Diversity in Spectral Learning for Natural Language Parsing

Shashi Narayan and Shay B. Cohen
School of Informatics, University of Edinburgh

Engineering and Physical Sciences Research Council

Introduction

Diversity in syntactic parsing: A diverse set of solutions from a decoder can be reranked or recombined in order to improve the accuracy in syntactic parsing (e.g., MaxEnt discriminative reranking [3], Parser combination by reparsing [1], Products of latent-variable PCFG (L-PCFG) [2] and Fusion [4]).

Contributions of this study: We describe an approach to create a diverse set of predictions with *spectral learning of L-PCFGs*. Our contributions are two fold:

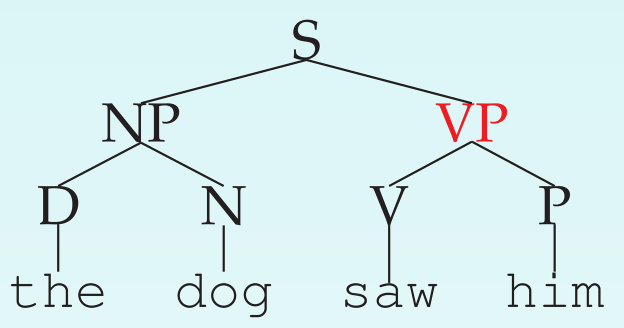
- ✓ we present a simple algorithm, a variant of the spectral estimation algorithm for L-PCFGs, which is also more accessible and
- ✓ we improve parsing with a family of noisy spectral algorithms for L-PCFGs.

For English, we achieve the F1 score of 90.18, and for German we achieve the F1 score of 83.38.

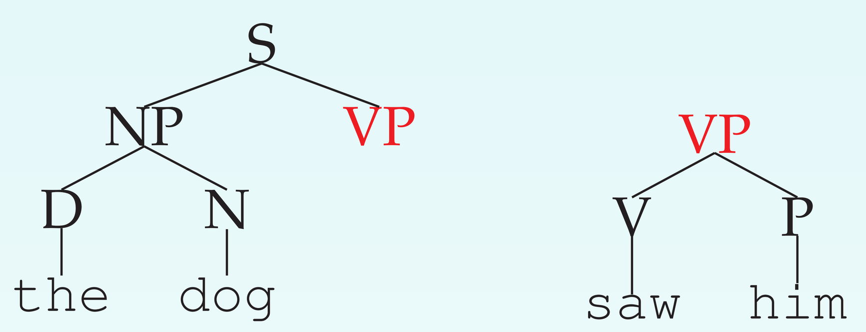
Basic Concepts [5]

Spectral algorithm for estimating L-PCFGs

At the node **VP**



Outside tree o **Inside tree t**



Feature functions ψ and ϕ

$\psi(o) = [0, 1, \ldots, 1] \in \mathbb{R}^{d'}$ $\phi(t) = [1, 0, \ldots, 0] \in \mathbb{R}^{d}$

SVD of Cross-covariance matrix Ω^a for a nonterminal a

$$\Omega^a = \frac{\sum_{i=1}^{M} [[a^{(i)} = a]] \phi(t^{(i)}) (\psi(o^{(i)}))^\top}{\sum_{i=1}^{M} [[a^{(i)} = a]]}$$

Projection functions Z and Y

$Z(o) = [1, 0.4, \ldots, 2] \in \mathbb{R}^k$ $Y(t) = [-2, \ldots, 0.5] \in \mathbb{R}^k$

This leads to a **faster training algorithm than the EM algorithm**.

Contributions

Our **clustering algorithm for estimating L-PCFGs** proceeds in following two steps:

✓ **Cluster projections using K-means:** For each nonterminal $a \in N$, we cluster all of its treebank instances into m states using projections $Z(o) \in \mathbb{R}^k$ and $Y(t) \in \mathbb{R}^k$ as features.

✓ **Compute Final Parameters:** We annotate each node in the treebank with its cluster count (latent state). We subsequently follow up with frequency count *maximum likelihood estimate* to estimate the probabilities of each parameter in the L-PCFG.

Clustering Algorithm Vs Spectral Algorithm

✓ Compact model ($97K$ Vs $54M$ nonzero parameters with $m = 24$)

✓ Lower parsing accuracy (86.48% Vs 88.53% on Section 22)

Final Results: Test Set

English: The Penn WSJ Treebank (Section 23)

	Method	F_1
Best	Spectral	89.21
	Clustering	89.25
Hier	Spectral	89.09
	Clustering	**90.18**

German: The NEGRA corpus standard test data

	Method	F_1
Best	Spectral	80.88
	Clustering	81.94
Hier	Spectral	80.64
	Clustering	**83.38**

✓ We applied our noise scheme to both *clustering* and *spectral* algorithms.

