Unsupervised Spectral Learning of WCFG as Low-rank Matrix Completion

Raphaël Bailly† Xavier Carreras† Franco M. Luque‡ Ariadna Quattoni†

† Universitat Politècnica de Catalunya ‡ Universidad Nacional de Córdoba y CONICET

— Abstract —

- Goal: induce a WCFG from observed strings (e.g. language modeling)
- Contribution: a spectral method for unsupervised WCFG learning
- Ingredients:
  - Convex optimization of a Hankel matrix for WCFG
  - Linear contraints to characterize WCFG Hankel matrices
  - Low-rank objective, derived from the spectral approach

— Weighted Context-Free Grammars —

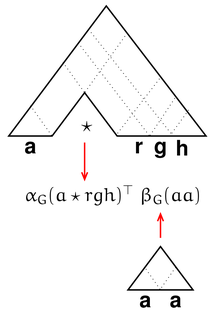
- Definition: WCFG $G = \langle \Sigma, n, \boldsymbol{\alpha}_\star, \{\boldsymbol{\beta}_\sigma\}, \mathbf{A} \rangle$
  - Alphabet Σ, n states (states = non-terminals)
  - Initial vector $\boldsymbol{\alpha}_\star \in \mathbb{R}^n$
  - Terminal vectors $\boldsymbol{\beta}_\sigma \in \mathbb{R}^n$ for $\sigma \in \Sigma$
  - Bilinear operator $\mathbf{A} \in \mathbb{R}^{n \times n^2}$
- Grammar function $g_G : \Sigma^\star \to \mathbb{R}$

$$g_G(x) = \boldsymbol{\alpha}_\star^\top \beta_G(x)$$

- Inside function $\beta_G : \Sigma^+ \to \mathbb{R}^n$

$$\beta_G(\sigma) = \boldsymbol{\beta}_\sigma$$

$$\beta_G(x) = \sum_{\substack{x_1, x_2 \in \Sigma^+ \\ x = x_1 x_2}} \mathbf{A}(\beta_G(x_1) \otimes \beta_G(x_2))$$



- Outside function $\alpha_G : \Sigma^\star \times \Sigma^\star \to \mathbb{R}^n$

$$\alpha_G(\langle \star \rangle) = \boldsymbol{\alpha}_\star$$

$$\alpha_G(\langle x \star z \rangle)^\top = \sum_{\substack{x_1 \in \Sigma^\star, x_2 \in \Sigma^+ \\ x = x_1 x_2}} \alpha_G(\langle x_1 \star z \rangle)^\top \mathbf{A}(\beta_G(x_2) \otimes \mathbf{I}_n) + \sum_{\substack{z_1 \in \Sigma^+, z_2 \in \Sigma^\star \\ z = z_1 z_2}} \alpha_G(\langle x \star z_2 \rangle)^\top \mathbf{A}(\mathbf{I}_n \otimes \beta_G(z_1))$$

— The Spectral Method —

1. Compute Hankel matrix $\mathbf{H}$ using training data
   - Supervised: training data contains derivations, count events
   - Unsupervised (this paper): induce $\mathbf{H}$ from plain strings
2. Compute the SVD of $\mathbf{H}_{\mathcal{O}^1,\mathcal{I}^1} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^\top$
3. Create a rank-n factorization of $\mathbf{H}_{\mathcal{O}^1,\mathcal{I}^1} \approx \mathbf{F}_n \mathbf{B}_n$

$$\mathbf{F}_n = \mathbf{U}_n \boldsymbol{\Lambda}_n \qquad \mathbf{B}_n = \mathbf{V}_n^\top$$

