World Wide Web
https://doi.org/10.1007/s11280-021-00963-7

Hierarchical neural topic modeling with manifold regularization

Ziye Chen[1] · Cheng Ding[1] · Yanghui Rao[1] · Haoran Xie[2] · Xiaohui Tao[3] · Gary Cheng[4] · Fu Lee Wang[5]

Received: 7 January 2021 / Revised: 23 August 2021 / Accepted: 4 October 2021

Abstract

Topic models have been widely used for learning the latent explainable representation of documents, but most of the existing approaches discover topics in a flat structure. In this study, we propose an effective hierarchical neural topic model with strong interpretability. Unlike the previous neural topic models, we explicitly model the dependency between layers of a network, and then combine latent variables of different layers to reconstruct documents. Utilizing this network structure, our model can extract a tree-shaped topic hierarchy with low redundancy and good explainability by exploiting dependency matrices. Furthermore, we introduce manifold regularization into the proposed method to improve the robustness of topic modeling. Experiments on real-world datasets validate that our model outperforms other topic models in several widely used metrics with much fewer computation costs.

Keywords Neural topic modeling · Hierarchical structure · Tree network · Manifold regularization

1 Introduction

As one of the most successful and prevalent language models, topic modeling can learn the latent explainable representation of documents automatically. Traditional topic models often utilize directed probability graph to describe their generative processes. However, as the expressiveness and structure of generative processes grows, the deviation of parameters tends to be tough and complicated, which also hinders the model's efficiency when it is trained on a large scale dataset [17]. Recently, many studies focus on utilizing neural networks [20, 28] to extract the topic information, and these neural topic models can easily scale to a larger amount of training data than classical probabilistic models like the

This article belongs to the Topical Collection: *Special Issue on Explainability in the Web*
Guest Editors: Guandong Xu, Hongzhi Yin, Irwin King, and Lin Li

✉ Yanghui Rao
raoyangh@mail.sysu.edu.cn

Extended author information available on the last page of the article.

Q1

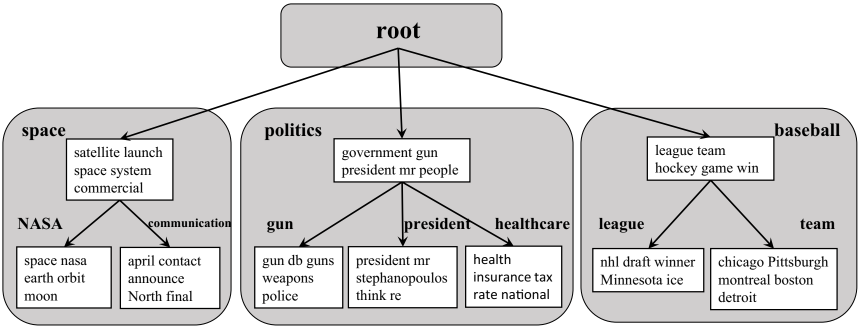
root
space
satellite launch space system commercial
NASA
communication
space nasa earth orbit moon
april contact announce North final
politics
government gun president mr people
gun
president
healthcare
gun db guns weapons police
president mr stephanopoulos think re
health insurance tax rate national
baseball
league team hockey game win
league
team
nhl draft winner Minnesota ice
chicago Pittsburgh montreal boston detroit
• • •

Figure 1 Topics inferred by our model from the 20NEWS dataset [20]. We present five most representative words for each topic and manually label these topics

latent Dirichlet allocation (LDA) [4] and its extensions. But most of the current neural topic models are flat models, which means the extracted topics are at the same level. This is a significant limitation because in many domains, topics can be naturally organized into hierarchies, where the root of each hierarchy represents the most general topic, and the topics become more specific toward the leaf nodes. For instance, when we want to post a review of a laptop, we may first determine its overall topic/aspect using words such as “cost performance” and “quality”. Then, we select the “appearance”, “hardware”, and other sub-topics to write the review in detail.

In probabilistic topic models, a hierarchical topic structure has been proven as useful for many tasks, including text categorization, text summarization, and aspect extraction [3, 12, 18, 22], because such a model can provide much explainable information with desirable granularity. Furthermore, explicitly modeling the hierarchical patterns allows us to learn more interpretable topics and clearly show the main topics of a corpus in a hierarchical structure, rather than the traditional word cloud. An example of topic hierarchy is shown in Figure 1. Such a hierarchy can be used to sharpen a user’s understanding of the text content.

Although several probabilistic topic models have been proposed to extract the hierarchical topic structure of a corpus [3, 12], the Markov chain Monte Carlo (MCMC) method [25] they employed for inference is quite time-consuming and is impractical to train for a large-scale dataset. Recently, TSNTM [11] is developed to model the topic hierarchy based on the neural variational inference (NVI) framework with good scalability, but the topic hierarchy extracted by TSNTM is not reasonable enough because the DRNN it applied is unsuitable to discover hierarchical semantics.

In this paper, we also focus on grouping documents into reasonable hierarchies based on NVI. With the rapid development of neural networks, it is possible to employ multi-level latent variables and obtain a hierarchical model. But few methods explicitly model the dependency among different layers and get interpretable hierarchical latent variables, e.g., topics, which is largely due to the weak interpretability of neural networks. Latent variables inside the network can hardly be displayed explicitly, so modeling the hierarchy of them is very difficult. To address this issue, we propose a novel NVI based method called hierarchical neural topic model (HNTM)[1] for hierarchical topic modeling with a pyramid-shaped structure. The model can also extract a tree-shaped structure by adding two constraints.

[1] The code of our model is available in public at: https://github.com/hostnlp/HNTM.

To enhance the robustness of our HNTM, we also incorporate a manifold regularization term to the NVI framework. Generally, manifold learning assumes that the points connected to each other should be as close as possible after dimensionality reduction. As a result, we introduce Laplacian Eigenmap [1] as a regularization term to make the related documents as similar as possible in the topic distribution at the document level. To summarize, our main contributions are as follows:

- We propose HNTM, a novel NVI based model for hierarchical topic modeling, which outperforms the existing models in several widely adopted metrics with much fewer computation costs.
- We introduce the manifold regularization into the NVI framework with the aim of making nearby document pairs have similar latent topic representations, which reduces the impact of noisy words and enhances the robustness of HNTM.

The rest of this paper is organized as follows. In Section 2, we introduce related work about hierarchical topic models and neural topic models. In Section 3, we present our model, introduce the network structure, and describe the regularization terms. In Section 4, we present empirical results and compare HNTM with baseline methods. In Section 5, we conclude the paper with discussions and future directions.

2 Related work

After proposing the classical LDA model [4], Blei et al. [3] extended it to a hierarchical version called HLDA by introducing a nested Chinese restaurant process (nCRP). Given a certain depth, HLDA constructs a topic tree through Gibbs sampling. However, each document in HLDA is generated by the topics along a single path of the tree, so the ancestor topic and its offspring topic generate the document together, making the hierarchical relation unclear. To overcome the weakness of single path sampling, Kim et al. [12] proposed a recurrent CRP (rCRP), which enables a document to have a distribution over the entire topic tree with unbounded depth and width. Experiments indicated that rCRP achieved remarkable performance in hierarchical topic modeling. However, the aforementioned sampling based methods suffer from the limitation of data scalability.

Mimno et al. [22] used a directed acyclic graph (DAG) structure and proposed the hierarchical pachinko allocation model (hPAM). The model includes a root topic, in addition to several super-topics and sub-topics. The root topic and other topics are connected to lower-level topics by multinomial distributions. A document can be generated by every topic in the DAG. Liu et al. [18] proposed the hierarchical latent tree analysis (HLTA), which iteratively employed the Bridged-Islands algorithm to cluster words and topics. However, the model failed to deal with polysemous words, which is one of the major contributions of topic modeling over text.

With the popularity of neural networks, many researchers aimed at addressing the drawbacks of traditional topic models by NVI. Miao et al. [21] assumed that topic distributions in documents can be represented by hidden variables sampled from multiple Gaussian distributions, and they used the variational lower bound as the objective function of their proposed model named NVDM. Since NVDM did not explicitly model the word distributions, Miao et al. [20] extended it to several models including GSM which conforms to the assumption of topic models with multinomial distributions over both topics and words. Srivastava and Sutton [28] employed the Gaussian distribution to approximate the Dirichlet distribution, which further improved the variational auto-encoder and LDA accordingly. Based on the Wasserstein autoencoders framework, Nan et al. [24] proposed the WLDA, which applied a suitable kernel in minimizing the Maximum Mean Discrepancy to perform distribution mapping. Burkhardt et al. [5] used the Dirichlet distribution as a prior and meanwhile decoupled sparsity and smoothness. Wu et al. [29] utilized Negative-Binomial process and Gamma Negative-Binomial process to improve the sparsity of topic distributions. For short texts, Wu et al. [30] proposed a new topic distribution quantization approach in the auto-encoder framework to generate peakier distributions, as well as a negative sampling decoder to avoid generating repetitive topics. Unfortunately, these neural topic models can not model the topic hierarchy.

