Investigation of the Effect of Word Embedding on Topic Models for Short Texts: A Review

Habbat Nassera, *Member, IAENG* and Nouri Hicham, *Member, IAENG*

***Abstract*— Topic modeling has been a successful text analysis method for almost two decades, with over a hundred models produced and various applications and techniques in neural language analysis, namely text generation, summarization, and language models. Whereas topic modeling met deep neural networks, neural topic models emerged as a new and increasingly prominent study area. To achieve this, it is necessary to summarize recent research findings and explore open issues and prospects. This paper briefly reviews various topic models, including probabilistic and neural ones. We also investigated the word embedding techniques for text representation that might affect the performance of topic models. These findings should spur additional topic modeling..**

***Index Terms*— Deep Learning, Neural Topic Model, Topic Modeling, Word embedding.**

I. Introduction

With so many online comments and reviews on e-commerce sites, social media sites, and other websites displaying user opinions directly, it is clear that text is a potent and extensively used form of human thought. Many industries stand to benefit from text-based data mining, and there are a variety of tasks for text categorization [1] in natural language processing (NLP), namely topic modeling, discussed in this work.

Based on recent breakthroughs in machine learning (ML) [2], it includes techniques and methodology for discovering patterns in the words of a collection of texts and categorizing them into distinct themes or topics. Topic modeling accomplishes this by analyzing every document and identifying the hidden subjects (topics) that run across it, resulting in a more comprehensive knowledge of each document. Every document, it claims, contains a collection of underlying topics, and a set of terms defines each topic. The topic model is an ensemble of algorithms for revealing a document's hidden thematic structure. It automatically captures the subjects in a corpus's text documents [3]. The concept behind the topic model is that every document or text contains several "topics," and the "topic" is a word collection that reflects the entire topic.

We highlight different topic modeling techniques used to figure out a text's topic. We use this unsupervised classification method, a type of statistical modeling, to identify abstract topics within a text collection. There are a lot of reviews summarizing research developments in this task; some works present the existing topic modeling methods [4]–[6]. Other works also discussed the applications of topic modeling [7], for example, mining software repositories [8], social network analysis [9]–[12], [13], [14]; other studies concentrated on the usage of topic models with different sort of data, for example, short text [15], video analysis [16], and so on. However, Latent Dirichlet Allocation (LDA) [17] is the standard method in this field [5], [18] However, with the latest progress in deep generative models and deep neural networks (DNNs), a new research direction called neural topic model (NTM) [19] has emerged to use DNNs to improve the performance, usability, and efficiency of topic modeling.

Some data preparation techniques, such as word embedding, may affect how well NLP tasks are performed. In this context, neural model-based word embeddings, such as word2vec, are more popular than matrix factorization-based word embeddings, despite the common use of word embedding to convert words into vectors [20]. Researchers in this field have investigated the influence of combining diverse methodologies with word embeddings to execute NLP tasks, yielding encouraging findings. BERT [21], GPT [22], and XLNet [23] are instances of contextualized word embeddings (CWEs) that consider the meaning of words and have recently increased in popularity.

The authors of this work conducted a review of the topic modeling literature in various languages. We provide a comprehensive assessment of the tools, methodologies, mechanisms, and results, focusing on deep learning (DL) techniques [24], which are highly popular in those domains and discussed in detail. We attempt to offer a comprehensive overview of relevant deep-learning-based topic modeling models by employing various word embedding techniques and DL algorithms. This study aims to provide a concise assessment of the current status of topic modeling research and identify potential areas for additional exploration.

We use the following sets of keywords to query several databases, including Springer, Elsevier, IEEE, ACM, and IEEE, to obtain reviewed papers (the majority from 2019): "Topic modeling" can also refer to "topic extraction", "LDA", "LSA", "LSI", "PLSA", "CTM", "NMF", "neural topic model", "deep learning", and "topic modeling".

We selected 78 publications for this review after applying the above queries. Figure 1 shows the distribution of papers chosen per publisher. During our research, we discovered that news media and social media were the most active domains used by the topic modeling methods.

Manuscript received September 25, 2023; revised January 14, 2025.

Habbat Nassera is a Professor at Faculty of Science and Technology of Settat, Hassan First University, Settat, Morocco. (Corresponding author e-mail: nassera.habbat@gmail.com).

Nouri Hicham is a PhD candidate of Research Laboratory on New Economy and Development (LARNED), Faculty of Legal Economic and Social Sciences AIN SEBAA, Hassan II University of Casablanca, Morocco. (e-mail: nhicham191@gmail.com).

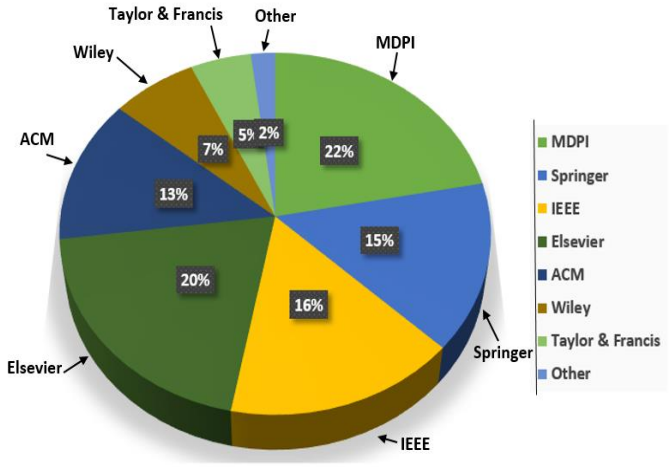

Fig. 1. The count of papers per publisher.

II. The Paper Composition and Main Contribution

We will organize the remainder of this article as follows: The third section introduces some of the most widely used word embedding models. Section 5, which discusses various related works about topic modeling, presents topic modeling, its applications, and its different models, Section 6 concludes the paper.

This paper's main contributions are: (1) analyzing different word embedding models used with different tasks, especially topic modeling; and (2) exploring different techniques and models used for topic modeling in combination with word embedding.

III. Word Embedding

In recent years, text data mining tools have depended mainly on ML and DL, which have aided in the advancement of NLP. To transform NLP issues into DL problems, one must first encode symbols such as text, generating a word embedding or text representation in the process. Glove [25] and Word2vec [20] are two-word embedding methods previously created to build a global word representation, taking the word into consideration over all of the sentences that use it, whereas FastText [26] is a well-known subword embedding model that stores and encodes every word as an n-gram of characters. Many works have recently attached great importance to the multiple meanings of a word in various situations; Doc2vec [27], for example, is based on Word2vec, which considers semantic and contextual information, and in the classification test, it performs only marginally better than the document's simple average word vectors. ELMO (Embeddings from Language Model) [28] takes a bidirectional language model (biLM) and extracts context-sensitive features. Next, there is OpenAI-GPT as the Transformer network, the generative pre-training Transformer [22], and BERT as a large-scale pre-training language model that uses the bidirectional Transformer [21]; On a number of tasks, the pre-trained language models from Transformers did much better than ever at getting contextual word embeddings. However, XLNet [23] blends the principles of autoregressive (AR) models, namely OpenGPT and bi-directional context modeling from BERT. During pre-training, the XLNet uses a permutation operation to integrate sentences from the right and left context sides to create a universal order-conscious AR language model.

