Scalable Deep Poisson Factor Analysis for Topic Modeling
Zhe Gan, Changyou Chen, Ricardo Henao, David Carlson and Lawrence Carin
Duke University, Durham NC 27708, USA

Problem of interest: Developing deep generative models for documents that are represented in bag-of-words form.
Main idea: • Poisson Factor Analysis (PFA) + Deep SiGmoid Belief Network (SBN) or Restricted Boltzmann Machine (RBM).
• PFA is employed to interact with data at the bottom layer.
• Deep SBN or RBM serves as a flexible prior for revealing topic structure.

Challenge: Designing scalable Bayesian inference algorithms.

Solutions: stochastic algorithms.
• Applying Bayesian conditional density filtering algorithm.
• Extending recently proposed work on stochastic gradient thermostats.

MODEL FORMULATION

Poisson Factor Analysis: We represent a discrete matrix $X \in \mathbb{Z}^{p \times N}$ containing counts from $N$ documents and $p$ words as $X = \text{Pois}(\Phi \Theta \cdot H^0)$, \hspace{1cm} (1)

• $\Phi$: factor loadings; $\Theta$: factor scores; $H^0$: latent binary features.
• We construct PFA by placing Dirichlet priors on $\phi_{h}$ (one column of $\Phi$), and gamma priors on $\theta_{l}$ (one column of $\Theta$), i.e., $\phi_{h} \sim \text{Dir}(\alpha_{h}, \ldots, \alpha_{h})$, and $\theta_{l} \sim \text{Ga}(\gamma_{l}, 1/\theta_{l})$.

The novelty in our model comes from the prior for $H_{1}$ (one column of $H^0$).

Structured Priors on the Latent Binary Matrix

Modeling with the SBN: Assume $h_{n}^{0} \in \{0, 1\}^{K_i}$, we define $h_{n}^{1} \in \{0, 1\}^{K_{i}}$ placed at a layer "above" $h_{n}^{0}$. An SBN model has the generative process, $p(h_{n}^{1} | h_{n}^{0} = 1) = \sigma(c_{n}^{1})$, $p(h_{n}^{1} = 1 | h_{n}^{0} = c_{n}^{1}) = \sigma(w_{c_{n}^{1}}^{1} + c_{n}^{1})$.

Modeling with the RBM: An RBM is a Markov random field with the same bipartite structure as the SBN. The energy function of an RBM is defined as $E(h_{n}^{0}, h_{n}^{1}) = -(h_{n}^{0} \cdot c^{0} - h_{n}^{0} \cdot W^{0} h_{n}^{1} - h_{n}^{1} \cdot c^{1})$.

Deep Architecture: We can add multiple layers of SBNs or RBMs to obtain a deep architecture, $p(h_{n}^{1}, \ldots, h_{n}^{K}) = \prod_{k=1}^{K} p(h_{n}^{k} | h_{n}^{k+1} \ldots h_{n}^{K})$.

SCALABLE INFERENCE

Bayesian conditional density filtering (BCDF):
• Repeatedly updating the surrogate conditional sufficient statistics (SCSS) using the current mini-batch.
• Drawing samples from the conditional posterior distributions of model parameters, based on SCSS.

Stochastic Gradient Néron-Hoover Thermostats (SGNHT):
• Extending Hamiltonian Monte Carlo using stochastic gradient.
• Introducing thermostat to maintain system temperature.
• Adaptively absorbing stochastic gradient noise.

Algorithms: (Left) BCD, (Right) SGNHT.

Problem of interest: Interpreting the output of our deep machine learning model. From the model output, the posterior distribution over the topics is obtained. We can then use these topics to explore the data.

Discussion:
• BCD: ease of implementation, but prefers the conditional densities for all the parameters.
• SGNHT: more general and robust, fast convergence.

EXPERIMENTS

Datasets:
• 20 Newsgroups: 20K documents with a vocabulary size of 2K.
• RCV1-v2: 800K documents with a vocabulary size of 10K.
• Wikipedia: 10M documents with a vocabulary size of 8K.

Quantitative Evaluation

Figure: Graphical model for the Deep Poisson Factor Analysis with three layers of hidden binary hierarchies. The directed binary hierarchy may be replaced by a deep Boltzmann machine.

Figure: The true posterior distribution.

Table: 20 Newsgroups.

Table: RCV1-v2 & Wikipedia.

Visualization: Sports, Computers, and Politics/Law.

Figure: Graphs induced by the correlation structure learned by DPFA for the 20 Newsgroups.

ACKNOWLEDGEMENTS

This research was supported by ARO, DARPA, DOE, NGA and ONR.