

## Contributions

- A deep architecture for topic models based entirely on Poisson Factor Analysis (PFA) modules.
- Inherent shrinkage in all layers, thanks to the DP-like formulation of PFA.
- Block updates for binary units improve mixing.
- PFA modules can be easily used to build discriminative topic models.
- Efficient MCMC inference scales as function of the number of non-cens in data and binary units.
- Scalable Bayesian inference algorithm based on Stochastic Variational Inference (SVI).

### Poisson factor analysis as a module

Assume \( \mathbf{x}_N \) is an \( M \)-dimensional vector containing word counts for the \( i \)-th of \( N \) documents, where \( M \) is the vocabulary size. We impose the model

\[
\mathbf{x}_N \sim \text{Poisson}(\mathbf{\Phi} \mathbf{\theta}_N + \mathbf{h}_N),
\]

where

- \( \mathbf{\Phi} \in \mathbb{R}_{+}^{M \times K} \), factor loadings matrix with \( K \) factors.
- \( \mathbf{\theta}_N \in \mathbb{R}_{+}^{K} \), factor intensities.
- \( \mathbf{h}_N \in \{0,1\}^K \), binary units indicating which factors are active for observation \( n \).
- Symbol \( \odot \) denotes element-wise (Hadamard) product.

Prior specification \( [\gamma] \)

\[
x_{MN} \sim \sum_{k=1}^{K} \gamma_{MN} & \text{Poisson}(\lambda_{MN}), \\
\lambda_{MN} \sim \text{Vstat}\phi \mathbf{h}_N & \text{Beta}(1,1), \\
\gamma_{MN} \sim \text{Beta}(1,1). \]

Note that \( \eta \) controls the sparsity of \( \mathbf{\Phi} \), while \( \gamma \) accommodates for over-dispersion in \( \mathbf{x}_N \) via \( \mathbf{\theta}_N \).

### PFA modules

**PFA modules for discriminative tasks**

Assume that there is a label \( \eta \in \{1, \ldots, C\} \) associated with document \( n \). We impose the model

\[
y_n \sim \text{Multinomial}(1, \mathbf{\lambda}_N), \quad \mathbf{\lambda}_N = \eta_0 \odot \sum_{k=1}^{K} \lambda_{MN},
\]

where

- \( \eta \) is represented as a \( C \)-dimensional one-hot vector, \( \eta_{\bullet} \).
- \( \mathbf{\lambda}_N = \mathbf{\lambda}(\mathbf{\theta}_N, h_N) \) and \( \mathbf{\lambda}_N \) is element of \( \mathbf{\lambda}_N \).
- \( \mathbf{h}_N \in \mathbb{R}_{+}^{K} \), matrix of nonnegative classification weights.
- \( \mathbf{h}_N \sim \text{Dirichlet}(1,C) \), for \( \mathbf{h}_N \) column of \( \mathbf{B} \).

### Deep representations with PFA modules

Develop a deep prior specification for \( \mathbf{h}_N \) as

\[
\mathbf{x}_N \sim \text{Poisson}(\mathbf{\Phi} \mathbf{\theta}_N + \mathbf{h}_N), \quad \mathbf{h}_N \sim \text{Dirichlet}(\mathbf{\lambda}_N), \quad \mathbf{\lambda}_N = \mathbf{\lambda}(\mathbf{\theta}_N, h_N), \quad \mathbf{\theta}_N \sim \text{Gamma}(1,1), \quad \mathbf{h}_N \sim \text{Bernoulli}(\mathbf{\gamma}_N).
\]

**Figure 1:** Graphical models. (a) PFA module. Nodes (\( \mathbf{h}_N \), \( \mathbf{\theta}_N \) and \( y_n \)) and edges drawn with dashed lines correspond to the discriminative PFA. (b) DPFM.

### Experiments

**Benchmark corpora**

- **Data:** 20 Newsgroups (20 News): 2,000 words, 11,315/7,531 training/test documents.
- **Models:** LDA, FTM, RSM, nHDP, DPFA-SBN, DPFA-RBM and DPFM.
- **Performance:** held-out perplexity on 20% of test set.
- **Models:** LDA, FTM, RSM, nHDP, DPFA-SBN, DPFA-RBM and DPFM.
- **Runtime:** one iteration of the two-layer DPFM on 20 News takes approx. 32/2 secs, for MCMC/SVI.

**Table 1:** Held-out perplexity on 20 News, RCV1 and Wiki. Subset of topics, and binary units.

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Size</th>
<th>20 News</th>
<th>RCV1</th>
<th>Wiki</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPFM</td>
<td>MCMC</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
<tr>
<td>DPFM</td>
<td>DPFM</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
<tr>
<td>DPFM</td>
<td>DocsDADE</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
<tr>
<td>DPFA-SBN</td>
<td>DPFA-RBM</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
<tr>
<td>DPFA-SBN</td>
<td>DocsDADE</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
<tr>
<td>DPFA-RBM</td>
<td>DocsDADE</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
</tbody>
</table>

**Classification**

- **Data:** 20 News for document classification.
- **Models:** LDA, DocNABE, BSM, OSM and DPFM.

**Table 2:** Test accuracy on 20 News. Subset accompanying model names indicate their size.

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Size</th>
<th>20 News</th>
<th>RCV1</th>
<th>Wiki</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>DocsDADE</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
<tr>
<td>FTM</td>
<td>DocsDADE</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
<tr>
<td>BSM</td>
<td>DocsDADE</td>
<td>128</td>
<td>584</td>
<td>394</td>
<td>751</td>
</tr>
</tbody>
</table>

**Medical records**

- **Duke University 5-year dataset (2007–2011):** 240,000 patients, 4.2M visits.
- **34,000 medication mapped to 1,691 pharmaceutical active ingredients (AI).**
- **Dataset:** 1,638,131,296 counts of AIs vs. patients.
- **MCMC-based DPFM of size 64-12.**

**Table 3:** Graph representation obtained from 20 News. Meta-topics are denoted by circles and layer-1 topics as boxes, with word lists corresponding to the top four words in layer-1 topics, \( \psi_{\bullet}^{(1)} \).

**References**


---

**Deep Poisson Factor Modeling**

Ricardo Henao, Zhe Gan, James Lu and Lawrence Carin

Department of Electrical and Computer Engineering, Duke University, Durham, NC