Figure 2 shows the development of the most popular word embedding models since 2013, which the following sections will split and present as static and contextualized word embedding models.

A. Static word embedding

It's a method for transforming each word into a single vector that's usually dense and has a lot lower dimensionality than the vocabulary size; such a function usually neglects the case that the same string of characters can have different meanings (restaurant table vs. table of contents) or fragments of speech (to run fast vs. the run). It assumes a vocabulary with a specific size. We'll define the most common static word embeddings in this section.

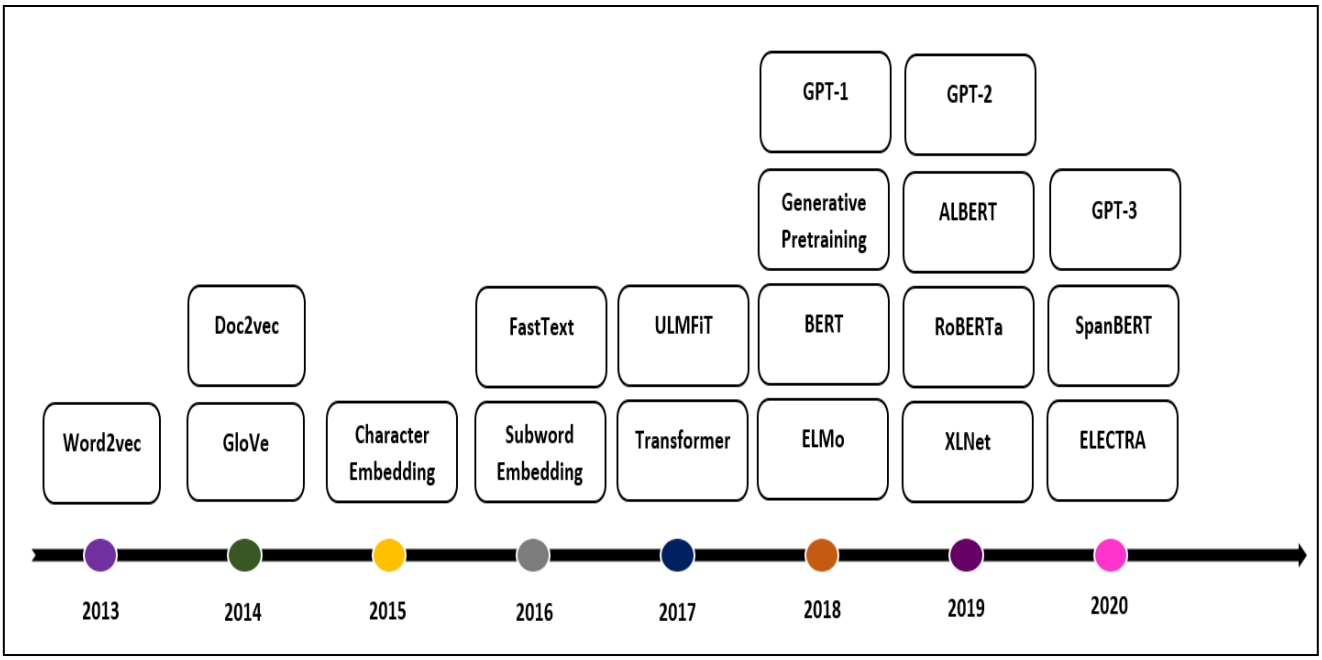

Fig. 2. Some popular word embedding models since

Word2vec

As illustrated in Figure 3, the word2vec [29] approach contains two techniques: continuous skip-gram and continuous bag-of-words (CBOW) models. The CBOW predicts a current word based on the average or sum of context words provided as input. The skip-gram model makes educated assumptions about each contextual word based on the present word input. Word2vec has fewer sizes than prior embedding approaches, which makes it quicker, more adaptable, and more appropriate for many applications for natural language processing. However, despite its broad applicability, it cannot flexibly adjust for specific jobs or overcome the polysemy issue.

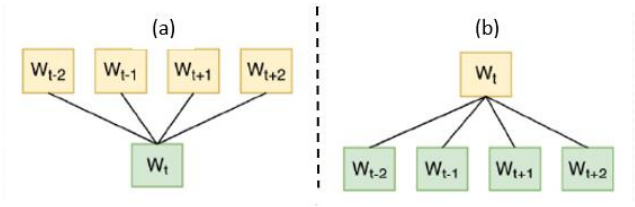

Fig. 3. The two models of word2vec (a) Skip-Gram (b) continuous bag-of-words (CBOW).

Glove

Pennington and al. [25] defined GloVe as a word representation technique using statistics and count. With a set-size context window, it first makes global co-occurrence statistics. Then, it uses stochastic gradient descent to lower its least-squares objective function, which factors the log co-occurrence matrix. It is faster than word2vec and allows parallelization; nonetheless, it consumes more memory than word2vec.

Doc2vec

Doc2Vec or Paragraph Vector (PV) [27], is intended to expand Word2Vec, in which Word2Vec learns to cast words into a latent d-dimensional space while Doc2Vec learns to launch a text into a hidden d-dimensional space.

Doc2vec is an unsupervised method for learning defined length embeddings from changeable text parts, like sentences, sections, and documents. Figure 4 depicts the diagram of doc2vec, which bears a significant modification from the CBOW model [20]; it incorporates an additional paragraph token, transmitted to a PV via a matrix.

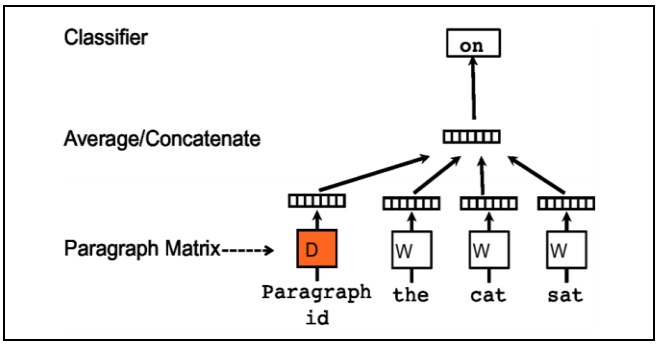

Fig. 4. The diagram of doc2vec.

In doc2vec, the PV serves as a memory for the paragraph's topic, representing information missing from the context. Once trained, the classifier receives the PV and uses it as a feature for the paragraph, either in addition to or instead of BoW.

FastText

The FastText [26] library enables the management of Out-Of-Vocabulary (OOV) words by providing the corresponding word vectors utilizing trained character n-gram embedding. Ignoring the low training time and not sharing the parameters, it performs poorly in large output spaces.

B. *Contextualized word embedding*

In contextualized embedding, a word is represented as a fixed dense vector, whereas in static embedding, contextual information is ignored. The word representation vector is a flexible vector, which means the same word's representation vector occurs in both situations in the pretraining model, but the portrayal is different in each. In this part, we'll go over the most common CWE models.

ULMFit

ULMFit [27] is a transfer learning technique for various NLP tasks. It comprises two pre-trained models, one taught from left to right and vice versa..

ELMo

ELMo [25] is a bi-LSTM-based deep contextualized word embedding solution that is possible to address the polysemy issue. This method can train a vast text corpus and model polysemy, but it takes a long time and cannot handle long-distance dependency.