A few researches concentrated on modeling the hierarchical structure among latent variables based on NVI. Goyal et al. [9] combined nCRP with variational auto-encoder to enable infinite flexibility of the latent representation. Their approach was applied in video representation learning and the joint training limited the efficiency. Isonuma et al. [11] incorporated a doubly-recurrent neural network (DRNN) into NVI and proposed a tree-structured neural topic model (TSNTM). The model greatly improved the computational efficiency compared with hLDA. However, the adopted DRNN was only used to generate topic representations, rather than taking document representations as input. Such an issue makes TSNTM fail to extract a reasonable topic hierarchy. Moreover, the topic hierarchy constructed by DRNN needs to be updated frequently via a heuristic method. This motivates us to propose HNTM, which extracts a explainable topic hierarchy via a feedforward decoder automatically with much fewer computation costs. Notice that the recent work of Chen et al. [7] also employs NVI with a feedforward decoder to extract the topic hierarchy, but the proposed nTSNTM is quite different from our HNTM. First, nTSNTM was a non-parametric model that used a stick-breaking process as prior, while HNTM adopts Gaussian distribution as prior. Second, nTSNTM used a softmax function with low temperature to ensure a tree-shaped structure, but it did not consider the balance of the topic tree. For HNTM, two regularization terms and manifold learning are applied to guarantee a balanced topic tree. To the best of our knowledge, this is the first study on tackling the issue of imbalance by introducing the manifold regularization into NVI based hierarchical topic modeling.

3 Hierarchical neural topic model

In this section, we first introduce the modeling of topic hierarchy based on the NVI framework and then describe the details of our HNTM.

3.1 Topic hierarchy

Previous hierarchical topic models mainly take a tree-shaped structure, but they have a difference in how to generate a document from the hierarchical topics. Figure 2 shows the tree structure of different models and topic distributions of a simulated document. Particularly, HLDA considers that a document is generated by topics of only one path, which violates the multi-topic assumption of topic models (i.e., a document may span several topics). Considering this issue, rCRP generates a document by all topics in the tree. We follow rCRP to develop a tree structure that a document is generated by all layers of the topic tree cooperatively.

Based on the framework of NVI, we reconstruct the input document by multiple layers of latent variables. Layers are connected with dependency matrices $\mathbf{D}$, where $\mathbf{D}_l$ means the

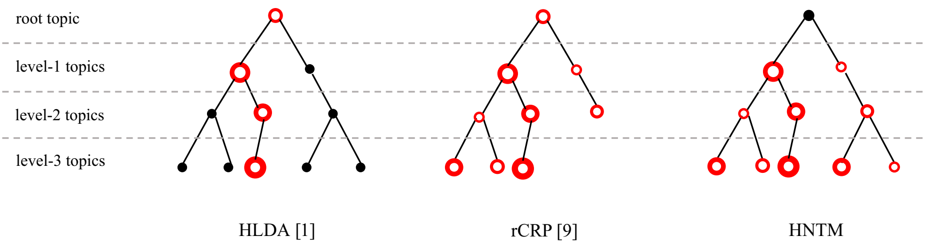

Figure 2 Tree structures and topic distributions of a simulated document for our HNTM and other models. Each node represents a topic with its own word distribution except for the root node in HNTM. Red node means that the topic is activated in the current document and the size of nodes represents the proportion

dependency matrix between layers l and $l+1$. To estimate $\mathbf{D}_l$ (i.e., the dependency strength between the super-topics at level l and the sub-topics at level $l+1$), we introduce super-topic vectors p_l and sub-topic vectors b_l, as follows:

$$D_{l,k} = softmax(p_l * b_{l,k}^T). \tag{1}$$

In the above, $D_{l,k}$, which represents the dependency vector of sub-topic k, approximates a discrete one-hot vector after using the softmax function. The super-topic vectors $p_l \in \mathbb{R}^{K_l * H}$, and the sub-topic vectors $b_l \in \mathbb{R}^{K_{l+1} * H}$, where H is the length of each topic vector, K_l and K_{l+1} represent the numbers of topics at level l and level $l+1$. To construct a pyramid-shaped topic tree, the topic number K_l is incremental from level 1 to level L.

3.2 Network structure

As in probabilistic topic models, we use the latent variables θ_d and z_n to capture the topic proportion of document d and the topic assignment for the observed word w_n, respectively. To learn the hierarchical structure, sub-topics are generated using multinomial distributions through dependency matrices $\mathbf{D}$. The topic distribution of level L can be generated by:

$$\theta_{d,L} \sim G(\mu_0, \sigma_0^2), \tag{2}$$

where $G(\mu_0, \sigma_0^2)$ is composed of a multi-layer perceptron (MLP) $\theta_L = f(x)$ conditioned on an isotropic Gaussian $x \sim N(\mu_0, \sigma_0^2)$, and L is the number of topic levels. Given $\theta_{d,l}$ which represents the topic distribution of document d at level l, the topic distribution at the upper level $l-1$ can be inferred by:

$$\theta_{d,l-1} = D_{l-1}\theta_{d,l} \qquad (l = 2...L). \tag{3}$$

Then the generative process of each word is described as follows:

$$z_{l,n} \sim Multi(\theta_{d,l}) \qquad (l = 1...L), \tag{4}$$

$$t \sim Multi(c_d), \tag{5}$$

$$w_n \sim Multi(\beta_{t,z_{t,n}}), \tag{6}$$

where $z_{l,n}$ and $w_{l,n}$ represent the topic assignment and word assignment of token n in document d generated by level l. $\beta_{t,z_{t,n}}$ is the word distribution of topic $z_{t,n}$ at level t. $c_{d,l}$ denotes the reconstruction weight of level l. Finally, the marginal likelihood of document d is:

$$p(d|\mu_0, \sigma_0, \beta) = \int_{\theta_{d,1}} p(\theta_{d,1}|\mu_0, \sigma_0^2) \prod_n \sum_l c_{d,l} \sum_{z_{l,n}} p(w_n|\beta_{l,z_{l,n}}) p(z_{l,n}|\theta_{d,l}) d\theta_{d,1}, \tag{7}$$

where θ_l can be calculated by Equation (2).

Following [20], we construct an inference network $q(\theta|\mu(d), \sigma(d))$ to approximate the posterior $p(\theta|d)$, and employ the reparameterization trick [13] for parameter update. Figure 3 shows the network structure of our HNTM. To explicitly model the word distribution of each topic, topic-word matrices $\boldsymbol{\beta}$ are constructed as similar to dependency matrices **D**.

We introduce topic vectors $t_l \in \mathbb{R}^{K_l * H}$ for each level and word vectors $v \in \mathbb{R}^{V*H}$, and generate the topic distributions over words at level l by:

$$\beta_{l,k} = softmax(v * t_{l,k}{}^T). \tag{8}$$

Given such word distributions and a sampled $\hat{\theta}_l$, layer l reconstructs document d by:

$$p(w_n|\beta_l, \hat{\theta}_l) = \sum_{z_n}[p(w_n|\beta_{l,z_n})p(z_n|\hat{\theta}_l)] = \hat{\theta}_l * \beta_l. \tag{9}$$

In fact, some documents may focus on general topics, which means topics from the high level are more often used, while some documents talk about more specific topics. Considering this, our model learns the weight **c** of topic levels from the original document, which will affect the reconstruction process. Finally, the variational lower-bound is defined as:

$$L_d = \mathbb{E}_{q(\theta|d)}\Bigg[\sum_n \log\Bigg(\Bigg[\sum_l c_l p(w_n|\beta_l, \theta_l)\Bigg)\Bigg]\Bigg] - D_{KL}\left[q(\theta|d)||p(\theta)\right]. \tag{10}$$

Level weight **c** can be obtained from a latent document embedding with a fully connected layer and softmax function. With the help of **c**, our model allocates the words of a document to different topic levels flexibly. Topics at higher levels learn more general words, while topics at lower levels learn more specific words.

3.3 Generating a tree-shaped structure

By training the dependency matrices between different layers, we can learn the latent relations of topics. The topic relations constitute a DAG, where the directed edges in the graph point from the ancestor topics to the sub-topics. Every two adjacent layers are fully

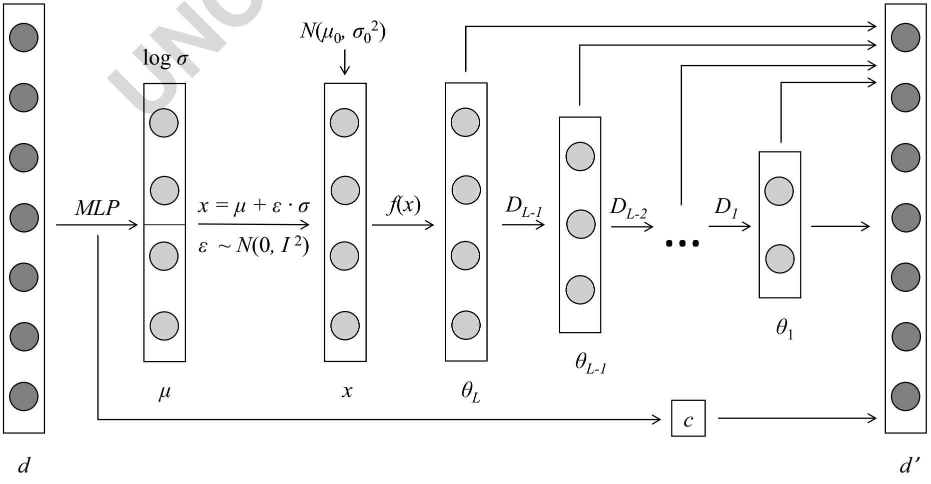

Figure 3 Network structure of an L-level HNTM

connected, which means a sub-topic may belong to several super-topics. To make the hierarchical affiliation obvious, we tend to organize topics to a tree structure. In this way, we can clearly know which sub-topics are included in a field.