4. Compute G:

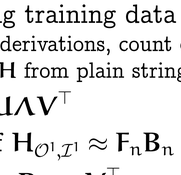
$$\boldsymbol{\alpha}_\star^\top = \mathbf{H}_{\star,\mathcal{I}^1} \mathbf{B}_n^+$$

$$\boldsymbol{\beta}_\sigma = \mathbf{F}_n^+ \mathbf{H}_{\mathcal{O}^1,\sigma}$$

$$\mathbf{A} = \mathbf{F}^+ \mathbf{H}_{\mathcal{O}^1,\mathcal{I}^2} (\mathbf{B}_n \otimes \mathbf{B}_n)^+$$

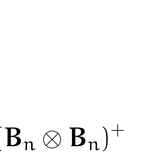
— Hankel Matrices for WCFG —

- Definition: $\mathbf{H}_g \in \mathbb{R}^{\mathcal{O} \times \mathcal{I}}$ with $\mathbf{H}_g(o, i) = g(o \cdot i)$
- Inside strings $\mathcal{I}$



(plain) (composed)

- Outside contexts $\mathcal{O}$



(plain) (composed)

- Theorem:

g computed by G with n states

$$\Leftrightarrow$$

$$\operatorname{rank}(\mathbf{H}_g) = n$$

Intuition:

$\mathbf{H}_{\mathcal{O}^1,\mathcal{I}^1} = \mathbf{F}\,\mathbf{B}$

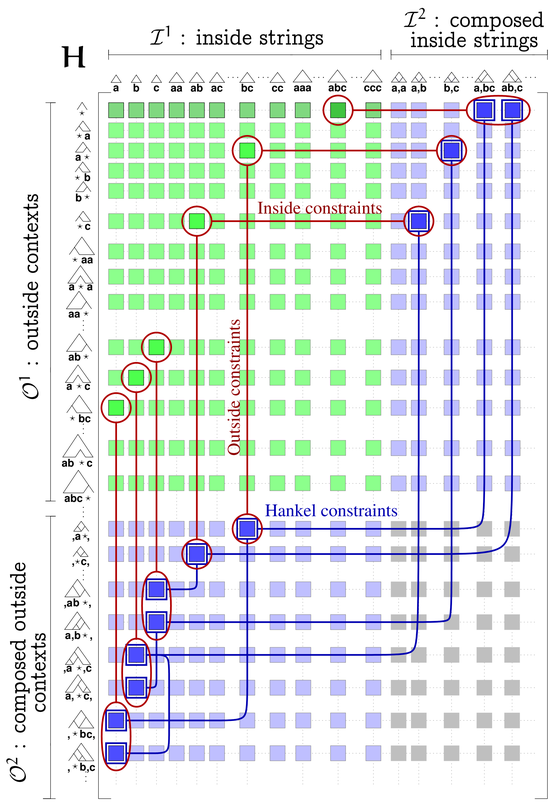
$\mathbf{F}(x \star z, \cdot) = \alpha(x \star z) \in \mathbb{R}^n$

$\mathbf{B}(\cdot, x) = \beta(x) \in \mathbb{R}^n$

$\mathbf{H}_{\mathcal{O}^1,\mathcal{I}^2} = \mathbf{F}\,\mathbf{A}(\mathbf{B} \otimes \mathbf{B})$

$\mathbf{H}_{\star,\mathcal{I}^1} = \boldsymbol{\alpha}_\star^\top \mathbf{B}$

$\mathbf{H}_{\mathcal{O}^1,\sigma} = \mathbf{F}\,\boldsymbol{\beta}_\sigma$



— Hankel Induction —

- Observable statistics (for full strings):

$$\mathbf{H}(\langle \star \rangle, (x)) = p(x)$$

- WCFG constraints:

  Hankel constraints:

$$\mathbf{H}(\langle x \star z \rangle, (y_1, y_2)) = \mathbf{H}(\langle x, y_1 \star, z \rangle, (y_2)) = \mathbf{H}(\langle x, \star y_2, z \rangle, (y_1))$$

