Towards the TOPMOST: A Topic Modeling System Toolkit

Xiaobao Wu   **Fengjun Pan**   **Anh Tuan Luu**

Nanyang Technological University

xiaobao002@e.ntu.edu.sg   fengjun001@e.ntu.edu.sg   anhtuan.luu@ntu.edu.sg

Abstract

Topic models have a rich history with various applications and have recently been reinvigorated by neural topic modeling. However, these numerous topic models adopt totally distinct datasets, implementations, and evaluations. This impedes quick utilization and fair comparisons, and thereby hinders their research progress and applications. To tackle this challenge, we in this paper propose a Topic Modeling System Toolkit (TOPMOST). Compared to existing toolkits, TOPMOST stands out by supporting more extensive features. It covers a broader spectrum of topic modeling scenarios with their complete lifecycles, including datasets, preprocessing, models, training, and evaluations. Thanks to its highly cohesive and decoupled modular design, TOP-MOST enables rapid utilization, fair comparisons, and flexible extensions of diverse cutting-edge topic models. These improvements position TOPMOST as a valuable resource to accelerate the research and applications of topic models. Our code, tutorials, and documentation are available at https://github.com/bobxwu/topmost. Our demo video is at https://youtu.be/9bN-rs4Gu3E?si=LunquCRhBZwyd1Xg.

1 Introduction

Topic models have been a fundamental and prevalent research area for decades. They aim to understand documents in an unsupervised fashion by discovering latent topics from them and inferring their topic distributions (Churchill and Singh, 2022b). Besides the basic topic modeling scenario (Blei et al., 2003), various other scenarios have been explored, *e.g.,* hierarchical, dynamic, and cross-lingual topic modeling (Griffiths et al., 2003; Blei and Lafferty, 2006; Mimno et al., 2009). Current topic models can be categorized into two types. The first type is conventional topic models which follow either non-negative matrix factorization (Lee and Seung, 2000; Kim et al., 2015; Shi

Topic Modeling Scenario		OCTIS	TOPMOST
Basic topic modeling	Datasets	✓	✓
	Models	✓	✓
	Evaluations	✓	✓
Hierarchical topic modeling	Datasets	✓	✓
	Models	✓	✓
	Evaluations	✗	✓
Dynamic topic modeling	Datasets	✗	✓
	Models	✗	✓
	Evaluations	✗	✓
Cross-lingual topic modeling	Datasets	✗	✓
	Models	✗	✓
	Evaluations	✗	✓

Table 1: Comparison between the latest OCTIS (Terragni et al., 2021) and TOPMOST. Our TOPMOST covers more topic modeling scenarios and their corresponding datasets, models, and evaluations.

et al., 2018) or probabilistic graphical models via Markov Chain Monte Carlo (Steyvers and Griffiths, 2007) or Variational Inference (Blei et al., 2017). The second type is recently popular neural topic models, learned through gradient back-propagation (Zhao et al., 2021a; Wu et al., 2024b). Thus they can avoid the laborious model-specific derivations of conventional models, attracting more research attention. Due to the effectiveness and interpretability of topic models, they have inspired various downstream tasks and applications, *e.g.,* text analysis and content recommendation (Boyd-Graber et al., 2017). Despite these significant achievements, quick utilization and fair comparisons of various topic models remain a formidable challenge. The reason lies in their unsystematic model implementations as well as inconsistent dataset and evaluation settings across papers, even within a paper (Hoyle et al., 2021).

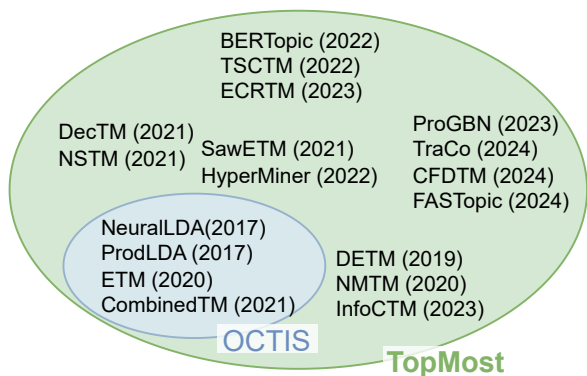Several topic modeling toolkits emerge in response to this challenge by integrating different

Figure 1: Comparison of neural topic models in OCTIS and our TOPMOST. Our TOPMOST covers more latest neural topic models than OCTIS.
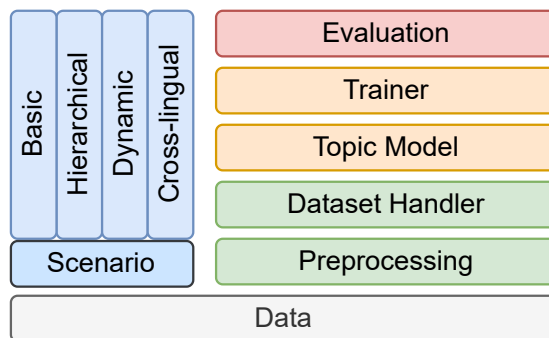