BERT

BERT [19] is a language representation architecture that uses transformers to represent bidirectional encoders. Fine-tuning and pre-training are the two sections. Its purpose is to precondition left and right contexts from an unlabeled text at all levels to pre-train deep bidirectional representations. Context-aware representation of bidirectional features is possible. Despite the integration of pre-training and generation processes, it leads to subpar generation performance. Additionally, it requires more processing capacity than other current models.

OpenAI- GPT

For language understanding problems, the OpenAI-GPT [20] method is a semi-supervised method that combines supervised fine-tuning with unsupervised pre-training. A language modeling approach first learns the fundamental parameters of the neural network model on an unlabeled dataset. The supervised goal can then adjust these parameters to meet the requirements of a specific task. Because it is a one-way AR language model, it cannot represent context-sensitive properties.

XLNet

Permutation language models like XLNet [21] combine autoregressive and self-encoding models' advantages while eliminating their disadvantages. XLNet shuffles the input sequence's order before estimating the prediction order using the original position-coding order. The matching context, on

the other hand, is dependent on the shuffled order's context. When predicting the target word, you can randomly look at the materials above and below. Furthermore, XLNET does not directly employ Transformer as a feature extractor, but rather as a feature generator. Instead, it uses Transformer-XL, an improved type that can capture longer-range word dependencies than regular Transformers.

C. Discussion

According to the papers presented in this review, word embeddings (WEs) are a commonly used approach in topic modeling and NLP in general. Articles that either introduced a novel theoretical approach for producing WEs or utilized WEs in combination with ML or DL techniques for topic modeling were considered relevant. This section showcases various applications of the WEs concept.

Mikolov et al. [20] defined Word2vec as a moniker for two language models that employ neural networks to build word representations as dense vectors. The Continuous Skip Gram and CBOW are two model designs that have been discussed in detail. Compared to Glove and Latent Semantic Analysis (LSA), Word2Vec is the finest tool for subject segmentation when vectorizing English words, according to [30]. Furthermore, when it comes to predicting the compositionality of multiword expressions [31], the word2vec model outperforms pre-trading models such as FastText, Glove, ELMO, and Doc2Vec using the datasets RAMISH and REDDY, which contain scores for the compositionality of overall multiword expressions ("MWEs"). FastText surpasses Word2Vec and GloVe methods for Indian languages on the POS tagging test [32].

Traditional WE methods assign a fixed representation to each word in the vocabulary collection. While the static word representation is common in NLP, it presents different drawbacks in modeling background data. To begin with, it is incapable of dealing with polysemy, and second, it does not alter the word's meaning depending on its context. To overcome the limitations of static WEs, there is a growing trend toward shifting from simplistic to deep contextualized representations. As shown in Table 1, contextualized WE produces the best results. We can fine-tune pre-trained WE models, such as ELMo, GPT, and BERT, for specific NLP tasks.

TABLE I
A SUMMARY OF THE RESEARCH CONDUCTED ON WORD EMBEDDING TECHNIQUES.

Year	Ref	Proposed approach	Language treated	Dataset	Compared word embeddings	Results
2023	[33]	Using a hybrid model of BERT and LDA in topic extraction and clustering.	English	COVID-19 dataset	LDA BERT BERT+LDA	The combination of LDA and BERT gives the best results.
2023	[34]	Using a novel approach to cognitive text categorization comprised of MTBERT-Attention, a distinct model based on multi-task learning (MTL), BERT, and the co-attention mechanism.	English	-Question dataset - A dataset comprises 2400 training targets.	BERT using different classifiers (CNN, LSTM, BiGRU, attention mechanism and BiGRU)	MTBERT-Attention has the best precision of 100%.
2022	[35]	Using deep learning-based language understanding for patent document multilabel categorization.	English	-M-patent dataset -USPTO-2M patent classification benchmark dataset	word2vec, BERT, RoBERTa, ELECTRA, and XLNet.	At 82.29, 3.635, and 0.850, respectively, XLNet has the best precision, convergent error, and label ranking average precision.
2021	[36]	LDA topical distributions are combined with XLNet contextualized representations.	English	The dataset, which is associated with COVID-19, contains 10,700 social media and news article postings.	USE (Universal sentence encoder) with SVM, BERT, BERT with topic distributions and XLNet	XLNet + Topic Distributions was the best, with an accuracy of 0.968, a f1 score, and a recall of 0.967.
2021	[37]	The application of a BERT-based model to Twitter data in order to analyze sentiment and recognize emotions.	English	6755 tweets.		a 90% for emotion analysis in terms of accuracy. SA has 92% accuracy.
2021	[36]	SRXLNet: syntactic relevance XLNet word embedding.	Mongolian-Chinese, Tibetan-	A total of 710 thousand multilingual	XLNet	By incorporating a dynamical word embedding with context

			Chinese, and Uyghur-Chinese	sentences.		representation, the model's performance was improved.
2021	[38]	The integration of XLNet with the capsule network for personality categorization using textual posts in a deep learning environment (XLNet-Caps).	English	Personality-related data set that includes the Big Five and the Myers Briggs Type Indicator (MBTI).	Glove with Logistic Regression, Glove + SVM, Glove + CNN, Glove + LSTM, BERT, BERT + CNN, BERT + capsule network, ALBERT + capsule network, RoBERTa + capsule network.	XLNet-Caps performed the best, with 0.680 and 0.682 for Macro-F1 and Micro-F1, respectively.
2020	[39]	Investigate the use of pre-trained word embeddings such as BERT, DistilBERT, and RoBERTa, as well as XLNet to determine how people feel about texts.	. English	Cross-cultural questionnaire research in 37 countries generated the ISEAR dataset. The dataset includes 7666 sentences, categorized into seven main emotion categories.	BERT, DistilBERT, RoBERTa, and XLNet	For RoBERTa, its model accuracy is 0.7431. The value for XLNet is 0.7299. The value for BERT is 0.7009, and the value for DistilBERT is 0.6693.
2020	[40]	BERT-based word models are analyzed geometrically to find new ways to embed sentences. The new method looks at the space occupied by a word representation and how it fits into that space.	English	Semantic textual similarity (STS) datasets from 2012 to 2016 with a total of 8,628 sentences.	BERT, SBERT, XLNET.	Without any fine-tuning, SBERT-WK retains a high level of performance.
2020	[41]	The proposition of the MAG (Multimodal Adaptation Gate) attachment for XLNet and BERT that allows them to receive multimodal nonverbal data while they are being fine-tuned.	English	2199 videos were collected from 93 YouTube movie reviews.	MAG-BERT, BERT, XLNet, MAG-XLNet.	MAG-BERT and MAG-XLNet are other models that utilize BERT.
2020	[42]	A comparative study of different word embedding models	English	-EnglishWikipedia, -CC-News, -Wikitext-103, -Stories, -OpenWebText. -BooksCorpus, -WebText,	ULMIFiT, ALBERT. BERT-LARGE, BERT-BASE, XLNet-LARGE based on ELMo, RoBERTa, XLNet-BASE, GPT, RoBERTa-BASE,	XLNet-LARGE and BERT-LARGE are not significantly better than ULMFiT in terms of performance.
2020	[43]	A proposal for estimating sentiment in code-mixed tweets using a methodology that combines four models: BERT, ALBERT, MultiFiT, and XLNET (English-Hindi).	Hindi, English	20,000 tweets.	BERT, ALBERT, MultiFiT, XLNET, an ensemble of word embeddings.	The proposed approach received an F1 score of 72.7%, 72.3% on accuracy, 72.6% on recall, and 72.29% on precision.
2020	[32]	Using several existing techniques,	Bengali, Odiya,	A lot of different sources have been	GloVe, word2vec, FastText, BERT,	FastText performs better than Glove and

		numerous word embeddings were created for 14 Indian languages.	Gujarati, Konkani, Assamese, Hindi, Malayalam, Kannada, Punjabi, Marathi, Nepali, Telugu, Sanskrit, and Tamil,	used to get the data in English and 14 Indian languages.	ELMO, and XLNet.	word2vec on the POS tagging test.
2018	[31]	A comparison of embedding models that can be bought off the shelf for predicting the compositionality of multiword expressions ("MWEs").	English	The REDDY and RAMISH data sets include scores indicating the quality of the MWEs as a whole.	FastText, Glove, ELMO, Doc2Vec.	With a score of 0,880 in the RAMISH data set and 0.710 in the REDDY data set, Word2vec achieved the best performance.

