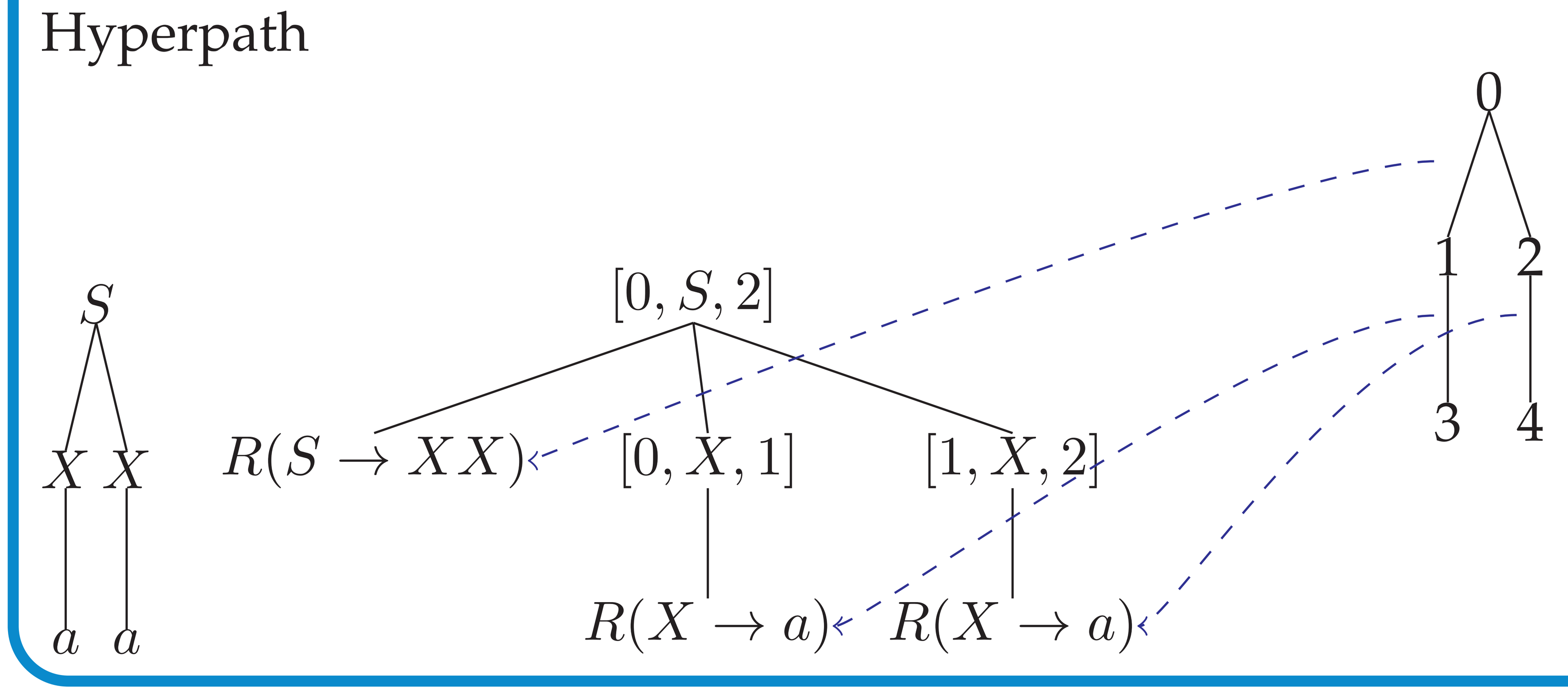
$$V(D) = \bigotimes_{i=1}^{j} R(d_i)$$

- where D is an item derivation
- $\text{yield}(D) = d_1 \ldots d_j$ is a grammar derivation
- $R(r)$ is the value of grammar rule r

$$V(D) = \begin{cases} R(r) & \text{if } D = \langle r \rangle \\ \bigotimes_{i=1}^{k} V(D_i) & \text{if } D = \langle b : D_1 \ldots D_k \rangle \end{cases}$$

