Modeling Dynamic Topics in Chain-Free Fashion by Evolution-Tracking Contrastive Learning and Unassociated Word Exclusion

Xiaobao Wu[1] **Xinshuai Dong**[2] **Liangming Pan**[3]
Thong Nguyen[4] **Anh Tuan Luu**[1]
[1]Nanyang Technological University [2]Carnegie Mellon University
[3]University of California, Santa Barbara [4]National University Singapore
xiaobao002@e.ntu.edu.sg xinshuad@andrew.cmu.edu
liangmingpan@ucsb.edu e0998147@u.nus.edu anhtuan.luu@ntu.edu.sg

Abstract

Dynamic topic models track the evolution of topics in sequential documents, which have derived various applications like trend analysis and opinion mining. However, existing models suffer from repetitive topic and unassociated topic issues, failing to reveal the evolution and hindering further applications. To address these issues, we break the tradition of simply chaining topics in existing work and propose a novel neural Chain-Free Dynamic Topic Model. We introduce a new evolution-tracking contrastive learning method that builds the similarity relations among dynamic topics. This not only tracks topic evolution but also maintains topic diversity, mitigating the repetitive topic issue. To avoid unassociated topics, we further present an unassociated word exclusion method that consistently excludes unassociated words from discovered topics. Extensive experiments demonstrate our model significantly outperforms state-of-the-art baselines, tracking topic evolution with high-quality topics, showing better performance on downstream tasks, and remaining robust to the hyperparameter for evolution intensities. Our code is available at https://github.com/bobxwu/CFDTM.

1 Introduction

Dynamic topic models seek to discover the evolution of latent topics in sequential documents divided by time slice. For example, Figure 1 illustrates Topic#1 about Covid-19 and Topic#20 about Ukraine evolve from 2020 to 2022 in the documents divided by year. This evolution reveals how topics emerge, grow, and decline due to trends and events. Such evolution has been employed in various downstream applications, *e.g.,* opinion mining, trend tracking, and sentiment analysis (Wu and Li, 2019; Sha et al., 2020; Li et al., 2020; Hu et al., 2015; Greene and Cross, 2017; Li et al., 2021; Murakami et al., 2021; Churchill and Singh, 2022).

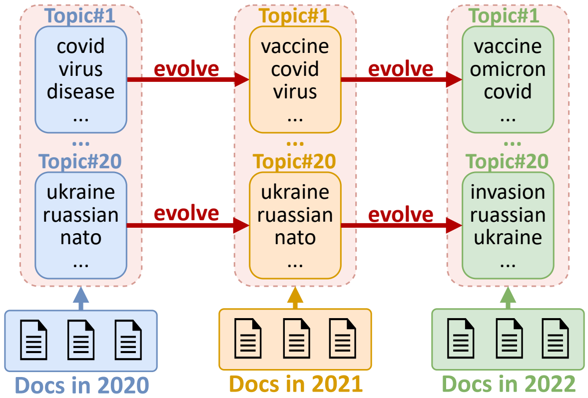


Figure 1: Illustration of dynamic topic modeling. Every time slice (year here) has certain latent topics, interpreted as related words. Each topic evolves across time slices.

Repetitive Topics			*Unassociated Topics*	
Time: **2015**			Time: **2017**	Time: **2018**
Topic#21	**Topic#30**	**Topic#40**	**Topic#2**	**Topic#2**
house	president	trump	health	*coronavirus*
impeachment	trump	president	said	health
capitol	biden	investigation	*coronavirus*	said
president	said	said	virus	virus
committee	white	report	people	new
senate	house	white	new	people
trump	administration	department	cases	cases
republicans	washington	officials	disease	*covid*
republican	office	intelligence	*covid*	disease
congress	american	house	deaths	deaths

Table 1: Illustration of repetitive topics and unassociated topics (from NYT). Each column is the top related words of a topic. Repetitive words are underlined. Unassociated words are in *red*.

Existing dynamic topic models can be classified into two types: (i) *probabilistic dynamic topic models* (Blei and Lafferty, 2006; Wang et al., 2008), learning through Variational Inference (Blei et al., 2017) or Gibbs sampling (Griffiths and Steyvers, 2004), and (ii) *neural dynamic topic models* (Dieng et al., 2019; Zhang and Lauw, 2022; Miyamoto

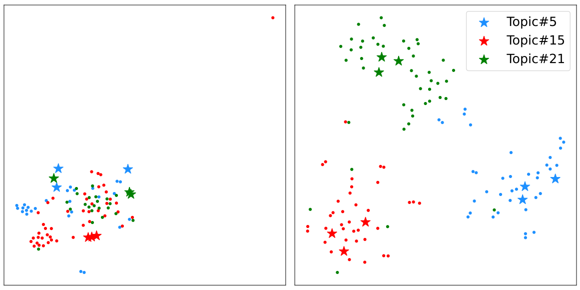
Findings of the Association for Computational Linguistics ACL 2024, pages 3088–3105
August 11-16, 2024

et al., 2023; Wu et al., 2024a), learning via gradient back-propagation.

However, these existing models mostly chain topics via Markov chains to capture topic evolution, suffering from two vital issues: **(i)** ***Repetitive Topics***: topics within a time slice are repetitive with similar semantics. Table 1 exemplifies that Topic#21, #30, and #40 in the year 2015 all include repeating words like "president" and "trump". As illustrated in Figure 2, simply chaining topics via Markov chains pushes topics to gather together and *cannot* separate them. These topics are less distinguishable and fail to uncover the complete semantics of their time slices. **(ii)** ***Unassociated Topics***: topics are unassociated with their corresponding time slices. Table 1 shows that Topic#2 in 2017 and 2018 refer to Covid-19 with words "covid" and "coronavirus", but Covid-19 does *not* outbreak in 2017 or 2018. This is because simply chaining topics via Markov chains may force topics across time slices to be overly related. These topics become less associated with their time slices, and the real topics in these slices are unrevealed. Consequently, these two issues bring about low-quality dynamic topics and hinder tracking topic evolution, which thus impairs downstream applications.

To solve the above issues, we break the tradition of chaining topics via Markov chains and propose a novel neural **Chain-Free Dynamic Topic Model** (**CFDTM**). First, we propose **Evolution-Tracking Contrastive learning** (**ETC**) to address the repetitive topic issue. ETC adaptively builds positive and negative relations among dynamic topics. Building positive relations tracks the topic evolution with different intensities. More importantly, building negative relations encourages topics within each slice to be distinct, maintaining topic diversity and thus alleviating the repetitive topic issue.

Second, we propose a new **Unassociated Word Exclusion** (**UWE**) method to solve the unassociated topic issue. UWE finds the top related words of topics at each time slice and identifies which of them do *not* belong to this slice as unassociated words. Then UWE explicitly excludes these unassociated words from topics to refine topic semantics, such as excluding the unassociated word "covid" from Topic#2 in 2017 in Table 1. This mitigates the unassociated topic issue and also enhances the robustness of our model to the hyperparameter for evolution intensities. Figure 2b visualizes the effectiveness of our CFDTM. Our contributions can be concluded as follows:



(a) DETM (b) **CFDTM**

Figure 2: t-SNE visualization with stars (★) as topic embeddings and circles (●) as word embeddings. Their time slice annotations are omitted for brevity. While baseline DETM mingles all word embeddings together, our CFDTM properly groups and separates them by topic and time slice (See their top words in Figure 5).

- To our best knowledge, we are the first to propose and investigate both the repetitive and unassociated topic issues in dynamic topic modeling.
- We propose a novel chain-free dynamic topic model with a new evolution-tracking contrastive learning method that tracks topic evolution and avoids producing repetitive topics.
- We further propose a new unassociated word exclusion method that excludes unassociated words from topics, which effectively alleviates the unassociated topic issue.
- We conduct extensive experiments on benchmark datasets and show our model consistently outperforms baselines, capturing topic evolution with high-quality topics and achieving higher downstream performance, with robustness to the hyperparameter for evolution intensities.

2 Related Work

Probabilistic Dynamic Topic Modeling Topic modeling aims to understand documents in unsupervised fashion with latent topics, deriving various text analysis (Nguyen et al., 2023; Liu et al., 2023; Mao et al., 2023, 2024a,b). Blei and Lafferty (2006) first propose Dynamic Topic Model (DTM) based on LDA (Blei et al., 2003). DTM adopts state space models to chain the natural parameters of latent topics with Gaussian noise and uses Kalman filter and wavelet regression as variational approximations. Afterward, Wang et al. (2008) introduce DTM under continuous time settings; Caron et al. (2012) extend DTM to nonparametric settings. Several other extensions are also proposed (Wang and

McCallum, 2006; Iwata et al., 2010; Bhadury et al., 2016; Jähnichen et al., 2018; Hida et al., 2018). They use Variational Inference or Gibbs sampling to optimize model parameters.

