Neural Sparse Topical Coding

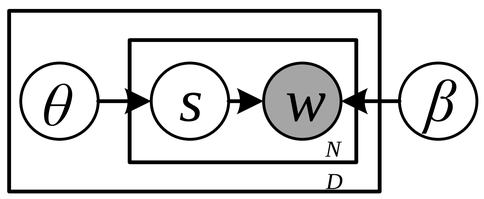
Min Peng[1], Qianqian Xie[1], Yanchun Zhang[2], Hua Wang[2], Xiuzheng Zhang[3]
[1]School of Computer Science, Wuhan University
[2]Centre for Applied Informatics, Victoria University
[3]School of Science, RMIT University

Abstract

- The poor scalability has become a bottleneck for the application and extension of the traditional topic models
- We propose Neural Sparse Topical Coding (***NSTC***) via adopting neural network to model the generation process of traditional sparsity-enhanced topic model ***STC***, thus can improve its flexibility
- To illustrate the flexibility, we present three extensions base on ***NSTC*** without re-deduced inference algorithms

Sparse Topical Coding

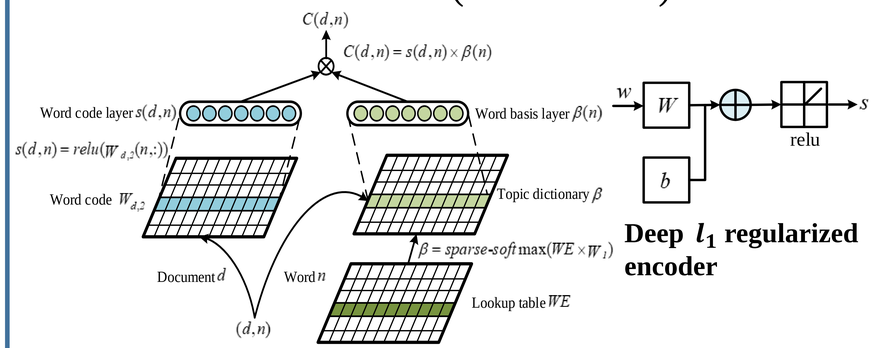
- Document code $\boldsymbol{\theta}$: the document representation of a document in topic space **k**
- Word code **s**: the word representation of a word in topic space **k**
- Topic dictionary $\boldsymbol{\beta}$: a global topic dictionary with **K** bases



$$\min_{\boldsymbol{\theta},\boldsymbol{s},\boldsymbol{\beta}} \sum_{d,n} \boldsymbol{l}(\boldsymbol{w}_{d,n}; \boldsymbol{s}_{d,n}, \boldsymbol{\beta}) + \lambda_1 \sum_{\boldsymbol{d}} ||\boldsymbol{\theta}_{\boldsymbol{d}}||_1 + \sum_{d,n} (\boldsymbol{\gamma}||\boldsymbol{s}_{d,n} - \boldsymbol{\theta}_{\boldsymbol{d}}||_2 + \boldsymbol{\rho}||\boldsymbol{s}_{d,n}||_1$$

Our Method

- **NSTC**: adopting neural network to model the generation process of STC
- The **cost function**: $\boldsymbol{L} = \boldsymbol{l}\left(\boldsymbol{w}_{d,n}, \boldsymbol{c}(\boldsymbol{d},\boldsymbol{n})\right) + ||\boldsymbol{s}_{d,n}||_1$



Deep l_1 regularized encoder

- **NSTCE**: imposing deep l_1 regularized encoder on word code

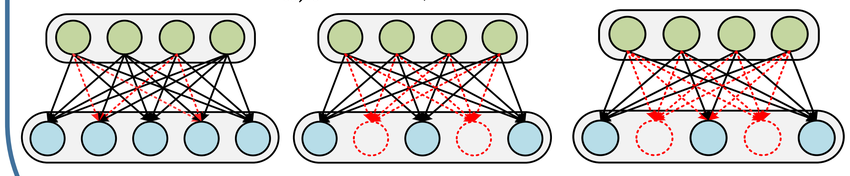
$$\boldsymbol{L} = \boldsymbol{l}\left(\boldsymbol{w}_{d,n}, \boldsymbol{c}(\boldsymbol{d},\boldsymbol{n})\right) + ||\boldsymbol{s}_{d,n}||_1 + \boldsymbol{\alpha} \sum_{d,n} ||\boldsymbol{s}_{d,n} - \boldsymbol{F}(\boldsymbol{w}_{d,n}; \boldsymbol{W}, \boldsymbol{b})||_2$$

- **NGSTC**： imposing group sparse regularization on word code

$$\boldsymbol{L} = \boldsymbol{l}\left(\boldsymbol{w}_{d,n}, \boldsymbol{c}(\boldsymbol{d},\boldsymbol{n})\right) + \sum_{d,k} \lambda ||\boldsymbol{s}_{d,.k}||_2$$

- **NSTCSG**: imposing sparse group lasso regularization on word code

$$\boldsymbol{L} = \boldsymbol{l}\left(\boldsymbol{w}_{d,n}, \boldsymbol{c}(\boldsymbol{d},\boldsymbol{n})\right) + \sum_{d,n} \lambda_1 ||\boldsymbol{s}_{d,n}||_1 + \sum_{d,k} \lambda_2 ||\boldsymbol{s}_{d,.k}||_2$$



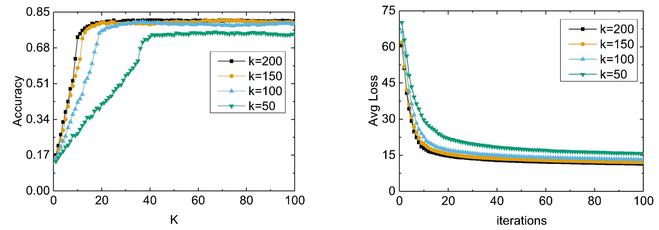
Lasso　　Group Lasso　　Sparse Group Lasso

Experiments

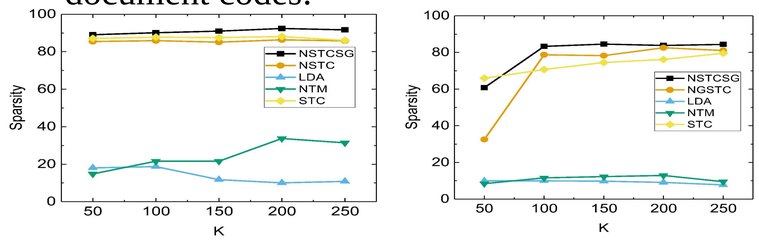
- **Datasets**: 20Newsgroups, web snippet

Data	label	docs	avg words	vocab
20NG	20	18775	135	60698
snippet	8	12265	10.72	5581

- The loss and accuracy curves:



- The average sparsity ratio of word and document codes:



- The perplexity on 20NG test data:

Model	LDA	STC	DNADE	Topicvec	NSTC
ppl	1091	611	896	650	517

- t-SNE projection of the estimated document codes.: