西安電子科技大學
XIDIAN UNIVERSITY

International Conference on Learning Representations (ICLR 2022)

ICLR

Representing Mixtures of Word Embeddings With Mixtures of Topic Embeddings

Dongsheng Wang[1,*] Dandan Guo[2,*] He Zhao[3] Huangjie Zheng[4]

Korawat Tanwisuth[4] Bo Chen[1] Mingyuan Zhou[4]

1 National Laboratory of Radar Signal Processing, Xidian University, Xi'an, China

2 School of Data Science and AIRS, The Chinese University of Hong Kong, Shenzhen

3 Department of Data Science and Artificial Intelligence, Monash University, Australia

4 McCombs School of Business, The University of Texas at Austin

Background & Motivation

- ➢ Bayesian probabilistic topic models (BPTMs)
  - ★ Usually under a bayesian generative model framework and based on GIbbs sampling or variational inference
    - ❏ Inflexible to additional inputs and downstream tasks
    - ❏ Unfriendly to large-scale text collections
    - ❏ inconvenient to integrate BPTMs with deep neural networks (DNNs)
- ➢ Embedded topic models (ETMs)
  - ★ Usually under the variational autoencoders (VAEs) framework and achieve appealing flexibility and scalability
    - ❏ Compromised reparameterization trick in VAEs
    - ❏ Fail to achieve coherent topics and good document representation at the same time, empirically.

Method

➢ From **W**ord **E**mbeddings to **T**opic **E**mbeddings (WeTe, this paper)

★ View document as a the set of word embeddings P, and a set of topic embeddings Q

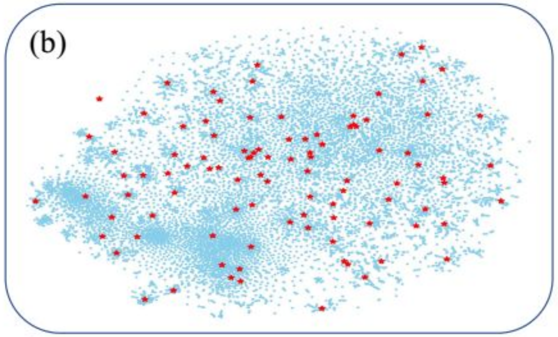
★ learn topic embeddings by minimizing the bidirectional conditional transport cost (CT) between P and Q over all documents

$$L = \frac{1}{J}\sum_{j=1}^{J}[L_{Q_j \to P_j} + L_{P_j \to Q_j}]$$
$$= \frac{1}{J}\sum_{j=1}^{J}[\sum_{k=1}^{K}\tilde{\theta}_{jk}\sum_{i=1}^{N_j} c(\boldsymbol{w}_{ji}, \boldsymbol{\alpha}_k)\pi(\boldsymbol{w}_{ji} \,|\, \boldsymbol{\alpha}_k) + \sum_{i=1}^{N_j}\frac{1}{N_j}\sum_{k=1}^{K} c(\boldsymbol{w}_{ji}, \boldsymbol{\alpha}_k)\pi(\boldsymbol{\alpha}_k \,|\, \boldsymbol{w}_{ji})]$$

● cost function $c(\boldsymbol{w}_{ji}, \boldsymbol{\alpha}_k)$

The closer the hidden space, the lower the transport cost, e.g.,

$$c(\boldsymbol{w}_{ji}, \boldsymbol{\alpha}_k) = e^{-\boldsymbol{w}_{ji}^{\mathrm{T}}\boldsymbol{\alpha}_k}.$$



Motivation of WeTe

Method

➢ From **W**ord **E**mbeddings to **T**opic **E**mbeddings (WeTe, this paper)

★ View document as a the set of word embeddings P, and a set of topic embeddings Q

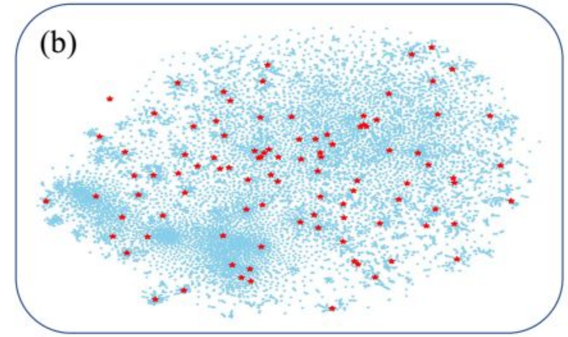
★ learn topic embeddings by minimizing the bidirectional conditional transport cost (CT) between P and Q over all documents

$$L = \frac{1}{J}\sum_{j=1}^{J}[L_{Q_j \to P_j} + L_{P_j \to Q_j}]$$
$$= \frac{1}{J}\sum_{j=1}^{J}[\sum_{k=1}^{K}\tilde{\theta}_{jk}\sum_{i=1}^{N_j} c(\boldsymbol{w}_{ji}, \boldsymbol{\alpha}_k)\pi(\boldsymbol{w}_{ji} \mid \boldsymbol{\alpha}_k) + \sum_{i=1}^{N_j}\frac{1}{N_j}\sum_{k=1}^{K} c(\boldsymbol{w}_{ji}, \boldsymbol{\alpha}_k)\pi(\boldsymbol{\alpha}_k \mid \boldsymbol{w}_{ji})]$$