A straightforward method is to constrain the dependency matrices so that the topic hierarchy can approximate a tree structure. We apply a negative L_2 normalization to the dependency matrices $\mathbf{D}$ as follows:

$$R_V = -\sum_{l}^{L-1} \sum_{i,j} {D_{l,i,j}}^2, \tag{11}$$

where $D_{l,i,j}$ represents the probability that the i-th sub-topic at level l+1 belongs to the j-th super-topic at level l. The negative L_2 normalization constrains the row vectors in each matrix to be discrete because the softmax function forces the vector sum up to 1, while traditional positive L_2 normalization forces the row vectors to be smooth. With such a constraint, every topic under level 1 belongs to only one parent topic, while parent topics can own several child topics.

However, a major problem of only using the above constraint term to generate a tree-shaped structure is that the model may learn very few super-topics from the bottom topics at level L, because most sub-topics are gathered under one super-topic. To avoid this issue, we further introduce a regularization term to adjust the number of children for each parent topic as follows:

$$R_N = \sum_{l}^{L-1} \sum_{j} \left(\sum_{i} D_{l,i,j} \right)^2. \tag{12}$$

Note that $\sum_i \sum_j D_{l,i,j} = K_{l+1}$, so reducing R_N can adjust the total amount of sub-topics for each super-topic. The above two terms work together to generate an effective and balanced topic tree.

3.4 Manifold regularization

Although HNTM with R_V and R_N can learn effective hierarchical relations between topics, they do not consider the impact of noisy words (i.e. non-topic words). In order to enhance the robustness of our model, we introduce Laplacian Eigenmap as a regularization term into our loss function with the aim of making the related texts as similar as possible in the topic distribution at the document level, and reducing the impact of noisy words. Laplacian Eigenmap is one of the famous methods in manifold learning for dimensionality reduction [1], which operates on a manifold, aiming to construct a representation for data sampled from a low dimensional manifold embedded in a higher dimensional space. Generally, manifold learning assumes that the learned representation should be smooth, which means that the points connected to each other should be as close as possible after dimensionality reduction. As an effective regularization term, manifold learning has been widely used in various algorithms, such as semi-supervised models [2, 10] and the Dirichlet Multinomial Mixture model [15].

Suppose that each document d in the corpus is regarded as a node in the graph, and for every two documents $d_i, d_j \in \boldsymbol{B}$, the adjacency matrix between documents d_i and d_j is defined as follows:

$$W_{i,j} = \begin{cases} 1, & \text{if } d_i \in \boldsymbol{\Delta}(d_j) \text{ or } d_j \in \boldsymbol{\Delta}(d_i); \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

In the above, $\boldsymbol{B}$ denotes a Batch in the neural network, and $\boldsymbol{\Delta}(d)$ denotes the document set of the R nearest neighbors of document d. Particularly, we employ the cosine distance of bag of words to measure the similarity of two documents to obtain the R nearest neighbors. The manifold regularization term is defined by:

$$R_M = \sum_{i,j=1}^{D} \sum_{k=1}^{K} W_{i,j} |\theta_{i,k} - \theta_{j,k}|, \tag{14}$$

where D is the number of documents in $\boldsymbol{B}$, K is the number of topics, $\theta_{i,k}$ and $\theta_{j,k}$ are the kth items in the topic distributions of documents d_i and d_j, respectively.

3.5 Loss function

Considering all regularization terms discussed above, the loss function of the model is defined as:

$$L = L_d + \lambda_V R_V + \lambda_N R_N + \lambda_M R_M, \tag{15}$$

where λ_V, λ_N, and λ_M are the weights of R_V, R_N, and R_M with respect to L_d, respectively. By incorporating these three regularization terms, our proposed model can extract an effective hierarchical tree structure of latent topics. In the following, we denote HNTM with R_V as HNTM-R_V, HNTM with R_V and R_N as HNTM-R_V+R_N, HNTM with R_M as HNTM-R_M, HNTM with R_V, R_N and R_M as HNTM-all. Since R_N is used to alleviate the issue of only using R_V as the constraint, we do not consider HNTM with R_N alone and other model variants for simplicity.

3.6 Computational complexity

For the feedforward propagation in our HNTM, the computational complexity is:

$$\mathcal{O}\left(nt\left(VH + (r-1)H^2 + HK_L + \sum_{l=1}^{L-1} K_l * K_{l+1} + \sum_{l=1}^{L} K_l V\right)\right), \tag{16}$$

where n is the number of training samples, t is the number of epochs, V is the vocabulary size, r is the number of fully connected layers in the encoder, H is the hidden size, K_l is the number of topics at level l, and L is the depth of the topic hierarchy. Note that V is much larger than H, r, and K_l generally, so the computational complexity will be:

$$\mathcal{O}\left(nt\left(V(H + \sum_{l=1}^{L} K_l)\right)\right). \tag{17}$$

The computational complexity of back propagation in our HNTM is exactly the same. Though the complexity is similar to that of TSNTM [11], our HNTM does not need another heuristic process to update the topic hierarchy in the training process of TSNTM, which will influence the training speed greatly.

4 Empirical results

In this section, we first describe the datasets and the experimental settings. Then, we evaluate the effectiveness of our method on the topic interpretability, hierarchical properties, data scalability, and the quality of topic words.

4.1 Datasets

We conduct experiments on three widely used benchmark datasets: 20NEWS [21], Reuters [29], and Wikitext-103 [19]. For 20NEWS, we use the same version as Miao et al. [21] which consists of 18,845 news articles under 20 categories. The news articles are divided into 11,314 training documents and 7,531 testing documents. The Reuters dataset contains 7,769 training documents and 3,019 testing documents. The Wikitext-103 [19] dataset is extracted from Wikipedia. It contains 28,472 training documents and 60 testing documents. Furthermore, the Wikitext-103 dataset has 20,000 words in the vocabulary to preserve enough information. Following Wu et al. [29], the first two datasets both have vocabularies with 2,000 most frequent words after stemming and stop words filtering.

4.2 Experimental setup

For hierarchical topic models, we use rCRP [12], HLDA [3], TSNTM [11], and nTSNTM [7] as our baselines. The other two models, i.e., hPAM [22] and HLTA [18], are not adopted for the following reasons. First, hPAM assumes that the hierarchy contains a root topic, super-topics, and sub-topics. The fixed depth setting limits the model's flexibility. Second, HLTA actually is more like a word clustering model, because it assumes that each word only belongs to one topic and fails to deal with polysemous words. For completeness, we also compare our model with several popular NVI-based flat topic models, including GSM [20], DVAE [5], NB-NTM & GNB-NTM [29].

For the aforementioned baseline models, the publicly available codes of rCRP[2], HLDA[3], TSNTM[4], nTSNTM[5], DVAE[6], NB-NTM & GNB-NTM[7] are directly used. As an extended model of NVDM, the baseline of GSM is implemented by us based on the code of NVDM[8]. For NVI based models, the number of hidden variables at each layer is set to 256 and we use the single sample by following [20]. For other model parameters such as λ_V, λ_N, and λ_M, grid search is carried out on the training set to determine their optimal values and achieve the held-out performance. Training is stopped when the performance on the validation set is not improved for 10 consecutive iterations.

We observe that hierarchical baselines can get relatively good performance when given $100 \sim 150$ topics for these three datasets. To generate a pyramid-shaped topic tree, we develop a three-level structure for HNTM with 10 level-1 topics, 30 level-2 topics, and 90 level-3 topics. The number of topics for GSM is set to 130 for fair comparison. In the training stage, we observe that KL-divergence quickly converges at the beginning, resulting the problem of component collapsing [5]. To avoid such a problem, we first give KL-divergence a small coefficient u, and increase the coefficient to 1 gradually by $u = u + 0.003 * epochs$.

[2] https://github.com/uilab-github/rCRP
[3] https://github.com/joewandy/hlda
[4] https://github.com/misonuma/tsntm
[5] https://github.com/hostnlp/nTSNTM
[6] https://github.com/sophieburkhardt/dirichlet-vae-topic-models
[7] https://github.com/mxiny/NB-NTM
[8] https://github.com/ysmiao/nvdm

4.3 Quantitative results

Perplexity is a traditional metric used to evaluate the goodness-of-fit of a model. The perplexity of each model on a testing set $\tilde{D}$ is calculated by:

$$Perplexity(\tilde{D}) = exp\left(\frac{-1}{|\tilde{D}|}\sum_{d}\frac{1}{N_d}\log p(d)\right), \tag{18}$$

where $\log p(d)$ is the log-likelihood on document d, and N_d is the number of words in d. For all neural topic models, the variational lower bound, which is proven as the upper bound of perplexity [23], is used to calculate the perplexity by following [21].

Several studies [6, 26] pointed that perplexity is not suitable for evaluating topic interpretability, and Lau et al. [14] showed that the normalized point-wise mutual information (NPMI), which evaluates the topic coherence, closely corresponds to the ranking of topic interpretability by human annotators. NPMI measures the relation between two words w_1 and w_2 as follows [14]:

$$NPMI(w_1, w_2) = \log\frac{P(w_1, w_2)}{P(w_1)P(w_2)}/(-\log P(w_1, w_2)). \tag{19}$$

The higher the value of NPMI, the more explainable the topic is. Note that topic coherence can not reveal the quality of all extracted topics, because high redundancy is not conflict with high coherence. Thus, we further adopt topic uniqueness (TU) by following [24] to evaluate the redundancy of topics. The TU for topic k is

$$TU(k) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{cnt(m,k)}, k = 1, ..., K, \tag{20}$$

where $cnt(m, k)$ is the total number of times the m^{th} top word in topic k appears in the top M words across all topics, and K is the number of topics. The final TU is computed as $TU = \frac{1}{K}\sum_{k=1}^{K} TU(k)$. Topics with both high TU and high NPMI are considered as well extracted. For NPMI and TU, we compute the average of three scores based on 5, 10, and 15 top words.

Table 1 shows the NPMI and TU of topics learned by each model respectively. All of our models except for HNTM-R_V outperform the other four hierarchical baselines on NPMI, while achieve the second highest TU for each dataset. Without the help of R_N, the constraint term R_V might cause the issue of imbalance, which has been discussed in Section 3.3, and HNTM-R_V performs worse on Reuters. With a similar Gaussian softmax framework, HNTM and its extensions perform better than GSM, which validates that hierarchical modeling can help extract more explainable topics with a low topic redundancy. Q2

Though it has been shown that perplexity is not a good metric for qualitative evaluation of topics [26], this metric can still reveal the fitting ability. According to Table 2, our models achieve competitive perplexity in comparison with other models except for rCRP. Previous studies [11, 28] also reported that sampling-based models always achieve lower perplexity when compared with NVI-based models.

To evaluate the impact of manifold regularization on the proposed method, we present our models' perplexity, NPMI and TU with different manifold regularization term coefficients (i.e., $\lambda_M = 0, 0.3, 1,$ and 3) in Figures 4 and 5. For Reuters and 20NEWS, HNTM-R_M with $\lambda_M = 0.3$ and $\lambda_M = 1$ achieve better NPMI and TU scores than HNTM to a certain extent while HNTM-R_M with $\lambda_M = 3$ performs worse than HNTM. This suggests that the

Table 1 NPMI and TU of different models, where the best results are bolded. For clarity, we present the ranking of each method on these two metrics in parenthesis

Model	20NEWS		Reuters		Wikitext-103	
	NPMI	TU	NPMI	TU	NPMI	TU
GSM	0.193 (13)	0.353 (11)	0.155 (13)	0.199 (13)	0.217 (8)	0.512 (12)
DVAE	0.263 (2)	0.404 (10)	0.357 (2)	0.413 (5)	**0.423 (1)**	0.584 (9)
NB-NTM	0.234 (8)	0.424 (8)	0.269 (3)	0.351 (8)	0.159 (13)	0.606 (7)
GNB-NTM	**0.269 (1)**	0.406 (9)	**0.368 (1)**	0.315 (9)	0.193 (11)	0.558 (10)
HLDA	0.210 (10)	0.497 (7)	0.207 (9)	0.366 (7)	0.180 (12)	0.586 (8)
rCRP	0.198 (12)	0.299 (13)	0.198 (10)	0.237 (12)	0.202 (10)	0.358 (13)
TSNTM	0.210 (10)	0.320 (12)	0.179 (11)	0.253 (11)	0.215 (9)	0.531 (11)
nTSNTM	0.227 (9)	**0.705 (1)**	0.229 (5)	**0.524 (1)**	0.237 (2)	**0.670 (1)**
HNTM	0.244 (5)	0.600 (5)	0.217 (8)	0.395 (6)	0.231 (6)	0.608 (4)
HNTM-R_V	0.238 (7)	0.614 (3)	0.176 (12)	0.300 (10)	0.227 (7)	0.608 (4)
HNTM-R_V+R_N	0.245 (4)	0.605 (5)	0.228 (6)	0.420 (4)	0.235 (5)	0.610 (3)
HNTM-R_M	0.243 (6)	0.616 (2)	0.223 (7)	0.446 (3)	0.237 (2)	0.612 (2)
HNTM+all	0.247 (3)	0.614 (3)	0.243 (4)	0.486 (2)	0.237 (2)	0.608 (4)

constraints of the characteristics of the data on the manifold can indeed improve the performance of HNTM, but too strong constraints will also make the model hard to converge. For Wikitext-103, the manifold regularization term has no obvious effect on the improvement of HNTM. This might be due to the sparse connections caused by the large scale of Wikitext-103.

Table 2 Perplexity of different models, where the best results are bolded and the ranking of each method is presented in parenthesis for clarity

Model	20NEWS	Reuters	Wikitext-103
GSM	1080.2 (11)	270.8 (10)	1869.5 (2)
DVAE	5131.6 (12)	5296.4 (12)	3461.9 (12)
NB-NTM	**811.3 (1)**	209.4 (2)	2214.5 (8)
GNB-NTM	871.6 (3)	221.4 (7)	2382.2 (10)
rCRP	811.5 (2)	**181.8 (1)**	**1722.3 (1)**
TSNTM	973.2 (9)	248.3 (9)	2267.7 (9)
nTSNTM	1000.3 (10)	357.2 (11)	2525.2 (11)
HNTM	883.8 (5)	217.3 (6)	2122.9 (4)
HNTM-R_V	890.2 (7)	223.3 (8)	2200.7 (7)
HNTM-R_V+R_N	898.1 (8)	212.8 (3)	2145.5 (6)
HNTM-R_M	884.7 (6)	215.4 (5)	2133.4 (5)
HNTM-*all*	880.5 (4)	214.5 (4)	2114.6 (3)

For HLDA, note that the inference of held-out documents will change the structure of topic trees, which involves another training process, thus we do not present its perplexity

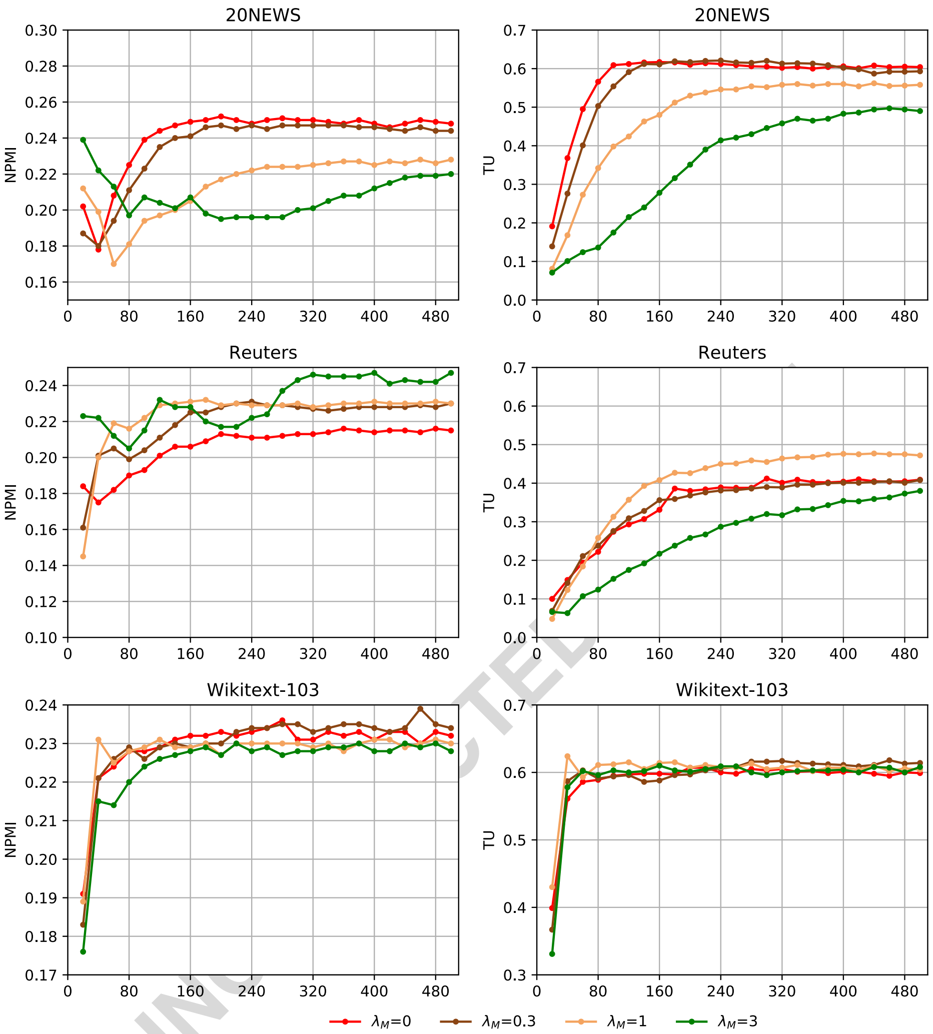

Figure 4 NPMI and TU for HNTM-*all* with various manifold regularization coefficients

4.4 Topical hierarchy analysis

In this part, we adopt topic specialization as an indicator of topical hierarchy [12]. An important feature of the tree structure is that the topics close to the root are more general, while the topics close to the leaves are more specific. Following [12], we calculate the cosine *similarity* of the word distribution between the corpus topic and all topics at each level of the topic tree, and measure the specialization score by $1 - similarity$. The corpus topic is defined as the word distribution of the entire corpus. A higher score indicates that the topic has drifted farther away from the entire corpus, which implies that the topic has become more specialized. Figure 6 presents the average topic specialization scores for each model. Though the scores of HLDA rise sharply, the topics are too general at level 1 and level 2, especially for 20NEWS. This is because the words of a document are divided into very few