  Inside constraints:

$$\mathbf{H}(o, (x)) = \sum_{x = x_1 x_2} \mathbf{H}(o, (x_1, x_2))$$

  Outside constraints:

$$\mathbf{H}(\langle x \star z \rangle, i) = \sum_{x = x_1 x_2} \mathbf{H}(\langle x_1, x_2 \star, z \rangle, i) + \sum_{z = z_1 z_2} \mathbf{H}(\langle x, \star z_1, z_2 \rangle, i)$$

- Property: any $\mathbf{H}$ satisfying constraints corresponds to a WCFG

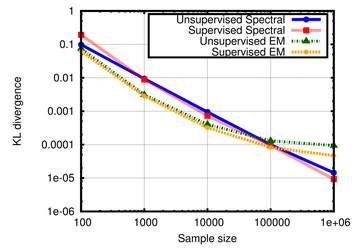
— Convex Optimization —

$$\begin{aligned} \underset{\mathbf{H}}{\text{minimize}} \quad & \|\mathbf{H}\|_* \\ \text{subject to} \quad & \|\mathbf{O} \cdot \operatorname{vec}(\mathbf{H}) - \mathbf{z}\|_2 \le \mu \\ & \mathbf{K} \cdot \operatorname{vec}(\mathbf{H}) = \mathbf{0} \\ & \|\mathbf{H}\|_2 \le 1. \end{aligned}$$

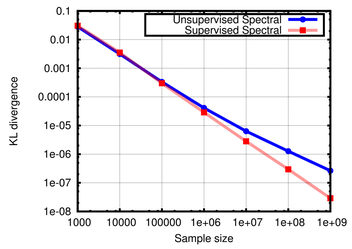
Optimization based on projections for $\|\cdot\|_*$, $\mathbf{O}$ and $\mathbf{K}$.

— Synthetic Experiments —

2 symbols, 2 states

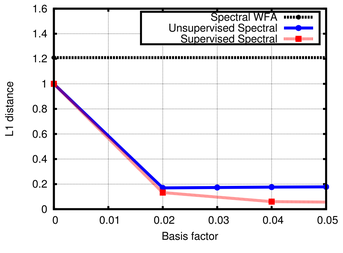


3 symbols, 6 states



- Random synthetic targets
- Fixed inside-outside patterns in Hankel
- Comparison to EM and a supervised spectral method

— Experiments with Dyck Languages —



- The target grammar is:

$$S \to \; S\,S\;(0.2) \mid a\,S\,b\;(0.4) \mid a\,b\;(0.4)$$

- e.g.: $p(aabb) = 0.16$, $p(aaa) = 0$
- We vary the size of the Hankel
- Comparison to spectral algorithm for Weighted Automata

— On WSJ-10 data —

basis	size of $\mathbf{H}$	obs.	i/o ctr.	basis	size of $\mathbf{H}$	obs.	i/o ctr.
1×11	39×159	34	162	36×34	$27{,}989 \times 11{,}682$	916	156,690
6×14	$1{,}163 \times 764$	146	6,360	42×37	$3{,}638 \times 15{,}026$	1,035	200,346
12×18	$4{,}462 \times 2{,}239$	322	25,374	48×41	$45{,}192 \times 18{,}235$	1,157	244,398
18×22	$9{,}124 \times 4{,}149$	479	52,524	54×45	$53{,}741 \times 21{,}196$	1,281	284,466
24×26	$15{,}755 \times 6{,}858$	657	89,718	60×48	$60{,}844 \times 23{,}890$	1,382	318,354
30×29	$19{,}801 \times 8{,}545$	769	112,37				

Acknowledgements: This work was supported by a Google Research Award, and by projects XLike (FP7-288342), ERA-Net CHISTERA VISEN, TACARDI (TIN2012-38523-C02-02), BASMATI (TIN2011-27479-C04-03), SGR-GPLN (2009-SGR-1082) and SGR-LARCA (2009-SGR-1428). Xavier Carreras was supported by the Ramón y Cajal program of the Spanish Government (RYC-2008-02223). Franco M. Luque was supported by a postdoctoral fellowship from CONICET of the Argentinian Government.