Unseen Patterns: Using Latent-Variable Models for Natural Language

Shay Cohen
Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

July 13, 2017
Thanks to...

JOINT WORK WITH ALL MY COLLABORATORS – SEE REFERENCES
Natural Language Processing

Translate
Understand
Search
Organise
Extract information
Interact
Main Challenge: Ambiguity

Ambiguity: Natural language utterances have many possible analyses.

Need to prune out thousands of interpretations even for simple sentences (for example: parse trees).
Variability

Many surface forms for a single meaning:

There is a bird singing
A bird standing on a branch singing
A bird opening its mouth to sing
A black and yellow bird singing in nature
A Rufous Whistler singing
A bird with a white patch on its neck
Approach to NLP

1980s - rule based systems

1990s and onwards - data-driven (machine learning)
Approach to NLP

1980s - rule based systems

1990s and onwards - data-driven (machine learning)

Challenge: The labeled data bottleneck
Labeled Data Bottleneck

Approach to NLP since 1990s: use labeled data. Leads to the labeled data bottleneck – never enough data

How to solve the labeled data bottleneck?

- Ignore it
- Unsupervised learning
- Latent-variable modelling
Seeking Life’s Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here, two genomic researchers with radically different approaches presented complementary views of the basic genes needed for life.

One research team, using computer analyses to compare known genomes, concluded that today’s organisms can be sustained with just 250 genes. And that the earliest life form required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 80 genes are plenty to do the job—but that anything short of 128 wouldn’t be enough.

Although the numbers don’t match precisely, those predictions are not all that far apart, especially in comparison to the 75,000 genes in the human genome, notes Stephen A. Johnston, University of British Columbia, in Vancouver, said the Sanger Center, but coming up with a consensus number may be more than just a guess. More and more genomes are being reported and sequenced. It may be a way of organizing any nearly sequenced genome,” explains Arvind Muneghat, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Computing an


Stripping down. Computer analyses yields an estimate of the minimum modern and ancient genomes.

(SCIENCE • VOL. 272 • 14 MAY 1996)

(Image from Blei, 2011)
**Machine Translation**

- Alignment is a hidden variable in translation models
- With deep learning, this is embodied in “attention” models
With Bayesian inference, the parameters are a “latent” variable:

\[ p(\theta, h \mid x) = \frac{p(\theta, h, x)}{\int_{\theta} \sum_{h} p(\theta, h, x)} \]

- Popularized latent-variable models (where structure is missing as well)
- Has been used for problems in morphology, word segmentation, syntax, semantics and others
This Talk in a Nutshell

How do we learn from incomplete data?

- The case of syntactic parsing
- Other uses of grammars for learning from incomplete data
- The canonical correlation principle and its uses
Do we need to work on parsing when we can build direct “transducers?” (such as with deep learning)
Do we need to work on parsing when we can build direct “transducers?” (such as with deep learning) **Yes!**

- We develop algorithms that generalize to structured prediction

- We see recent results that even with deep learning, incorporating parse structures can help applications such as machine translation (Bastings et al., 2017; Kim et al., 2017)

- We develop theories for syntax in language and test them empirically

- One of the classic problems that demonstrates so well ambiguity in natural language
In a general way such speculation is epistemologically relevant, as suggesting how organisms maturing and evolving in the physical environment we know might conceivably end up discoursing of abstract objects as we do

(Quine, 1960, p. 123)

- Should be interpreted: organisms might end up ...
In a general way such speculation is epistemologically relevant as suggesting how organisms maturing and evolving in the physical environment we know objects as we do conceivably end up discoursing of abstract
Latent-State Syntax (Matsuzaki et al., 2005; Prescher, 2005; Petrov et al., 2006)

Improves the accuracy of a PCFG model from $\sim 70\%$ to $\sim 90\%$. 
Latent-State Syntax (Matsuzaki et al., 2005; Prescher, 2005; Petrov et al., 2006)

Improve the domain of locality. PCFG – domain of locality is just one rule. Latent-variable PCFGs – latent states extend the domain of locality.