Figure 2: Overall architecture of TOPMOST. It covers the most common topic modeling scenarios and decouples data loading, model constructions, model training and evaluations in topic modeling lifecycles.

topic models and evaluations. However, they fail to fully meet practical requirements due to lacking certain essential features. Early toolkits (McCallum, 2002; Rehurek and Sojka, 2011; Qiang et al., 2020; Lisena et al., 2020) often lack the support for neural topic models or necessary steps in the topic modeling lifecycle, *e.g.,* data preprocessing and evaluations. The latest toolkit OCTIS (Terragni et al., 2021) is more comprehensive, but as shown in Table 1 and Figure 1, it solely considers basic and hierarchical topic modeling scenarios and overlooks the latest advancements of neural topic models, offering only two neural topic models introduced after 2018. As a consequence, these issues pose hurdles to the comparisons, developments, and applications of topic models.

To resolve these issues, we in this paper introduce **Top**ic **Mo**deling **S**ystem **T**oolkit (**TOPMOST**), which supports extensive features. In contrast to existing toolkits, TOPMOST thoroughly incorporates the most prevalent topic modeling scenarios: **basic, hierarchical, dynamic, and cross-lingual topic modeling**, as well as the latest neural topic models as detailed in Table 1 and Figure 1. It covers the entire lifecycles of these scenarios, including datasets, preprocessing, models, training, and evaluations. More importantly, TOPMOST adheres to an object-oriented paradigm with a highly cohesive and decoupled modular design. This enhances the readability and extensibility of TOPMOST, enabling users to flexibly customize their own datasets, models, and evaluations for their diverse research or application purposes. As a result, TOPMOST excels in fulfilling the practical requirements of topic modeling. We conclude the advantages of our TOPMOST as follows:

- TOPMOST provides handy and complete cutting-edge topic models for various scenarios;

- TOPMOST allows users to effortlessly and fairly compare topic models through comprehensive evaluation metrics;

- TOPMOST with better readability and extensibility facilitates the smooth development of new topic models and downstream applications.

2 Related Work

Throughout the long history of topic modeling, numerous toolkits have emerged and gained widespread adoption. The earliest among those include Mallet (McCallum, 2002) [1] and gensim (Rehurek and Sojka, 2011) [2]. While these fundamental frameworks sufficiently embrace conventional topic models, they generally overlook the recent advancements in neural topic models. STTM (Qiang et al., 2018) particularly focuses on probabilistic short text topic models, like BTM (Yan et al., 2013) and DMM (Yin and Wang, 2014). A more recent entrant, OCTIS (Terragni et al., 2021), integrates both conventional and neural topic models. Nevertheless, it merely covers basic and hierarchical topic modeling scenarios and neglects the latest neural topic models developed after 2018. Moreover, OCTIS couples the implementations of model construction and training, exacerbating the challenges of toolkit maintenance. Different from these existing work, our TOPMOST extensively incorporates a spectrum of popular topic modeling scenarios and the latest developments in neural topic models. In addition, our TOPMOST clearly

---

[1] https://mimno.github.io/Mallet/topics.html
[2] https://radimrehurek.com/gensim/

Topic Modeling Scenarios	Topic Models	Evaluation Metrics	Datasets
Basic topic modeling	LDA (Blei et al., 2003) NMF (Lee and Seung, 2000) NeuralLDA (Srivastava and Sutton, 2017) ProdLDA (Srivastava and Sutton, 2017) ETM (Dieng et al., 2020) DecTM (Wu et al., 2021) NSTM (Zhao et al., 2021b) CombinedTM (Bianchi et al., 2021) BERTopic (Grootendorst, 2022) TSCTM (Wu et al., 2022) ECRTM (Wu et al., 2023b) FASTopic (Wu et al., 2024c)	TC TD Classification Clustering	20NG IMDB Wikitext-103 NeurIPS ACL NYT
Hierarchical topic modeling	HDP (Teh et al., 2006) SawETM (Duan et al., 2021) HyperMiner (Xu et al., 2022) ProGBN (Duan et al., 2023) TraCo (Wu et al., 2024d)	TC over levels TD over levels Classification over levels Clustering over levels	
Dynamic topic modeling	DTM (Blei and Lafferty, 2006) DETM (Dieng et al., 2019) CFDTM (Wu et al., 2024a)	TC over time slices TD over time slices Classification Clustering	NeurIPS ACL NYT
Cross-lingual topic modeling	NMTM (Wu et al., 2020a) InfoCTM (Wu et al., 2023a)	TC (CNPMI) TD over languages Classification Clustering	ECNews Amazon Review Rakuten

Table 2: Summary of topic modeling scenarios, topic models, evaluation metrics, and datasets covered by TopMost.

decouples each step (data, models, and training) in the topic modeling lifecycles, resulting in neat code structures and simplified maintenance.