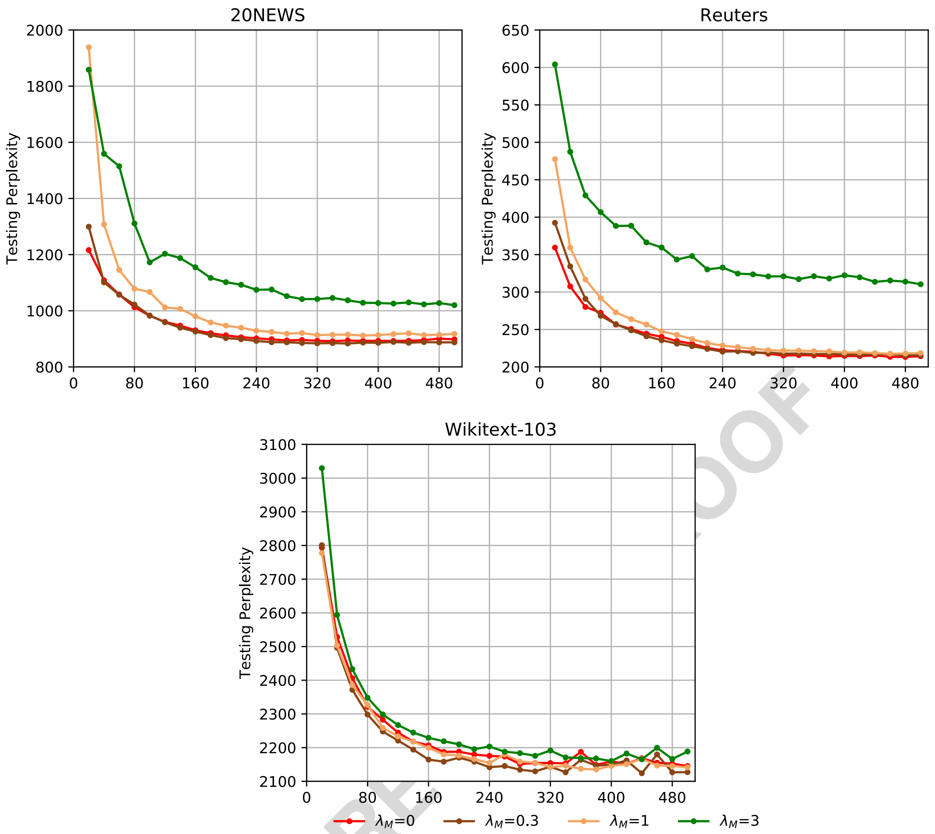

Figure 5 Perplexity for HNTM-*all* with various manifold regularization coefficients

topics, and the general words are concentrated at shallower levels. We observe that TSNTM achieves higher specialization scores at level 1 than deeper levels for all datasets, which means the topics at level 1 are more specific than their offspring topics and it indicates a bad topical hierarchy. nTSNTM obtains the highest specialization scores at every level for each dataset, indicating a poor progressive semantic structure. Our proposed model performs the best in topic specialization scores because it can learn general topics from the bottom topics flexibly.

A problem of topic specialization score is that it can not reflect the relations between parent topics and their children. In addition, since NPMI can only measure the relation between words inside the topic, we thus compute the cross-level NPMI (CLNPMI) [7] to measure the relation of top words between two connected topics by calculating the average NPMI value of every two different words from an ancestor topic and its sub-topic. The CLNPMI is defined by:

$$CLNPMI(W_p, W_c) = \frac{1}{N^2} \sum_{w_i \in W_p} \sum_{w_j \in W_c} [NPMI(w_i, w_j) \frac{\mathbb{I}(w_i \neq w_j)}{\mathbb{I}(w_j \in W_p) + 1}], \qquad (21)$$

where W_p and W_c denote the top N words of a parent topic and one of its children. The words that appear in both topics will bring a penalty to the value of CLNPMI. We also

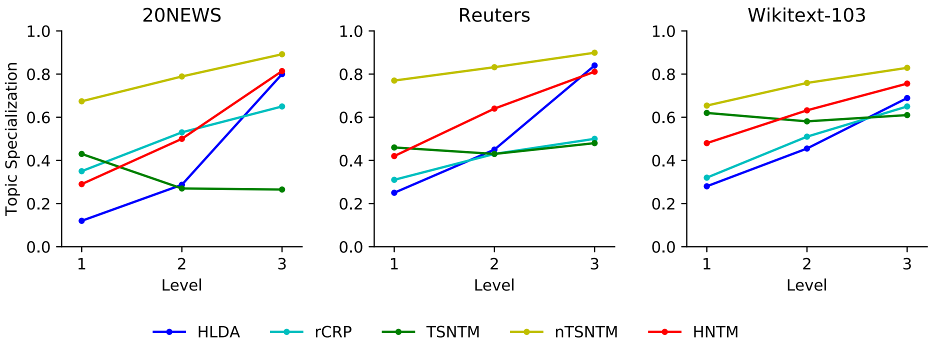

Figure 6 Topic specialization of different models at each level. Since the results of all our models are quite similar, we here present the result of HNTM for simplicity

compute the averaged overlap rate (OR) [7] to measure the repetitions between parent topics and their children. OR is defined as:

$$OR(W_p, W_c) = \frac{|W_p \cap W_c|}{N}. \tag{22}$$

As shown in Table 3, although HLDA achieves the lowest OR scores, the poor CLNPMI indicates that the relation between parents topics and their children are not very close. rCRP seriously suffered from the high topic redundancy, since it achieves high OR scores and high TU scores as aforementioned. HNTM with all regularization terms (i.e., HNTM-*all*) achieves the best CLNPMI in all datasets, with relative low OR scores. The improvement from HNTM-R_V and HNTM-R_V+R_N validates that the manifold regularization term can help extract the topic relations. In detail, Figure 7 explores the impact of different weights of manifold regularization on these two measurements. To validate the effect of R_N, we display the distributions over different numbers of children for all parent topics in Figure 8. The results indicate that our model with R_N has more proper distributions over numbers of children. Considering the poor results of HNTM-R_V presented in previous tables, the

Table 3 CLNPMI and OR of hierarchical topic models, in which, a higher CLNPMI and a lower OR indicate better performance

Model	20NEWS		Reuters		Wikitext-103	
	CLNPMI	OR	CLNPMI	OR	CLNPMI	OR
HLDA	0.084 (7)	**0.020 (1)**	0.065 (6)	**0.034 (1)**	0.083 (6)	**0.045(1)**
rCRP	0.114 (4)	0.317 (7)	0.079 (5)	0.528 (7)	0.107 (5)	0.436 (7)
TSNTM	0.115 (3)	0.289 (6)	0.081 (4)	0.181 (6)	0.083 (6)	0.132 (6)
nTSNTM	0.114 (4)	0.061 (4)	0.106 (3)	0.102 (3)	0.117 (2)	0.111 (5)
HNTM-R_V	0.110 (6)	0.022 (2)	0.020 (7)	0.070 (2)	0.109 (4)	0.062 (2)
HNTM-R_V+R_N	0.124 (2)	0.054 (5)	0.110 (2)	0.102 (3)	0.115 (3)	0.082 (4)
HNTM-*all*	**0.125 (1)**	0.043 (3)	**0.114 (1)**	0.112 (5)	**0.120 (1)**	0.078 (3)

Since R_V is necessary to generate a tree structure for HNTM, we mainly compare the performance of HNTM-R_V, HNTM-R_V+R_N, and HNTM-*all* with other tree-structured baselines. For clarity, the best results are bolded and the ranking of each method is presented in parenthesis