Neural Dynamic Topic Modeling Due to the success of neural topic modeling (Miao et al., 2016; Srivastava and Sutton, 2017; Wu et al., 2020a,b, 2021, 2024c,b), neural dynamic topic modeling has attracted more attention (Balepur et al., 2023). Dieng et al. (2019) first propose DETM in the framework of VAE (Kingma and Welling, 2014; Rezende et al., 2014). Later, Zhang and Lauw (2022) capture evolution from temporal document networks; Cvejoski et al. (2023) focus on modeling topic activities over time; Rahimi et al. (2023) cluster documents to find dynamic topics but cannot infer their topic distributions. Miyamoto et al. (2023) model the dependencies among topics with an attention mechanism. They all chain topics via Markov chains following probabilistic models. Differently instead of Markov chains, we propose the novel evolution-tracking contrastive learning and unassociated word exclusion methods to track topic evolution and address the repetitive topic and unassociated topic issues.

Contrastive Learning The goal of contrastive learning is to learn the similarity relations among samples (Hadsell et al., 2006; Oh Song et al., 2016; Van den Oord et al., 2018; Frosst et al., 2019; He et al., 2020; Nguyen et al., 2022, 2024a). It has become a prevalent self-supervised fashion in vision and textual fields (Chen et al., 2020; Xie et al., 2021; Gao et al., 2021; Zhao et al., 2021). Several studies apply contrastive learning in static topic models (Wu et al., 2022, 2023a; Nguyen and Luu, 2021; Nguyen et al., 2024b; Zhou et al., 2023; Wang et al., 2023; Han et al., 2023). Rather than these static topic models, we focus on dynamic topic models, which motivates our new evolution-tracking contrastive learning.

3 Methodology

We recall the problem setting of dynamic topic modeling and present our novel evolution-tracking contrastive learning and unassociated word exclusion. Finally we introduce the Chain-Free Dynamic Topic Model (CFDTM) on these two methods.

3.1 Problem Setting and Notations

Following DTM (Blei and Lafferty, 2006), we introduce the problem setting of dynamic topic mod-

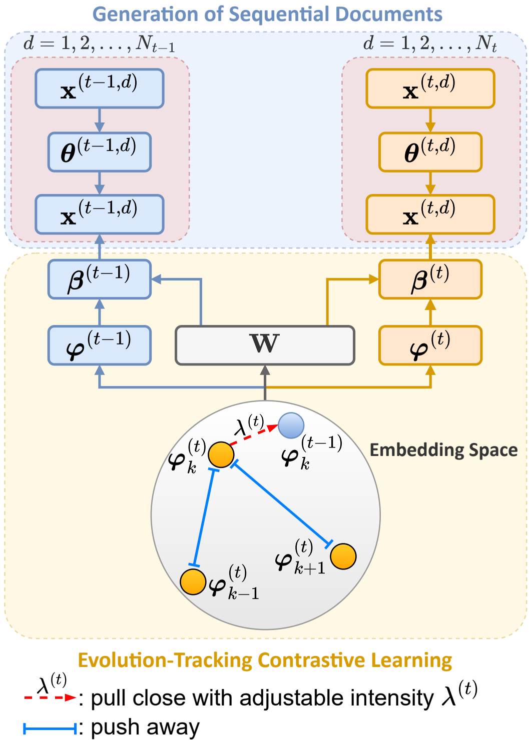


Figure 3: Illustration of the generation of sequential documents (following VAE) and Evolution-Tracking Contrastive learning (ETC). For $\boldsymbol{\varphi}_k^{(t)}$ (topic embedding of Topic#k at slice t), ETC adaptively pulls it close to $\boldsymbol{\varphi}_k^{(t-1)}$ by adjustable intensity hyperparameter $\lambda^{(t)}$, and pushes it away from $\boldsymbol{\varphi}_{k'}^{(t)}$ $(k' \neq k)$, for instance $k'=k-1$ and $k+1$ here.

eling. As shown in Figure 1, consider sequential document collections divided by T time slices, for example by year. Slice t has N_t documents with $\mathbf{x}^{(t,d)}$ as the d-th document at slice t. The entire vocabulary set is $\mathcal{V}$, and the vocabulary set of slice t is $\mathcal{V}^{(t)}$. We aim to discover K latent topics at each slice, and topics at slice t evolve from topics at slice $t-1$, *e.g.,* Topic#1 in 2022 evolves from Topic#1 in 2021 in Figure 1. **We generally preset the evolution intensity through hyperparameters**: we use strong intensity if documents evolve dramatically and weak otherwise, like the Gaussian variance in DTM (Blei and Lafferty, 2006). Following LDA (Blei et al., 2003), each topic is defined as a distribution over all words (topic-word distribution): Topic#k at slice t is defined as $\boldsymbol{\beta}_k^{(t)} \in \mathbb{R}^{|\mathcal{V}|}$. Then $\boldsymbol{\beta}^{(t)} = (\boldsymbol{\beta}_1^{(t)}, \ldots, \boldsymbol{\beta}_K^{(t)}) \in \mathbb{R}^{|\mathcal{V}| \times K}$ is the topic-word distribution matrix of slice t. The same topic at different slices reveals the evolution. We also infer the topic distribution of document $\mathbf{x}^{(t,d)}$ (doc-topic distribution), denoted as $\boldsymbol{\theta}^{(t,d)} \in \Delta_K$, where Δ_K is a probability simplex.

3.2 Parameterizing Topics as Embeddings

We begin by parameterizing topics. Following Miao et al. (2017); Dieng et al. (2020), we project words in the vocabulary into a D-dimensional embedding space as $|\mathcal{V}|$ word embeddings: $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_{|\mathcal{V}|}) \in \mathbb{R}^{D \times |\mathcal{V}|}$. We reuse word embeddings for different slices for parameter efficiency. Similarly, we project topics at each time slice into the same space: slice t has K topic embeddings $\boldsymbol{\varphi}^{(t)} = (\boldsymbol{\varphi}_1^{(t)}, \dots, \boldsymbol{\varphi}_K^{(t)}) \in \mathbb{R}^{D \times K}$, and $\boldsymbol{\varphi}_k^{(t)}$ denotes the topic embedding of Topic#k at slice t. Each topic (word) embedding represents the semantics of the topic (word). Following Wu et al. (2023b), we formulate $\boldsymbol{\beta}^{(t)}$, topic-word distribution matrix at slice t, as

$$\beta_{k,i}^{(t)} = \frac{\exp(-\|\boldsymbol{\varphi}_k^{(t)} - \mathbf{w}_i\|^2/\pi)}{\sum_{k'=1}^{K} \exp(-\|\boldsymbol{\varphi}_{k'}^{(t)} - \mathbf{w}_i\|^2/\pi)}. \quad (1)$$

Here $\beta_{k,i}^{(t)}$ models the correlation between word i and Topic#k at slice t with a scale hyperparameter π, calculated as the Euclidean distance between their embeddings with normalization along all topics. As such, a word relates to a topic if their embeddings are close and away from other topic embeddings, which cooperates with our next evolution-tracking contrastive learning.

3.3 Evolution-Tracking Contrastive Learning

To track topic evolution, we abandon traditionally chaining topics via Markov chains and propose the novel Evolution-Tracking Contrastive learning (ETC) to avoid repetitive topics. Figure 3 illustrates our ETC method.

Positive Relations among Dynamic Topics We first build positive relations among dynamic topics from a contrastive learning perspective to track their evolution. Recall that topics at slice t evolve from topics at slice $t-1$ via different intensities (Blei and Lafferty, 2006). Hence we model their topic embeddings $(\boldsymbol{\varphi}_k^{(t)}, \boldsymbol{\varphi}_k^{(t-1)})$ as positive pairs and **build their positive relations with evolution intensity hyperparameter** $\lambda^{(t)}$:

$$\mathcal{L}_{\text{pos}} = \sum_{t=2}^{T} \sum_{k=1}^{K} -\lambda^{(t)} g(\boldsymbol{\varphi}_k^{(t)}, \boldsymbol{\varphi}_k^{(t-1)}) \quad (2)$$

where a positive pair is the embeddings of Topic#k at slice t and Topic#k at slice t–1. Here following InfoNCE (Van den Oord et al., 2018), $g(\cdot, \cdot)$ measures the similarity between two embeddings, modeled as a scaled cosine function (Wu et al., 2018): $g(a, b) = \cos(a, b)/\tau$ with τ as a temperature hyperparameter (Wu et al., 2022). Eq. (2) pulls these topic embeddings (positive pairs) close to each other in the semantic space. This encourages these topics to cover related semantics, so we can track their evolution with the later topic modeling objective in Sec. 3.5.

Following the common practice (Blei and Lafferty, 2006; Dieng et al., 2019; Miyamoto et al., 2023), **we use hyperparameter $\lambda^{(t)}$ to adaptively adjust evolution intensities between time slices**. If topics evolve slightly between slice $t-1$ and t, we use a large $\lambda^{(t)}$; otherwise a small $\lambda^{(t)}$ if they evolve dramatically (*e.g.,* new events emerge). **We later demonstrate the strong robustness of our model to this hyperparameter in Sec. 4.3**.

