Admixture of Poisson MRFs: A New Topic Model with Word Dependencies

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon

April 30, 2015

* Presenter
Analyzing Large Collections of Documents

Collection of Documents

Examples:
1. Research papers
2. News articles
3. Twitter posts
Analyzing Large Collections of Documents

Collection of Documents

Digital Representation

Examples:
1. Research papers
2. News articles
3. Twitter posts

Bag of Words Matrix:
- Removes order and syntax information
- Unrealistic but powerful
Analyzing Large Collections of Documents

Examples:
1. Research papers
2. News articles
3. Twitter posts

Bag of Words Matrix:
- Removes order and syntax information
- Unrealistic but powerful

Collection of Documents

Digital Representation

Model Computation

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon
Admixture of Poisson MRFs
Analyzing Large Collections of Documents

Collection of Documents

Examples:
1. Research papers
2. News articles
3. Twitter posts

Digital Representation

Bag of Words Matrix:
- Removes order and syntax information
- Unrealistic but powerful

Model Computation

Summary

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon
Admixture of Poisson MRFs
Example of Research Paper - Top Words

**Collection of Documents**

**Digital Representation**

```
Doc 1
Doc 2
Doc 3
Doc 4
...
```

```
"networks"
"learning"
"programming"
```

**Model Computation**

**Summary**

Top Words
- networks
- learning
- based
- using
- analysis
- network
- wireless
- data
- model
- multi
- control
- efficient
- systems
- models
- time

**Titles of Research Papers:**
1. Machine Learning (ICML, NIPS)
2. Communication Networks (INFOCOM)
3. Programming Languages (PLDI, CAV, POPL, OOPSLA)
Research Paper Example - Topic Modeling

Collection of Documents → Digital Representation → Model Computation → Summary

Titles of Research Papers:
1. Machine Learning (ICML, NIPS)
2. Communication Networks (INFOCOM)
3. Programming Languages (PLDI, CAV, POPL, OOPSLA)

Example:
1. Research papers
2. News articles
3. Twitter posts

Bag of Words Matrix:
- Removes order and syntax information
- Unrealistic but powerful

Examples:
- Topic 1: networks, wireless, network, control
- Topic 2: learning, using, models, neural, model
- Topic 3: analysis, model, object, programming, language

Admixture of Poisson MRFs
Applications for Topic Modeling

► Applications

1. Summarize/Visualize [Hall et al. 2008]
2. Word sense disambiguation [Boyd-Graber et al. 2007]
4. Information retrieval [Wei & Croft 2006]

► Different domains

1. Genetics [Pritchard et al. 2000 (14,000 citations)]
2. Computer vision [Li et al. 2010]
4. Social science surveys [Roberts et al. 2014]
5. Social E-commerce [Hu et al. 2014]
Brief History - Latent Semantic Analysis (LSA)

Digital Representation

<networks>
<learning>
<programming>
...

Doc 1
Doc 2
Doc 3
Doc 4
...

p

n

Digital Representation

“networks”
“learning”
“programming”

...
Digital Representation

Singular Value Decomposition

Brief History - Latent Semantic Analysis (LSA)

Low Dimensional Document Representation

“Latent Topic”

Doc 1
Doc 2
Doc 3
Doc 4
...

“networks”
“learning”
“programming”
...

Digital Representation

Singular Value Decomposition

U
Σ
V^T

Positive and negative values difficult to interpret

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon
Admixture of Poisson MRFs
Digital Representation

Singular Value Decomposition

Low Dimensional Document Representation

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon

Admixture of Poisson MRFs
Brief History - Latent Semantic Analysis (LSA)

Digital Representation

Singular Value Decomposition

Positive and negative values difficult to interpret
Probability vectors are much easier to interpret

LDA - Added Bayesian priors for regularization

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon

Admixture of Poisson MRFs
Comparison of 2D Projections

- SVD dimensions are difficult to interpret
- APM has smooth distribution compared to LDA

SVD

LDA

APM

- comm-net.6978
- mach-learn.8925
- prog-lang.6618

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon

Admixture of Poisson MRFs
1. Add time information [Blei & Lafferty 2006]
2. Add author information [Rosen-Zvi et al. 2004]
3. Add document category information [Mcauliffe & Blei 2008]
4. Automatically discover number of topics [Teh et al. 2006]
5. Model correlation between topics [Blei & Lafferty 2006]
6. . . .
7. . . .
Brief History - Extensions/Variants

1. Add time information [Blei & Lafferty 2006]

2. Add author information [Rosen-Zvi et al. 2004]

3. Add document category information [Mcauliffe & Blei 2008]

4. Automatically discover number of topics [Teh et al. 2006]

5. Model correlation between topics [Blei & Lafferty 2006]

6. . . .