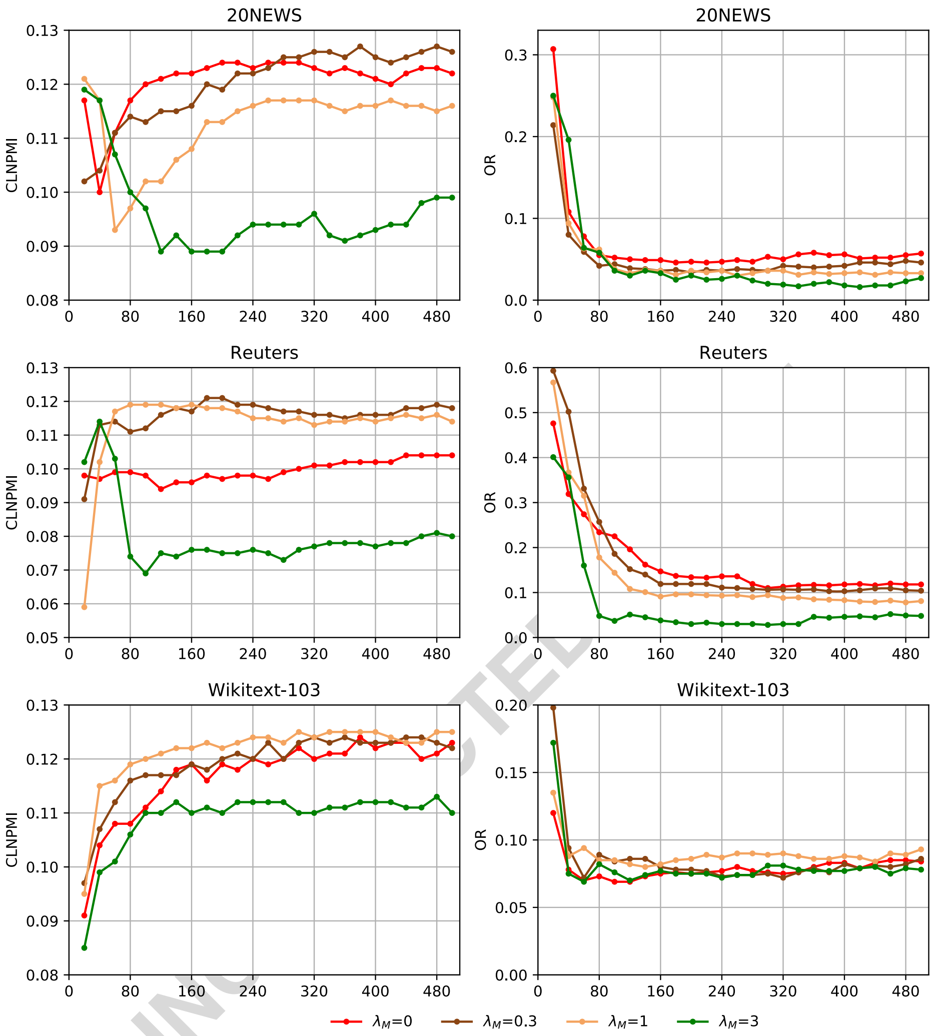

Figure 7 CLNPMI and OR for HNTM-*all* with various manifold regularization coefficients

regularization term R_N could indeed help avoid the problem of failing to extract high level topics. 377 378

We also demonstrate the discretization of the row vectors in dependency matrices **D**. As shown in Figure 9, most of the maximum elements in the row vectors are larger than 0.95 with regularization term R_V, which means these sub-topics largely belong to one super-topic. In other words, this term makes sure that the hierarchical topic structure extracted by our HNTM is a tree. 379 380 381 382 383

4.5 Data scalability 384

To evaluate the efficiency of our method, we randomly sample several numbers of documents (1,000, 2,000, 4,000, 8,000, 16,000, and all) from the training set of Wikitext-103. 385 386

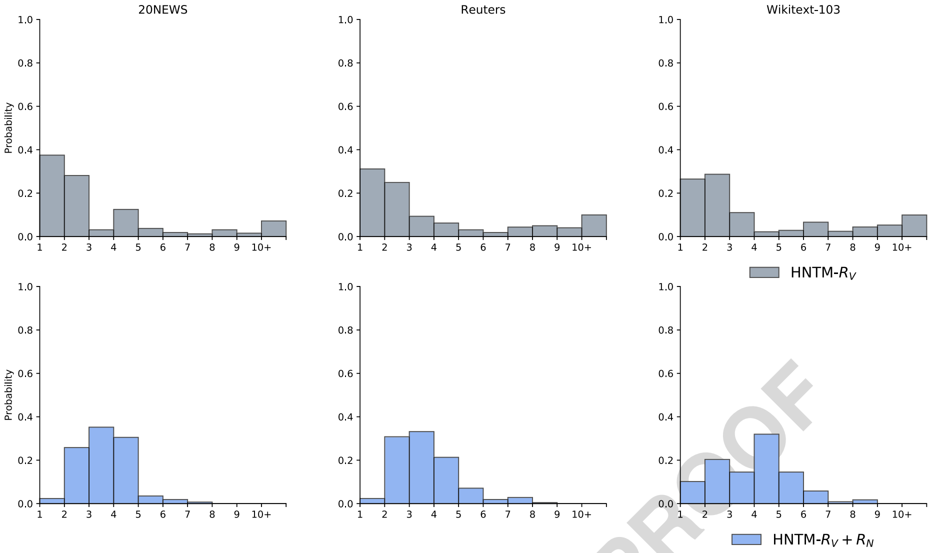

Figure 8 Distributions over the amounts of children for HNTM-R_V and HNTM-R_V+R_N

387 Figure 10 shows the training time of all hierarchical topic models, in which, the experiments are conducted on an Intel Xeon Skylake 6146 CPU with 8 cores and an Nvidia Tesla P4 GPU. Sampling-based models are run on CPU, and NVI-based models are tested on GPU. HNTM shows an advantage in time cost when compared with all these baselines. Different from flat sampling-based topic models, HLDA and rCRP spend considerable computation time on path sampling, which is much more serious when dealing with a large-scaled dataset. Additionally, these two sampling-based models are serial, which means they

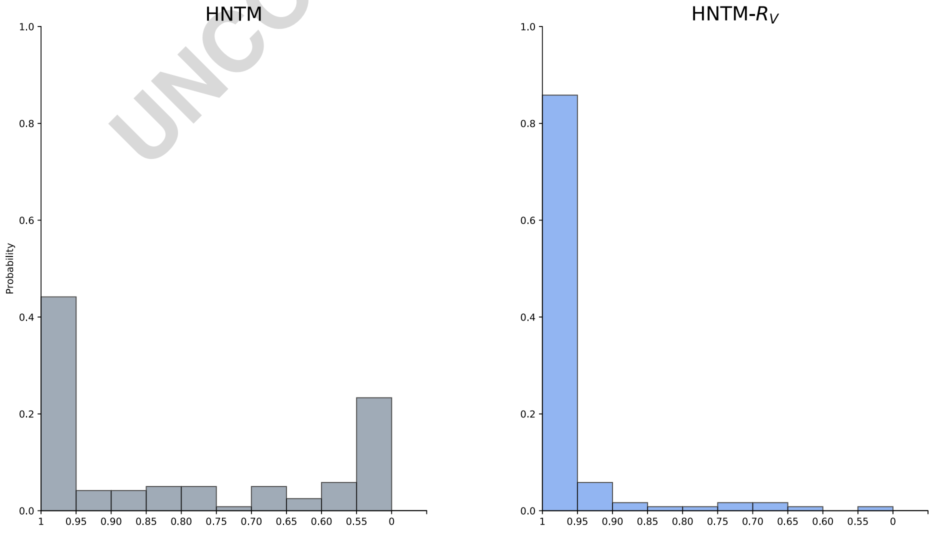

Figure 9 Distributions over the value of the maximum elements in matrices **D** for HNTM and HNTM-R_V

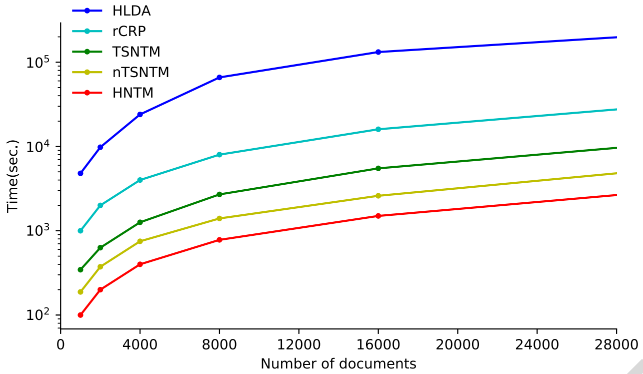

Figure 10 Training time of different models on various numbers of documents. Since the time costs of all our models are nearly the same, we here present the result of HNTM for simplicity

can only utilize one core of the CPU. TSNTM and nTSNTM respectively apply a doubly-recurrent network and a stick-breaking prior, which largely slow down the speed of both models. HNTM can be trained around 1.8 times faster than nTSNTM, 3.6 times faster than TSNTM, 10.4 times faster than rCRP, and 74 times faster than HLDA with all 28,372 documents.

4.6 Evaluation on the topic words

Figures 11, 12, 13, 14, 15 show some representative military-related branches generated by hierarchical topic models on Wikitext-103. Top 5 words are shown for each topic, and red marked topics with italic words are irrelevant to military by manually checking. Topics are truncated from level 1 to level 3.

