Convolutional Poisson Gamma Belief Network

Chaojie Wang♖ Bo Chen♖ Sucheng Xiao♖ Mingyuan Zhou♘

♖ *National Laboratory of Radar Signal Processing, Xidian University, Xi'an, Shaanxi, China*

♘ *McCombs School of Business, The University of Texas at Austin, Austin, TX, USA*

西安電子科技大學
XIDIAN UNIVERSITY

THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

Motivation

Document Representation

- **Basic Lossless Representation**
  - **A sequence of one-hot vectors**
    - ✓ Preserve all textual information
    - ※ Extremely large and sparse matrices
    - ※ Burdens of calculation and storage
    - ※ Difficult to model directly

Simplified

- **Simplified Lossy Representation**
  - **Bag-of-words**
    - ✓ Term-document frequency count matrix
    - ※ Lose word order
  - **Word embeddings**
    - ✓ Project words to low-dimensional vectors
    - ※ Require additional large corpora

Document

"I love it"

One-hot Sequence

$$
\begin{array}{c|c}
 & \begin{array}{ccc} \textit{I} & \textit{love} & \textit{it} \end{array} \\
\begin{array}{c} \textit{don't} \\ \textit{hate} \\ \textit{I} \\ \textit{it} \\ \textit{love} \end{array} & \begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix}
\end{array}
$$

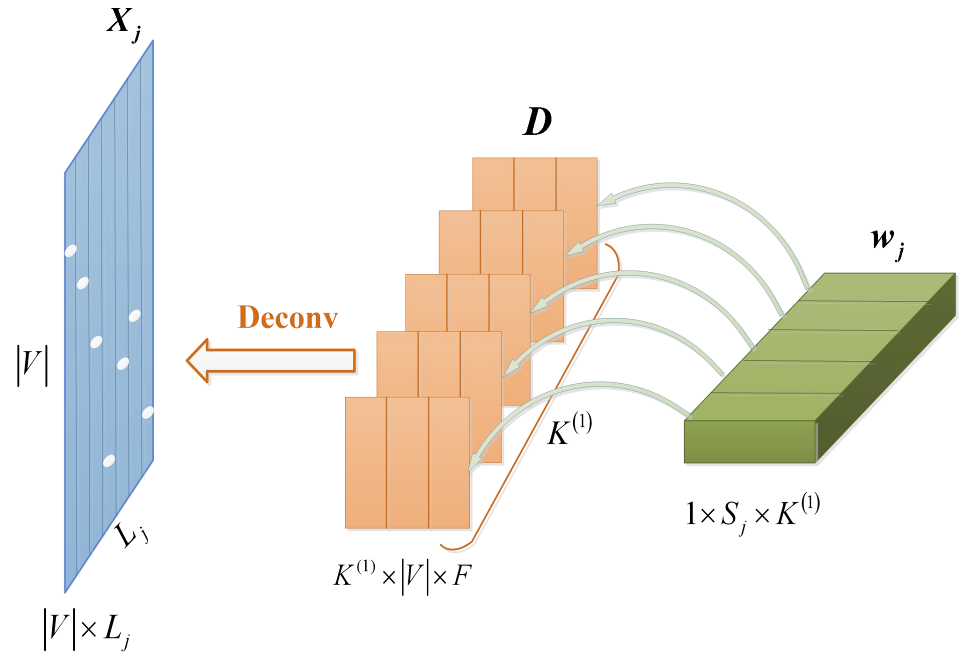
Challenge

Most basic representation

Our Contribution

Convolutional Poisson Factor Analysis

- **Generative model of CPFA**



$$X_j = \mathbf{1}(M_j > 0),\ M_j \sim \mathrm{Pois}(\textstyle\sum_{k=1}^{K} D_k * w_{jk}),$$
$$w_{jk} \sim \mathrm{Gam}(r_k, 1/c_j),\ \ D_k(:) \sim \mathrm{Dir}(\eta \mathbf{1}_{|V|F}),$$

- ✓ Preserve word order information
- ✓ Directly model sparse matrices
- ✓ Take advantages of the sparsity
- ✓ Support parallel computation
- ✓ Capture pharse-level topics

