Different from conventional topic models, NTMs directly optimize parameters without requiring model-specific derivations. This endows NTMs with better scalability and flexibility, resulting in significant research attention and plentiful new methods and applications.

Unlike conventional topic models, NTMs can efficiently and flexibly infer model parameters through automatic gradient back-propagation by adopting deep neural networks to model latent topics, such as the popular Variational AutoEncode (VAE, Kingma and Welling, 2014; Rezende et al., 2014).

FALTAAAA **VAE**

This flexibility enables researchers to tailor model structures to fit diverse application scenarios. In addition, NTMs can seamlessly handle largescale datasets by harnessing parallel computing facilities like GPUs. Owing to these advantages, NTMs have witnessed the exploration of numerous new methods and applications.

*Problem Setting and Notations*

We introduce the problem setting and notations of topic modeling following LDA (Blei et al., 2003). Consider a collection of N documents with V unique words (vocabulary size), and a document is denoted as x. As illustrated in Figure 2, topic models aim to discover K latent topics from this ollection. The number of topics K is a hyperparameter, usually determined by researchers manually according to the characteristics of datasets and their target tasks. Each topic is defined as a distribution over the vocabulary, i.e., topic-word distribution, βk ∈ R V . Then the topic-word distribution matrix of all topics is β = (β1 , . . . , βK) ∈ R V ×K. (1) In addition, topic models also infer the topic distribution of a document (doc-topic distribution): θ ∈ ∆K, implying what topics a document contains. Here θk refers to the proportion of Topic#k in the document, and ∆K denotes a probability simplex ∆K = {θ ∈ R K + | PK k=1 θk = 1}.

***Dynamic NTMs***

Dynamic NTMs are explored following dynamic topic models (Blei and Lafferty, 2006b; Wang et al., 2008). Previous static topic models implicitly assume that documents are exchangeable. However, this assumption is inappropriate since documents are produced sequentially, such as scholarly journals, emails, and news articles. As such, dynamic topic models are proposed. While topics in previous methods are all static, dynamic topic models allow topics to shift over time to capture the topic evolution in sequential documents.

To be specific, dynamic topic models assume that documents are divided by time slice, for example by year, and each time slice has K latent topics. The topics associated with slice t evolve from the topics associated with slice t − 1. As the example in Figure 6, Topic#1 about Ukraine and Russia evolves from the year 2020 to 2022. Due to the emergence of the word “invasion”, we see Topic#1 captures the Ukraine-Russia war that exploded in 2022. Similarly, Topic#K about Covid-19 evolves from the year 2020 to 2022 with the explosion of the Omicron variant. These topic evolution reveals how topics emerge, grow, and vanish, which has been applied for trend analysis and opinion mining.

Estado del arte

Dieng et al. (2019) first propose a neural dynamic topic model, DETM (Dynamic Embedding Topic Model). It uses word and topic embeddings to interpret latent topics following Dieng et al. (2020) and chains topic embeddings at slice t with topic embeddings at slice t − 1 by Markov chains. Besides, it uses a LSTM to learn temporal priors of doc-topic distributions Rahimi et al. (2023) discover topic evolution by clustering documents but cannot infer doc-topic distributions as required. Zhang and Lauw (2022) focus on the dynamic topics of temporal document networks and incorporate the linking information between documents. Following DETM, Miyamoto et al. (2023) propose to employ a self-attention mechanism to model the dependencies among dynamic topics. Wu et al. (2024a) focus on the unassociated topic and repetitive topic issues. Instead of the previous Markov hains fashion, they propose CFDTM with a contrastive learning method to resolve these issues and track topic evolution. Rather than modeling topic evolution, Cvejoski et al. (2023) model the activities of topics over time. Note that the activities of topics evolve over time but their topics are invariant.

NTMs with Embeddings

Alternative to directly modeling topics, Miao et al. (2017) propose to decompose topics as two embedding parameters: β = W⊤T. (16) Here W ∈ R D×V denotes V word embeddings, and T ∈ R D×K denotes K topic embeddings, where D is the dimension of embedding space. Then Dieng et al. (2020) follow this setting and propose ETM (Embedding Topic Model). ETM facilitates topic learning by initializing W with pretrained word embeddings like Word2Vec (Mikolov et al., 2013) or GloVe (Pennington et al., 2014). This approach also confers flexibility and efficiency to other topic modeling scenarios. For instance, it is much cheaper to repeat topic embeddings for each time slice in dynamic topic modeling than repeating the entire topic-word distribution matrix (Dieng et al., 2019).

