Identifiability and Unmixing of Latent Parse Trees

Daniel Hsu
Microsoft Research
Sham M. Kakade
Microsoft Research
Percy Liang
Stanford University

Overview

Model: \( P_\theta(x, z) \) over parse trees \( z \) and sentences \( x \)

Goal: given \( n \) sentences \( x^{(1)}, \ldots, x^{(n)} \), produce parameter estimate \( \hat{\theta} \)

The dog barked
Congress passed the bill

Challenge: tree topology is unobserved and varies across sentences

Two questions:
• Which model families \( P_\theta(x, z) \) are identifiable?
  Our result: PCFG is not identifiable.
• How to estimate parameters without local optima issues?
  Our result: new unmixing technique works for restricted PCFGs.

Probabilistic Context-Free Grammars (PCFG)

For \( L = 3 \) words:

<table>
<thead>
<tr>
<th>Topology(( z ))</th>
<th>Parameters ( \theta = (T, B, \eta, \pi) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topology(( z )) = 1</td>
<td>( \pi(x) ) unidentifiable, initial state ( \pi(x) ) is probability of generating ( x )</td>
</tr>
<tr>
<td>Topology(( z )) = 2</td>
<td>( \pi(x) ) is probability of generating ( x )</td>
</tr>
</tbody>
</table>

Restricted PCFG

Generate left and right children independently from state transitions \( T \in \mathbb{R}^{k \times k} \)

\( \theta = (T, B, \eta, \pi) \)

Parameter estimation

Standard approach (maximum likelihood):
Estimator: \( \hat{\theta} = \arg \max x P_\theta(x) \)
Intractable, EM algorithm gets stuck in local optima [Lari & Young, 1990]

Our strategy (method of moments):
Moment function: \( \phi(x) = \mathbb{E}_z \phi(z|x) = \mathbb{E}_z [g(z)] \) \( \in \mathbb{R}^{d \times d} \)
Estimator: \( \hat{\theta} = \sum_{i=1}^n \phi(x^{(i)}) \)

Challenge: tree topology is unobserved and varies across sentences

Conclusion

Identifiability from moments

Definition (global identifiability): model family \( \Theta = [0, 1]^p \) is identifiable from a moment function \( \phi(x) \) if \( \mathbb{E}_z \phi(z|x) = \mathbb{E}_z \phi(z) \in \mathbb{R}^{d \times d} \)

General identifiability checker:
1. Choose a single \( \theta \in \Theta \) uniformly at random.
2. Compute Jacobian matrix \( J(\theta) = \frac{\partial}{\partial \theta_1} \mathbb{E}_z \phi(z) \in \mathbb{R}^{d \times p} \).
3. Return identifiable iff \( J(\theta) \) is full rank.

Result: PCFG is not identifiable from any moments of \( x \) and \( L \leq 5 \).

Unmixing

Known tree structure (for \( L = 3 \) words):
\( \Psi_{x,y} = \mathbb{E}_z \phi(x, y|z) \) \( \text{Topologies (2)} = \mathbb{E}_T \mathbb{E}_O \mathbb{E}_\eta \mathbb{E}_x \mathbb{E}_y \) \( M = \mathbb{E}_\tau \mathbb{E}_\pi \mathbb{E}_z \mathbb{E}_\theta \)

Compute \( \Psi_{x,y} \) for two different \( \eta \), apply Decompose to recover \( M \).
Apply simple matrix algebra to extract all parameters \( \theta = (\pi, T, O) \).

Unknown tree structure (for \( L = 3 \) words):
Strategy: reduce to the known tree structure case

Result: for restricted PCFG, \( e_2 \) in row space of \( M \), can recover \( \Psi_{x,y} \).

Other results

Restricted PCFG

Unidentifiable

PCFG

Identifiable

Dependency parsing models:

Result: identifiable, unmixing works for restricted version

Related work on spectral methods:
HMMs [Hsu/Kakade/Zhang 2008]
Latent tree models with known structure [Panchanet/Song/Xiang 2014]
Unknown fixed structure [Anandkumar/Chandramouli/Hsu/Kakade/Song/Zhang 2014]
PCFGs with known tree structure [Colmenarez/Stratos/Collins/Foster/Uигар 2012]
Recover parameters for HMMs [Anandkumar/Hsu/Kakade 2012]

This work: recover parameters, unknown random structure
Two contributions:
• Identifiability checker: easy method to see if model family identifiable
• Unmixing technique: consistent parameter recovery with random structures

Conclusion