Probabilistic Convolutional Layer

Our Contribution

Convolutional Poisson Gamma Belief Network

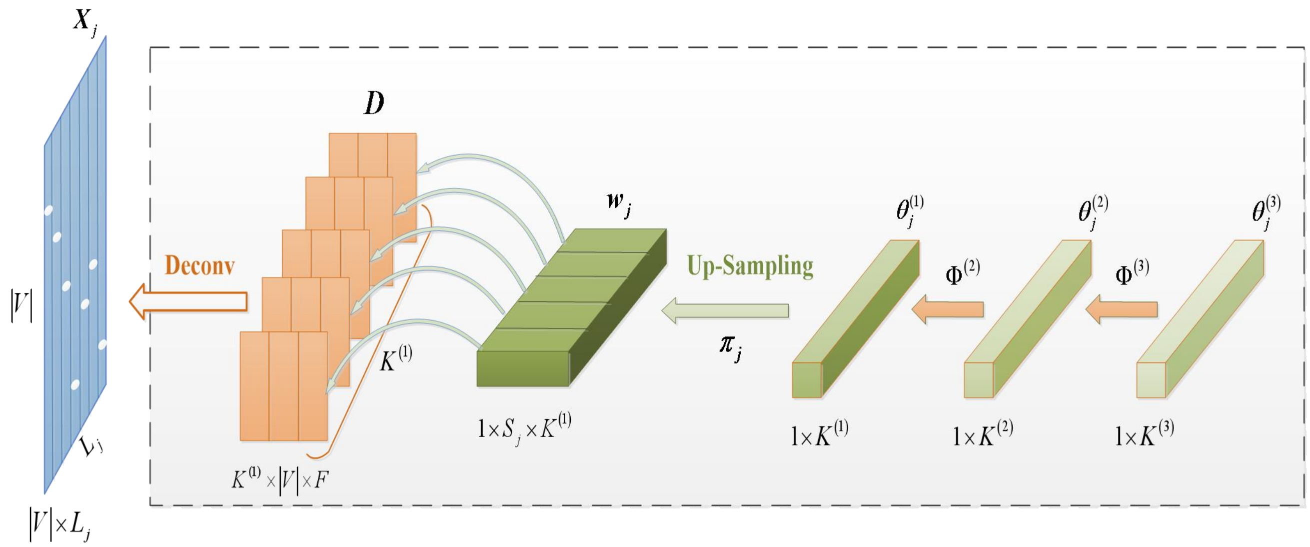
Probabilistic Pooling Layer

$$w_{jks} \sim \mathrm{Gam}(\boldsymbol{\Phi}_{k:}^{(2)} \boldsymbol{\theta}_j^{(2)} / S_j, 1/c_j^{(2)})$$

Equivalent

$$\theta_{jk}^{(1)} = \sum_{s=1}^{S_j} w_{jks} \sim \mathrm{Gam}(\boldsymbol{\Phi}_{k:}^{(2)} \boldsymbol{\theta}_j^{(2)}, 1/c_j^{(2)})$$

$$\boldsymbol{w}_{jk} = \boldsymbol{\pi}_{jk} \theta_{jk}^{(1)}, \ \boldsymbol{\pi}_{jk} \sim \mathrm{Dir}(\boldsymbol{\Phi}_{k:}^{(2)} \boldsymbol{\theta}_j^{(2)} / S_j \mathbf{1}_{S_j})$$