Negative Relations among Dynamic Topics The above positive relations only track topic evolution, but cannot prevent topics within a slice from being similar, which leads to repetitive topics as shown in Figure 2 (See empirical results in Sec. 4.2 and 4.4). To address this issue, we further build negative relations among dynamic topics. Specifically, we require a topic to be different from others at each time slice to avoid repetitive topics. Similar to Eq. (2), we model the embeddings of different topics at a time slice as negative pairs and build their negative relations as

$$\mathcal{L}_{\text{neg}} = \gamma \sum_{t=1}^{T} \sum_{k=1}^{K} \log \sum_{k' \neq k} \exp(g(\boldsymbol{\varphi}_k^{(t)}, \boldsymbol{\varphi}_{k'}^{(t)})) \quad (3)$$

where γ is a weight hyperparameter. Eq. (3) pushes these topic embeddings (negative pairs) away from each other in the semantic space, forcing them to cover distinct semantics. Hence topics at each slice become different from each other, enhancing topic diversity and alleviating the repetitive topic issue.

Objective for Evolution-Tracking Contrastive Learning Combining the objectives for positive and negative relations (Eq. (2) and (3)), we formulate the objective for Evolution-Tracking Contrastive learning (ETC) as

$$\mathcal{L}_{\text{ETC}} = \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{neg}}. \quad (4)$$

As illustrated in Figure 2, this objective builds positive and negative relations among dynamic topics, which captures topic evolution and mitigates the repetitive topic issue as well.

3.4 Unassociated Word Exclusion

To address the unassociated topic issue, we further break the tradition of simply chaining topics and propose the Unassociated Word Exclusion (UWE) as illustrated in Figure 4.

What Causes Unassociated Topics? As aforementioned, previous dynamic topic models produce unassociated topics (See examples in Table 1). We believe the reason is that they only chain topics across time slices through Markov chains to track topic evolution (Blei and Lafferty, 2006; Dieng et al., 2019; Miyamoto et al., 2023). This pushes the chained topics at different slices to include similar words. In consequence, topics may be overly related, where a topic could be polluted by the words that do *not* belong to its slice but to other slices. Thus these words are unassociated with its slice, which incurs the unassociated topic issue.

To solve this issue, one may wonder if we can simply weaken the actual required evolution intensity between chained topics. Unfortunately, this is challenging and may fail to track topic evolution, missing the goals of dynamic topic modeling.

Excluding Unassociated Words To this end, we propose the Unassociated Word Exclusion (UWE) method, which refines topic semantics by excluding unassociated words. We first sample unassociated words from the top related words of topics. To be specific, we denote $\mathcal{V}^{(t)}_{\text{top}}$ as the set of top words of all topics at time slice t, which is sampled as

$$\mathcal{V}^{(t)}_{\text{top}} = \bigcup_{k=1}^{K} \text{topK}(\boldsymbol{\beta}^{(t)}_k, N_{\text{top}}). \tag{5}$$

Here we use topK function to sample N_{top} top words from discovered topics according to their topic-word distribution $\boldsymbol{\beta}^{(t)}_k$. Then $\mathcal{V}^{(t)}_{\text{top}}$ is the union of the selected top words from all topics. Later as illustrated in Figure 4, we formulate the set of unassociated words $\mathcal{V}^{(t)}_{\text{UW}}$ as

$$\mathcal{V}^{(t)}_{\text{UW}} = \mathcal{V}^{(t)}_{\text{top}} \setminus \mathcal{V}^{(t)}, \tag{6}$$

where we identify the top words that do *not* exist in $\mathcal{V}^{(t)}$, the vocabulary set of slice t. These words are unassociated since they do *not* belong to slice t, such as the "covid" of Topic#2 in 2017 in Table 1.

Then we aim to exclude the unassociated words from discovered topics. For this purpose, we might directly mask these words in each topic, but unfortunately this simple way cannot fully improve

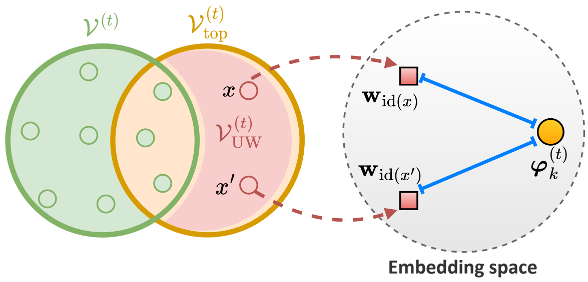


Figure 4: Illustration of Unassociated Word Exclusion (UWE). Unassociated word set $\mathcal{V}^{(t)}_{\text{UW}}$ contains words in the top word set $\mathcal{V}^{(t)}_{\text{top}}$ but *not* in the vocabulary set $\mathcal{V}^{(t)}$. UWE pushes topic embedding $\boldsymbol{\varphi}^{(t)}_k$ away from the embeddings of unassociated words, $\mathbf{w}_{\text{id}(x)}$ and $\mathbf{w}_{\text{id}(x')}$.

dynamic topic quality as it leaves topic semantics unrefined (See ablation studies in Sec. 4.4). Thus to refine topic semantics, we propose to model topic embeddings and the embeddings of unassociated words as negative pairs: $(\boldsymbol{\varphi}^{(t)}_k, \mathbf{w}_{\text{id}(x)})$ where $x \in \mathcal{V}^{(t)}_{\text{UW}}$ and $\text{id}(x)$ returns the index of word x. Similar to Eq. (3), we formulate the objective for UWE with negative pairs as

$$\mathcal{L}_{\text{UWE}} = \sum_{t=1}^{T} \sum_{k=1}^{K} \log \sum_{x \in \mathcal{V}^{(t)}_{\text{UWE}}} \exp(g(\boldsymbol{\varphi}^{(t)}_k, \mathbf{w}_{\text{id}(x)})). \tag{7}$$

It refines topic semantics by pushing topic embeddings away from the embeddings of unassociated words. This excludes unassociated words from topics and hence mitigates the unassociated topic issue. This also keeps topics associated to their time slices even if they are overly related by large evolution intensity hyperparameter $\lambda^{(t)}$, which increases the robustness of our model to the hyperparameter $\lambda^{(t)}$ (See experiment supports in Sec. 4.3).

3.5 Chain-Free Dynamic Topic Model

In this section, we combine the above ETC and UWE methods with the generation of sequential documents to formulate our CFDTM.

Generation of Sequential Documents As illustrated in Figure 3, our generation process follows VAE as Dieng et al. (2019). Specifically for document $\mathbf{x}^{(t,d)}$, we use a latent variable $\mathbf{r}^{(t,d)}$ following a logistic normal prior: $p(\mathbf{r}^{(t,d)}) = \mathcal{LN}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the mean the diagonal covariance matrix. We model its variational distribution as $q_\Theta(\mathbf{r}^{(t,d)}|\mathbf{x}^{(t,d)}) = \mathcal{N}(\boldsymbol{\mu}^{(t,d)}, \boldsymbol{\Sigma}^{(t,d)})$. To model parameters $\boldsymbol{\mu}^{(t,d)}, \boldsymbol{\Sigma}^{(t,d)}$, we adopt a neural network encoder f_Θ parameterized by Θ with the Bag-

Model	NeurIPS		ACL		UN		NYT		WHO	
	TC	TD	TC	TD	TC	TD	TC	TD	TC	TD
DTM	‡0.452	‡0.456	‡0.451	‡0.395	‡0.434	‡0.238	‡0.464	‡0.632	‡0.397	‡0.343
NDTM	‡0.423	‡0.641	‡0.456	‡0.539	‡0.438	‡0.588	‡0.473	‡0.446	‡0.302	‡0.448
NDTM-b	‡0.455	‡0.622	‡0.444	‡0.651	‡0.400	‡0.606	‡0.424	‡0.479	‡0.270	‡0.471
DETM	‡0.424	‡0.324	‡0.422	‡0.279	‡0.361	‡0.130	‡0.462	‡0.368	‡0.351	‡0.349
BERTopic	‡0.427	‡0.432	‡0.428	‡0.429	‡0.329	‡0.203	‡0.412	‡0.597	‡0.227	‡0.292
DSNTM	‡0.427	‡0.685	‡0.370	‡0.609	‡0.376	‡0.593	‡0.374	‡0.414	‡0.385	‡0.479
CFDTM	**0.581**	**0.846**	**0.571**	**0.879**	**0.502**	**0.799**	**0.538**	**0.732**	**0.589**	**0.675**

Table 2: Topic quality results of Topic Coherence (TC) and Topic Diversity (TD). The best are in **bold**. The superscript ‡ means the gains of CFDTM are statistically significant at 0.05 level.

of-Words of $\mathbf{x}^{(t,d)}$ as inputs. Here we reuse this encoder for documents at different time slices for parameter efficiency as Dieng et al. (2019). Through the reparameterization trick of VAE (Kingma and Welling, 2014), we sample $\mathbf{r}^{(t,d)}$ as

$$\mathbf{r}^{(t,d)} = \boldsymbol{\mu}^{(t,d)} + (\boldsymbol{\Sigma}^{(t,d)})^{1/2}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (8)$$

We model the doc-topic distribution via a softmax function as $\boldsymbol{\theta}^{(t,d)} = \mathrm{softmax}(\mathbf{r}^{(t,d)})$. Then we generate words in $\mathbf{x}^{(t,d)}$ by sampling from a multinomial distribution: $x \sim \mathrm{Mult}(\mathrm{softmax}(\boldsymbol{\beta}^{(t)}\boldsymbol{\theta}^{(t,d)}))$ following Srivastava and Sutton (2017). See more details about the generation process in Appendix B.