The branches extracted by HLDA contain many irrelevant topics, while rCRP, TSNTM, and HNTM-R_V+R_N produce relatively clean branches. Furthermore, rCRP mixes topics of "military", "royalty", and "religion" into a large topic, while TSNTM and HNTM-R_V+R_N concentrate on "military". Unfortunately, TSNTM also bring in some irrelevant topics. This

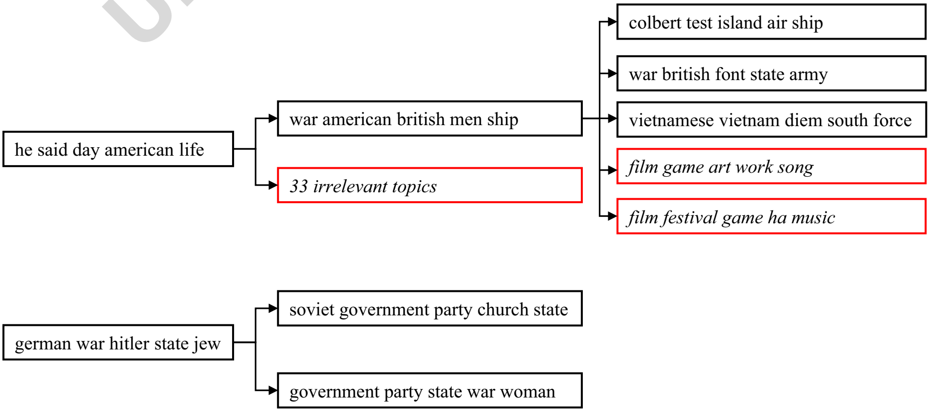
colbert test island air ship
war british font state army
war american british men ship
vietnamese vietnam diem south force
he said day american life
film game art work song
33 irrelevant topics
film festival game ha music
soviet government party church state
german war hitler state jew
government party state war woman

Figure 11 Topic branches extracted by hLDA on Wikitext-103

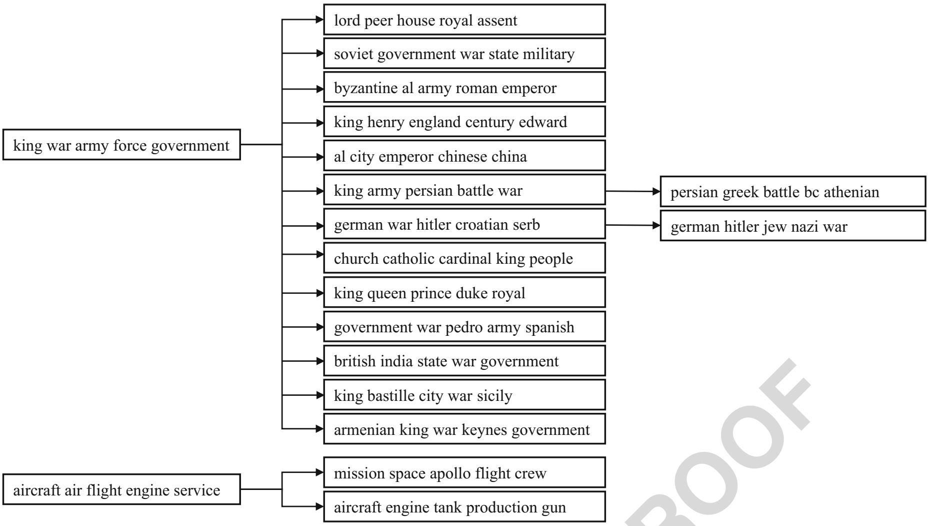
king war army force government
lord peer house royal assent
soviet government war state military
byzantine al army roman emperor
king henry england century edward
al city emperor chinese china
king army persian battle war
german war hitler croatian serb
church catholic cardinal king people
king queen prince duke royal
government war pedro army spanish
british india state war government
king bastille city war sicily
armenian king war keynes government
persian greek battle bc athenian
german hitler jew nazi war
aircraft air flight engine service
mission space apollo flight crew
aircraft engine tank production gun

Figure 12 Topic branches extracted by rCRP on Wikitext-103

result validates that the single path assumption of HLDA may be inappropriate for modeling the topic hierarchy. In addition, rCRP gets few level-3 topics in the branches, because the probability of producing deeper topics decreases exponentially. Compared to HLDA and rCRP, the hierarchical relation of topic branches obtained by HNTM-R_V+R_N is clearer and the performance is remarkable. The level-1 topic consists of general words about "military", which contains four level-2 topics including "government", "battle", "death", and "colony", each of which can be further divided into several level-3 topics. We also present the results of HNTM to verify the impact of these two regularization terms. Without the constraint of the tree structure, the topic hierarchy of HNTM is more like a DAG. Though we connect the topics by max-probability, the affiliation is still not obvious, resulting some irrelevant topics. With R_V and R_N together, our model can extract an effective and explainable topic tree. Since manifold regularization has little influence on topic words, we do not present the results of our models with R_M.

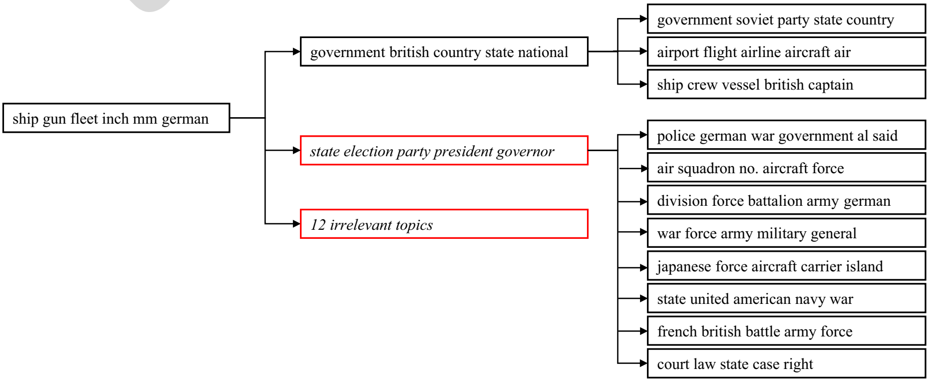
ship gun fleet inch mm german
government british country state national
state election party president governor
12 irrelevant topics
government soviet party state country
airport flight airline aircraft air
ship crew vessel british captain
police german war government al said
air squadron no. aircraft force
division force battalion army german
war force army military general
japanese force aircraft carrier island
state united american navy war
french british battle army force
court law state case right

Figure 13 Topic branch extracted by TSNTM on Wikitext-103

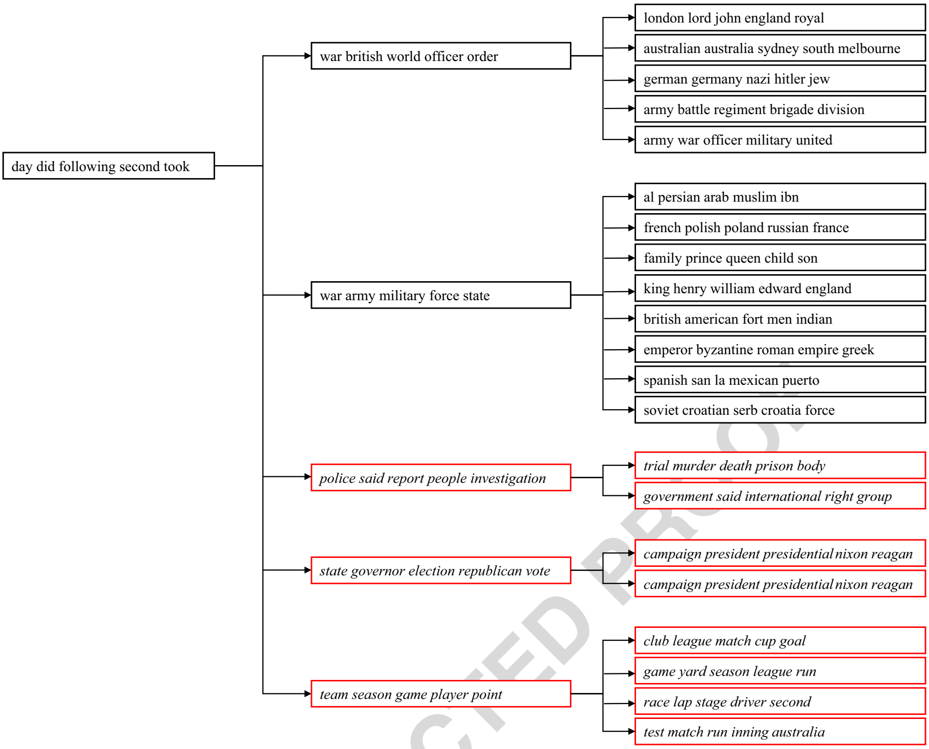
day did following second took
war british world officer order
london lord john england royal
australian australia sydney south melbourne
german germany nazi hitler jew
army battle regiment brigade division
army war officer military united
war army military force state
al persian arab muslim ibn
french polish poland russian france
family prince queen child son
king henry william edward england
british american fort men indian
emperor byzantine roman empire greek
spanish san la mexican puerto
soviet croatian serb croatia force
police said report people investigation
trial murder death prison body
government said international right group
state governor election republican vote
campaign president presidential nixon reagan
campaign president presidential nixon reagan
team season game player point
club league match cup goal
game yard season league run
race lap stage driver second
test match run inning australia

Figure 14 Topic branch extracted by HNTM on Wikitext-103

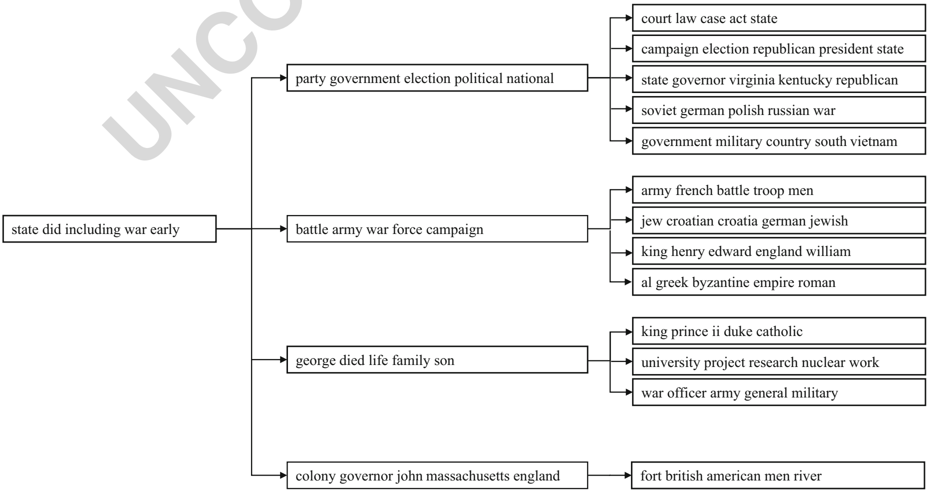
state did including war early
party government election political national
court law case act state
campaign election republican president state
state governor virginia kentucky republican
soviet german polish russian war
government military country south vietnam
battle army war force campaign
army french battle troop men
jew croatian croatia german jewish
king henry edward england william
al greek byzantine empire roman
george died life family son
king prince ii duke catholic
university project research nuclear work
war officer army general military
colony governor john massachusetts england
fort british american men river

Figure 15 Topic branch extracted by HNTM-R_V+R_N on Wikitext-103

Although the hierarchical baselines can automatically adjust the number of topics, the effects are severely affected by multiple hyper-parameters, and the resulting hierarchy is not satisfactory. HNTM predetermines suitable numbers of nodes, and can adjust the granularity of each layer according to a held-out document set, so as to obtain an effective topic hierarchy.