Since 2020, CWEs have been the most popular, especially XLNet and BERT (as well as their parameterized variations), which have performed well. We should fine-tune word embedding methods for a specific task, such as topic modeling, to determine their efficacy.

IV. Topic Modeling

The topic model is a type of probabilistic generative method that finds hidden topics in documents using unsupervised learning. Recently,

Informatics has heavily utilized a topic model, with a focus on text mining and knowledge discovery. This model has received a lot of attention and has piqued the interest of researchers from a variety of fields.

Latent Semantic Analysis (LSA) [44], also known as Latent Semantic Indexing (LSI) [45], serves as the foundation for the development of a topic model. However, LSA maps documents from a sparse high-dimensional vocabulary space to a low-dimensional vector space, referred to as the latent semantic space. LSA is a method that utilizes the singular value decomposition (SVD) technique to get the implicit semantic structure from a group of documents. After LSA, Probabilistic LSA (PLSA) [46] is a different version of the LSA model that is based on a statistical model of hidden classes.

Blei et al. [47] created latent Dirichlet allocation (LDA) as an extension of PLSA; it is a more thorough generative probabilistic model that is currently the most used and the foundation for many other approaches. Today, researchers are developing various probabilistic models based on LDA for precise applications. The community of text analysis initially presented all of the topic models listed above for unsupervised document discovery. Many extensions have emerged from a foundational topic model. For example, HLDA [48] establishes the hierarchical link between topics and automatically determines the topic number. A correlated topic model (CTM) [49] overcomes LDA's inability to predict the importance of subjects in documents. The logistic normal distribution is chosen as the document-topic distribution, it opts for the logistic normal distribution. After that, Arora et al. [50] propose non-negative matrix factorization (NMF) because they believe each topic has an anchor word that distinguishes it from others. The machine learning community has been studying this feature of separability using NMF for at least a decade. Note that NMF comes to solve the limitations of SVD, which consists of only one subject per document or only recovers topic vectors' spans rather than the vectors themselves.

We need increasingly complicated inference algorithms as topic models get more expressive. With DL's rapid growth, neural networks have emerged as valuable methods for topic models that learn complex patterns. However, variational inference uses the alternative distribution to estimate the posterior distribution, whereas neural variational inference employs neural networks to parameterize the alternate distribution [51, p.]. Neural topic models have just become available; ProdLDA [52] is one of the new forms of LDA that changes LDA by mixing a model with a product of experts, resulting in significantly more interpretable subjects. Figure 5 depicts the evolution of some of the most popular topic models since 1998.

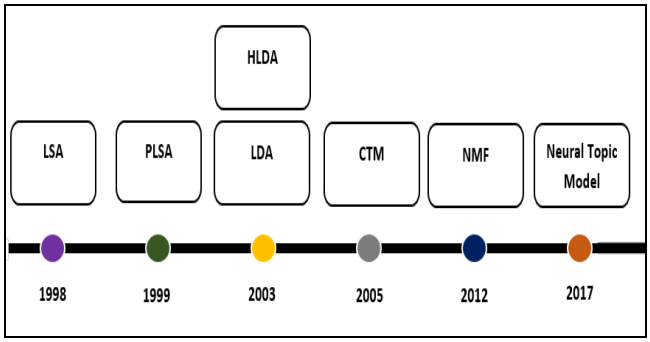

Fig. 5. Some of the most well-known topic models.

A. Topic modeling and its applications

Many applications, such as text mining, machine learning, information retrieval, text classification, and recommendation engines, use topic modeling, leading to the improvement and invention of supervised,

unsupervised, and semi-supervised algorithms. Topic modeling has a variety of uses, as listed below.

• System Recommendations: Many systems, such as those that recommend jobs to interested individuals, map the appropriate work based on their knowledge of the subject matter, sociology, history, location, media theory, and other factors to make employment recommendations.

• Financial Analysis: It is common practice to utilize information about the value of stocks to persuade people to engage in different forms of trading on the stock exchange, among other things.

• Bioinformatics: Examine patient-related literature based on clinical data to learn about the field's knowledge architecture.

• Manufacturing Applications: It is widely used in online advertisement networks and social networking sites.

• Information Technology: Annotating photographs with text and retrieving relevant data from image processing are all instances of data mining.

• Analyzing Social Networks: Social web platforms mine information about the natural world, including crucial characteristics about services and users.

• Software Engineering: Engineers use unstructured repositories in the software business to assist a wide range of technical functions, including program comprehension and location.

• Discovering Themes in Texts: Useful for spotting trends in online publications, such as news headlines and forum highlights.

• Document Summaries: Topic models can better analyze and summarize scientific articles, allowing for more efficient research and development. Historical documents, blogs, newspapers, and event fiction are examples.

B. *Models of topic modeling*

Latent Semantic Analysis

To create semantic information, LSA focuses on developing vector-based representations of texts. LSA uses vector representations to calculate text similarity and identify the most relevant related phrases. LSA was formerly known as latent semantic indexing (LSI), although it was subsequently modified for information retrieval purposes. It is possible to find a few documents from many records that are comparable to the supplied query. Many aspects of LSA should be examined, including keyword matching, width keyword matching, and vector representations corresponding to different document occurrences. LSA also makes use of SVD to restructure the data it processes.

SVD is a technique that utilizes a matrix to reorganize and compute the total reduction in vector space. Furthermore, we will calculate and group the punishments in vector space, ranging from the most important to the least important. In LSA, we base our assumptions on the most crucial assertion; otherwise, we disregard the least essential hypothesis. If those wards have comparable vectors, searching for words with a high similarity rate will happen. To summarize the most critical processes in LSA, start by gathering a large amount of relevant text and then dividing it into papers. Secondly, construct a co-occurrence matrix for both documents and terms, using cell names such as x for documents, n for terms, an n-dimensional vector for documents, and m for a dimensional value for terms. Lastly, add the calculated finals to each cell. SVD will play a significant role in calculating all the diminutions and constructing three matrices.

Probalistic Latent Semantic Analysis

PLSA is a generative method that addresses LSA's statistical flaw. Therefore, we created PLSA to tackle some of the shortcomings of LSA. The primary purpose of PLSA is to detect and distinguish among words utilized in various contexts without the use of a dictionary or thesaurus. It draws two conclusions. i) disambiguates polysemy, allowing the use of terms with multiple meanings. Clustering words [53] with similar meanings conveys similarity.