Topic Modeling Objective Based on the above generation process of sequential documents, we write the topic modeling objective following the ELBO of VAE (Dieng et al., 2019) as

$$\mathcal{L}_{\mathrm{TM}} = \sum_{t=1}^{T}\sum_{d=1}^{N_t} -(\mathbf{x}^{(t,d)})^\top \log(\mathrm{softmax}(\boldsymbol{\beta}^{(t)}\boldsymbol{\theta}^{(t,d)})) + \mathrm{KL}\left[q_\Theta(\mathbf{r}^{(t,d)}|\mathbf{x}^{(t,d)}) \| p(\mathbf{r}^{(t,d)})\right] \quad (9)$$

where we model $\boldsymbol{\beta}^{(t)}$ as Eq. (1). The first term is the reconstruction error between the input and reconstructed document with $\boldsymbol{\beta}^{(t)}$ in Eq. (1). The second term is the KL divergence between the prior and variational distributions.

Overall Objective for CFDTM We formulate the overall objective function for our CFDTM by combining Eq. (4), (7) and (9) as

$$\min_{\Theta, \mathbf{W}, \{\boldsymbol{\varphi}^{(t)}\}_{t=1}^{T}} \mathcal{L}_{\mathrm{TM}} + \mathcal{L}_{\mathrm{ETC}} + \lambda_{\mathrm{UWE}}\mathcal{L}_{\mathrm{UWE}} \quad (10)$$

where λ_{UWE} is a weight hyperparameter of $\mathcal{L}_{\mathrm{UWE}}$. Note the weight parameters of $\mathcal{L}_{\mathrm{ETC}}$ are included in Eq. (4). In sum, $\mathcal{L}_{\mathrm{TM}}$ learns doc-topic distributions and latent topics from sequential documents; $\mathcal{L}_{\mathrm{ETC}}$ tracks topic evolution and avoids repetitive topics by building similarity relations among dynamic topics; $\mathcal{L}_{\mathrm{UWE}}$ mitigates the unassociated topic issue by excluding unassociated words from topics.

4 Experiment

In this section, we conduct comprehensive experiments and show that our model achieves state-of-the-art performance and effectively solves the repetitive topic and unassociated topic issues.

4.1 Experiment Setup

Datasets We experiment with the following benchmark datasets: (i) **NeurIPS** [1] contains papers published between 1987 and 2017 at the NeurIPS conference. (ii) **ACL** (Bird et al., 2008) is an article collection between 1973 and 2006 from ACL Anthology. (iii) **UN** (Baturo et al., 2017) [2] includes the statement transcriptions at the United Nations from 1970 to 2015. (iv) **NYT** [3] contains news articles in the New York Times from 2012 to 2022 with 12 categories, like "Arts", "Business", and "Health". (v) **WHO** (Li et al., 2020) contains articles about non-pharmacological interventions from the World Health Organization, divided by week from January to May 2020. We follow the toolkit TopMost (Wu et al., 2023c) [4] to preprocess these datasets. See also Appendix A for details. Note that NYT and WHO have relatively shorter documents than other datasets.

Baseline Models We consider the following dynamic topic models as baselines: (i) **DTM** (Blei and Lafferty, 2006), a widely-used probabilistic

[1] https://www.kaggle.com/datasets/benhamner/nips-papers

[2] https://www.kaggle.com/unitednations/un-general-debates

[3] https://huggingface.co/datasets/Matthewww/nyt_news

[4] https://github.com/bobxwu/topmost

Model	NeurIPS		ACL		UN		NYT		WHO	
	TC	TD	TC	TD	TC	TD	TC	TD	TC	TD
w/o ETC	‡0.503	‡0.466	‡0.550	‡0.473	‡0.432	‡0.334	‡0.519	‡0.662	‡0.542	‡0.422
w/o negative	‡0.534	‡0.361	‡0.423	‡0.249	‡0.439	‡0.267	‡0.456	‡0.536	‡0.381	‡0.266
w/o UWE	‡0.483	‡0.739	‡0.484	‡0.790	‡0.417	‡0.699	‡0.496	‡0.709	‡0.278	‡0.493
masking	‡0.544	‡0.830	‡0.545	‡0.867	‡0.472	‡0.772	‡0.518	‡0.671	‡0.564	‡0.649
CFDTM	**0.581**	**0.846**	**0.571**	**0.879**	**0.502**	**0.799**	**0.538**	**0.732**	**0.589**	**0.675**

Table 3: Ablation study. **w/o ETC**: without evolution-tracking contrastive learning. **w/o negative**: without building negative relations in ETC. **w/o UWE**: without unassociated word exclusion. **masking**: directly masking unassociated words. The best are in **bold**. The superscript ‡ means the gains of CFDTM are statistically significant at 0.05 level.

dynamic topic model; (ii) **NDTM** (Dieng et al., 2019), a neural model extending DTM with neural variational inference; (iii) **NDTM-b**, a variant of NDTM using our Eq. (1) to model topic-word distributions. We propose this baseline for fair comparisons. (iv) **DETM** (Dieng et al., 2019), a neural dynamic topic model with pre-trained word embeddings. (v) **BERTopic** (Grootendorst, 2022), a clustering-based topic discovery model via document embeddings. (vi) **DSNTM** (Miyamoto et al., 2023), the latest neural model using attention mechanism to model dependencies among dynamic topics. Some studies are inapplicable for comparison since they focus on different problem settings: Zhang and Lauw (2022) deal with temporal document networks instead of sequential documents; Cvejoski et al. (2023) models the activities of static topics instead of the evolution of dynamic topics. (Rahimi et al., 2023) only cluster words to form topics and cannot specify the number of topics or infer doc-topic distributions. We fine-tune the hyperparameters of these baselines like the Gaussian variance of DTM and the min similarity of BERTopic. See implementation details of our model in Appendix B.

4.2 Dynamic Topic Quality

Evaluation Metrics To compare dynamic topic models, we evaluate the quality of topics from two aspects following Dieng et al. (2019): (i) **Topic Coherence** (**TC**), referring to the coherence of top words in a topic. We employ the popular C_V as the coherence metric (Röder et al., 2015), which outperforms early metrics like NPMI (Lau et al., 2014). To measure the association between a topic and its time slice as well, we compute its TC with the documents of its slice as a reference corpus to estimate the word occurrence probabilities. As such, higher TC means this topic is more coherent within the documents of the slice, indicating a stronger association with the slice. Therefore, TC can confirm if the unassociated topic issue happens. (ii) **Topic Diversity** (**TD**), referring to difference between discovered topics (Dieng et al., 2020). For a topic at a slice, we compute the proportion of its top words that only occur once and also exist at that slice. Higher TD indicates topics are more distinct from others, so TD can verify if the repetitive topic issue exists. We take the average TC and TD over all time slices. We set the number of topics (K) as 50 following Dieng et al. (2019) and report the average results of 5 runs.

Result Analysis Table 2 summarizes the topic coherence (TC) and diversity (TD) results on benchmark datasets. We have the following observations: **(i) CFDTM significantly outperforms baselines in terms of TD.** We see baseline models commonly generate repetitive topics as indicated by their low TD scores. For instance, the TD score of CFDTM is 0.879 on ACL whereas the highest baseline only is 0.651. This demonstrates that CFDTM produces more diverse topics than baselines. The improvements result from our evolution-tracking contrastive learning that encourages topics to be distinct from each other. Hence it mitigates the repetitive topic issues while tracking topic evolution. **(ii) CFDTM achieves the best TC scores with significant improvements.** For instance, CFDTM has a TC score of 0.581 on NeurIPS while the runner-up is only 0.455. This implies the topics of CFDTM are more coherent and associated with their time slices than these baselines. The reason lies in that CFDTM adopts our unassociated word exclusion method, which excludes unassociated words from topics. Thus it alleviates the unassociated topic issue and better captures topic evolution. We mention that merely modeling doc-topic distributions as Eq. (1) cannot result in these improvements since NDTM-b is incomparable to CFDTM.

The above results manifest that our CFDTM ef-

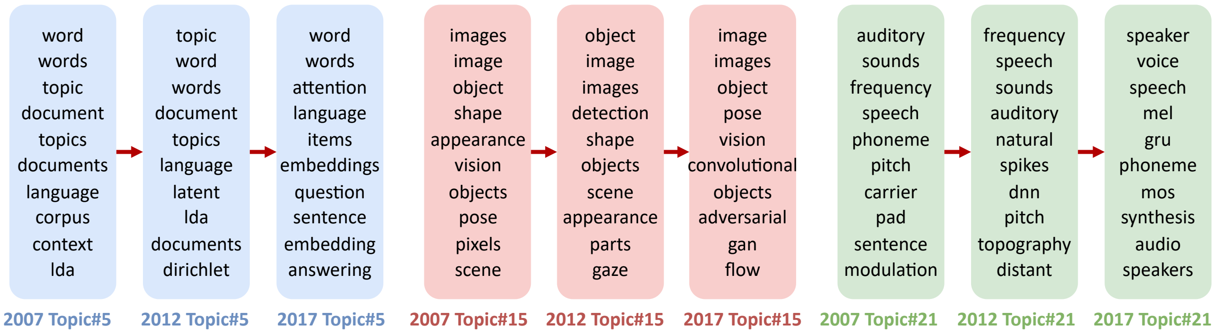


Figure 5: Case study. Top related words of discovered topics in 2007, 2012, and 2017 from the NeurIPS dataset.

