NEURAL INFORMATION PROCESSING SYSTEMS

西安電子科技大學 XIDIAN UNIVERSITY

TEXAS The University of Texas at Austin

Knowledge-Aware Bayesian Deep Topic Model

Dongsheng Wang[1] Yishi Xu[1] Miaoge Li[1] Zhibin Duan[1] Chaojie Wang[1]

Bo Chen[1] Mingyuan Zhou[2]

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

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

Knowledge-Aware Bayesian Deep Topic Model

西安電子科技大學 XIDIAN UNIVERSITY

❖ Background And Motivation

★ Most embedded topic models(ETMs) are learned by maxizing the likelihood

- Purely data-driven, ignoring the easily accessible knowledge graph
- The learned topics sometimes are unfriendly to users

★ While several knowledge-based ETMs have recently been proposed

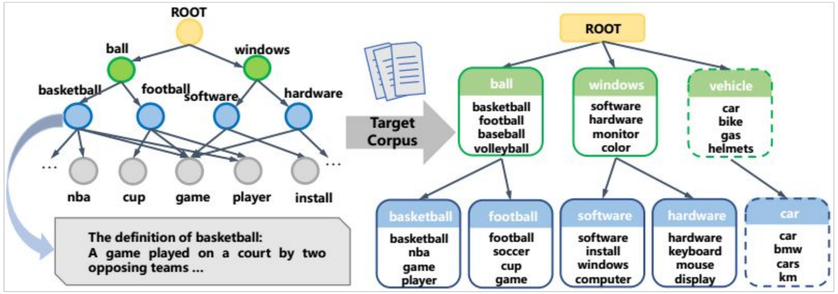
- Shallow topic structures
- Or ignoring the mismatch issue of the provided knowledge graph between the target corpus

NEURAL INFORMATION PROCESSING SYSTEMS

TEXAS
The University of Texas at Austin

Knowledge-Aware Bayesian Deep Topic Model

❖ Background And Motivation



★ We first propose a Bayesian generative framework (TopicKG) for incorporating domain knowledge into deep topic modeling

★ TopicKG is extended to TopicKGA that allows the given domain knowledge to-be-finetuned according to the target corpus

NEURAL INFORMATION PROCESSING SYSTEMS

TEXAS The University of Texas at Austin

Knowledge-Aware Bayesian Deep Topic Model

❖ Topic-KG

★ The generative model of corpus x, and the knowledge graph S and C

$$\boldsymbol{\theta}_j^{(L)} \sim \mathrm{Gam}(\boldsymbol{\gamma}, 1/c_j^{(L+1)}), \quad \left\{\boldsymbol{\theta}_j^{(l)} \sim \mathrm{Gam}(\boldsymbol{\Phi}^{(l+1)}\boldsymbol{\theta}_j^{(l+1)}, 1/c_j^{(l+1)})\right\}_{l=1}^{L-1},$$

$$\boldsymbol{x}_j \sim \mathrm{Pois}(\boldsymbol{\Phi}^{(1)}\boldsymbol{\theta}_j^{(1)}) \quad \left\{\boldsymbol{\Phi}_k^{(l)} = \mathrm{Softmax}(\boldsymbol{\Psi}_k^{(l)})\right\}_{l=1}^{L}, \quad \left\{\Psi_{k_1k_2}^{(l)} = \boldsymbol{e}_{k_1}^{(l-1)^T}\boldsymbol{e}_{k_2}^{(l)},\right\}_{k_1=1,k_2=1,l=1}^{K_{l-1},K_l,L},$$

$$\left\{S_{k_1,k_2}^{(l)} \sim \mathrm{Bern}(\sigma(\boldsymbol{e}_{k_1}^{(l-1)^T}\boldsymbol{W}\boldsymbol{e}_{k_2}^{(l)})),\right\}_{k_1=1,k_2=1,l=1}^{K_{l-1},K_l,L}, \left\{C_{vk}^{(l)} \sim \mathrm{Bern}(\sigma(\boldsymbol{e}_v^{(0)^T}\boldsymbol{e}_k^{(l)})),\right\}_{v=1,k=1,l=1}^{V,K_l,L},$$