In this approach, the Latent variable $v_k \in \{ v_1, v_2, \dots v_k \}$, which is equivalent to a possible semantic layer, is introduced. The three stages outlined below can be used to summarize the generative model for each phrase in each corpus and the Figure 8 presents the graphical representation of PLSA

1) Choose a document a_i that has a probability of $p(a_i)$
2) Choose a v_k topic with $P(v_k|a_i)$ probability.
3) Produce a term w_j with $P(w_j|v_k)$ probability.

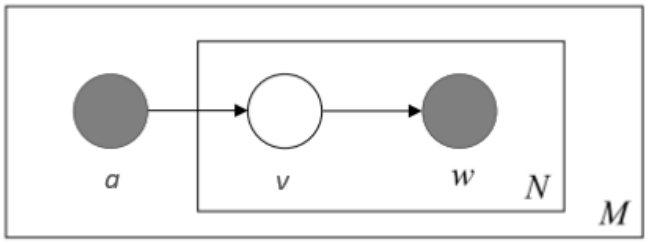

Fig. 6. The PLSA graphical representation.

Latent Dirichlet Allocation

People commonly use the LDA generative probabilistic method as a basic topic model. LDA's appearance aims to improve how mixture models capture word and text interchangeability better than LSA and PLSA. We now create data in various formats such as web pages, news, blogs, social media [54], publications, and other documents. As a result, there is a rising demand for an automatic process or approach to organize, interpret, and synthesize these documents. Only recently established techniques for latent topic modeling provide entirely unsupervised methods for discovering themes from massive data sets [55].

Each According to LDA, a variety of subjects make up each document. LDA describes every topic as a collection of words with a probability, indicating the likelihood of a phrase appearing in the topic. It employs the BoW technique, in which each document is a word collection with no discernible organization other than word and subject statistics. The basic idea behind LDA is that it tries to mimic the writing process. As a result, it only creates a paper on the provided topic. It reveals a corpus's underlying theme or topic.

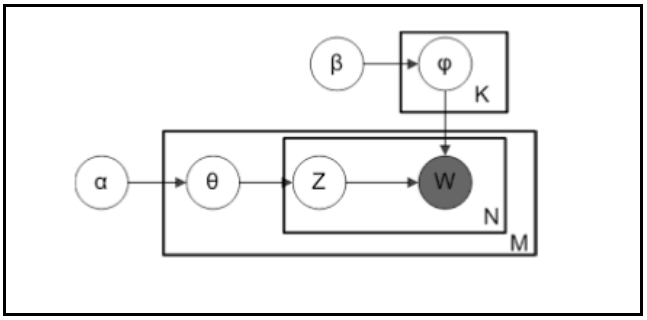

Fig. 7. The LDA graphical representation.

Figure 7 depicts a graphical description of the LDA model. A multinomial distribution of N words characterizes each M element in LDA as a mixture of K hidden topics. The core LDA's creative process is as follows:

In j document, for each $N_j$word:

1) Determine a topic $z_{ij} \sim Mult(\theta_j)$
2) Determine a word $w_{ij} \sim Mult(\varphi_{z_{ij}})$

Where the Dirichlet priors are the multinomial parameters for themes in a document θ_j and words in a topic φ_k.

Hierarchical LDA

Another version of the original LDA is a hierarchical LDA (HLDA), instead of picking a set number of subjects. It's a broader and more adaptable model for topic trees that can aid with expanding data. HLDA finds a tree-like structure of topics inside a corpus, with each extra level being more particular than the preceding level in the hierarchy.

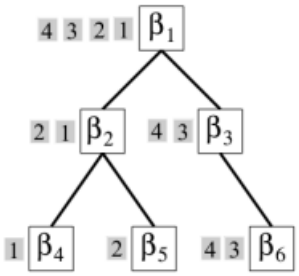

Fig. 8. HLDA process in the form of a tree.

This method learns topic hierarchies using a nested Chinese restaurant process (nCRP) [48], [56]. Each restaurant is a topic in this paradigm, and each document's topics correspond to a path. Figure 8 illustrates this with four documents $(1,\ldots,4)$ and six themes$\{\beta_1,\beta_2,\ldots\beta_6\}$. The root topic 1 will be shared by all documents, and they will each take their own path based on the nCRP. A document is created in the following way, as shown in Figure 9:

1) Let c_1 be the main topic.
2) Using CRP, for every level $l \in \{2,\ldots,L\}$, $c\{l-1\}$. At this level, set c_l to the root topic.
3) From $Dir(\alpha)$, create an L-dimension topic ratio vector τ.
4) For every word $n \in \{1,\ldots,N\}$, create $z \in \{1,\ldots,L\}$ according to multinomial distribution $Mult(\tau)$. Create w_n from the topic associated with topic c_z based on z.

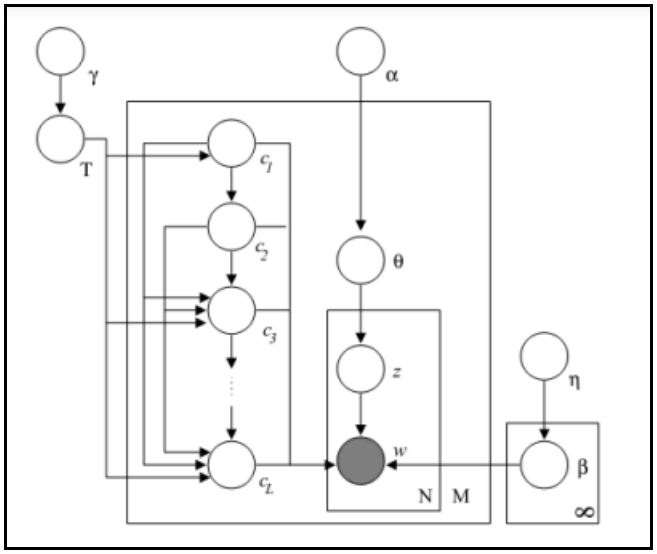

Fig. 9. The HLDA graphical representation.

Correlated Topic Modeling

CTM is a statistical method that illustrates the relationships between topics. For example, the topic "life" is more likely to be associated with "DNA" than with "aviation." LDA does not show the relationship between topics. Blei and Lafferty [49] presented correlated topic modeling as an addition to LDA. We use CTM to identify the subjects represented in a collection of documents.

In the CTM method, it is permissible for one latent subject to be associated with another latent topic. The logistic normal distribution yields a covariance matrix that describes the interdependency. Figure 10 depicts the CTM model in graphical form.

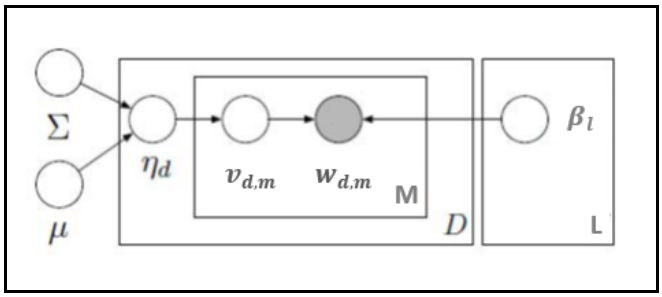

Fig. 10. The CTM graphical representation.