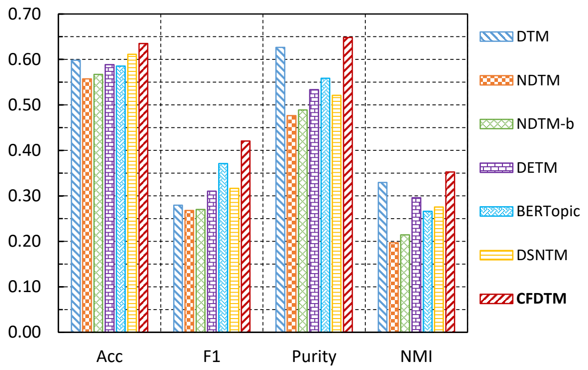


Figure 6: Text classification (Acc and F1) and clustering results (Purity and NMI).

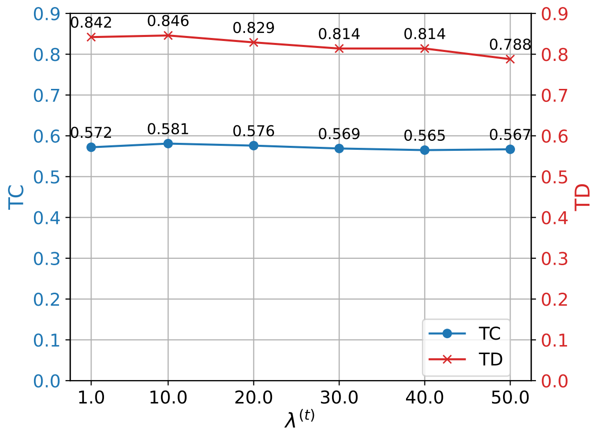


Figure 7: Influence of evolution intensity hyperparameter $\lambda^{(t)}$. TC and TD scores of our CFDTM remain relatively stable along with varying $\lambda^{(t)}$.

fectively tracks topic evolution with more diverse, coherent, and associated topics, which can benefit downstream tasks and applications. **See case studies of word and topic evolution in Sec. 4.6 and appendix D.**

4.3 Robustness to Evolution Intensity Hyperparameter

We demonstrate the robustness of our model to the evolution intensity hyperparameter $\lambda^{(t)}$. Figure 7 reports the topic coherence (TC) and diversity (TD) results with varying $\lambda^{(t)}$. We see that TC and TD scores remain relatively stable with larger $\lambda^{(t)}$. This can be attributed to our UWE method which excludes unassociated words from topics even if they are overly related by large $\lambda^{(t)}$. These validate that our method is robust to the evolution intensity hyperparameter. This advantage is vital as it effectively reduces the efforts to find suitable intensity hyperparameters, overcoming a difficult challenge in previous work (Dieng et al., 2019).

4.4 Ablation Study

We conduct ablation studies to show the necessity of our proposed two methods: Evolution-Tracking Contrastive learning (ETC) and Unassociated Word Exclusion (UWE). From Table 3, we have the following discoveries: **(i) ETC effectively mitigates the repetitive topic issue.** Compared to CFDTM, TD greatly degrades if without ETC (w/o ETC). For example, the TD decreases from 0.846 to 0.466 on NeurIPS. Besides, TD reduces more if building positive relations without negative relations (w/o negative). This implies only building positive relations worsens the repetitive topic issue. **(ii) UWE can alleviate the unassociated topic issue.** Table 3 shows the TC of CFDTM are significantly better than the case without UWE (w/o UWE), *e.g.,* 0.581 vs. 0.483 on NeurIPS. More importantly, CFDTM reaches higher TC than directly masking unassociated words (w/ masking). This is because direct masking cannot leverage these words to refine topic semantics. In summary, these results demonstrate the necessity of our ETC and UWE for improving dynamic topic modeling performance.

4.5 Text Classification and Clustering

Apart from dynamic topic quality, we compare the quality of learned doc-topic distributions through downstream tasks: text classification and clustering. In detail, we leverage the document categories of the NYT dataset for evaluation. For text classification, we train SVM classifiers with learned doc-topic distributions as features and predict the categories of testing documents following Wu et al. (2023b). This performance is evaluated by Accuracy and F1. For text clustering, we use the most significant topics in the doc-topic distributions as clustering assignments and evaluate this performance by widely-used metrics Purity and NMI following Wu et al. (2023b). Note that our purpose is not to achieve state-of-the-art classification or clustering performance but to compare the quality of doc-topic distributions.

Figure 6 shows that our CFDTM consistently surpasses all baselines in terms of both classification and clustering performance. We mention that the improvements of CFDTM over baselines are statistically significant at 0.05 level. These results manifest that our model generates more accurate doc-topic distributions for downstream classification and clustering tasks.

4.6 Case Study: Evolution of Topics

Furthermore, we report case studies to show our model captures the evolution of topics. Figure 5 illustrate some topics discovered by our model from NeurIPS, which evolve from the year 2007 to 2017. We see Topic#5 tracks the trend of word embeddings and attention mechanism in the NLP field. In addition, Topic#15 recognizes the popularity of convolutional neural networks and adversarial training in computer vision, indicated by the appearances of the words “convolutional”, “adversarial”, and “gan”. Topic#21 focuses on the evolution of speech processing, which also captures the application of neural networks as implied by the words “dnn” and “gru”. See full topic lists of different models in Appendix E.

5 Conclusion

In this paper, we propose CFDTM, a new chain-free neural dynamic topic model. To break the tradition of simply chaining topics in previous work, CFDTM employs the novel evolution-tracking contrastive learning and unassociated word exclusion methods. Experiments demonstrate that our CFDTM consistently outperforms baselines and effectively mitigates the repetitive topic and unassociated topic issues. CFDTM tracks topic evolution with higher coherence and diversity, and achieves better performance on downstream tasks. Our model also shows robustness to the hyperparameter for evolution intensities.

Limitations

Our proposed method has achieved promising improvements by mitigating the repetitive topic and unassociated topic issues in dynamic topic modeling, but we consider the following limitations as future work:

- **Extend our model to discover dynamic topics from multilingual corpora.** Numerous multilingual corpora are prevalent in the current Internet (Zosa and Granroth-Wilding, 2019). It would inspire more applications to discover and compare historical trends from different cultures and languages by extending our method to the multilingual setting.
- **Leverage Large Language Models (LLMs) for dynamic topic modeling.** Popular LLMs retain rich knowledge from pretraining corpora (Petroni et al., 2019; Pan et al., 2023; Wu et al., 2024d), but they cannot directly discover topics from a large dataset. Alternatively, we may leverage the prior knowledge in LLMs to further improve the performance of dynamic topic modeling.

Acknowledgements

This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-TC-2022-005).

References

Nishant Balepur, Shivam Agarwal, Karthik Venkat Ramanan, Susik Yoon, Diyi Yang, and Jiawei Han. 2023. Dynamite: Discovering explosive topic evolutions with user guidance. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 194–217.

Alexander Baturo, Niheer Dasandi, and Slava J Mikhaylov. 2017. Understanding state preferences with text as data: Introducing the un general debate corpus. *Research & Politics*, 4(2):2053168017712821.

Arnab Bhadury, Jianfei Chen, Jun Zhu, and Shixia Liu. 2016. Scaling up dynamic topic models. In *Proceedings of the 25th International Conference on World Wide Web*, pages 381–390.

Steven Bird, Robert Dale, Bonnie J Dorr, Bryan R Gibson, Mark Thomas Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir R Radev, Yee Fan Tan, et al. 2008. The acl anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. In *LREC*.

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. 2017. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877.

David M Blei and John D Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

François Caron, Manuel Davy, and Arnaud Doucet. 2012. Generalized polya urn for time-varying dirichlet process mixtures. *arXiv preprint arXiv:1206.5254*.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR.

Rob Churchill and Lisa Singh. 2022. Dynamic topic-noise models for social media. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 429–443. Springer.

Kostadin Cvejoski, Ramsés J Sánchez, and César Ojeda. 2023. Neural dynamic focused topic model. *arXiv preprint arXiv:2301.10988*.

Adji B Dieng, Francisco JR Ruiz, and David M Blei. 2019. The dynamic embedded topic model. *arXiv preprint arXiv:1907.05545*.

Adji B Dieng, Francisco JR Ruiz, and David M Blei. 2020. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453.

Nicholas Frosst, Nicolas Papernot, and Geoffrey Hinton. 2019. Analyzing and improving representations with the soft nearest neighbor loss. In *International conference on machine learning*, pages 2012–2020. PMLR.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.

Derek Greene and James P Cross. 2017. Exploring the political agenda of the european parliament using a dynamic topic modeling approach. *Political Analysis*, 25(1):77–94.

Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1):5228–5235.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE.