3 Overview of Toolkit Design and Architecture

In this section, we delineate the overview of our toolkit design and architecture. We build Top-Most with Python and use PyTorch (Paszke et al., 2019) as the neural network framework for neural topic models. Figure 2 illustrates the overall architecture of TopMost.

3.1 Topic Modeling Scenarios and Topic Models

As summarized in Table 2, TopMost reaches a wider coverage by involving the 4 most popular topic modeling scenarios and their corresponding conventional or neural topic models.

Basic Topic Modeling discovers a number of latent topics from normal documents like news articles and web snippets, as the most common scenario (Blei et al., 2003). For basic topic models, TopMost supports conventional LDA (Blei et al., 2003), NMF (Lee and Seung, 2000), and most of the mainstream neural models such as

ProdLDA (Srivastava and Sutton, 2017), ETM (Dieng et al., 2020), CombinedTM (Bianchi et al., 2021), BERTopic (Grootendorst, 2022), TSCTM (Wu et al., 2022), ECRTM (Wu et al., 2023b), and FASTopic (Wu et al., 2024c).

Hierarchical Topic Modeling organizes topics into a tree structure instead of flat topics in the basic topic modeling (Griffiths et al., 2003; Isonuma et al., 2020). Topics at each level of the structure involve different semantic granularity: child topics are more specific to their parent topics. This provides more desirable granularity for downstream applications. Hierarchical topic models in Top-Most include conventional HDP (Teh et al., 2006) and recently popular neural hierarchical topic models, *e.g.,* HyperMiner (Xu et al., 2022), ProGBN (Duan et al., 2023), and TraCo (Wu et al., 2024d).

Dynamic Topic Modeling discovers the evolution of topics in sequential documents, such as the conference papers published by year (Blei and Lafferty, 2006). This discloses how topics emerge, grow, and decline over time due to real-world trends and events, which has derived applications like trend analysis and public opinion mining (Li et al., 2020; Churchill and Singh, 2022a). For dynamic topic models, we provide the conventional

DTM (Blei and Lafferty, 2006) and its neural variant, DETM (Dieng et al., 2019). We also cover recent CFDTM (Wu et al., 2024a).

Cross-lingual Topic Modeling discovers aligned cross-lingual topics from bilingual corpora (Mimno et al., 2009). These reveal the commonalities and differences across languages and cultures, enabling cross-lingual text analysis without supervision (Yuan et al., 2018; Yang et al., 2019). Cross-lingual topic models in TOPMOST include NMTM (Wu et al., 2020a) and InfoCTM (Wu et al., 2023a).

We carefully adapt the original implementations of these topic models and unify their APIs of initialization, training, and testing, ensuring that our toolkit remains user-friendly, readable, and extendable. Note that we will constantly update TOP-MOST to include more newly released models.

3.2 Datasets and Preprocessing

TOPMOST contains extensive benchmark datasets for the involved topic modeling scenarios, as reported in Table 2. We summarize the statistics of these datasets in Tables 3 to 5.

For basic and hierarchical topic modeling, we have the following datasets: (i) **20NG** (20 News Groups, Lang, 1995) is one of the most widely used datasets for evaluating topic models, including news articles with 20 labels. (ii) **IMDB** [3] (Maas et al., 2011) is the movie reviews from the IMDB website, containing two sentimental labels, positive and negative. (iii) **Wikitext-103** [4] (Merity et al., 2016) includes Wikipedia articles (Nguyen and Luu, 2021).

For dynamic topic modeling, TOPMOST provides the datasets as (i) **NeurIPS** [5] includes the published papers at the NeurIPS conference from 1987 to 2017. (ii) **ACL** (Bird et al., 2008) is an article collection between 1973 and 2006 from ACL Anthology [6]. (iii) **NYT** [7] contains the news articles in the New York Times, from 2012 to 2022, with 12 categories, like "Arts", "Business", and "Health".