Implements the accuracy of a PCFG model from $\sim 70\%$ to $\sim 90\%$. 
Generative Process

$S^1$
Generative Process

S¹

NP  VP
Generative Process
Generative Process

S

NP

VP

D N
Generative Process

S
   /\      /
  NP  VP
  /   / \  /
D   N  D  N
Generative Process

S
  / \  
NP3   VP2
   / \   /  
D1  N2 V  P
Generative Process

S¹

NP³

D¹ N²

VP²

V⁴ P¹
Generative Process

- Derivational process is similar to that of PCFG together with contextual information

- We read the grammar off the treebank, but not the latent states
Evolution of L-PCFGs

1997 - Charniak, Collins
Lexicalized grammars
1998 – Johnson
Treebank transformations
1999 – Eisner
Bilexical grammars
Evolution of L-PCFGs

2004 - Klein and Manning
Linguistic annotations

1997 - Charniak, Collins
Lexicalized grammars
1998 – Johnson
Treebank transformations
1999 – Eisner
Bilexical grammars
Evolution of L-PCFGs

2004 - Klein and Manning
Linguistic annotations

1997 - Charniak, Collins
Lexicalized grammars
1998 - Johnson
Treebank transformations
1999 - Eisner
Bilexical grammars

2005 - Matsuzaki et al.
2005 - Prescher
Latent annotations

S
  NP
    D
   the
  N
  V
  P
  the dog
  saw
  him
Evolution of L-PCFGs

1997 - Charniak, Collins
Lexicalized grammars
1998 – Johnson
Treebank transformations
1999 – Eisner
Bilexical grammars

2004 - Klein and Manning
Linguistic annotations

2005 - Matsuzaki et al.
2005 - Prescher
Latent annotations

2006 - Petrov et al.
Coarse-to-fine methods
Goal: Given a treebank, estimate rule probabilities, including for latent states.

Traditional way: use the expectation-maximization (EM) algorithm:

- E-step - infer values for latent states using dynamic programming
- M-step - re-estimate the model parameters based on the values inferred
Local Maxima with EM

Convex optimization

Non-convex optimization

EM finds a local maximum of a non-convex objective

Especially problematic with unsupervised learning
How Problematic are Local Maxima?

For unsupervised learning, local maxima are a very serious problem:

For deep learning, can also be a problem. For L-PCFGs, variability is smaller.

Depends on the problem and the model.
At node VP:

Outside tree $o =$

Inside tree $t =$

Conditionally independent given the label and the hidden state

$$p(o, t|VP, h) = p(o|VP, h) \times p(t|VP, h)$$
Cross-Covariance Matrix

Create a cross-covariance matrix and apply singular value decomposition to get the latent space:

\[
\begin{pmatrix}
\begin{array}{cccc}
\text{outside tree 1} & \text{outside tree 10} \\
\text{inside tree 1} & 1 & 0 & \ldots & 1 \\
\text{inside tree 10} & 1 & 0 & \ldots & 1 \\
\end{array}
\end{pmatrix}
\]

Based on the method of moments – set up a set of equations that mix moments and parameters and have a unique solution
The idea of using a co-occurrence matrix to extract latent information is an old idea. It has been used for:

- Learning hidden Markov models and finite state automata (Hsu et al., 2012; Balle et al., 2013)
- Learning word embeddings (Dhillon et al., 2011)
- Learning dependency and other types of grammars (Bailly et al., 2010; Luque et al., 2012; Dhillon et al., 2012)
- Learning document-topic structure (Anandkumar et al., 2012)

Much of this work falls under the use of **canonical correlation analysis** (Hotelling, 1935)
Feature Functions

Need to define feature functions for inside and outside trees

\[ \phi \left( \begin{array}{c}
V \\
\text{saw}
\end{array} \right) \left( \begin{array}{c}
P \\
\text{him}
\end{array} \right) = (0, \ldots, 1, \ldots, 1, 0, 1, 0) \]

\[ \psi \left( \begin{array}{c}
S \\
\text{NP}
\end{array} \right) \left( \begin{array}{c}
\text{VP}
\end{array} \right) = (0, \ldots, 1, \ldots, 0, 0, 0, 1) \]
Consider the VP node in the following tree:

The inside features consist of:

- The pairs \((VP, V)\) and \((VP, NP)\)
- The rule \(VP \rightarrow V \ NP\)
- The tree fragment \((VP (V \ saw) \ NP)\)
- The tree fragment \((VP \ V (NP \ D \ N))\)
- The pair of head part-of-speech tag with VP: \((VP, V)\)
- The width of the subtree spanned by VP: \((VP, 2)\)
Consider the D node in the following tree:

The outside features consist of:
- The fragments
- The pair \((D, \text{NP})\) and triplet \((D, \text{NP}, \text{VP})\)
- The pair of head part-of-speech tag with D: \((D, \text{N})\)
- The widths of the spans left and right to D: \((D, 3)\) and \((D, 1)\)
Consider the D node in the following tree:

As such, the algorithm is another step in the evolution of L-PCFGs – throw in all the information and local context from previous work and automatically distill it into latent states!