Sungwon Han, Mingi Shin, Sungkyu Park, Changwook Jung, and Meeyoung Cha. 2023. Unified neural topic model via contrastive learning and term weighting. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1794–1809.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738.

Philipp Hennig, David Stern, Ralf Herbrich, and Thore Graepel. 2012. Kernel topic models. In *Artificial Intelligence and Statistics*, pages 511–519.

Rem Hida, Naoya Takeishi, Takehisa Yairi, and Koichi Hori. 2018. Dynamic and static topic model for analyzing time-series document collections. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 516–520.

Jiajun Hu, Xiaobing Sun, David Lo, and Bin Li. 2015. Modeling the evolution of development topics using dynamic topic models. In *2015 IEEE 22nd international conference on software analysis, evolution, and reengineering (SANER)*, pages 3–12. IEEE.

Tomoharu Iwata, Takeshi Yamada, Yasushi Sakurai, and Naonori Ueda. 2010. Online multiscale dynamic topic models. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 663–672.

Patrick Jähnichen, Florian Wenzel, Marius Kloft, and Stephan Mandt. 2018. Scalable generalized dynamic topic models. In *International Conference on Artificial Intelligence and Statistics*, pages 1427–1435. PMLR.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *The International Conference on Learning Representations (ICLR)*.

Jey Han Lau, David Newman, and Timothy Baldwin. 2014. 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*, pages 530–539.

Ke Li, Hubert Naacke, and Bernd Amann. 2021. An analytic graph data model and query language for exploring the evolution of science. *Big Data Research*, 26:100247.

Yue Li, Pratheeksha Nair, Zhi Wen, Imane Chafi, Anya Okhmatovskaia, Guido Powell, Yannan Shen, and David Buckeridge. 2020. Global surveillance of covid-19 by mining news media using a multi-source dynamic embedded topic model. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–14.

Chaoqun Liu, Wenxuan Zhang, Guizhen Chen, Xiaobao Wu, Anh Tuan Luu, Chip Hong Chang, and Lidong Bing. 2023. Zero-shot text classification via self-supervised tuning. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1743–1761.

Rui Mao, Kelvin Du, Yu Ma, Luyao Zhu, and Erik Cambria. 2023. Discovering the cognition behind language: Financial metaphor analysis with MetaPro. In *2023 IEEE International Conference on Data Mining (ICDM)*, pages 1211–1216, Shanghai, China. IEEE.

Rui Mao, Qika Lin, Qiawen Liu, Gianmarco Mengaldo, and Erik Cambria. 2024a. Understanding public perception towards weather disasters through the lens of metaphor. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, Jeju, South Korea.

Rui Mao, Tianwei Zhang, Qian Liu, Amir Hussain, and Erik Cambria. 2024b. Unveiling diplomatic narratives: Analyzing United Nations Security Council debates through metaphorical cognition. In *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)*, Rotterdam, the Netherlands.

Yishu Miao, Edward Grefenstette, and Phil Blunsom. 2017. Discovering discrete latent topics with neural variational inference. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2410–2419. JMLR. org.

Yishu Miao, Lei Yu, and Phil Blunsom. 2016. Neural variational inference for text processing. In *International conference on machine learning*, pages 1727–1736.

Nozomu Miyamoto, Masaru Isonuma, Sho Takase, Junichiro Mori, and Ichiro Sakata. 2023. Dynamic structured neural topic model with self-attention mechanism. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5916–5930.

Riki Murakami, Basabi Chakraborty, and Yukari Shirota. 2021. Dynamic topic tracking and visualization using covid-19 related tweets in multiple languages. In *2021 International Conference on Artificial Intelligence and Big Data Analytics*, pages 16–21. IEEE.

Cong-Duy Nguyen, Thong Nguyen, Xiaobao Wu, and Anh Tuan Luu. 2024a. Kdmcse: Knowledge distillation multimodal sentence embeddings with adaptive angular margin contrastive learning. *arXiv preprint arXiv:2403.17486*.

Thong Nguyen and Anh Tuan Luu. 2021. Contrastive learning for neural topic model. *Advances in Neural Information Processing Systems*, 34.

Thong Nguyen, Xiaobao Wu, Anh-Tuan Luu, Cong-Duy Nguyen, Zhen Hai, and Lidong Bing. 2022. Adaptive contrastive learning on multimodal transformer for review helpfulness predictions. *arXiv preprint arXiv:2211.03524*.

Thong Thanh Nguyen, Xiaobao Wu, Xinshuai Dong, Cong-Duy T Nguyen, Zhen Hai, Lidong Bing, and Anh Tuan Luu. 2023. Gradient-boosted decision tree for listwise context model in multimodal review helpfulness prediction. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.

Thong Thanh Nguyen, Xiaobao Wu, Xinshuai Dong, Cong-Duy T Nguyen, See-Kiong Ng, and Anh Tuan Luu. 2024b. Topic modeling as multi-objective optimization with setwise contrastive learning. In *The Twelfth International Conference on Learning Representations*.

Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. 2016. Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4004–4012.

Liangming Pan, Xiaobao Wu, Xinyuan Lu, Anh Tuan Luu, William Yang Wang, Min-Yen Kan, and Preslav Nakov. 2023. Fact-checking complex claims with program-guided reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6981–7004.

Nanyun Peng, Yiming Wang, and Mark Dredze. 2014. Learning polylingual topic models from code-switched social media documents. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 674–679.

Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*.

Hamed Rahimi, Hubert Naacke, Camelia Constantin, and Bernd Amann. 2023. Antm: An aligned neural topic model for exploring evolving topics. *arXiv preprint arXiv:2302.01501*.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. 2014. Stochastic backpropagation and approximate inference in deep generative models. *In Proceedings ofthe 31th International Conference on Machine Learning*.

Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408. ACM.

Hao Sha, Mohammad Al Hasan, George Mohler, and P Jeffrey Brantingham. 2020. Dynamic topic modeling of the covid-19 twitter narrative among us governors and cabinet executives. *arXiv preprint arXiv:2004.11692*.

Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. In *ICLR*.

Aaron Van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv e-prints*, pages arXiv–1807.

Boyu Wang, Linhai Zhang, Deyu Zhou, Yi Cao, and Jiandong Ding. 2023. Neural topic modeling based on cycle adversarial training and contrastive learning. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9720–9731.

Chong Wang, David Blei, and David Heckerman. 2008. Continuous time dynamic topic models. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 579–586.

Xuerui Wang and Andrew McCallum. 2006. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 424–433.

Xiaobao Wu, Xinshuai Dong, Thong Nguyen, Chaoqun Liu, Liang-Ming Pan, and Anh Tuan Luu. 2023a. Infoctm: A mutual information maximization perspective of cross-lingual topic modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13763–13771.

Xiaobao Wu, Xinshuai Dong, Thong Nguyen, and Anh Tuan Luu. 2023b. Effective neural topic modeling with embedding clustering regularization. In *International Conference on Machine Learning*. PMLR.

Xiaobao Wu and Chunping Li. 2019. Short Text Topic Modeling with Flexible Word Patterns. In *International Joint Conference on Neural Networks*.

Xiaobao Wu, Chunping Li, and Yishu Miao. 2021. Discovering topics in long-tailed corpora with causal intervention. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 175–185, Online. Association for Computational Linguistics.

Xiaobao Wu, Chunping Li, Yan Zhu, and Yishu Miao. 2020a. Learning Multilingual Topics with Neural Variational Inference. In *International Conference on Natural Language Processing and Chinese Computing*.

Xiaobao Wu, Chunping Li, Yan Zhu, and Yishu Miao. 2020b. 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 (EMNLP)*, pages 1772–1782, Online.

Xiaobao Wu, Anh Tuan Luu, and Xinshuai Dong. 2022. Mitigating data sparsity for short text topic modeling by topic-semantic contrastive learning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2748–2760, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Xiaobao Wu, Thong Nguyen, and Anh Tuan Luu. 2024a. A survey on neural topic models: Methods, applications, and challenges. *Artificial Intelligence Review*.

Xiaobao Wu, Thong Nguyen, Delvin Ce Zhang, William Yang Wang, and Anh Tuan Luu. 2024b. Fastopic: A fast, adaptive, stable, and transferable topic modeling paradigm. *arXiv preprint*.

Xiaobao Wu, Fengjun Pan, and Anh Tuan Luu. 2023c. Towards the topmost: A topic modeling system toolkit. *arXiv preprint arXiv:2309.06908*.

Xiaobao Wu, Fengjun Pan, Thong Nguyen, Yichao Feng, Chaoqun Liu, Cong-Duy Nguyen, and Anh Tuan Luu. 2024c. On the affinity, rationality, and diversity of hierarchical topic modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Xiaobao Wu, Liangming Pan, William Yang Wang, and Anh Tuan Luu. 2024d. Updating language models with unstructured facts: Towards practical knowledge editing. *arXiv preprint arXiv:2402.18909*.

Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. 2018. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742.

Enze Xie, Jian Ding, Wenhai Wang, Xiaohang Zhan, Hang Xu, Peize Sun, Zhenguo Li, and Ping Luo. 2021. Detco: Unsupervised contrastive learning for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8392–8401.