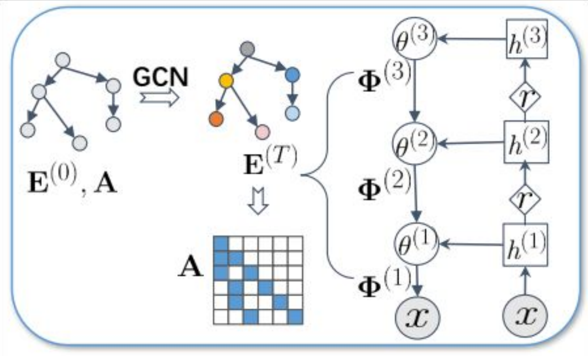


Fig1. Overview of TopicKG

★ The ELBO

$$\mathcal{L} = \sum_{j=1}^{J} \mathrm{E}_{q(\boldsymbol{\theta}|\boldsymbol{x}_j)}[\log p(\boldsymbol{x}_j|\boldsymbol{\Phi}^{(1)}, \boldsymbol{\theta}_j^{(1)})] + \beta \sum_{l=1}^{L}\sum_{k_2=1}^{K_l}(\sum_{k_1=1}^{K_{l-1}} \log p(S_{k_1k_2}^{(l)})|\boldsymbol{e}_{k_1}^{(l-1)}, \boldsymbol{e}_{k_2}^{(l)})$$

$$+ \sum_{v=1}^{V} \log p(C_{vk_2}^{(l)}|\boldsymbol{e}_v^{(0)}, \boldsymbol{e}_{k_2}^{(l)})) - \sum_{j=1}^{J}\sum_{l=1}^{L} \mathrm{KL}(q(\boldsymbol{\theta}_j^{(l)})||p(\boldsymbol{\theta}_j^{(l)}|\boldsymbol{\Phi}^{(l+1)}, \boldsymbol{\theta}_j^{(l+1)})))$$

Knowledge-Aware Bayesian Deep Topic Model

❖ Topic-KGA

★ TopicKG requires a perfect topic tree that matches the target corpos, however, the knowledge graph may be

(1) noisy
(2) built on an ad hoc basis
(3) not closely related to the topic discovering task

★ Graph adaptive technique

$$\tilde{\boldsymbol{A}} = \tilde{\boldsymbol{A}} + \tilde{\boldsymbol{A}}^{\text{ada}}$$

$$\tilde{\boldsymbol{A}}^{\text{ada}} = \text{Softmax}(k(\boldsymbol{E}_A, \boldsymbol{E}_A))$$

Knowledge-Aware Bayesian Deep Topic Model

西安電子科技大學
XIDIAN UNIVERSITY

❖ Experiment

★ Topic quality

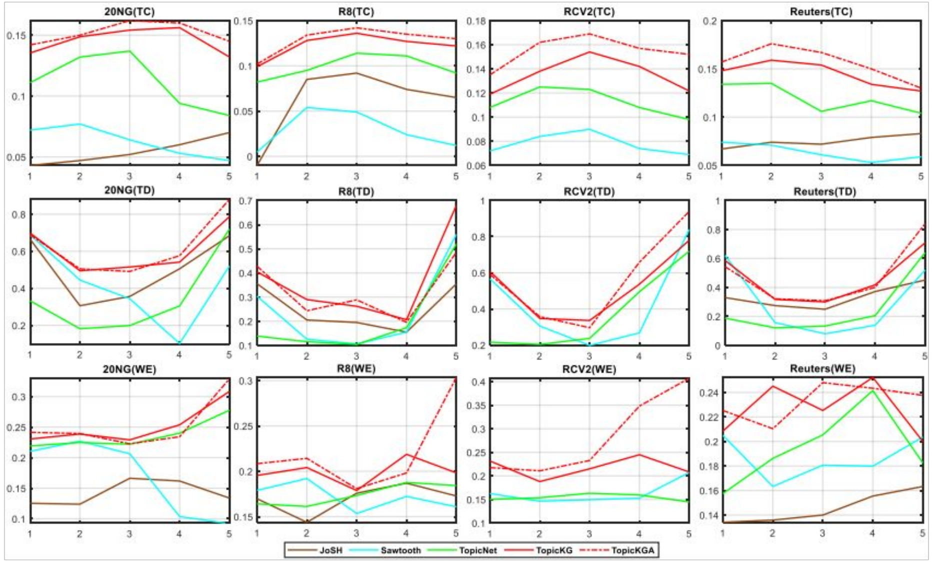


Figure 3: Topic coherence (TC, top row), topic diversity (TD, middle row), and word embedding coherence (WE, bottom row) results for various deep topic models on four datasets. In each subfigure, the horizontal axis indicates the layer index of the topics. For all metrics, higher is better.