A process of an M-word document generative is as follows:

1) Choose $\theta \mid \{\mu,\Sigma\}$
2) For $m \in \{1,2,\ldots N\}$:
   a) Choose topic assignment $v_m|\theta$ from $Mult(f(\theta))$.
   b) Determine word $w_m|\{v_m,\beta_L\}$ from $Mult(f(\beta_{v_m}))$.

A K-dimensional mean and covariance matrix is represented by$\{\mu,\Sigma\}$. The CTM model's generating process is like that of the LDA model, except that the CTM model's

topic proportions are derived from a logistic normal distribution.

Non-Negative Matrix factorization

NMF [50] is a linear algebra optimization procedure for integrating high-dimensional data into a low-dimensional representation using a non-negative hidden framework, which is then presented as coordinate axes in the converted space using geometric views. In a nutshell, NMF describes a large input matrix as the product of two smaller matrices.

Assume that M and N are factorizations of the X matrix, such that X ≈ MN. To transmit information about X, M and N, use NMF's intrinsic clustering feature as follows:

- X represents the document-word matrix, the input containing the words that appear in particular documents.
- N represents the basic vectors, or clusters (topics) detected in the documents.
- M represents the coefficient matrix, the membership weights for each document's topics.

Optimization via an objective function (e.g., the EM algorithm [57]) can measure M and N, and then iteratively update both M and N to convergence. Using Euclidean distance, the given objective function calculates the rebuilding error between X and the product of its components N and M:

$$\frac{1}{2}\|X - MN\|_F^2 = \sum_{i=1}^{k}\sum_{j=1}^{h}\left(X_{ij} - (MN)_{ij}\right)^2$$

The values of M and N are obtained by deriving via the update rules using the objective function:

$$M_{ic} \leftarrow M_{ic}\frac{(XN)_{ic}}{(MNN)_{ic}} \qquad N_{cj} \leftarrow H_{cj}\frac{(MX)_{cj}}{(MMN)_{cj}}$$

The reconstruction error is recalculated, and the method is continued until convergence is reached with the updated M and N. Parallel procedures compute these updated values.

Neural Topic Models

A new field of study, the Neural Topic Model (NTM), emerged when topic modeling and deep neural networks intersected. A deep generative model, a variational autoencoder (VAE) [51], and amortised variational inferences (AVI) [58] were the most common things used in NTMs. Figure 11 shows that VAE-based NTMs are made up of a decoder and an encoder that are both based on neural networks. These represent the inference and generative mechanisms, respectively. The computational complexity of neural-topic models is lower than that of traditional Bayesian probabilistic topic models (BPTM); their deployment is more accessible thanks to several advanced deep learning architectures; and NTMs for prior-knowledge acquisition are simple to deploy using pre-trained word embeddings.

Some examples are the Neural Variational Document Model (NVDM) [59], the Dirichlet Variational Autoencoder topic model (DVAE) [60], the Neural Variational Latent Dirichlet Allocation (NVLDA) [52], the Dirichlet Variational Autoencoder (DirVAE) [61], iTM-VAE [62] and the Gaussian Softmax Model (GSM) [63]. This is not an exhaustive list, and it is still developing.

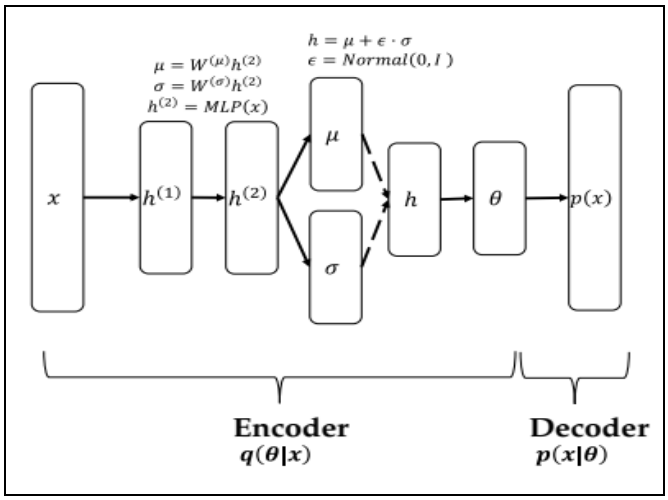

Fig. 11. Model based on the VAE.

In the broader design of the Variational autoencoder (VAE)-based NTMs, as shown in Figure 13, the encoder is a network component that produces θ, and the decoder is the part that takes θ and outputs. The encoder transforms the input BoW to a latent document–topic vector, and the decoder changes the document–topic vector to a given distribution of vocabulary words. The decoder seeks to recreate the input word distribution, so they are called autoencoders. VAE is formed by applying a transformation on a Gaussian distribution sampled by it.

C. Discussion

Natural language processing (NLP) widely employs topic modeling methods to uncover latent semantic patterns buried in large-scale corpora. Hirchoua et al. [64] designed the revolutionary ranking paradigm known as the ordered biterm topic model (OBTM). OBTM represents a semantic relationship between any two words or terms, regardless of whether they appear in the same brief text, enhancing the corpus' ability to discover actual semantic instances. They tested their model on three datasets: the question dataset, Arabic tweets, and the review collection. Cui et al. [65] presented a feature extractor for intrusive short text using a similar model (BTM) and developed a sensitive word characteristic data source from descriptions of Internet applications (aBTMd). Afterwards, they built a standard judgment for choosing the K value of the BTM topic model, which can automatically trigger self-adaptation. Also, Shah et al. [66] used BTM to group the comments on Instagram into specific topic clusters.

Researchers have refined the LDA model, effectively utilizing it in a variety of applications [67]. For instance, Liu [68] utilized LDA to generate a short text feature vector representation, taking into account the topic distribution. In addition, Chang et al. [69] provided a method for determining hotness based on the frequency of topic tags using an enhanced LDA topic model for detecting latent themes, which effectively solves the text sparsity problem. However, Chandrasekaran et al. [70] introduced the CARTMAN approach, which employs LDA to generate features from sensor data, thereby improving classification performance through machine learning algorithms for complex activity recognition.

Many LDA variants exist; for instance, Steuber et al. [71] used transfer learning to apply pre-trained word embedding to hashtags, which they then fed into the archetypal LDA (A-LDA) as supervising information or seed subjects. Yu et al. [72], on the other hand, showed the hierarchical sLDA model. This uses a better four-layer supervised LDA (sLDA) to group patient traits in HDRs as patient feature-value pairs in a one-hot way that is based on doctors' suggestions for a CHD lab test. This paper by Wang et al.[73] adds Neural Labeled LDA (NL-LDA), a new supervised topic modeling strategy for classification problems, to sLDA for semi-supervised document classification. This process is based on the VAE architecture and creates a unique generative network that incorporates prior knowledge. As a result, the suggested approach provides considerable advantages for semi-supervised document categorization with minimally labeled data.

The interaction between topic modeling and deep neural networks gave rise to a new field of study known as neural topic models. TopNet is one of NTM's algorithms proposed by Yang et al. [74], which takes advantage of recent improvements in neural topic models to generate skeletal words with high quality to supplement the brief input. Instead of immediately constructing a narrative, the model learns to transform the short input text into a low-dimensional topic distribution. AdditionalFurthermore, Qin et al. [75] introduced a method known as TACN (Topical Adversarial Capsule Network), which they divided into three The first part extracts the embedding representation from the topological structure, vertex characteristics, and document-topic distributions. We use the neural topic model with Gaussian Softmax construction to produce document topic distributions, ensuring a consistent back-propagation training procedure. The second section employs a prediction model to leverage the labels of vertices. The third section employs an adversarial capsule model to identify latent representations from the graph architecture domain, document-topic distribution domain, and vertex attribute domain.