The outside features used:

- The fragments $\langle D, NP \rangle$ and $\langle D, NP, VP \rangle$
- The pair of head part-of-speech tag with D: $\langle D, N \rangle$
- The widths of the spans left and right to D: $\langle D, 3 \rangle$ and $\langle D, 1 \rangle$
Final Results on Multilingual Parsing

Narayan and Cohen (2016):

<table>
<thead>
<tr>
<th>Language</th>
<th>Berkeley</th>
<th>Spectral</th>
<th>Spectral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster</td>
<td>SVD</td>
<td></td>
</tr>
<tr>
<td>Basque</td>
<td>74.7</td>
<td>81.4</td>
<td>80.5</td>
</tr>
<tr>
<td>French</td>
<td>80.4</td>
<td>75.6</td>
<td>79.1</td>
</tr>
<tr>
<td>German</td>
<td>78.3</td>
<td>76.0</td>
<td>78.2</td>
</tr>
<tr>
<td>Hebrew</td>
<td>87.0</td>
<td>87.2</td>
<td>89.0</td>
</tr>
<tr>
<td>Hungarian</td>
<td>85.2</td>
<td>88.4</td>
<td>89.2</td>
</tr>
<tr>
<td>Korean</td>
<td>78.6</td>
<td>78.4</td>
<td>80.0</td>
</tr>
<tr>
<td>Polish</td>
<td>86.8</td>
<td>91.2</td>
<td>91.8</td>
</tr>
<tr>
<td>Swedish</td>
<td>80.6</td>
<td>79.4</td>
<td>80.9</td>
</tr>
</tbody>
</table>

Parsing is far from being solved in the multilingual setting
Closed-word tags essentially do lexicalization:

<table>
<thead>
<tr>
<th>State</th>
<th>Frequent words</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>of $\times 323$</td>
</tr>
<tr>
<td>1</td>
<td>about $\times 248$</td>
</tr>
<tr>
<td>2</td>
<td>than $\times 661$, as $\times 648$, because $\times 209$</td>
</tr>
<tr>
<td>3</td>
<td>from $\times 313$, at $\times 324$</td>
</tr>
<tr>
<td>4</td>
<td>into $\times 178$</td>
</tr>
<tr>
<td>5</td>
<td>over $\times 122$</td>
</tr>
<tr>
<td>6</td>
<td>Under $\times 127$</td>
</tr>
</tbody>
</table>
# What Do We Learn?

<table>
<thead>
<tr>
<th>State</th>
<th>Frequent words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>DT (determiners)</strong></td>
</tr>
<tr>
<td>0</td>
<td>These $\times 105$</td>
</tr>
<tr>
<td>1</td>
<td>Some $\times 204$</td>
</tr>
<tr>
<td>2</td>
<td>that $\times 190$</td>
</tr>
<tr>
<td>3</td>
<td>both $\times 102$</td>
</tr>
<tr>
<td>4</td>
<td>any $\times 613$</td>
</tr>
<tr>
<td>5</td>
<td>the $\times 574$</td>
</tr>
<tr>
<td>6</td>
<td>those $\times 247$, all $\times 242$</td>
</tr>
<tr>
<td>7</td>
<td>all $\times 105$</td>
</tr>
<tr>
<td>8</td>
<td>another $\times 276$, no $\times 211$</td>
</tr>
</tbody>
</table>
## What Do We Learn?

<table>
<thead>
<tr>
<th>State</th>
<th>Frequent words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>CD (numbers)</strong></td>
</tr>
<tr>
<td>0</td>
<td>$8 \times 132$</td>
</tr>
<tr>
<td>1</td>
<td>million $\times 451$, billion $\times 248$</td>
</tr>
<tr>
<td></td>
<td><strong>RB (adverb)</strong></td>
</tr>
<tr>
<td>0</td>
<td>up $\times 175$</td>
</tr>
<tr>
<td>1</td>
<td>as $\times 271$</td>
</tr>
<tr>
<td>2</td>
<td>not $\times 490$, n’t $\times 2695$</td>
</tr>
<tr>
<td>3</td>
<td>not $\times 236$</td>
</tr>
<tr>
<td>4</td>
<td>only $\times 159$</td>
</tr>
<tr>
<td>5</td>
<td>well $\times 129$</td>
</tr>
</tbody>
</table>
## What Do We Learn?