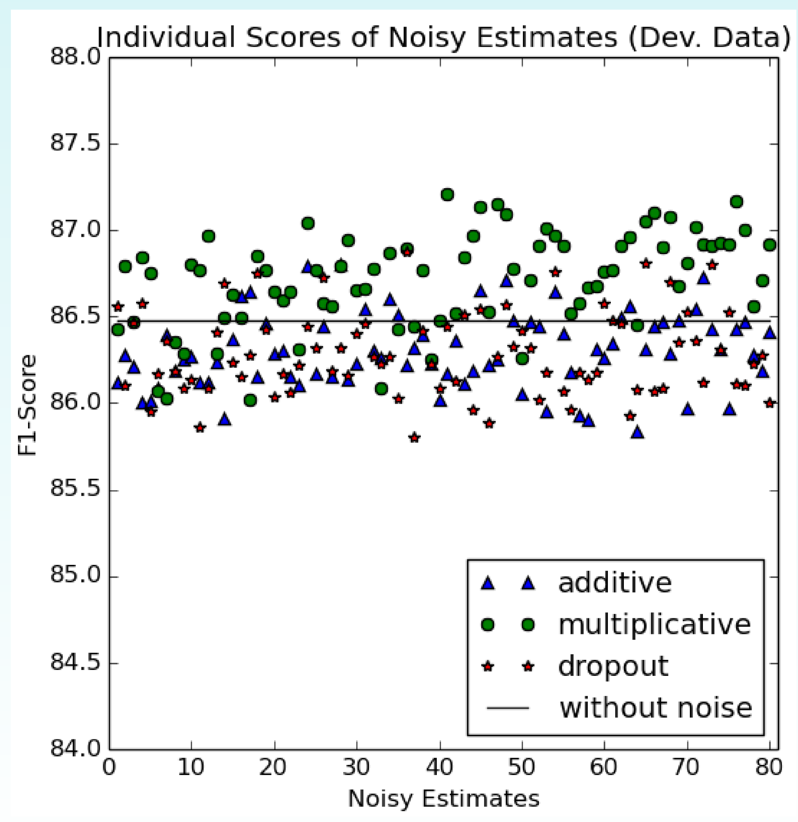
✓ *Best* is the MaxEnt decoder with *all* noisy models. *Hier* is MaxTree decoder over all MaxEnt outputs.

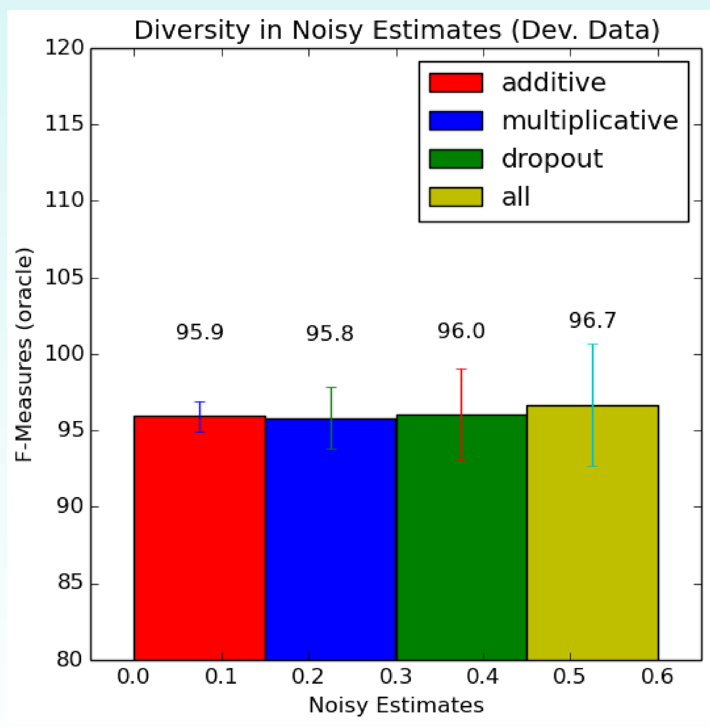
Contributions: Spectral Estimation with Noise

Gaussian (additive): For each element x in $Z(o)$ and $Y(t)$, $x \leftarrow x + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

Gaussian (multiplicative): For each element x in $Z(o)$ and $Y(t)$, $x \leftarrow x \otimes (1 + \varepsilon)$ where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

Dropout noise: For each element x in the feature vectors $\phi(t)$ and $\psi(o)$, $x \leftarrow 0$ with probability $\sigma \in [0, 1]$.