Alternatively, Zhao et al. (2021b) propose NSTM. It also models topics as embeddings, but uses the optimal transport distance between doctopic distributions and input documents to measure the reconstruction error. Wang et al. (2022a) share the same idea and instead use conditional transport distance. Duan et al. (2022) learn a group of global topic embeddings for task-specific adaptations. Xu et al. (2022) propose HyperMiner, using hyperbolic embeddings to model topics. Due to the tree-likeness property of hyperbolic space, they can capture the hierarchy among topics. Differently, Wu et al. (2023b) propose ECRTM, which models the topic-word distribution matrix as βjk = exp(−∥wj − tk∥ 2/τ ) PK k ′=1 exp(−∥wj − tk ′∥ 2/τ ). Here βjk denotes the correlation between j-th word and k-th topic with τ as a temperature hyperparameter; wj is the j-th word embedding in W, and tk is the k-th topic embedding in T. It computes the Euclidean distance between topic and word embeddings and normalizes overall topics in a softmax manner. This works together with a clustering regularization method. The regularization considers topic embeddings as cluster centers and word embeddings as cluster samples; then it forces topic embeddings to be the centers of separately aggregated word embeddings by optimal transport. This effectively avoids the topic collapsing issue where topics are repetitive to each other.

NTMs with Metadata

While common NTMs learn topics in an unsupervised manner (only using documents), NTMs can also leverage the metadata of documents to guide topic modeling, similar to supervised LDA (Mcauliffe and Blei, 2007). In detail, Card et al. (2018) introduce SCHOLAR, a NTM that can incorporate various metadata of documents. It encodes a document together with its labels (e.g., sentiment) and covariates (e.g., publication year), and generates the document conditioned on the covariates. Korshunova et al. (2019) model the generation of documents and labels together in a discriminative way; then train their model with mean-field variational inference. They can also incorporate a variety of data modalities like images. Wang and Yang (2020) jointly model topics and train a RNN classifier to predict document labels. They are connected by an attention mechanism. Wang et al. (2021a) incorporate document networks in a NTM and jointly reconstruct documents and networks.

NTMs with Pre-trained Language Models

Researchers frequently combine NTMs with pretrained language models. Pre-trained language models based on Transformers (Vaswani et al., 2017) have been prevalent in NLP fields, which are pre-trained on large-scale corpora to capture contextual linguistic features. Multiple studies leverage contextual features from these pre-trained models to provide richer information than conventional BoW. For instance, Bianchi et al. (2021a) input the concatenation of BoW and the contextual document embeddings from Sentence-BERT (Reimers and Gurevych, 2019), and then reconstruct BoW as previous work. Hoyle et al. (2020) propose to distill knowledge from BERT (Devlin et al., 2018) to NTMs. In detail, they produce pseudo BoW from the predictive word probability of BERT. Then their NTM reconstructs both the real and pseudo BoW. Bianchi et al. (2021b); Mueller and Dredze (2021) employ multilingual BERT to infer cross-lingual doc-topic distributions for zero-shot learning but they cannot discover aligned cross-lingual topics.

Contrstive Learning

As a self-supervised learning fashion, contrastive learning has been employed to facilitate NTMs (Hadsell et al., 2006; Nguyen et al., 2022, 2024a). The idea of contrastive learning is to measure the similarity relations among sample pairs in a representation space (Van den Oord et al., 2018). Nguyen and Luu (2021) propose the contrastive learning on doc-topic distributions where they build positive and negative pairs by sampling salient words of documents. Differently, Wu et al. (2022) directly sample positive and negative pairs based on the topic semantics of documents to capture relations among samples. Specifically, they quantize doc-topic distributions following Wu et al. (2020b) and then sample documents with the same quantization indices as positive pairs and different indices as negative pairs. Their method can also capture the similarity relations among additional augmented data. Zhou et al. (2023) improve topic disentanglement with contrastive learning on word and topic embeddings. Han et al. (2023) cluster documents, compute term weights, and make NTMs reconstruct salient words. They also use contrastive learning to refine doc-topic distributions where positive samples come from pre-trained language models. Besides document-level contrastive learning, Nguyen et al. (2024b) also consider topic-level and propose a multi-objective contrastive learning method.