- transport plan $\pi(\boldsymbol{\alpha}_k \mid \boldsymbol{w}_{ji})$ and $\pi(\boldsymbol{w}_{ji} \mid \boldsymbol{\alpha}_k)$

The more relevant the semantics, the higher the transfer probability

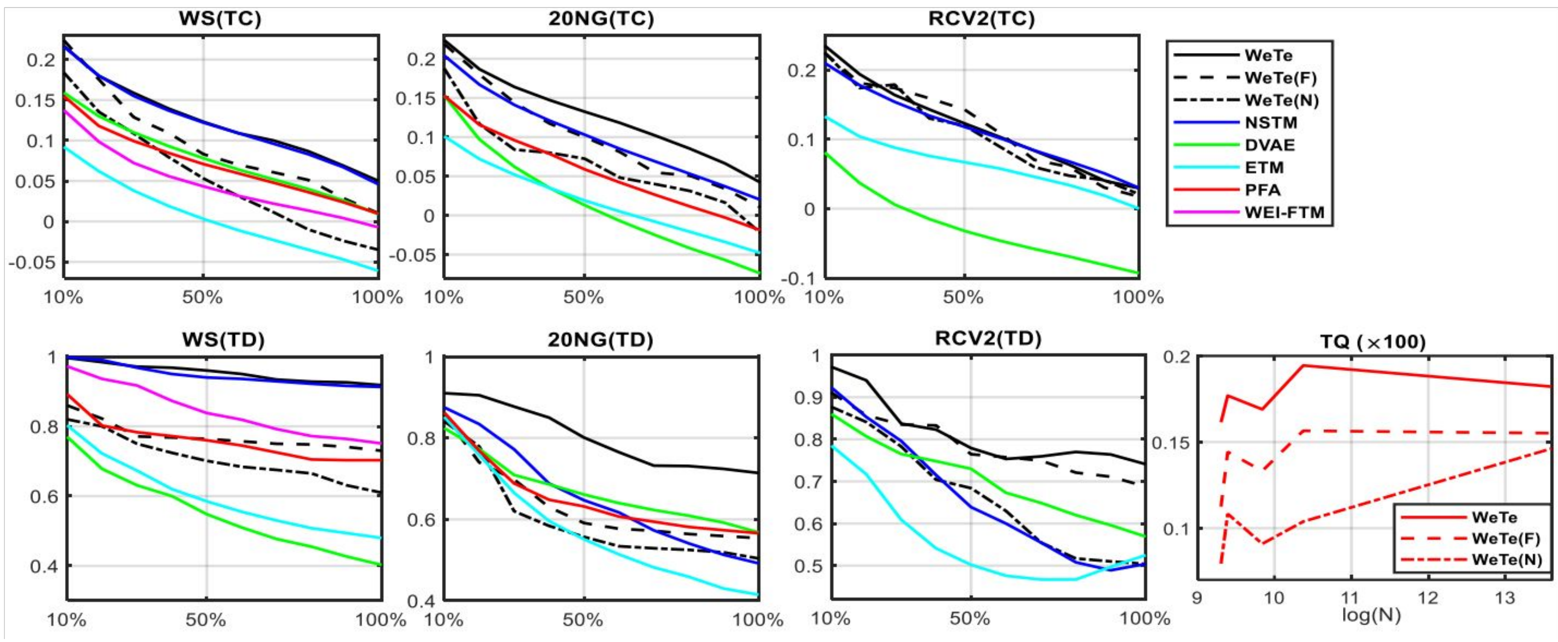
$$\pi_{N_j}(\boldsymbol{w}_{ji} \mid \boldsymbol{\alpha}_k) = \frac{P_j(\boldsymbol{w}_{ji})e^{-d(\boldsymbol{w}_{ji}, \boldsymbol{\alpha}_k)}}{\sum_{i'=1}^{N_j} P_j(\boldsymbol{w}_{ji'})e^{-d(\boldsymbol{w}_{ji'}, \boldsymbol{\alpha}_k)}} = \frac{e^{-d(\boldsymbol{w}_{ji}, \boldsymbol{\alpha}_k)}}{\sum_{i'=1}^{N_j} e^{-d(\boldsymbol{w}_{ji'}, \boldsymbol{\alpha}_k)}}, \quad \boldsymbol{w}_{ji} \in \{\boldsymbol{w}_{j1}, \ldots, \boldsymbol{w}_{jN_j}\},$$