7. . . .

- Previous models - topics only have weights for single words
- Our model - topics have weights for pairs of words
Interpreting Topics

**LDA 3 topics**

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>networks</td>
<td>learning</td>
<td>analysis</td>
</tr>
<tr>
<td>wireless</td>
<td>using</td>
<td>model</td>
</tr>
<tr>
<td>network</td>
<td>models</td>
<td>object</td>
</tr>
<tr>
<td>based</td>
<td>neural</td>
<td>programming</td>
</tr>
<tr>
<td>control</td>
<td>model</td>
<td>language</td>
</tr>
<tr>
<td>routing</td>
<td>based</td>
<td>using</td>
</tr>
<tr>
<td>performance</td>
<td>bayesian</td>
<td>oriented</td>
</tr>
<tr>
<td>analysis</td>
<td>classification</td>
<td>systems</td>
</tr>
</tbody>
</table>
### Interpreting Topics

#### LDA 3 topics

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>networks</td>
<td>learning</td>
<td>analysis</td>
</tr>
<tr>
<td>wireless</td>
<td>using</td>
<td>model</td>
</tr>
<tr>
<td>network</td>
<td>models</td>
<td>object</td>
</tr>
<tr>
<td>based</td>
<td>neural</td>
<td>programming</td>
</tr>
<tr>
<td>control</td>
<td>model</td>
<td>language</td>
</tr>
<tr>
<td>routing</td>
<td>based</td>
<td>using</td>
</tr>
<tr>
<td>performance</td>
<td>bayesian</td>
<td>oriented</td>
</tr>
<tr>
<td>analysis</td>
<td>classification</td>
<td>systems</td>
</tr>
</tbody>
</table>

#### LDA 6 topics

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
<th>Topic 5</th>
<th>Topic 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>networks</td>
<td>learning</td>
<td>learning</td>
<td>networks</td>
<td>networks</td>
<td>object</td>
</tr>
<tr>
<td>wireless</td>
<td>neural</td>
<td>models</td>
<td>performance</td>
<td>control</td>
<td>programming</td>
</tr>
<tr>
<td>sensor</td>
<td>using</td>
<td>data</td>
<td>algorithm</td>
<td>oriented</td>
<td>oriented</td>
</tr>
<tr>
<td>network</td>
<td>networks</td>
<td>using</td>
<td>algorithms</td>
<td>analysis</td>
<td>analysis</td>
</tr>
<tr>
<td>routing</td>
<td>model</td>
<td>data</td>
<td>algorithm</td>
<td>language</td>
<td>language</td>
</tr>
<tr>
<td>multi</td>
<td>inference</td>
<td>model</td>
<td>decision</td>
<td>model</td>
<td>model</td>
</tr>
<tr>
<td>data</td>
<td>clustering</td>
<td>network</td>
<td>based</td>
<td>traffic</td>
<td>using</td>
</tr>
<tr>
<td>mobile</td>
<td>sparse</td>
<td>sparse</td>
<td>using</td>
<td>atm</td>
<td>using</td>
</tr>
<tr>
<td>recognition</td>
<td>gaussian</td>
<td>gaussian</td>
<td>vector</td>
<td>based</td>
<td>programs</td>
</tr>
</tbody>
</table>
Interpreting Topics

**LDA 3 topics**

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>networks</td>
<td>learning</td>
<td>analysis</td>
</tr>
<tr>
<td>wireless</td>
<td>using</td>
<td>model</td>
</tr>
<tr>
<td>network</td>
<td>models</td>
<td>object</td>
</tr>
<tr>
<td>based</td>
<td>neural</td>
<td>programming</td>
</tr>
<tr>
<td>control</td>
<td>model</td>
<td>language</td>
</tr>
<tr>
<td>routing</td>
<td>based</td>
<td>using</td>
</tr>
<tr>
<td>performance</td>
<td>bayesian</td>
<td>oriented</td>
</tr>
<tr>
<td>analysis</td>
<td>classification</td>
<td>systems</td>
</tr>
</tbody>
</table>