NEURAL INFORMATION PROCESSING SYSTEMS

TEXAS
The University of Texas at Austin

❖ Experiment

★ Document representation

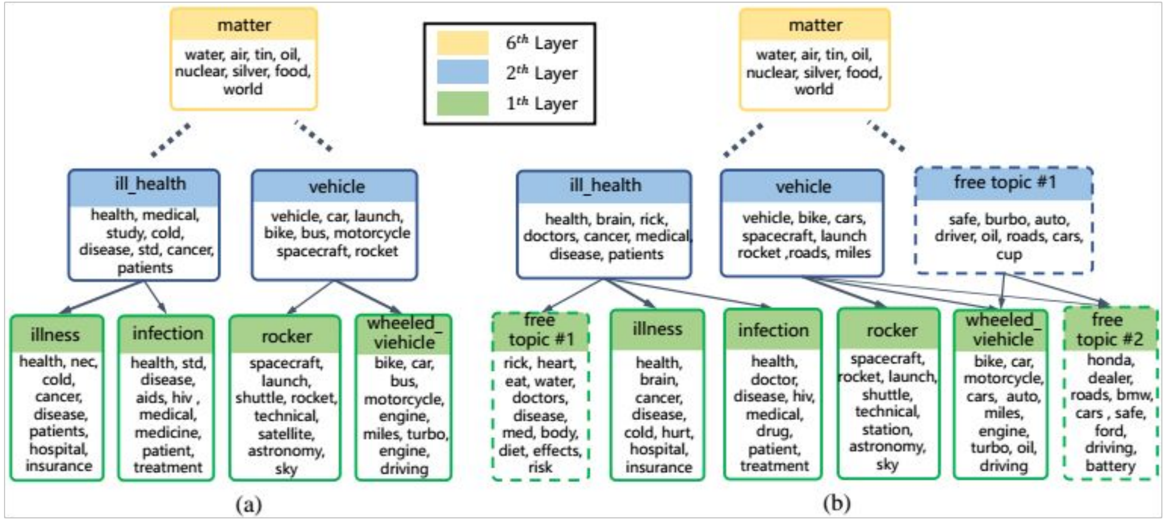
Table 1: Micro F1 and Macro F1 score of different models on three datasets. The digits in brackets indicate the number of layers. Micro F1 /Macro F1.

Model	20NG	R8	RCV2
ETM	50.25 ±0.42 / 47.44 ±0.21	88.10 ±0.45 / 59.67 ±0.24	68.63 ±0.15 / 24.40 ±0.11
CombinedTM	56.43 ±0.14 / 54.95 ±0.11	93.69 ±0.09 /84.14 ±0.10	84.85 ±0.11 / 51.47 ±0.21
Sawtooth(3)	52.41 ±0.08 / 51.53 ±0.10	90.04 ±0.15 / 78.84 ±0.21	82.54 ±0.11 / 49.25 ±0.10
TopicNet(3)	55.16 ±0.22 / 54.78 ±0.34	89.95 ±0.17 / 64.15 ±0.16	84.15 ±0.25 / 50.37 ±0.22
TopicKG(3)	55.73 ±0.15 / 54.48 ±0.08	93.6 ±0.05 / 83.32 ±0.07	84.75 ±0.16 / 50.51 ±0.41
TopicKGA(3)	**58.63** ±0.15 / **57.90** ±0.10	**93.70** ±0.52 / **84.50** ±0.11	**85.34** ±0.14 / **52.35** ±1.10
ETM	47.79 ±0.12 / 44.19 ±1.01	86.54 ±0.84 / 59.88 ±1.11	63.77 ±0.14 / 21.44 ±1.04
CombinedTM	58.16 ±0.15 / 58.10 ±0.10	93.50 ±0.13 /84.84 pm0.11	82.91 ±0.11 / 48.17 ±0.05
Sawtooth(7)	53.71 ±0.11 / 53.02 ±0.47	92.86 ±0.07 / 82.54 ±0.41	82.46 ±0.15 / 49.34 ±0.34
TopicNet(7)	56.13 ±0.19 / 55.41 ±0.39	90.65 ±0.00 / 66.57 ±0.00	82.81 ±0.00 / 49.44 ±0.00
TopicKG(7)	56.32 ±0.12 / 57.35 ±0.04	94.04 ±0.12 / 85.04 ±0.11	82.48 ±0.11 / 48.24 ±0.09
TopicKGA(7)	**60.04** ±0.34 / **59.12** ±0.13	**94.10** ±0.08 / **85.50** ±0.10	**83.08** ±0.23 / **50.50** ±0.08

Knowledge-Aware Bayesian Deep Topic Model

西安电子科技大学 XIDIAN UNIVERSITY

- ❖ Experiment
  - ★ The learned topic hierarchies



Thank you for listening