5 Conclusion

In this paper, we have proposed a hierarchical neural topic model named HNTM. The network structure of HNTM explicitly models the dependency of latent variables at different layers, and combines them to reconstruct the input. We further introduce manifold regularization into the proposed method to improve its robustness on noisy words. Extensive experiments validate that our network structure can extract a reasonable topic hierarchy with high topic interpretability and low topic redundancy. Compared with the existing NVI based nTSNTM, our HNTM has better data scalability because it can be trained in parallel completely. Particularly, HNTM can be trained 1.8 times faster than nTSNTM on the Wikitext-103 dataset. This makes our method possible to deal with the ever-increasing scale of data on the Internet. The multiple explainable latent variables with optional granularity extracted by our HNTM can be also used in many downstream tasks, like information retrieval and text summarization. Furthermore, our model is not limit to text. A suitable dataset might be a collection of images, a collection of DNA sequences or other collections. Modeling hierarchical latent patterns with interpretability from these data is also meaningful.

However, HNTM still has some limitations. For instance, the numbers of topics at each layer must be preset. Though other models [3, 7, 11, 12] can adjust the numbers of topics dynamically, they still have to preset the hyper-parameters which control the numbers of topics. A method for deciding the appropriate numbers of topics is very important. In addition, this study only explores the Gaussian prior, while various priors have been proposed for neural topic modeling in recent years. It follows that adopting other priors deserves further research. With the rapid development of cloud storage e-commerce platforms [27], cloud computing [8, 31] and edge computing [16] services, we also plan to deploy our model efficiently by these platforms or services.

Acknowledgements We are grateful to the reviewers for their constructive comments and suggestions on this study. This work has been supported by the National Natural Science Foundation of China (61972426), Guangdong Basic and Applied Basic Research Foundation (2020A1515010536), and Research Grants Council of Hong Kong SAR, China (UGC/FDS16/E01/19).

References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 585–591 (2001)
2. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. J. Mach. Learn. Res. **7**, 2399–2434 (2006)
3. Blei, D.M., Griffiths, T.L., Jordan, M.I., Tenenbaum, J.B.: Hierarchical topic models and the nested chinese restaurant process. In: Proceedings of the 16th International Conference on Neural Information Processing Systems, pp. 17–24 (2003)

4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
5. Burkhardt, S., Kramer, S.: Decoupling sparsity and smoothness in the dirichlet variational autoencoder topic model. J. Mach. Learn. Res. **20**(131), 131:1–131:27 (2019)
6. Chang, J., Gerrish, S., Wang, C., Boyd-graber, J., Blei, D.M.: Reading tea leaves: How humans interpret topic models. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 288–296 (2009)
7. Chen, Z., Ding, C., Zhang, Z., Rao, Y., Xie, H.: Tree-structured topic modeling with nonparametric neural variational inference. In: Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pp. 2343–2353 (2021)
8. Fang, W., Yao, X., Zhao, X., Yin, J., Xiong, N.: A stochastic control approach to maximize profit on service provisioning for mobile cloudlet platforms. IEEE Trans. Syst. Man Cybern. Sys. **48**(4), 522–534 (2018)
9. Goyal, P., Hu, Z., Liang, X., Wang, C., Xing, E.P.: Nonparametric variational auto-encoders for hierarchical representation learning. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5094–5102 (2017)
10. Hu, W., Zhu, J., Su, H., Zhuo, J., Zhang, B.: Semi-supervised max-margin topic model with manifold posterior regularization. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 1865–1871 (2017)
11. Isonuma, M., Mori, J., Bollegala, D., Sakata, I.: Tree-structured neural topic model. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 800–806 (2020)
12. Kim, J.H., Kim, D., Kim, S., Oh, A.H.: Modeling topic hierarchies with the recursive chinese restaurant process. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 783–792 (2012)
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: Proceedings of the 2nd International Conference on Learning Representations (2014)
14. Lau, J.H., Newman, D., Baldwin, T.: Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, pp. 530–539 (2014)
15. Li, X., Zhang, J., Ouyang, J.: Dirichlet multinomial mixture with variational manifold regularization: Topic modeling over short texts. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence, vol. 33, pp. 7884–7891 (2019)
16. Lin, B., Zhu, F., Zhang, J., Chen, J., Chen, X., Xiong, N.N.: A time-driven data placement strategy for a scientific workflow combining edge computing and cloud computing. IEEE Trans. Industr. Inform. **15**(7), 4254–4265 (2019)
17. Liu, L., Huang, H., Gao, Y., Zhang, Y., Wei, X.: Neural variational correlated topic modeling. In: Proceedings of the World Wide Web Conference, pp. 1142–1152 (2019)
18. Liu, T., Zhang, N.L., Chen, P.: Hierarchical latent tree analysis for topic detection. In: Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 256–272 (2014)
19. Merity, S., Xiong, C., Bradbury, J., Socher, R.: Pointer sentinel mixture models. In: Proceedings of the 5th International Conference on Learning Representations (2017)
20. Miao, Y., Grefenstette, E., Blunsom, P.: Discovering discrete latent topics with neural variational inference. In: Proceedings of the 34th International Conference on Machine Learning, pp. 2410–2419 (2017)
21. Miao, Y., Yu, L., Blunsom, P.: Neural variational inference for text processing. In: Proceedings of the 33nd International Conference on Machine Learning, pp. 1727–1736 (2016)
22. Mimno, D.M., Li, W., McCallum, A.: Mixtures of hierarchical topics with pachinko allocation. In: Proceedings of the 24th International Conference on Machine Learning, pp. 633–640 (2007)
23. Mnih, A., Gregor, K.: Neural variational inference and learning in belief networks. In: Proceedings of the 31st International Conference on Machine Learning, pp. 1791–1799 (2014)
24. Nan, F., Ding, R., Nallapati, R., Xiang, B.: Topic modeling with wasserstein autoencoders. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 6345–6381 (2019)
25. Neal, R.M.: Probabilistic inference using markov chain monte carlo methods. Department of Computer Science, University of Toronto Toronto, Ontario, Canada (1993)
26. Newman, D., Lau, J.H., Grieser, K., Baldwin, T.: Automatic evaluation of topic coherence. In: Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 100–108 (2010)

27. Qu, Y., Xiong, N.: Rfh: A resilient, fault-tolerant and high-efficient replication algorithm for distributed cloud storage. In: Proceedings of the 41st International Conference on Parallel Processing, pp. 520–529 (2012)
28. Srivastava, A., Sutton, C.: Autoencoding variational inference for topic models. In: Proceedings of the 5th International Conference on Learning Representations (2017)
29. Wu, J., Rao, Y., Zhang, Z., Xie, H., Li, Q., Wang, F.L., Chen, Z.: Neural mixed counting models for dispersed topic discovery. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 6159–6169 (2020)
30. Wu, X., Li, C., Zhu, Y., Miao, Y.: Short text topic modeling with topic distribution quantization and negative sampling decoder. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 1772–1782 (2020)
31. Xiong, N., Vasilakos, A.V., Wu, J., Yang, Y.R., Rindos, A., Zhou, Y., Song, W.Z., Pan, Y.: A self-tuning failure detection scheme for cloud computing service. In: Proceedings of the 26th IEEE International Parallel and Distributed Processing Symposium, pp. 668–679 (2012)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Ziye Chen[1] · Cheng Ding[1] · Yanghui Rao[1] · Haoran Xie[2] · Xiaohui Tao[3] · Gary Cheng[4] · Fu Lee Wang[5]

Ziye Chen
chenzy35@mail2.sysu.edu.cn

Cheng Ding
dingch6@mail2.sysu.edu.cn

Haoran Xie
hrxie2@gmail.com

Xiaohui Tao
xiaohui.tao@usq.edu.au

Gary Cheng
chengks@eduhk.hk

Fu Lee Wang
pwang@ouhk.edu.hk

[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
[2] Department of Computing and Decision Sciences, Lingnan University, Tuen Mun, Hong Kong SAR
[3] School of Sciences, University of Southern Queensland, Toowoomba, Australia
[4] Department of Mathematics and Information Technology, The Education University of Hong Kong, Ting Kok, Hong Kong SAR
[5] School of Science and Technology, The Open University of Hong Kong, Ho Man Tin, Hong Kong SAR

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES:

Q1. Please check affiliations if correct.
Q2. Please check Tables if presented correctly.