**LDA 30 topics**

<table>
<thead>
<tr>
<th>Top 2 Words for 30 Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>bounds+maximum</td>
</tr>
<tr>
<td>learning+algorithms</td>
</tr>
<tr>
<td>analysis+approach</td>
</tr>
<tr>
<td>networks+network</td>
</tr>
<tr>
<td>networks+wireless</td>
</tr>
<tr>
<td>learning+based</td>
</tr>
<tr>
<td>object+oriented</td>
</tr>
<tr>
<td>data+cloud</td>
</tr>
<tr>
<td>detection+networks</td>
</tr>
<tr>
<td>analysis+data</td>
</tr>
</tbody>
</table>
## Interpreting Topics

### LDA 3 topics

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>networks</td>
<td>learning</td>
<td>analysis</td>
</tr>
<tr>
<td>wireless</td>
<td>using</td>
<td>model</td>
</tr>
<tr>
<td>network</td>
<td>models</td>
<td>object</td>
</tr>
<tr>
<td>based</td>
<td>neural</td>
<td>programming</td>
</tr>
<tr>
<td>control</td>
<td>model</td>
<td>language</td>
</tr>
<tr>
<td>routing</td>
<td>based</td>
<td>using</td>
</tr>
<tr>
<td>performance</td>
<td>bayesian</td>
<td>oriented</td>
</tr>
<tr>
<td>analysis</td>
<td>classification</td>
<td>systems</td>
</tr>
</tbody>
</table>

### LDA 30 topics

<table>
<thead>
<tr>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>bounds + maximum</td>
</tr>
<tr>
<td>learning + algorithms</td>
</tr>
<tr>
<td>analysis + approach</td>
</tr>
<tr>
<td>networks + network</td>
</tr>
<tr>
<td>networks + wireless</td>
</tr>
<tr>
<td>algorithm + fast</td>
</tr>
<tr>
<td>model + checking</td>
</tr>
<tr>
<td>memory + based</td>
</tr>
<tr>
<td>software + peer</td>
</tr>
<tr>
<td>high + performance</td>
</tr>
<tr>
<td>learning + clustering</td>
</tr>
<tr>
<td>detection + networks</td>
</tr>
<tr>
<td>code + java</td>
</tr>
</tbody>
</table>

### APM 3 topics

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>programs</td>
<td>learning</td>
<td>routing</td>
</tr>
<tr>
<td>program</td>
<td>monte + carlo</td>
<td>wireless</td>
</tr>
<tr>
<td>languages</td>
<td>gaussian</td>
<td>large + scale</td>
</tr>
<tr>
<td>java</td>
<td>support + vector</td>
<td></td>
</tr>
<tr>
<td>session + poster</td>
<td>decision</td>
<td>networks</td>
</tr>
<tr>
<td>software</td>
<td>neighbor + nearest</td>
<td></td>
</tr>
<tr>
<td>domain + specific</td>
<td>reinforcement</td>
<td>networks</td>
</tr>
<tr>
<td>typed</td>
<td>high + dimensional</td>
<td></td>
</tr>
<tr>
<td>object + oriented</td>
<td>neural</td>
<td>high + speed</td>
</tr>
<tr>
<td>workshop</td>
<td>message + passing</td>
<td></td>
</tr>
<tr>
<td>model + checking</td>
<td>least + squares</td>
<td></td>
</tr>
<tr>
<td>language</td>
<td>temporal + difference</td>
<td></td>
</tr>
<tr>
<td>high + level</td>
<td>nonparametric</td>
<td></td>
</tr>
<tr>
<td>order + higher</td>
<td>supervised + semi</td>
<td></td>
</tr>
</tbody>
</table>

---

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon

Admixture of Poisson MRFs
Overview of APM

LDA

Admixture of Poisson MRFS (APM)

Multinomial

Admixture

Mixture

Poisson MRF

Gaussian MRF

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon

Admixture of Poisson MRFs
Overview of APM

LDA

Multinomial

Admixture

Mixture

Admixture of Poisson MRFS (APM)