<table>
<thead>
<tr>
<th>State</th>
<th>Frequent words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CC</strong> (conjunction)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>But ×255</td>
</tr>
<tr>
<td>1</td>
<td>and ×101</td>
</tr>
<tr>
<td>2</td>
<td>and ×218</td>
</tr>
<tr>
<td>3</td>
<td>But ×196</td>
</tr>
<tr>
<td>4</td>
<td>or ×162</td>
</tr>
<tr>
<td>5</td>
<td>and ×478</td>
</tr>
</tbody>
</table>
What Do We Learn?

Example latent for NP:

- “James McCall, vice president, materials, at Battelle, a technology and management-research giant based in Columbus, Ohio”
- “Frank Kline Jr., partner in Lambda Funds, a Beverly Hills, Calif., venture capital concern”
- “Allen Hadhazy, senior analyst at the Institute for Econometric Research, Fort Lauderdale, Fla., which publishes the New Issues newsletter on IPOs”
- “Charles J. O’Connell, deputy district director in Los Angeles of the California Department of Transportation, nicknamed Caltrans”
- “Francis J. McNeil, who, as deputy assistant secretary of state for inter-American affairs, first ran across reports about Mr. Noriega in 1977”
What Do We Learn?

Example latent state for NP:

Another state:

- “AMERICAN BUILDING MAINTENANCE INDUSTRIES Inc., San Francisco, provider of maintenance services, annual revenue of $582 million, NYSE,”
- “DIASONICS INC., South San Francisco, maker of magnetic resonance imaging equipment, annual sales of $281 million, Amex,”
- “EVEREX SYSTEMS INC., Fremont, maker of personal computers and peripherals, annual sales of $377 million, OTC,”
- “ANTHEM ELECTRONICS INC., San Jose, distributor of electronic parts, annual sales of about $300 million, NYSE,”
If you are interested in further looking at such patterns for other languages (English, French, German, Hebrew, Hungarian, Korean, Polish, Swedish, Basque), consider visiting http://cohort.inf.ed.ac.uk/lpcfgviewer/index.php.
This Talk in a Nutshell

How do we learn from incomplete data?

- The case of syntactic parsing
- Other uses of grammars for learning from incomplete data
- The canonical correlation principle and its uses
A Different Perspective

S

NP       VP

D        N       V       P

the      dog     saw     him

(0.1, 0.8, 0.1)   (0.1, 0.7, 0.2)   (0.5, 0.4, 0.1)   (0.3, 0.2, 0.5)
A Different Perspective

S

NP [D, N]

VP [V, P]

the, dog, saw, him

(0.1, 0.8, 0.1) (0.1, 0.7, 0.2) (0.5, 0.4, 0.1) (0.3, 0.2, 0.5)
A Different Perspective

S

NP

D
the

(0.1, 0.8, 0.1)

N
dog

(0.1, 0.7, 0.2)

VP

V
saw

(0.5, 0.4, 0.1)

P
him

(0.3, 0.2, 0.5)
A Different Perspective

- Related to neural network models with grammars (Socher et al., 2010; Socher et al., 2013)
- Also related to compositional distributional semantics (Baroni and Bernardi, 2014; Grefenstette and Sadrzadeh, 2010; Coecke et al., 2010)
Question Answering (Narayan et al., 2016)
What language do people in Czech Republic speak?

Question Answering (Narayan et al., 2016)
PC Hardware

About This Forum  /  Real-Time Activity  /  My Tracked Discussions  /  FAQs  /  Policies  /  Moderators

Question

Cleaned computer, now runs games extremely slow?

by Naggles71 / June 3, 2015 5:45 AM PDT

Hi everyone, hope you can give me some insight to my issue because I have no idea what has happened.

Yesterday, I decided to clean the dust from my computer as well as check my CPU for any damage because lately it has been running very very hot (I didn’t remove the CPU, just the fan and checked around it and didn’t have any issues removing or putting the fan back in). Now once I started my computer again I tried running games such as BF4/Lol/HotS, and everytime a lot of action would occur my FPS would drop to 2-3, where I used to run everything at a solid 30-60 FPS.

Does anyone have any idea what could have happened? My specs show that everything is working as it should and I am not well-versed enough to tell a difference in the performance, nor do I have logs from before this happened using the 3DMark benchmark.

If you need more info just let me know I will be monitoring this very closely, thank you!