For cross-lingual topic modeling, we offer the

---

[3] http://ai.stanford.edu/~amaas/data/sentiment/aclImdb_v1.tar.gz
[4] https://www.salesforce.com/products/einstein/ai-research/the-wikitext-dependency-language-modeling-dataset/
[5] https://www.kaggle.com/datasets/benhamner/nips-papers
[6] https://aclanthology.org/
[7] https://huggingface.co/datasets/Matthewww/nyt_news

following bilingual datasets: (i) **ECNews** [8] (Wu et al., 2020a) is a collection of English and Chinese news with 6 categories like business, education, and entertainment. (ii) **Amazon Review** (Wu et al., 2020a) includes English and Chinese reviews from the Amazon website, where each review has a rating from one to five. We simplify it as a binary classification task by labeling reviews with ratings of five as "1" and the rest as "0" following Yuan et al. (2018). (iii) **Rakuten Amazon** (Wu et al., 2023a) contains Japanese reviews from Rakuten (a Japanese online shopping website, Zhang and LeCun, 2017), and English reviews from Amazon (Yuan et al., 2018). Similarly, it is also simplified as a binary classification task according to the ratings. Note that basic topic models can employ the datasets for dynamic topic modeling as well.

We preprocess these datasets with standard steps, such as removing stop words and punctuation, removing short tokens, and filtering low-frequency words (Card et al., 2018; Wu et al., 2020b). Users can directly download these off-the-shelf datasets for experiments through TOPMOST from our GitHub repository. See Appendix A for more details of these datasets. We also provide configurable preprocessing implementations, allowing users to flexibly customize their datasets.

3.3 Evaluation Metrics

TOPMOST provides sufficient evaluation metrics to evaluate topic models. We first evaluate the quality of discovered topics in terms of **topic coherence** (**TC**, Newman et al., 2010) and **topic diversity** (**TD**, Dieng et al., 2020). TC refers to the coherence between the top words of discovered topics, and TD measures the differences between topics. We consider different implementations of TC and TD, for example, NPMI (Lau et al., 2014), C_V (Röder et al., 2015), and TU (Nan et al., 2019), for extensive comparisons.

Then, we evaluate the quality of inferred doc-topic distributions via extrinsic tasks: **text classification** and **clustering** (Wu and Li, 2019; Zhao et al., 2021b; Nguyen et al., 2024). For classification, we train an ordinary classifier (*e.g.,* SVM) with doc-topic distributions as document features and predict the labels of others. For clustering, we use the most significant topics in doc-topic distributions as clustering assignments.

Apart from these fundamental ones, we addition-

---

[8] https://github.com/bobxwu/NMTM

ally include metrics for special scenarios. For cross-lingual topic modeling, we measure the average TD over all languages and evaluate the alignment between cross-lingual topics with cross-lingual NPMI (CNPMI, Hao and Paul, 2018). For hierarchical topic modeling, we evaluate the quality of discovered topic hierarchies, concerning the coherence and diversity between parent and child topics, the diversity between parent and non-child topics, and the diversity between sibling topics (Chen et al., 2021b,a; Wu et al., 2024d).

4 Comparison to Existing Toolkit

To highlight our significant strengths, we compare our TOPMOST with the latest counterpart, OCTIS (Terragni et al., 2021), which integrates more features than earlier toolkits. Our TOPMOST outperforms OCTIS in three key aspects:

(**i**) As detailed in Table 1, TOPMOST offers a broader coverage of topic modeling scenarios, accompanied by corresponding datasets, models, and evaluation metrics. This better fulfills the various requirements of researchers and developers.

(**ii**) TOPMOST provides a more extensive array of topic models compared to OCTIS. As reported in Figure 1 while OCTIS merely includes 4 neural topic models, TOPMOST incorporates 16 ones, including the latest NSTM (Zhao et al., 2021b), HyperMiner (Xu et al., 2022), and ECRTM (Wu et al., 2023b). These advanced models empower users with cutting-edge topic modeling techniques and simplify their comparisons and applications.

(**iii**) TOPMOST entirely decouples the implementations of data loading, model construction, model training, and evaluations, as illustrated in Figure 2. This design streamlines the code structure for high reusability and facilitates fair comparisons among diverse topic models. It aligns with prominent libraries such as Huggingface Transformers and PyTorch Lightning. See the code examples in Sec. 5.

5 Toolkit Usage

We showcase the simplicity and user-friendly design of our TOPMOST toolkit with code examples. Users can directly install our TOPMOST through pip [9]: `pip install topmost`.

Figure 3 shows how to quickly utilize TOPMOST to discovers topics from documents with a few

handy steps: dataset preprocessing, model construction (here ProdLDA (Srivastava and Sutton, 2017)), and training. We emphasize that our TOPMOST supports other languages besides English. We can simply employ different tokenizers in the preprocessing for other languages, for example, jieba [10] for Chinese and nagisa [11] for Japanese. Other preprocessing settings are also configurable, including maximum vocabulary size, stop words, and maximum or minimum document frequency. This allows users to flexibly apply our toolkit.

```
from topmost.data import RawDatasetHandler
from topmost.models import ProdLDA
from topmost.trainers import BasicTrainer

docs = [ "A document about space, satellite,
    launch, orbit.", # more example documents...
]
# build a dataset
dataset = RawDatasetHandler(docs)
# create a topic model
model = ProdLDA(dataset.vocab_size)
# create a trainer
trainer = BasicTrainer(model)
topic_top_words, doc_topic_dist = trainer.
    fit_transform(dataset)
```

Figure 3: A code example for quick start.