Pathak et al. [76] suggested a topic-level sentiment analysis [77] based on deep learning. The proposed method stands out because it uses online latent semantic indexing at the sentence level with regularization constraints to identify the subject, followed by a topic-level attention mechanism in a long short-term memory (LSTM) network for sentiment analysis. However, Van Linh et al. [78] chose to add graph convolutional networks (GCN) to the LDA topic model in order to create a new graph convolutional topic model (GCTM). This model learns both the topic model and the networks for streaming data at the same time.

Using word embedding with topic models can help make topics easier to understand. For example, Murakami et al. [79] find that adding an extra fine-tuning step to create more document-specific themes or topics from short texts improves the subject coherence of short texts using an extra word embedding with a large external dataset. In the same way, Liu et al. [80] proposed a method for unsupervised text representation that combines WEs and expanded topic information. This approach contains two parts: weighted word embeddings (WWE) and extended topic information (ETI). They use Word2Vec and TF-IDF as weighted word embeddings (WWE), while they use the LDA topic model and word sequence extension as extended topic information (ETI). Truica et al. [81] proposed DOCTOPIC2VEC, a novel document-topic embedding model. Secondly, they developed DOC2VECs for each study, enhancing them with the local context that the WEs provided. Thirdly, they built topic embeddings called TOPIC2VEC utilizing three topic models, namely, LSA, LDA, and NMF, to improve the overall context of the SA. Finally, they created the new DOCTOPIC2VEC for each document and its dominant topic by appending the DOC2VEC to the TOPIC2VEC, which they constructed using the same WE.

However, Kim et al. [82] suggested W2V-LSA, an LSA based on Word2vec, as a new topic modeling method that uses Word2vec and spherical k-means clustering to better extract and show the corpus context. However, Chang et al. [83] combine a contextual BERT WE method with a spherical k-means clustering algorithm.

Generally, contextualized WEs consider the word's meaning, thereby improving the efficiency of the topic modeling task. For example, Wenfu et al. [84] employed the ALBERT model and LDA topic model to construct the topic vector and contextual vector of each word, thereby obtaining the document's detailed topic and semantic representations. Similarly, Bianchi et al. [85] combine contextualized BERT document embeddings in an NTM to generate more consistent topics from various English datasets. Habbat et al.[86] used AraBERT as an Arabic contextualized WE model and CamemBERT [87] as a French CWE with an NTM named ProdLDA to extract underlying topics from a Facebook page. The following table presents a detailed comparison of existing topic models.

TABLE II
A SYNOPSIS OF THE WORK ACCOMPLISHED ON TOPIC MODELING TECHNIQUES.

Year	Reference	Target language	Dataset/domain	Used Word embedding	Compared Topic models	Metrics	Results
2023	[88]	English	- Multi-Xscience - TAD and TAS2	Word decoder	- LexRank and TextRank - HeterSumGraph - Ext-Oracle: - TransS2S -BART -…	ROUGE-1 and ROUGE-L	A hierarchical decoding strategy gives the best results
2023	[89]	English	Twitter dataset related to COVID-19.	BERT	LDA + BERTopic	-	Using the combination of LDA and BERTopic gives the best result.

2022	[65]	English	Application descriptions dataset.	-	-aBTMd: Adaptive undesirable short text filtering framework based on Biterm Topic Modeling (BTM) - LDA	-The receiver operating characteristic (ROC) curve and area under the ROC curve (AUC). -TThe false alarm rate (FR).	aBTMd is better than LDA.
2022	[90]	English	5514 articles indexed in Web of Science	-	NMF		
2022	[64]	English Arabic	- Questions dataset -Arabic tweets - Reviews collection	-	-OBTM: The ordered biterm topic model - LDA, - CTM, -WNTM, -BPDTM	-topic coherence, -UMass	OBTM is better than other baseline models.
2022	[72]	English	patient feature-value pairs: Real-world clinical datasets gathered from the Cardiology Department of Xinjiang Medical University's First Affiliated Hospital.	TF-IDF	Hierarchical sLDA: Supervised latent Dirichlet allocation - Multi-class sLDA model.	-Accuracy, -Macro-F1, -Training time, -Testing time..	Hierarchical sLDA is better.
2022	[91]	English	Twitter dataset	TF	MVCS-VAT: Visual techniques include visual cluster tendency evaluation, cosine-based, and cosine similarity characteristics VAT (visual assessment of tendency) based on multi-viewpoint. -NMF, -LDA, -LSI, -PLSI.	-The cluster accuracy (CA), - Normalized mutual information (NMI), - Precision, - Recall.	MVCS-VAT is better than other models.
2022	[79]	English	BBC_news 20NewsGroup SearchSnippets TrecTweet Biomedical GoogleNews DBLP PascalFlicker StackOverflow	Glove, Word2vec, FastText	Neural-Topic Models: -NVDM -NVLDA -ProdLDA -GSB -RSB -WLDA -NSTM	- NPMI, -Word Embeddings Topic Coherence (WETC), -Topic diversity	NVDM and GSM showed good performances.
2022	[80]	English	-IMDB dataset -20 Newsgroups dataset	Word2Vec, TF-IDF	-WWE based on TF-IWF and Word2Vec, -ETI (based on the LDA and word sequence extension) -LDA W2V GLV FPW	-Accuracy, -Precision, -Recall, -F1 score	ETI method is better than LDA
2021	[81]	English	Game reviews	Word2vec, FastText,	DOCTOPIC2VEC, for document-level	Accuracy	DOCTOPIC2VEC is better than other

				GloVe	polarity detection using Different Topic models: LDA, LSI, NMF.		baselines.
2021	[84]	English	- AAPD: The arXiv Academic Paper Dataset, - RCV1: Reuters Corpus Volume I. - IMDB: The Internet Movie Database,	ALBERT	-LDA -TextCNN -XML-CNN - DTFEM-ML_KNN	-Accuracy, -Precision, -Recall, -F1 score -Subset accuracy (SA), -Hamming loss (HL),	Using ABERT improves the performance.
2021	[74]	English	-ROCStories	Glove	Neural Topic Model -LDA -Inc-Seq2seq -Skeleton Model -Fusion Model	Perplexity	Neural Topic Model is better than other models.
2021	[70]	English	The Ubicomp 08 Complex Activity dataset		CARTMAN: It uses LDA topic model to obtain features from sensor informations. - XGBoost - Multi-Layer Perceptron - AROMA - DeepConvLSTM	-Precision, -Recall, -F1 score	CARTMAN is better than other models.
2021	[92]	English	-20 Newsgroups dataset - AFP News dataset - AG News dataset -BBC news dataset	-BoW	-CSTM :the Common Sense Topic Model, -LDA -NMF -K-means	-NPMI - Word Embeddings coherence	CSTM is better than other models.
2021	[93]	English	-MNIST dataset, -The IMDB dataset -, Reuters Corpus Volume (IRCV1) datasets -Wiki dataset.	-	-DATM :deep autoencoding topic model. -LDA -OR-softmax - DocNADE - DPFA - AVITM - DLDA-Gibbs	-Perplexity, -Test time	DATM is better than other models.
2021	[94]	English	-Weibo dataset -Twitter dataset	TF-IDF	-LDA -k-means		LDA is better than the baselines.
2021	[73]	-English -German -Spanish	News article		-PLTM: Polylingual Topic Model -LDA -STM	-Coherence metrics, -Consistency metrics	PLTM is better than the baselines.
2021	[83]	English	Abstracts of papers related to geospatial data	BERT	-LSA -PLSA	-UMass -NPMI	LSA+BERT is better than PLSA
2021	[86]	Arabic	Aljazeera Facebook page.	-AraBERT -ELMO	-ProdLDA -LDA	-NPMI -Topic cohenrece -Perplexity	ProdLDA + AraBERT is better than other models.
2021	[95]	English	- Lectures dataset - Textbook dataset -Introduction dataset -Wiki dataset	TF-IDF	-Biclustering Technique to Topic modeling and Segmentation –BATS -LDA -HDP -LSA	-NPMI - Topic diversity -UMass -Runtime	BATS is better than the baselines.