DERIVATIONS

Grammar derivation
Item derivation
Hyperpath



SLICE SAMPLING

Sample probabilistically from a distribution proportional to $f(\mathbf{d})$

- **non-negative** function over derivations $\mathbf{d} \in \mathcal{D}$
- $\langle \mathcal{D}, f(\mathbf{d}) \rangle$ is represented by a **hypergraph**
- a derivation $\mathbf{d} = \langle e_1, \ldots, e_m \rangle$ is a sequence of m steps (edges)
- $r_h \in \mathbf{d}$ refers to an edge headed by h whose underlying rule is r

$$f(\mathbf{d}) = \psi(\mathbf{d}) \times \theta(\mathbf{d}) = \psi(\mathbf{d}) \times \prod_{r_h \in \mathbf{d}} \theta_{r_h}$$

Data augmentation and *slice sampling*

$$f(\mathbf{d}, \mathbf{u}) = \psi(\mathbf{d}) \prod_{u_h : r_h \in \mathbf{d}} \delta_{(0, \theta_{r_h})}(u_h) \prod_{u_h \in r_h \notin \mathbf{d}} \phi(u_h; \boldsymbol{\alpha})$$

u: one auxiliary variable per item

Gibbs sampling

$$f(u_s | \mathbf{d}) = \begin{cases} \frac{\delta_{(0,\theta_{r_h})}(u_h)}{\theta_{r_h}} & \text{if } r_h \in \mathbf{d} \\ \phi(u_h; \boldsymbol{\alpha}) & \text{otherwise} \end{cases}$$

$$f(\mathbf{d}|\mathbf{u}) \propto \psi(\mathbf{d}) \prod_{r_h \in \mathbf{d}} \frac{\delta_{(0,\theta_{r_h})}(u_h)}{\phi(u_h; \boldsymbol{\alpha})}$$

Sampling from the $f(\mathbf{d}|\mathbf{u})$

- rescore the slice exactly (if $\psi(\mathbf{d})$ is simple enough)
- estimate an empirical distribution
  - uniform sampling
  - importance sampling

SLICE: SUBSET OF $\mathcal{D}$ WHERE $f(\mathbf{d}|\mathbf{u}) > 0$

ITEM FORMS	$R(X \to \alpha)$, $T\left(q \xrightarrow{x} r\right)$, $[X \to \alpha \bullet \beta, q, s]$ and $\langle X; q, r \rangle$
GOAL	$\langle S'; q, r \rangle \quad q \in I, r \in F$
AXIOMS	$\frac{}{[S' \to \bullet S, q, q]} \quad q \in I$
PREDICT	$\frac{R\left(Y \xrightarrow{\theta} \gamma\right)}{\left[Y \xrightarrow{\theta} \bullet\gamma, r, r\right]} \quad [X \to \alpha \bullet Y\beta, q, r]$
SCAN	$\frac{T\left(r \xrightarrow{x/\omega} s\right) \quad \left[X \xrightarrow{\theta} \alpha \bullet x\beta, q, r\right]}{\left[X \xrightarrow{\omega \otimes \theta} \alpha x \bullet \beta, q, s\right]}$
COMPLETE	$\frac{\left[X \xrightarrow{\theta} \alpha \bullet Y\beta, q, r\right] \quad \langle Y; r, s \rangle}{\left[X \xrightarrow{\theta} \alpha Y_{r,s} \bullet \beta, q, s\right]}$
CONVERT	$\frac{\left[X \xrightarrow{\theta} \alpha\bullet, q, r\right]}{\langle X; q, r \rangle} \quad \theta > u_{X_{q,r}}$

Key: auxiliary variables are **lazily sampled** $u_h \sim f(u_h|\mathbf{d})$

FEATURES

Grammar formalism	epsilon-free CFG
Weighted deduction	bottom-up and top-down (exact and sliced)
Forest rescoring	top-down (exact and sliced)
Real-valued semirings	BOOLEAN, COUNTING, VITERBI, INSIDE
Value recursion	robust to cycles
Derivation semirings	1-BEST, k-BEST, SAMPLE, FOREST, SLICE
Sampling algorithms	ancestral sampling, slice sampling
LM queries	kenlm
Applications	constituency parsing, decoding for hiero models

ACKNOWLEDGMENTS

This research is funded by The Netherlands Organisation for Scientific Research (NWO), NOW VICI grant nr. 277-89-002.