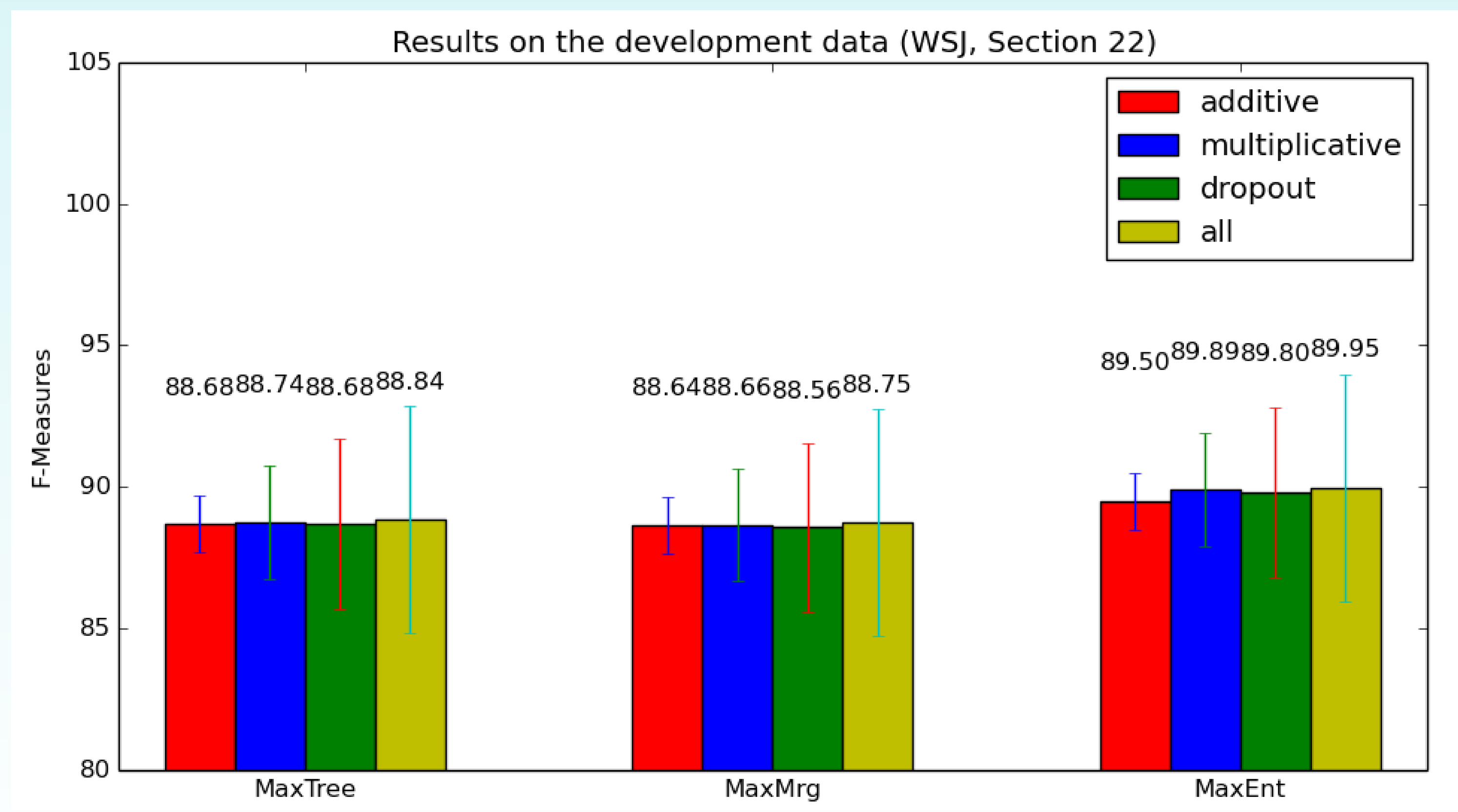


Contributions: Decoding with Multiple Models

Maximal tree coverage: Returns tree that maximizes its coverage with respect to all other trees that are decoded.

Maximal marginal coverage: Similar to maximal tree coverage, but softer and relies on the marginals that each model gives.

MaxEnt reranking: Returns a tree out of all other trees that are decoded using a MaxEnt discriminative reranker [3].



References

[1] Kenji Sagae and Alon Lavie. 2006. *Parser combination by reparsing*. In Proceedings of HLT-NAACL.
[2] Slav Petrov. 2010. *Products of random latent variable grammars*. In Proceedings of HLT-NAACL.
[3] Eugene Charniak and Mark Johnson. 2005. *Coarse-to-fine n-best parsing and maxent discriminative reranking*. In Proceedings of ACL.
[4] Do Kook Choe, David McClosky, and Eugene Charniak. 2015. *Syntactic parse fusion*. In Proceedings of EMNLP.
[5] Shay B. Cohen, Karl Stratos, Michael Collins, Dean P. Foster, and Lyle Ungar. 2013. *Experiments with spectral learning of latent-variable PCFGs*. In Proceedings of NAACL.

Acknowledgements

We would like to thank David McClosky for his help with running the BLLIP parser and the three anonymous reviewers for their helpful comments. This research was supported by an EPSRC grant (EP/L02411X/1).