Delvin Ce Zhang and Hady Lauw. 2022. Dynamic topic models for temporal document networks. In *International Conference on Machine Learning*, pages 26281–26292. PMLR.

Xiangyun Zhao, Raviteja Vemulapalli, Philip Andrew Mansfield, Boqing Gong, Bradley Green, Lior Shapira, and Ying Wu. 2021. Contrastive learning for label efficient semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10623–10633.

Xixi Zhou, Jiajun Bu, Sheng Zhou, Zhi Yu, Ji Zhao, and Xifeng Yan. 2023. Improving topic disentanglement via contrastive learning. *Information Processing & Management*, 60(2):103164.

Elaine Zosa and Mark Granroth-Wilding. 2019. Multilingual dynamic topic model. *RANLP 2019-Natural Language Processing a Deep Learning World*.

Model	K=60		K=80		K=100	
	TC	TD	TC	TD	TC	TD
DTM	‡0.444	‡0.461	‡0.439	‡0.436	‡0.436	‡0.418
NDTM	‡0.427	‡0.601	‡0.428	‡0.546	‡0.429	‡0.500
NDTM-b	‡0.443	‡0.598	‡0.421	‡0.534	‡0.433	‡0.377
DETM	‡0.409	‡0.284	‡0.386	‡0.218	‡0.377	‡0.192
BERTopic	‡0.436	‡0.406	‡0.438	‡0.356	‡0.438	‡0.328
DSNTM	‡0.429	‡0.656	‡0.427	‡0.600	‡0.430	‡0.556
CFDTM	**0.591**	**0.817**	**0.605**	**0.740**	**0.575**	**0.648**

Table 4: Topic quality results of coherence (TC) and diversity (TD) under topic number $K = 60, 80, 100$. The best scores are in **bold**. The superscript ‡ means the gains of CFDTM are statistically significant at 0.05 level.

Dataset	#doc	Average Length	Vocabulary Size	#time slices
NeurIPS	7,237	2,085.9	10,000	31
ACL	10,560	2,023.0	10,000	31
UN	7,507	1,421.6	10,000	46
NYT	9,172	175.4	10,000	11
WHO	12,145	41.3	10,000	15

Table 5: Statistics of datasets.

A Pre-processing Datasets

To pre-process datasets, we conduct the following the steps of Wu et al. (2023c) [5]: (i) tokenize documents and convert to lowercase; (ii) remove punctuation; (iii) remove tokens that include numbers; (iv) remove tokens less than 3 characters; (v) remove stopwords.

Table 5 summarizes the statistics of datasets after pre-processing.

B Model Implementation

Generation of Sequential Documents As mentioned in Sec. 3.5, the generation process of sequential documents in our model is in the framework of VAE (Kingma and Ba, 2014; Rezende et al., 2014). Following Srivastava and Sutton (2017); Wu et al. (2022, 2023b), the prior distribution is modeled with Laplace approximation (Hennig et al., 2012) to approximate a symmetric Dirichlet prior as $\mu_{0,k} = 0$ and $\Sigma_{0,kk} = (K-1)/K$. The encoder network f_Θ is a two-layer MLP with softplus as the activation function, concatenated with two single layers each for the mean and covariance matrix. The encoder takes the Bag-of-Words of documents as inputs and outputs the mean and covariance matrix parameters, $\boldsymbol{\mu}^{(t,d)}$ and $\boldsymbol{\Sigma}^{(t,d)}$. We reuse the encoder for all time slices to save model parameters as Dieng et al. (2019). The decoder generates documents with doc-topic distribution $\boldsymbol{\theta}^{(t,d)}$ and the topic-word distribution matrix of the corresponding time slice $\boldsymbol{\beta}^{(t)}$. We use pre-trained 200-dimensional GloVe (Peng et al., 2014) to initialize the word embeddings $\mathbf{W}$ in Eq. (1).

[5] https://github.com/bobxwu/topmost

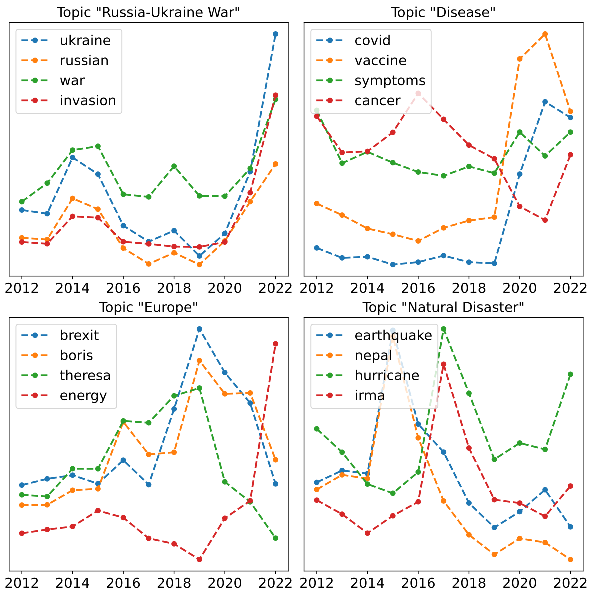


Figure 8: Case study. Evolution of word probability. X-axis denotes years; Y-axis denotes word probabilities.

Hyperparameter Selection We set π in Eq. (1) as 1.0, N_{top} in Eq. (5) as 15, and τ in Eq. (2), (3) and (7) as 0.1. We use Adam (Kingma and Ba, 2014) to optimize model parameters and run our model for 800 epochs with a learning rate as 0.002.

See our code for more implementation details.

C Robustness to the Number of Topics

Apart from the aforementioned results under $K = 50$ in Table 2, we experiment with different numbers of topics on NeurIPS to verify the robustness of our model. Table 4 reports the performance under $K = 60, 80, 100$. We observe that our CFDTM outperforms all baseline methods concerning both topic coherence and diversity. These results validate the robustness of our model to the number of topics.

D Case Study: Evolution of Word Probabilities

We conduct case studies to show our model captures the evolution of word probabilities of topics. Following Dieng et al. (2019), Figure 8 plots the word probability evolution in the topics (Eq. (1)) discovered by our CFDTM from the NYT dataset. In detail, we see that our model successfully captures the Crimea Crisis in 2014 and the Russia-Ukraine War in 2022 as shown by the probability increments of the words "war" and "invasion". For the topic "Disease", our approach discerns the explosion of Covid-19 in 2020, indicated by the increase of "covid" and "vaccine". We notice people focus on cancer before Covid-19. In addition, our model discovers Brexit around 2019 and the energy crisis after 2020 under the topic "Europe". Under the topic "Natural Disaster", our model reveals the Nepal earthquake in 2015 and Hurricane Irma in 2017 as implied by the word probabilities.

E Full Topic Lists

Here are the discovered topics of different models in 2017 of NeurIPS (the latest time slice).
DETM

Topic#1: data set problem learning algorithm number based using results one

Topic#2: networks neural network deep layer weights learning convolutional arxiv training

Topic#3: model data using figure learning approach one based used set

Topic#4: layer convolutional network layers image model cnn input feature networks

Topic#5: image object images segmentation detection flow ground network vision pose

Topic#6: model data figure one set based using used learning two

Topic#7: matrix rank matrices tensor low algorithm norm entries singular spectral

Topic#8: gan adversarial generator discriminator image images generative gans training samples

Topic#9: user items users item model revenue data recommendation ranking information

Topic#10: model learning set number function two one first using linear

Topic#11: learning data set problem two model also one algorithm systems

Topic#12: policy state learning function value reward optimal action algorithm probability

Topic#13: wasserstein transportation function model pipeline data earth nov coming one

Topic#14: neural brain time data information activity neurons model fig spike

Topic#15: model using set one algorithm learning results number based two

Topic#16: algorithm submodular set approximation function functions log theorem privacy problem

Topic#17: clustering algorithm cluster clusters points means algorithms graph let number

Topic#18: recurrent lstm memory rnn neural sequence model attention arxiv models

Topic#19: gaussian posterior inference bayesian model prior distribution log variational likelihood

Topic#20: log estimator theorem error estimation bound analysis probability results rate

Topic#21: model algorithm number data learning set time one given using

Topic#22: one set figure model learning function using given two data

Topic#23: using data method learning model function based one methods set

Topic#24: model data models group structure methods latent variables using analysis

Topic#25: kernel space kernels random function functions features approximation data learning

Topic#26: networks training deep neural layer network dropout layers learning arxiv

Topic#27: learning task tasks data transfer model multi performance features different

Topic#28: causal variables learning set data treatment discrimination fair model two

Topic#29: features feature classification accuracy prediction learning cost set training datasets

Topic#30: inference variational latent log models model distribution variables generative likelihood

Topic#31: loss learning risk function algorithm bounds bound generalization empirical regression

Topic#32: algorithm data time distributed algorithms parallel computation number machine computing

Topic#33: model training learning neural models task language arxiv output sequence

Topic#34: policy agent learning agents reward reinforcement state model policies environment

Topic#35: set problem algorithm one using data number method first function

Topic#36: communication workers worker protocol asynchronous decentralized server message attack attacks

Topic#37: regret algorithm bandit online bound arm problem log bandits optimal

Topic#38: distribution random samples distributions sample estimation probability density model models