Poisson MRF

Gaussian MRF
Mixtures

- Multiple sub-populations
- The sub-populations are usually unknown a priori
- Each individual from the population comes from exactly one subpopulation

Figure source: Kalai, Moitra, and Valiant. Disentangling Gaussians. Communications of the ACM. 2012.
Admixtures

- **Mixtures** - Draws from single component distribution. (Top)

- **Admixtures** - Draws from a distribution whose parameters are a convex combination of component parameters. (Bottom)
Overview of APM

LDA

Multinomial

Admixture

Mixture

Admixture of Poisson MRFS (APM)

Poisson MRF

Gaussian MRF

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon
Gaussian MRFs

- Allows for dependencies between variables

What if the data dimension is large?
- If dimension is 1000, $1000^2/2 = 500,000$ parameters
- Assume some conditional independence between variables.
1. Each **conditional** ("slice") of a PMRF is 1-D Poisson.
2. **Distinct** from Gaussian MRF
3. Positive dependencies can model **word co-occurrence**.
If we assume the node conditional distributions are Poisson, does there exist a joint MRF distribution that has these conditionals?

▶ Poisson MRF joint distribution:

$$\text{Pr}_{\text{PMRF}}(x | \theta, \Theta) \propto \exp \left\{ \theta^T x + x^T \Theta x - \sum_{s=1}^{p} \ln(x_s!) \right\}.$$

▶ Node conditionals are 1-D Poissons:

$$\text{Pr}(x_s | x_{-s}, \theta_s, \Theta_s) \propto \exp\left\{ (\theta_s + x_{-s}^T \Theta_s) x_s - \ln(x_s!) \right\}.$$

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon
Admixture of Poisson MRFs
Overview of APM

- LDA
- Admixture of Poisson MRFS (APM)
  - Multinomial
  - Admixture
  - Mixture
  - Poisson MRF
  - Gaussian MRF

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon
Admixture of Poisson MRFs (APM) [Inouye et al. 2014]

- APM replaces standard Multinomial with Poisson MRF

\[
\Pr_{\text{APM}} (x, w, \theta^{1...k}, \Theta^{1...k}) = \Pr_{\text{PMRF}} \left( x \mid \bar{\theta} = \sum_{j=1}^{k} w_j \theta^j, \bar{\Theta} = \sum_{j=1}^{k} w_j \Theta^j \right) \Pr_{\text{Dir}} (w) \prod_{j=1}^{k} \Pr (\theta^j, \Theta^j)
\]
APM Algorithm

1. Optimization problem is not convex

2. Want to exploit parallel computing

3. Large optimization problem: APM has $O(kp^2)$ parameters versus $O(kp)$ for LDA
   - LDA($k = 5, p = 1000$) ⇒ 5,000 parameters
   - APM($k = 5, p = 1000$) ⇒ 5,000,000 parameters
   - APM($k = 5, \text{NNZ}(\Theta) = 10 \text{ per word}$) ⇒ 50,000 free parameters
Parallel Alternating Newton-like Algorithm

1. Split the algorithm into alternating convex problems

\[
\begin{align*}
\arg \min_{\Phi^1, \Phi^2, \ldots, \Phi^p} & -\frac{1}{n} \sum_{s=1}^{p} \left[ \text{tr}(\Psi^s \Phi^s) - \sum_{i=1}^{n} \exp(z_i^T \Phi^s w_i) \right] + \sum_{s=1}^{p} \lambda \|\text{vec}(\Phi^s)\|_1 \\
\arg \min_{w_1, w_2, \ldots, w_n \in \Delta^k} & -\frac{1}{n} \sum_{i=1}^{n} \left[ \psi_i^T w_i - \sum_{s=1}^{p} \exp(z_i^T \Phi^s w_i) \right]
\end{align*}
\]

where \( z_i = [1 \ x_i^T]^T \)

\( \psi^s = f(X, W) \)

\( \phi^j_s = [\theta^j_s \ (\Theta^j_s)^T]^T \)

\( \psi_i = f(X, \Phi^{1\cdots k}) \)

\( \Phi^s = [\phi_s^1 \phi_s^2 \cdots \phi_s^k] \)