Figure 4 exemplifies how to train a topic model with preprocessed datasets. The training of other topic models follows similar steps.

```
from topmost.data import download_dataset,
    BasicDatasetHandler
from topmost.models import ProdLDA
from topmost.trainers import BasicTrainer

#download a dataset
download_dataset('20NG', cache_path='./datasets')
# load a dataset
dataset = BasicDatasetHandler("./datasets/20NG")

# create a topic model
model = ProdLDA(dataset.vocab_size)
# create a trainer
trainer = BasicTrainer(model)
# train the topic model
trainer.train(dataset)
```

Figure 4: A code example for training a topic model (ProdLDA (Srivastava and Sutton, 2017)).

Figure 5 shows how to fully evaluate the trained topic model with diverse metrics including topic coherence, topic diversity, text classification, and text clustering.

---

[9] https://pypi.org/project/topmost

[10] https://github.com/fxsjy/jieba
[11] https://github.com/taishi-i/nagisa

Topic Distribution of a Document

Input document:

How to plot lines between all points in vector in Matlab?

[upload]

Topic Distribution

```
1    0.8   0.6   0.4   0.2   0
```

Topic#1: scala actors implicit iterator java traits case iterable sugar trait
Topic#2: cocoa window icon builder nsview hide nsoutlineview nstextfield nstableview outline
Topic#3: menu item right navigation dropdown click rss autocomplete bar popup
Topic#4: data core type animation structures streaming filtering ms simulink storage
Topic#5: oracle cursor ref plsql table select trigger insert statements odpnet
Topic#6: expression naming convention tool instead cases included maven switching regular
Topic#7: searching keyword mysql capture dependencies real webservices comet sbt include
Topic#8: linq sql lambda group iqueryable ienumerable distinct query clause dataset
Topic#9: column calculated max range id count values unique number increment
Topic#10: message exception undefined found error unexpected failed cause errors fatal
Topic#11: ajax jquery response div json ie browser aspnet javascript calls
Topic#12: windows bit drop platform cross mobile drag ui linux instances
Topic#13: subversion merge branch trunk repositories revisions branches merging repository changes
Topic#14: svn copy revision working repository tortoise directory ignore externals recursively
Topic#15: matlab plot vectors plotting points figure equations axis two distribution
Topic#16: stored connection procedure transaction procedures jdbc connecting connections secure privileges
Topic#17: node drupal taxonomy views cck nodes module terms submission theme
Topic#18: script bash shell scripts ssh commands program process directory prompt
Topic#19: hibernate cache level lazy session loading second nd already entities
Topic#20: loop work snow ide leopard play wait code generation alternative
Topic#21: mac os terminal osx leopard iphone installing ruby snow sdk
Topic#22: virtual server testing software client host web unit service edition