Motivation of WeTe

Experiments

- Topic quality



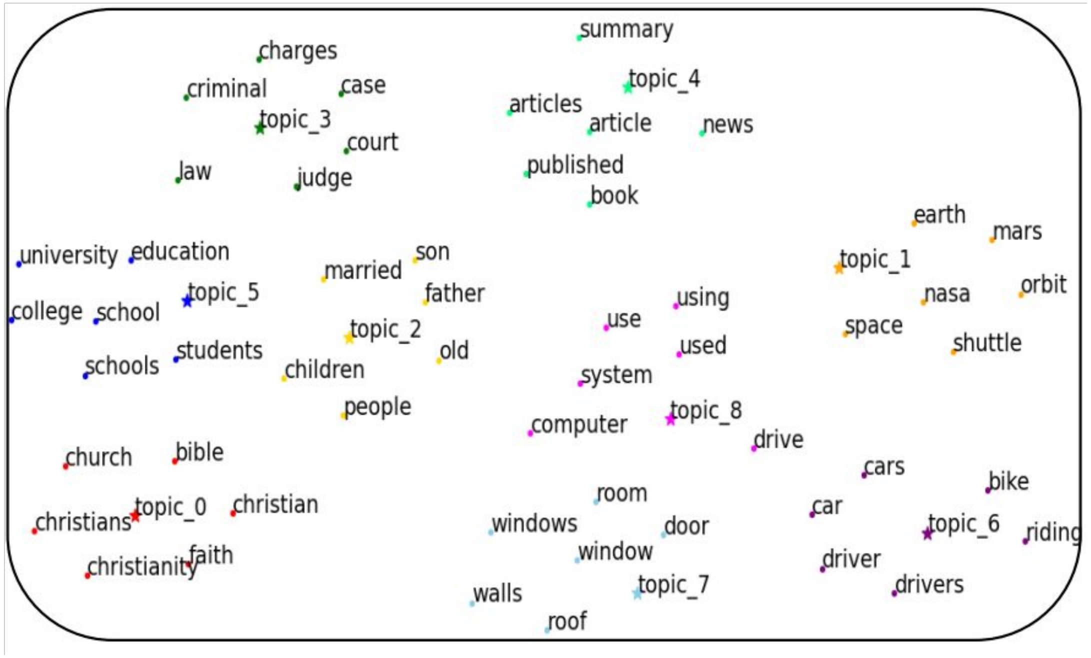
➢ document clustering

Table 1: Comparison of K-Means clustering purity (km-Purity) and NMI (km-NMI) for various methods. We use the 6 super-categories as 20NG's ground truth and denote it as 20NG(6). The best and second best scores of each dataset are highlighted in boldface and with an underline, respectively.

Method	**km-Purity(%)** WS	RCV2	DP	20NG(6)	**km-NMI(%)** WS	RCV2	DP	20NG(6)
LDA-Gibbs	46.4±0.6	52.4±0.4	60.8 ±0.5	59.2±0.6	25.1±0.4	38.2±0.5	54.7 ±0.3	32.4 ±0.4
PFA	55.7±0.4	-	64.6 ±0.7	61.2±0.6	31.1±0.3	-	55.4±0.5	32.7 ±1.1
PTM	33.2 ±1.1	-	56.3 ±1.7	-	7.9±1.4	-	45.2 ±1.5	-
WEI-FTM	54.6±1.5	-	65.3 ±2.4	-	32.4±1.5	-	59.7±1.6	-
DVAE	26.6±1.5	52.6±1.2	67.2 ±1.1	64.6 ±1.0	3.7 ±0.8	31.3±0.9	50.8 ±0.6	29.8 ±0.6
ETM	32.9±2.3	50.2±0.6	63.1 ±1.5	62.6 ±2.2	12.3±2.3	30.3±1.0	53.2 ±0.7	29.3 ±1.5
NSTM	42.1±0.6	53.8±1.0	20.2 ±0.7	62.6±1.2	17.4 ±0.6	36.8±0.3	6.63±0.11	31.1 ±1.2
WeTe	59.0±0.1	<u>59.2</u>±0.2	<u>75.8</u> ±0.8	67.3 ±0.6	<u>34.5</u>±0.1	40.3±0.4	<u>62.5</u>±0.8	<u>35.0</u> ±0.4
WeTe(N)	<u>59.7</u>±0.1	58.5±0.3	74.1 ±3.3	**70.2** ±1.0	34.1±0.1	<u>41.2</u>±0.1	60.1±1.1	34.3 ±0.8
WeTe(F)	**60.8**±0.2	**62.9**±0.5	**77.1** ±1.0	<u>68.5</u> ±0.2	**34.9**±0.4	**42.8**±0.3	**63.7**±0.4	**36.3** ±0.2

Experiments

➢ Qualitative analysis



NSTM	software, pc, desktop, server, computer, os, hardware
	computer, hardware, pc, software, user, interface, computers
	user, web, computer, server, software, internet, files
WeTe	system, use, computer, drive, chip, used, using
	file, web, graphics, files, server, data, fax
	sun, color, image, black, images, white, light

Representing Mixtures of Word Embeddings With Mixtures of Topic Embeddings

Dongsheng Wang [1,*] Dandan Guo [2,*] He Zhao [3] Huangjie Zheng [4]
Korawat Tanwisuth [4] Bo Chen [1] Mingyuan Zhou [4]

thank you

the code is available at https://github.com/BoChenGroup/WeTe