Topic#39: algorithm number problem set time one data algorithms given using

Topic#40: embedding embeddings neural word vectors words representations vector text learning

Topic#41: label learning labels data classifier class classification domain training labeled

Topic#42: graph node nodes graphs network networks tree edge edges set

Topic#43: gradient optimization stochastic methods sampling batch variance function method step

Topic#44: set one function two number size given also based used

Topic#45: convex optimization algorithm gradient convergence descent stochastic problem method problems

Topic#46: michel succinctly guo nielsen fulfilled jonas misspecified herein plethora gross

Topic#47: attention visual model question image object prediction video dataset models

Topic#48: time state process dynamics model transition states systems processes system

Topic#49: distance metric data hash points function embedding nearest point hashing

Topic#50: sparse linear solution problem convex lasso min local condition optimization

DSNTM

Topic#1: functions sessions repeated function constant potential behavior questions performances programming

Topic#2: used modified use using compact also algorithm encoding commonly improve

Topic#3: characters divide google ideas actors intersection texts differences divided text

Topic#4: functionals cns hlt torr multivariate marcel recursion lyapunov ell rna

Topic#5: networks programming network data telephone broadcast programs settings sites yahoo

Topic#6: capital sources statistical largest found popular large smaller bulk widely

Topic#7: multiplier optimization sensor occupancy damping compression kernel sigmoid disk segmentation

Topic#8: structures highest level status living class skills million built cognitive

Topic#9: absolute theory model practical dimension scope social political serious party

Topic#10: recovery georgia ari gallant sustained complicate undergoes welch efforts fluid

Topic#11: chart jia flip degree elevation deng illinois compression unsuccessful stacked

Topic#12: bracket completely category lasts cells status attains mediated tile flattened

Topic#13: features version music compositions versions feature experimental contains programming generic

Topic#14: algorithm algorithms detection adaptive inference recursive automatically mathematically simplest stimuli

Topic#15: performance performances inadequate objects evidence effectiveness immune abilities insufficient adequate

Topic#16: learning skills abilities hebbian tricks learn opportunities teach positioning methodologies

Topic#17: experience benefits form say law hard david says bill deal

Topic#18: period represented since december february proceedings october appearance january basis

Topic#19: carnegie summary programming grove news larry computing annual java ibm

Topic#20: data element contains strings sheet fiber graph component funding dramatic

Topic#21: high layers air volume science layer royal highest scientists filled

Topic#22: enhancement spontaneous associated segmentation pat approximations strategies substitutions term smooth

Topic#23: two several three kitchen including like one number little small

Topic#24: use appropriate equipment fairly manner fire clean buy measures necessary

Topic#25: information document details text provide matrix data clarity management false

Topic#26: algorithm gaussian algorithms mapped theme reality sequence structure format linear

Topic#27: isometry modulo vanilla multiplicative polynomial descending hamiltonian simoncelli leftmost diag

Topic#28: accuracy gauge align values coordinate footprint style size dimensions measurements

Topic#29: method convergence divergence quantum notation clustering stochastic sampling machines macroscopic

Topic#30: functions complementary optical units algorithms sharing software points systems discrete

Topic#31: optimization optimal variables analyze constraint algorithms behaviors dynamics heuristic equilibrium

Topic#32: physically images finite independent magnetic image web endpoints arrays site

Topic#33: algorithm algorithms estimator parameterized bayesian decoding viterbi encoding optimization asymptotically

Topic#34: satisfying richness refines nicely nervous formalizes carbonell diverse musical automating

Topic#35: unused pool property layer permanent reuse thumb unwanted potential plastic

Topic#36: subjective probability judgment methods race decisions spatial formula projects develop

Topic#37: hope believe impossible solution solve find cause still say important

Topic#38: elimination kyoto processing division bases seed reductions eliminating eliminated regulatory

Topic#39: residuals theoretical modeling solving mathematical empirical length numerical ratio approximate

Topic#40: stock character updates draw bid general force role jobs meeting

Topic#41: models problems term problem production forced companies company many names

Topic#42: model changed change changes models design almost different prediction new

Topic#43: treat want make able good throw think get achieve accurate

Topic#44: millions estimated families men knew compensation alarm discrimination execution magnitude

Topic#45: videos dirichlet eigenvectors oscillatory mathieu unbounded nicol distributions magnitudes sigmoid

Topic#46: affine maze heaviside computational singularities tensor logarithms gradients algebraic pdf

Topic#47: svd kernels decomposition bursting learner maximization auditory sift ocr predicate

Topic#48: function appropriate input reserve determine criteria parameters ensure information healthy

Topic#49: process application psychology community applications management overview language systems implementation

Topic#50: basic code tuning range base assembly comprises community consists ranges

CFDTM

Topic#1: policy reward agent reinforcement state policies value action actions planning

Topic#2: loss bounds bound complexity risk hypothesis online generalization upper empirical

Topic#3: sup lim inf strategy strategies lugosi ergodic feasible experts objectives

Topic#4: parameterizations munos convolved gaussianity toy nonasymptotic berkeley costly vedaldi partitioned

Topic#5: word words attention language items embeddings question sentence embedding answering

Topic#6: constraints springerverlag constraint entropy resulting mrf assignment configurations maximum segment

Topic#7: group interactions topic structure gibbs topics tree additive groups trees

Topic#8: hlog majorization citeseer lowrank afosr krizhevsky convnet salakhutdinov slices subnetwork

Topic#9: clustering clusters cluster algorithm queries algorithms query number means cost

Topic#10: overcomplete rmse osindero balcan lowdimensional rbms dbns kkt yannlecuncom exdb

Topic#11: variational posterior inference likelihood distribution latent bayesian sampling generative samples

Topic#12: fukumizu guestrin rkhs multiarmed symp steyvers dkl supx noising dbn

Topic#13: identifiability convolving bigram theart adaptivity rescale multiresolution scholkopf interpolate rbms

Topic#14: layer networks layers deep neural network convolutional activations weights batch

Topic#15: image images object pose vision convolutional objects adversarial gan flow

Topic#16: rule rules realization group music abstraction compositional raw lasso path

Topic#17: isomap roweis highdimensional parameterize whye lies saul regarded name put

Topic#18: libsvm approximator bfgs wwwcsientuedutw cjlin minf reparameterization olkopf overfits mdps

Topic#19: decorrelated denoise denoised indexhtml gehler tresp twodimensional centroid tsochantaridis loosely

Topic#20: lasso selection regularized screening regularization logistic argmin biometrika sparsity regression

Topic#21: speaker voice speech mel gru phoneme mos synthesis audio speakers

Topic#22: convergence games equilibrium nash equilibria game gan player descent divergence

Topic#23: kernel kernels space points invariant metric fourier quadrature embedding gaussian

Topic#24: components correlations mit component mixtures solid ica knn eigenvectors obtained

Topic#25: regret arm bandit arms bound reward bandits optimal rewards expected

Topic#26: estimator estimation density sample estimators statistics distributions statistical iid high

Topic#27: stimulus sensory information memory stimuli recall strength choice retrieval items

Topic#28: optimization convex gradient convergence descent stochastic iteration problems methods accelerated

Topic#29: warmuth vempala stochasticity littlestone servedio adversarially mehryar olkopf schapire yishay

Topic#30: andez nonstationary symmetrized comput zabih universit parametrize tardos auai submodularity

Topic#31: revenue discrimination resolving proxy price causal prices market race unresolved

Topic#32: transductive abbeel pomdp discretize maxj mdps discriminatively boykov tractably timesteps

Topic#33: tsitsiklis ihler componentwise hassibi parallelizable labelings isit willsky limt little

Topic#34: classification task label classifier learning tasks labels class dataset domain

Topic#35: pomdps maxk jmlr liblinear herbrich graepel tsochantaridis joachims vayatis puterman

Topic#36: tishby chervonenkis sompolinsky ijcai markram initialisation datapoints vapnik thresholded factorizes

Topic#37: rmax svms cristianini vectorized semisupervised tresp adaboost cholesky linearize regularizes

Topic#38: itti conll treebank compensate leading pontil alter great enhances enforced

Topic#39: matrix rank matrices low entries tensor norm singular completion sparse

Topic#40: unnormalized felzenszwalb sudderth buhmann ijcv perona groundtruth bmvc sublinear resampled

Topic#41: neurosci neurophysiol movshon spatio glms paninski schraudolph preserving exponentiated scholkopf

Topic#42: brain activity responses fig neurons neuroscience connectivity spike spikes neuron

Topic#43: invariances olshausen bergstra pretraining resample vedaldi loglikelihood convolved subgradients rescale

Topic#44: learnability hyperparameter heckerman timit universit borgwardt multiagent rithm inpainting downward

Topic#45: shadlen heskes knn xti axt devroye yedidia orfi kkt enyi

Topic#46: srebro elementwise regularizers ritov xxt nonnegativity juditsky negahban omnipress tsybakov

Topic#47: initialised probabilit vldb montr kernelized independencies auai delta things kinds

Topic#48: velocity described motion direction retina appropriate orientation system position cells

Topic#49: graph nodes node graphs edges edge vertices influence vertex set

Topic#50: sparse linear solution problem convex lasso min local condition optimization