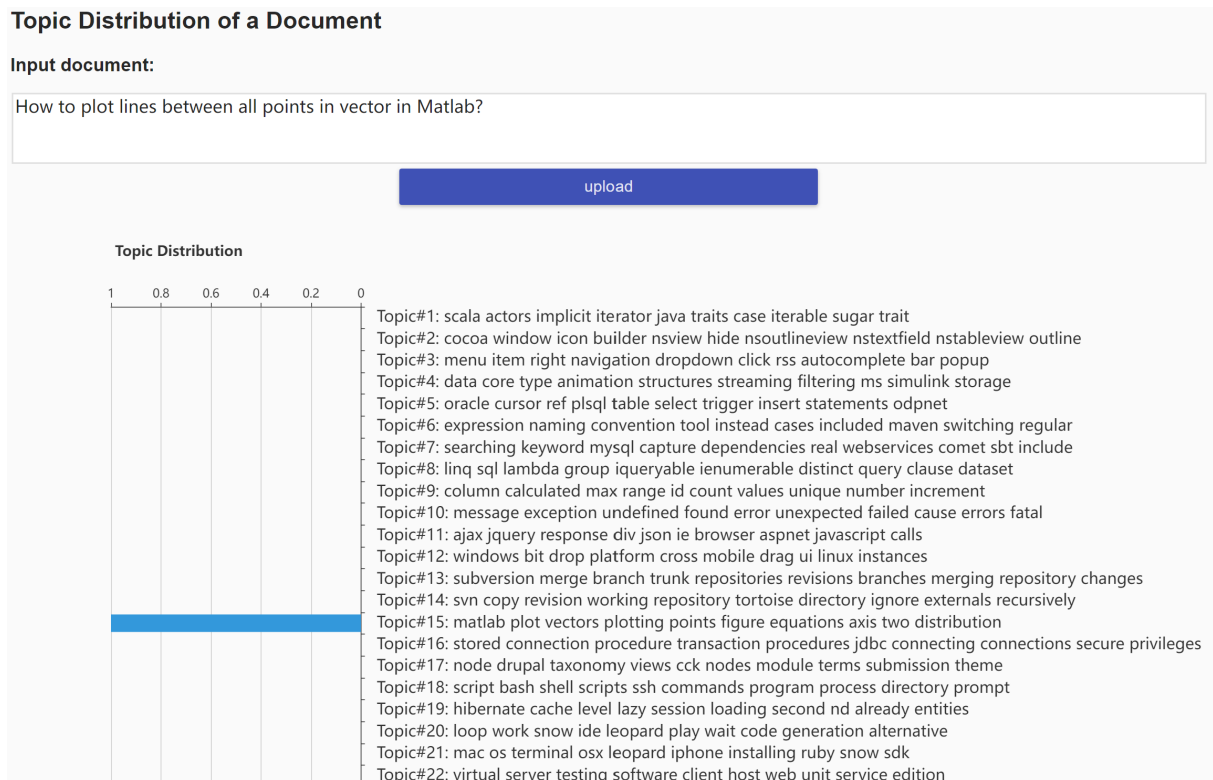Figure 6: Demonstration of testing new documents. It plots the inferred topic distribution of an input document from a trained topic model.

```python
from topmost.evaluations import (
    compute_topic_diversity,
    compute_topic_coherence, evaluate_clustering
    , evaluate_classification)

# doc-topic distributions
train_theta, test_theta = trainer.export_theta(
    dataset)
# top words of topics
topic_top_words = trainer.export_top_words(
    dataset.vocab)
# topic coherence
compute_topic_coherence(topic_top_words, dataset.
    train_texts)
# topic diversity
compute_topic_diversity(topic_top_words)
# text clustering
evaluate_clustering(test_theta, dataset.
    test_labels)
# text classification
evaluate_classification(train_theta, test_theta,
    dataset.train_labels, dataset.test_labels)
```

Figure 5: A code example for evaluating a topic model, including topic coherence, topic diversity, text classification, and clustering.

The above examples illustrate that TOPMOST clearly decouples the APIs of data, models, and training, so TOPMOST becomes more accessible, easy-to-use, and extendable to users. Due to limited page space, see more examples and tutorials on our GitHub project page, like data preprocessing and other topic modeling scenarios.
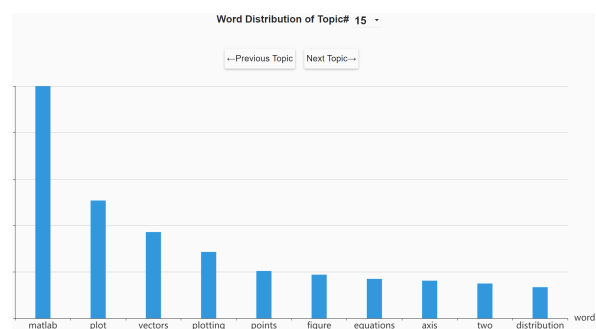
Word Distribution of Topic# 15

[←Previous Topic] [Next Topic→]

```
matlab  plot  vectors  plotting  points  figure  equations  axis  two  distribution   word
```

Figure 7: Visualization of discovered topics. It plots the top related words of each topic and the modeled word distributions.

6 Visualization Interfaces

TOPMOST furthermore provides visualization interfaces for topic models. We create a web demo system with Flask [12] as the server framework following Material design to visualize and test topic models. It is designed to be intuitive and user-friendly, enabling users to easily understand and leverage topic models.

Figure 7 shows the visualization of topics. We plot the top related words of discovered topics and the modeled probability of each word. For example,

[12] https://flask.palletsprojects.com/

Topic#15 in Figure 7 mostly relates to words like "matlab", "plot", and "vectors". By selecting the index or clicking the *Previous* and *Next* buttons, we can view the details of any topic.

Figure 6 demonstrates the interactive utilization of a trained topic model. Upon inputting a document, we can click the *upload* button to obtain the inferred topic distribution of the document. The horizontal bar chart in Figure 6 plots the distribution over all topics of the input *How to plot lines between all points in vector in Matlab?*. We see that the topic distribution mainly lies on Topic#15, which refers to Matlab.