Generative model of CPGBN

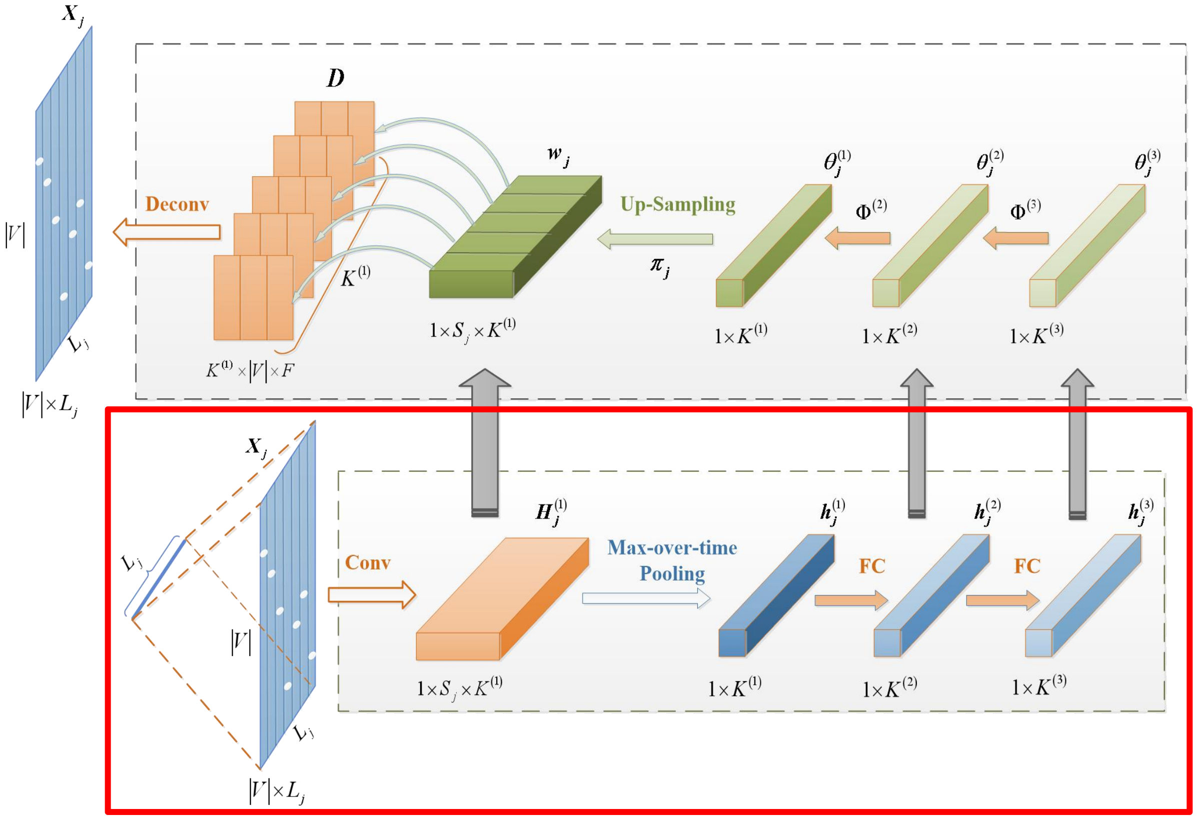
$$\boldsymbol{\theta}_j^{(T)} \sim \mathrm{Gam}(\boldsymbol{r}, 1/c_j^{(T+1)}),$$
$$\ldots,$$
$$\boldsymbol{\theta}_j^{(t)} \sim \mathrm{Gam}(\boldsymbol{\Phi}^{(t+1)} \boldsymbol{\theta}_j^{(t+1)}, 1/c_j^{(t+1)}),$$
$$\ldots,$$
$$\boxed{\begin{aligned} &\boldsymbol{\theta}_j^{(1)} \sim \mathrm{Gam}(\boldsymbol{\Phi}^{(2)} \boldsymbol{\theta}_j^{(2)}, 1/c_j^{(2)}), \\ &\boldsymbol{w}_{jk} = \boldsymbol{\pi}_{jk} \theta_{jk}^{(1)}, \ \boldsymbol{\pi}_{jk} \sim \mathrm{Dir}(\boldsymbol{\Phi}_{k:}^{(2)} \boldsymbol{\theta}_j^{(2)} / S_j \mathbf{1}_{S_j}), \end{aligned}}$$
$$\boldsymbol{M}_j \sim \mathrm{Pois}\left(\sum_{k=1}^{K^{(1)}} \boldsymbol{D}_k * \boldsymbol{w}_{jk}\right),$$

- ✓ Transfer the messages from deeper layers
- ✓ Jointly Train all the other layers
- ✓ Deep extention can boost performance
- ✓ Hierachical pharse-level topic

Our Contribution

Hybrid MCMC/Variational Inference

❑ Convolutional inference network



❑ Weibull Reparameterization

$$\text{Weibull PDF: } P(x\,|\,k,\lambda) = \frac{k}{\lambda^k}x^{k-1}e^{(x/\lambda)^k}$$

$$x = \lambda(-\ln(1-\epsilon))^{1/k},\ \epsilon \sim \text{Uniform}(0,1)$$

Approximate

$$\text{Gamma PDF: } P(x\,|\,\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$$

- ✓ Fast in out-of-sample prediction
- ✓ Parallel scalable inference
- ✓ Easy extension (e.g., modeling document labels)

Experiment

Phrase-level Topics Visualization

Table 3. Example phrases learned from TREC by CPGBN.

Kernel Index	Visualized Topic			Visualized Phrase
	1st Column	2nd Column	3rd Column	
192th Kernel	**how** cocktail stadium run	do many much long	you years miles degrees	**how** do you, **how** many years, **how** much degrees
80th Kernel	microsoft virtual answers.com softball	e-mail email ip brothers	address addresses floods score	microsoft e-mail address, microsoft email address, virtual ip address
177th Kernel	**who** willy bar hydrogen	created wrote fired are	maria angela snoopy caesar	**who** created snoopy, **who** fired caesar, **who** wrote angela
47th Kernel	dist all-time wheel saltpepter	**how** stock 1976 westview	far high tall exchange	dist **how** far, dist **how** high , dist **how** tall

Thank you !

Pacific Ballroom #237

Chaojie Wang♖ *Bo Chen*♖ *Sucheng Xiao*♖ *Mingyuan Zhou*♘

♖ *National Laboratory of Radar Signal Processing, Xidian University, Xi'an, Shaanxi, China*

♘ *McCombs School of Business, The University of Texas at Austin, Austin, TX, USA*