2. Subproblems in summation can be computed in parallel

3. Use fast Newton-like optimization method [Hsieh et al. 2014]
Timing Results on Wikipedia Dataset ($k = 5, \lambda = 0.5$)

- Algorithm scales approximately as $O(np^2)$
BNC dataset has $n = 4049$ and $p = 1646$

Speedup could be $O(\min(n, p))$ on distributed system
Evaluating APM: No Direct Evaluation of Edge Parameters

- Previous metrics evaluate the similarity of word pairs [Newman et al. 2010, Mimno et al. 2011, Aletras and Court 2013]
  - Averaged statistic for all $\binom{10}{2}$ pairs of top words computed
  - Attempted to correlate with human judgment

- Unlike previous topic models, APM explicitly models dependencies between words

- How can we semantically evaluate the parameters for these dependencies?
Evocation [Boyd-Graber et al. 2006]

- Evocation denotes the idea of which words “evoke” or “bring to mind” other words

- Different types of evocation:
  1. Rose - Flower (example)
  2. Brave - Noble (kind)
  3. Yell - Talk (manner)
  4. Eggs - Bacon (co-occurrence)
  5. Snore - Sleep (setting)
  6. Wet - Desert (antonymy)
  7. Work - Lazy (exclusivity)
  8. Banana - Kiwi (likeness)

- Distinctive from word similarity or synonymy

- Collected human scores for approximately 15% of word pairs
Evocation Metric Illustration

<table>
<thead>
<tr>
<th>Word Pair</th>
<th>(\mathcal{H})</th>
<th>(\mathcal{M})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(w_1 \leftrightarrow w_2)</td>
<td>?</td>
<td>0.01</td>
</tr>
<tr>
<td>(w_1 \leftrightarrow w_3)</td>
<td>23</td>
<td>0.1</td>
</tr>
<tr>
<td>(w_1 \leftrightarrow w_4)</td>
<td>0</td>
<td>0.001</td>
</tr>
<tr>
<td>(w_2 \leftrightarrow w_3)</td>
<td>?</td>
<td>12.4</td>
</tr>
<tr>
<td>(w_2 \leftrightarrow w_4)</td>
<td>60</td>
<td>0.67</td>
</tr>
<tr>
<td>(w_3 \leftrightarrow w_4)</td>
<td>5</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Rank by model weights \(\mathcal{M}\) and Sum top-\(m\) human scores \(\mathcal{H}\)

<table>
<thead>
<tr>
<th>Word Pair</th>
<th>(\mathcal{H})</th>
<th>(\mathcal{M})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(w_2 \leftrightarrow w_3)</td>
<td>?</td>
<td>23.4</td>
</tr>
<tr>
<td>(w_3 \leftrightarrow w_4)</td>
<td>5</td>
<td>1.1</td>
</tr>
<tr>
<td>(w_2 \leftrightarrow w_4)</td>
<td>60</td>
<td>0.67</td>
</tr>
<tr>
<td>(w_1 \leftrightarrow w_3)</td>
<td>23</td>
<td>0.1</td>
</tr>
<tr>
<td>(w_1 \leftrightarrow w_2)</td>
<td>?</td>
<td>0.01</td>
</tr>
<tr>
<td>(w_1 \leftrightarrow w_4)</td>
<td>0</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word Pair</th>
<th>(\mathcal{H})</th>
<th>(\mathcal{M})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(w_3 \leftrightarrow w_4)</td>
<td>5</td>
<td>1.1</td>
</tr>
<tr>
<td>(w_2 \leftrightarrow w_4)</td>
<td>60</td>
<td>0.67</td>
</tr>
<tr>
<td>(w_1 \leftrightarrow w_3)</td>
<td>23</td>
<td>0.1</td>
</tr>
<tr>
<td>(w_1 \leftrightarrow w_4)</td>
<td>0</td>
<td>0.001</td>
</tr>
</tbody>
</table>

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon
Admixture of Poisson MRFs
Models for Comparison

- APM: Admixture of Poisson MRFs
- APM-LowReg: Very small regularization parameter
- APM-HeldOut: Chooses $\lambda$ from held-out documents
- CTM: Correlated Topic Models
- HDP: Hierarchical Dirichlet Process (Non-parametric)
- LDA: Latent Dirichlet Allocation
- RSM: Replicated Softmax (Undirected Topic Model)
- RND: Random baseline
Evocation Metric Results