7 Conclusion and Future Work

In this paper, we present TOPMOST, an open-source, comprehensive, and up-to-date topic modeling system toolkit. TOPMOST provides complete lifecycles of various topic modeling scenarios, including datasets, preprocessing, models, training, and evaluations, which outperforms existing counterparts. TOPMOST allows users to smoothly explore topic models, verify their new ideas, and develop novel topic modeling applications. This benefits both the communities in academia and industry. In the future, we plan to keep TOPMOST updated to incorporate more latest topic models and support more features to facilitate the research and application of topic modeling.

Limitations

We consider the following limitations of TOP-MOST. First, TOPMOST only includes the mainstream evaluation metrics. Some less popular ones like perplexity are ignored. Second, TOPMOST does not cover the topic models based on prompting large language models (Pan et al., 2023; Wu et al., 2024e; Pham et al., 2023). Different from LDA-like models, they define a topic as a textual description, so we cannot assess them through existing evaluation metrics.

References

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.

Jordan L Boyd-Graber, Yuening Hu, David Mimno, et al. 2017. *Applications of topic models*, volume 11. now Publishers Incorporated.

Dallas Card, Chenhao Tan, and Noah A Smith. 2018. Neural Models for Documents with Metadata. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2031–2040.

Ziye Chen, Cheng Ding, Yanghui Rao, Haoran Xie, Xiaohui Tao, Gary Cheng, and Fu Lee Wang. 2021a. Hierarchical neural topic modeling with manifold regularization. *World Wide Web*, 24:2139–2160.

Ziye Chen, Cheng Ding, Zusheng Zhang, Yanghui Rao, and Haoran Xie. 2021b. Tree-structured topic modeling with nonparametric neural variational inference. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2343–2353.

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

Rob Churchill and Lisa Singh. 2022b. The evolution of topic modeling. *ACM Computing Surveys*, 54(10s):1–35.

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.

Zhibin Duan, Xinyang Liu, Yudi Su, Yishi Xu, Bo Chen, and Mingyuan Zhou. 2023. Bayesian progressive deep topic model with knowledge informed textual data coarsening process. In *International Conference on Machine Learning*, pages 8731–8746. PMLR.

Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2021. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 759–766.

Zhibin Duan, Dongsheng Wang, Bo Chen, Chaojie Wang, Wenchao Chen, Yewen Li, Jie Ren, and Mingyuan Zhou. 2021. Sawtooth factorial topic embeddings guided gamma belief network. In *International Conference on Machine Learning*, pages 2903–2913. PMLR.

Thomas Griffiths, Michael Jordan, Joshua Tenenbaum, and David Blei. 2003. Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems*, 16.

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

Shudong Hao and Michael Paul. 2018. Learning multilingual topics from incomparable corpora. In *Proceedings of the 27th international conference on computational linguistics*, pages 2595–2609.

Alexander Hoyle, Pranav Goel, Andrew Hian-Cheong, Denis Peskov, Jordan Lee Boyd-Graber, and Philip Resnik. 2021. Is automated topic model evaluation broken? the incoherence of coherence. In *Advances in Neural Information Processing Systems*.

Masaru Isonuma, Junichiro Mori, Danushka Bollegala, and Ichiro Sakata. 2020. Tree-structured neural topic model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 800–806.

Hannah Kim, Jaegul Choo, Jingu Kim, Chandan K Reddy, and Haesun Park. 2015. Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 567–576.

Ken Lang. 1995. Newsweeder: Learning to filter netnews. In *Proceedings of the Twelfth International Conference on Machine Learning*, pages 331–339.

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.

Daniel Lee and H Sebastian Seung. 2000. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13.

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.

Pasquale Lisena, Ismail Harrando, Oussama Kandakji, and Raphael Troncy. 2020. Tomodapi: a topic modeling api to train, use and compare topic models. In *Proceedings of second workshop for NLP open source software (NLP-OSS)*, pages 132–140.

Andrew Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, pages 142–150.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for languagetoolkit. *http://mallet. cs. umass. edu*.

Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. 2016. Pointer sentinel mixture models. *arXiv preprint arXiv:1609.07843*.

David Mimno, Hanna Wallach, Jason Naradowsky, David A Smith, and Andrew McCallum. 2009. Polylingual topic models. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, pages 880–889, Singapore. Association for Computational Linguistics.

Feng Nan, Ran Ding, Ramesh Nallapati, and Bing Xiang. 2019. Topic modeling with Wasserstein autoencoders. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6345–6381, Florence, Italy. Association for Computational Linguistics.

David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 100–108. Association for Computational Linguistics.

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

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

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.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style,

high-performance deep learning library. *Advances in neural information processing systems*, 32.