ANSWER THIS  ASK FOR CLARIFICATION

Track this discussion  Thread display: Collapse / Expand  8 total posts
Discussion Forums

OK, I have to guess that's the CPU thermal paste.
by R. Profit / June 3, 2015 6:12 AM PDT
In reply to: Cleaned computer, now runs games extremely slowly?
How about the GPU? For example those need love too. Example follows.
http://www.tomshardware.com/reviews/radeon-r9-290x-thermal-paste-efficiency.3678.html

GPU
by Napplastig! / June 3, 2015 10:19 AM PDT
In reply to: OK, I have to guess that's the CPU thermal paste.
I am going to check out everything in my PC and make sure it's all plugged in correctly. I could have bumped something without my knowledge but I feel like when I checked my information it would have told me that something was not working correctly whereas it says everything is working. I will give an update once I do this, thanks.

But Bob's Right
by itadigger / June 3, 2015 12:38 PM PDT
In reply to: GPU
If your GPU is old or even new, you should apply new thermal paste.
I have 2 brand new Nvidia 750 Ti GPU's and right out of the box I removed and replaced with new thermal paste as from the factory I could see that it was way too thick.

Digger
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>p₀</strong> Bob:</td>
<td>When I play a recorded video on my camera, it looks and sounds fine. On my computer, it plays at a really fast rate and sounds like Alvin and the Chipmunks!</td>
<td></td>
</tr>
<tr>
<td><strong>p₁</strong> Kate:</td>
<td>I’d find and install the machine’s latest audio driver.</td>
<td></td>
</tr>
<tr>
<td><strong>p₂</strong> Mary:</td>
<td>The motherboard supplies the clocks for audio feedback. So update the audio and motherboard drivers.</td>
<td></td>
</tr>
<tr>
<td><strong>p₃</strong> Chris:</td>
<td>Another fine mess in audio is volume and speaker settings. You checked these?</td>
<td></td>
</tr>
<tr>
<td><strong>p₄</strong> Jane:</td>
<td>Yes, under speaker settings, look for hardware acceleration. Turning it off worked for me.</td>
<td></td>
</tr>
<tr>
<td><strong>p₅</strong> Matt:</td>
<td>Audio drivers are at this link. Rather than just audio drivers, I would also just do all drivers.</td>
<td></td>
</tr>
<tr>
<td>Page</td>
<td>Character</td>
<td>Response</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>0</td>
<td>Bob</td>
<td>When I play a recorded video on my camera, it looks and sounds fine. On my computer, it plays at a really fast rate and sounds like Alvin and the Chipmunks!</td>
</tr>
<tr>
<td>1</td>
<td>Kate</td>
<td>I’d find and install the machine’s latest audio driver.</td>
</tr>
<tr>
<td>2</td>
<td>Mary</td>
<td>The motherboard supplies the clocks for audio feedback. So update the audio and motherboard drivers.</td>
</tr>
<tr>
<td>3</td>
<td>Chris</td>
<td>Another fine mess in audio is volume and speaker settings. You checked these?</td>
</tr>
<tr>
<td>4</td>
<td>Jane</td>
<td>Yes, under speaker settings, look for hardware acceleration. Turning it off worked for me.</td>
</tr>
<tr>
<td>5</td>
<td>Matt</td>
<td>Audio drivers are at this link. Rather than just audio drivers, I would also just do all drivers.</td>
</tr>
</tbody>
</table>
Bob: When I play a recorded video on my camera, it looks and sounds fine. On my computer, it plays at a really fast rate and sounds like Alvin and the Chipmunks!

Kate: I’d find and install the machine’s latest audio driver.

Mary: The motherboard supplies the clocks for audio feedback. So update the audio and motherboard drivers.

Chris: Another fine mess in audio is volume and speaker settings. You checked these?

Jane: Yes, under speaker settings, look for hardware acceleration. Turning it off worked for me.

Matt: Audio drivers are at this link. Rather than just audio drivers, I would also just do all drivers.
<table>
<thead>
<tr>
<th>Page</th>
<th>Name</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>Bob</td>
<td>When I play a recorded video on my camera, it looks and sounds fine. On my computer, it plays at a really fast rate and sounds like Alvin and the Chipmunks!</td>
</tr>
<tr>
<td>p1</td>
<td>Kate</td>
<td>I’d find and install the machine’s latest audio driver.</td>
</tr>
<tr>
<td>p2</td>
<td>Mary</td>
<td>The motherboard supplies the clocks for audio feedback. So update the audio and motherboard drivers.</td>
</tr>
<tr>
<td>p3</td>
<td>Chris</td>
<td>Another fine mess in audio is volume and speaker settings. You checked these?</td>
</tr>
<tr>
<td>p4</td>
<td>Jane</td>
<td>Yes, under speaker settings, look for hardware acceleration. Turning it off worked for me.</td>
</tr>
<tr>
<td>p5</td>
<td>Matt</td>
<td>Audio drivers are at this link. Rather than just audio drivers, I would also just do all drivers.</td>
</tr>
</tbody>
</table>
Conversation Trees

My audio plays like Alvin and Chipmunks!