- APM
- APM-LowReg
- APM-HeldOut
- CTM
- HDP
- LDA
- RSM
- RND

**Evocation (m= 50)**

**Evocation**

<table>
<thead>
<tr>
<th>k</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evoc-1 (Avg. Evoc. of Topics)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evoc-2 (Evoc. of Avg. Topic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**APM APM-LowReg APM-HeldOut CTM HDP LDA RSM RND**

David Inouye*, Pradeep Ravikumar, Inderjit Dhillon

Admixture of Poisson MRFs
### Table: Top 20 Word Pairs for Best LDA

<table>
<thead>
<tr>
<th>Human Score</th>
<th>Word Pair</th>
<th>Human Score</th>
<th>Word Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>run.v ↔ car.n</td>
<td>38</td>
<td>woman.n ↔ man.n</td>
</tr>
<tr>
<td>82</td>
<td>teach.v ↔ school.n</td>
<td>38</td>
<td>give.v ↔ church.n</td>
</tr>
<tr>
<td>69</td>
<td>school.n ↔ class.n</td>
<td>38</td>
<td>wife.n ↔ man.n</td>
</tr>
<tr>
<td>63</td>
<td>van.n ↔ car.n</td>
<td>38</td>
<td>engine.n ↔ car.n</td>
</tr>
<tr>
<td>51</td>
<td>hour.n ↔ day.n</td>
<td>35</td>
<td>publish.v ↔ book.n</td>
</tr>
<tr>
<td>50</td>
<td>teach.v ↔ student.n</td>
<td>32</td>
<td>west.n ↔ state.n</td>
</tr>
<tr>
<td>44</td>
<td>house.n ↔ government.n</td>
<td>32</td>
<td>year.n ↔ day.n</td>
</tr>
<tr>
<td>44</td>
<td>week.n ↔ day.n</td>
<td>25</td>
<td>member.n ↔ give.v</td>
</tr>
<tr>
<td>38</td>
<td>university.n ↔ institution.n</td>
<td>25</td>
<td>dog.n ↔ animal.n</td>
</tr>
<tr>
<td>38</td>
<td>state.n ↔ government.n</td>
<td>25</td>
<td>seat.n ↔ car.n</td>
</tr>
</tbody>
</table>

### Table: Top 20 Word Pairs for Best APM

<table>
<thead>
<tr>
<th>Human Score</th>
<th>Word Pair</th>
<th>Human Score</th>
<th>Word Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>telephone.n ↔ call.n</td>
<td>57</td>
<td>question.n ↔ answer.n</td>
</tr>
<tr>
<td>97</td>
<td>husband.n ↔ wife.n</td>
<td>57</td>
<td>prison.n ↔ cell.n</td>
</tr>
<tr>
<td>82</td>
<td>residential.a ↔ home.n</td>
<td>51</td>
<td>mother.n ↔ baby.n</td>
</tr>
<tr>
<td>76</td>
<td>politics.n ↔ political.a</td>
<td>50</td>
<td>sun.n ↔ earth.n</td>
</tr>
<tr>
<td>75</td>
<td>steel.n ↔ iron.n</td>
<td>50</td>
<td>west.n ↔ east.n</td>
</tr>
<tr>
<td>75</td>
<td>job.n ↔ employment.n</td>
<td>44</td>
<td>weekend.n ↔ sunday.n</td>
</tr>
<tr>
<td>75</td>
<td>room.n ↔ bedroom.n</td>
<td>41</td>
<td>wine.n ↔ drink.v</td>
</tr>
<tr>
<td>72</td>
<td>aunt.n ↔ uncle.n</td>
<td>38</td>
<td>south.n ↔ north.n</td>
</tr>
<tr>
<td>72</td>
<td>printer.n ↔ print.v</td>
<td>38</td>
<td>morning.n ↔ afternoon.n</td>
</tr>
<tr>
<td>60</td>
<td>love.v ↔ love.n</td>
<td>38</td>
<td>engine.n ↔ car.n</td>
</tr>
</tbody>
</table>
1. Visualization
2. Better inference of parameters
3. Extension to other domains