Chau Minh Pham, Alexander Hoyle, Simeng Sun, and Mohit Iyyer. 2023. Topicgpt: A prompt-based topic modeling framework. *arXiv preprint arXiv:2311.01449*.

Jipeng Qiang, Yun Li, Yunhao Yuan, Wei Liu, and Xindong Wu. 2018. Sttm: A tool for short text topic modeling. *arXiv preprint arXiv:1808.02215*.

Jipeng Qiang, Zhenyu Qian, Yun Li, Yunhao Yuan, and Xindong Wu. 2020. Short text topic modeling techniques, applications, and performance: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(3):1427–1445.

Radim Rehurek and Petr Sojka. 2011. Gensim – statistical semantics in python. *Retrieved from genism. org*.

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.

Tian Shi, Kyeongpil Kang, Jaegul Choo, and Chandan K Reddy. 2018. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In *Proceedings of the 2018 World Wide Web Conference*, pages 1105–1114. International World Wide Web Conferences Steering Committee.

Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Mark Steyvers and Tom Griffiths. 2007. Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7):424–440.

Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. 2006. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.

Silvia Terragni, Elisabetta Fersini, Bruno Giovanni Galuzzi, Pietro Tropeano, and Antonio Candelieri. 2021. Octis: Comparing and optimizing topic models is simple! In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 263–270.

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, Xinshuai Dong, Liangming Pan, Thong Nguyen, and Anh Tuan Luu. 2024a. Modeling dynamic topics in chain-free fashion by evolution-tracking contrastive learning and unassociated word exclusion. *arXiv preprint arXiv:2405.17957*.

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. 2024b. 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. 2024c. Fastopic: A fast, adaptive, stable, and transferable topic modeling paradigm. *arXiv preprint arXiv:2405.17978*.

Xiaobao Wu, Fengjun Pan, Thong Nguyen, Yichao Feng, Chaoqun Liu, Cong-Duy Nguyen, and Anh Tuan Luu. 2024d. 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. 2024e. Updating language models with unstructured facts: Towards practical knowledge editing. *arXiv preprint arXiv:2402.18909*.

Yishi Xu, Dongsheng Wang, Bo Chen, Ruiying Lu, Zhibin Duan, and Mingyuan Zhou. 2022. Hyperminer: Topic taxonomy mining with hyperbolic embedding. In *Advances in Neural Information Processing Systems*, volume 35, pages 31557–31570. Curran Associates, Inc.

Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1445–1456. ACM.

Weiwei Yang, Jordan Boyd-Graber, and Philip Resnik. 2019. A multilingual topic model for learning weighted topic links across corpora with low comparability. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1243–1248.

Jianhua Yin and Jianyong Wang. 2014. A dirichlet multinomial mixture model-based approach for short text clustering. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 233–242. ACM.

Michelle Yuan, Benjamin Van Durme, and Jordan L Ying. 2018. Multilingual anchoring: Interactive topic modeling and alignment across languages. *Advances in neural information processing systems*, 31.

Xiang Zhang and Yann LeCun. 2017. Which encoding is the best for text classification in chinese, english, japanese and korean? *arXiv preprint arXiv:1708.02657*.

He Zhao, Dinh Phung, Viet Huynh, Yuan Jin, Lan Du, and Wray Buntine. 2021a. Topic modelling meets deep neural networks: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4713–4720. International Joint Conferences on Artificial Intelligence Organization. Survey Track.

He Zhao, Dinh Phung, Viet Huynh, Trung Le, and Wray L. Buntine. 2021b. Neural topic model via optimal transport. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Dataset	Language	#docs	Vocabulary Size	Average length	#labels
ECNews	English	46,870	5,000	12.0	6
	Chinese	50,000	5,000	10.6	
Amazon Review	English	25,000	5,000	30.6	2
	Chinese	25,000	5,000	43.2	
Rakuten Amazon	English	25,000	5,000	30.6	2
	Japanese	25,000	5,000	22.5	

Table 3: Statistics of pre-processed datasets for cross-lingual topic modeling.

Dataset	#docs	Vocabulary Size	Average Length	#labels
20NG	18,846	5,000	110.5	20
IMDB	50,000	5,000	95.0	2
Wikitext-103	28,532	10,000	1,355.4	/

Table 4: Statistics of pre-processed datasets for basic and hierarchical topic modeling.

Dataset	#docs	Vocabulary Size	Average Length	#labels	#time slices
NeurIPS	7,237	10,000	2,085.9	/	31
ACL	10,560	10,000	2,023.0	/	31
NYT	9,172	10,000	175.4	12	11

Table 5: Statistics of pre-processed datasets for dynamic topic modeling.

A Datasets

Tables 3 to 5 report the statistics of datasets for different topic modeling scenarios after preprocessing. Users can directly download all these datasets via TOPMOST from our GitHub repository.