Bob: install audio driver.

Mary: update audio and motherboard drivers.

Chris: volume & speaker settings

Jane: hardware acceleration

(Louis and Cohen, 2015)
Conversation Trees

S

X-5

Bob: install audio driver.

X-2

Mary: update audio and motherboard drivers.

X-3

Chris: volume & speaker settings

X-3

Jane: hardware acceleration

My audio plays like Alvin and Chipmunks!

(Louis and Cohen, 2015)
How do we learn from incomplete data?

- The case of syntactic parsing
- Other uses of grammars for learning from incomplete data
- The canonical correlation principle and its uses
• Assume a “confounding” variable that explains two separate views

• Correlation between $x$ and $y$ gives $z$ – the two are independent given $z$

• In the case of L-PCFGs: $x$ and $y$ are inside and outside trees

• Where else can this principle be used?
Word Embeddings

\[
\begin{pmatrix}
\text{the} & \star & \text{chased} \\
\text{mouse} & 1 & 0 & \ldots & 1 \\
& 0 & 1 & \ldots & 0 \\
& \vdots & \vdots & \ddots & \vdots \\
\text{cat} & 1 & 0 & \ldots & 1 \\
\end{pmatrix}
\]

- Co-occurrence matrix of words and contexts ("the cat chased", "the mouse chased")
- Apply CCA on this matrix to get word embeddings (Dhillon et al., 2011)
- Inject prior knowledge into matrix (Osborne et al., 2016)
• Requires also generation (using sampling techniques)

• The probability of text we sample is proportional to the “similarity” of the text to the image
Example Predictions

Good predictions:
- mike and jenny are camping
- mike is holding a bat
- jenny is throwing the frisbee

Bad predictions:
- mike is kicking a blass
- jenny wants the bear
- the rocket is behind mike
The dog, true to form, chased the cat.

- Very challenging problem
- Sensitive to local maxima with existing techniques such as EM
- What if the tree for each pair of words in the sentence is a latent, confounding, hierarchical variable?

(Parikh et al., 2014)
Conclusion

Latent-variable grammars are useful for problems outside of syntax

- Their symbolic component is interpretable
- Their probabilistic component helps reasoning under uncertainty
- Latent variables help detect unseen patterns

I have shown you how grammars can be used for several problems, and also how the principle behind learning latent-variable grammars can be used for other problems.
References I

Steven Abney.
Statistical methods and linguistics.
*The balancing act: Combining symbolic and statistical approaches to language*, pages 1–26, 1996.

Anima Anandkumar, Dean P Foster, Daniel J Hsu, Sham M Kakade, and Yi-Kai Liu.
A spectral algorithm for latent dirichlet allocation.

R. Bailly, A. Habrard, and F. Denis.
A spectral approach for probabilistic grammatical inference on trees.
Borja Balle, Xavier Carreras, Franco M Luque, and Ariadna Quattoni.  
Spectral learning of weighted automata.  

Marco Baroni, Raffaela Bernardi, and Roberto Zamparelli.  
Frege in space: A program of compositional distributional semantics.  
*LiLT (Linguistic Issues in Language Technology)*, 9, 2014.

Graph convolutional encoders for syntax-aware neural machine translation.  


A. Louis and S. B. Cohen. 
Conversation trees: A grammar model for topic structure in forums. 

F. M. Luque, A. Quattoni, B. Balle, and X. Carreras. 
Spectral learning for non-deterministic dependency parsing. 

T. Matsuzaki, Y. Miyao, and J. Tsujii. 
Probabilistic CFG with latent annotations. 

S. Narayan and S. B. Cohen. 
Optimizing spectral learning for parsing. 
Paraphrase generation from latent-variable pcfgs for semantic parsing.

A. P. Parikh, S. B. Cohen, and E. Xing.
Spectral unsupervised parsing with additive tree metrics.

S. Petrov, L. Barrett, R. Thibaux, and D. Klein.
Learning accurate, compact, and interpretable tree annotation.

D. Prescher.
Inducing head-driven PCFGs with latent heads: Refining a tree-bank grammar for parsing.