			- Choi dataset -News dataset		-NMF		
2021	[78]	English	NYTtitle Yahoo-title TagMyNews-title Agnews-title, Twitter datasets	WordNet Word2vec	-GCTM: Novel graph convolutional topic model - Population variationalBayes - Streaming variationalBayes (SVB) - SVB Power prior -	- Log predictive probability -NPMI	GCTM is better than other models.
2021	[75]	English	Datasets of Scientific publications: - WebKB - Citeseer - Cora - Pubmed		-TACN: Topical Adversarial Capsule Network - M-NMF - LINE - LANE - DeepWalk - Node2vec - GraRep	-Average accuracy, - Precision, -Recall, -F1 score	TACN is better than the baselines.
2021	[69]	English	-Weibo -Tianya -China news		-An improved LDA - PLSA, -LDA, -BTM	- topic discovery	The improved LDA is better than other models.
2021	[66]	English	Instagram comments		BTM		
2021	[73]	English	- Yahoo Collection -20NewsGroups - IMDB dataset -AGNews		-NL-LDA: Neural Labeled LDA - Dependency-LDA, -TL-LDA, -SCHOLAR	-Tthe correct classificatio n rate (CCR) - Macro-F1, -Micro-F1.	NL-LDA is better than other models.
2020	[82]	English	-231 abstracts of blockchain-related papers.	Word2vec	-LSA -PLSA	- Accuracy, -Diversity of topics	LSA is better than PLSA.
2020	[85]	English	-20NewsGroup -Tweets2011 -Google News -StackOverflow dataset	SBERT	-ProdLDA, -Neural-ProdLDA, -NVDM, -LDA.	-NPMI -External word embeddings topic coherence, -Rankbiased overlap (RBO)	ProdLDA + SBERT is better than other models.
2020	[96]	English	-E-books dataset		-LSA -LDA	-UMass -UCI	LDA is better than LSA
2020	[97]	English	-StackOverflow -TagMyNews Title -Snippet dataset -Yahoo Answer		-NQTM: the Negative sampling and Quantization Topic Model, -LDA -BTM -DMM -SeaNMF -NVDM -GMS -ProdLDA -TMN	-NPMI -UMass	NQTM is better than other models.

Neural topic models (NTMs) have emerged and intend to use DNNs to improve the performance, efficiency, and usability of topic modeling. NTMs have ample research following due to their appealing versatility and scalability, generating over a hundred models and variants.

In addition, contextualized word embedding enhances the performance to produce more coherent topics.

V. Overall Discussion

This This paper thoroughly examines the development and expansion of topic modeling techniques, specifically in natural language processing (NLP). The article focuses on different models, their uses, and the creative methods

researchers have employed to analyze sentiment, represent documents, and extract topics from large-scale collections of texts.

1. Overview of Topic Modeling in Natural Language Processing: The paper initially discusses the importance of topic modeling in natural language processing (NLP), highlighting its ability to reveal hidden semantic patterns in large textual datasets.
2. Hirchoua et al. developed the Ordered Biterm Topic Model (OBTM), a groundbreaking ranking paradigm. It aims to depict semantic associations between words or phrases, regardless of whether they appear together in a text. This model is particularly notable for its ability to improve the identification of actual semantic occurrences in a variety of datasets, such as inquiries, Arabic tweets, and evaluations.
3. LDA variants: The paper examines the flexibility of the LDA model, demonstrating how various researchers have enhanced and modified it. Some examples of using LDA include generating concise text feature vectors, assessing popularity based on subject tags, and extracting features from sensor data for advanced activity recognition.
4. Neural Topic Models (NTM): We examine the combination of topic modeling and deep neural networks, leading to the creation of a novel discipline known as neural topic models. Within this paradigm, TopNet and TACN algorithms use neural topic models for a variety of purposes. These objectives include creating skeleton words and doing topic-level sentiment analysis.
5. The Fusion of Graph Convolutional Networks (GCN) with LDA: Researchers are investigating the integration of graph convolutional networks (GCN) and LDA to develop a new Graph Convolutional Topic Model (GCTM). This approach aims to acquire topic models and networks for simultaneously streaming data.
6. Word Embeddings with Topic Models: The paragraph examines the effectiveness of integrating word embeddings with topic models to produce easier-to-understand topics. Diverse methodologies are emphasized, such as incorporating a fine-tuning stage to address document-specific themes and leveraging unsupervised text representation with weighted word embeddings and expanded topic information.
7. Contextualized Word Embeddings: The exploration involves integrating contextualized word embeddings, specifically ALBERT and BERT, with topic models. This approach considers the contextual meaning of words, which enhances the effectiveness of topic modeling tasks and produces more comprehensive topic and semantic representations.
8. Innovative Models and Approaches: At the end of the article, a number of new methods are suggested. These include using contextualized Arabic word embeddings along with ProdLDA to pull topics out of social media content and using contextualized BERT document embeddings in neural topic models.

Fundamentally, this review presents an exhaustive examination of a wide range of topic modeling methodologies. It shows how basic models, like LDA, have given way to more complex and unified methods that use contextualized word embeddings and neural networks to better pull out topics in a wide range of linguistic and cultural settings.

Neural topic models (NTMs) exhibit the potential to capture intricate patterns and representations in textual data, although they encounter several obstacles. Neural topic models present several significant challenges:

1) Interpretability: The concern for interpretability is among the most significant obstacles presented by neural topic models. Conventional topic models, including LDA, provide topics that are easily understandable by their word distribution representations. On the contrary, neural network representations may acquire incredible intricacy and need more transparency, presenting a formidable obstacle in interpreting topics produced by NTMs.

2) Computing complexity: Neural topic models frequently incorporate intricate architectures, including deep neural networks. Training and inference in these models can be computationally intensive and time-consuming when dealing with large datasets. This complexity may result in scalability limitations for NTMs for parsing large corpora.

3) Data Efficiency: For practical training, neural networks and those used in topic modeling frequently require large quantities of labeled data. Lack of available data can hinder NTM efficacy, especially in domains where acquiring labeled data is challenging or expensive.

4) Sensitivity to hyperparameters: To achieve optimal performance, neural networks, including NTMs, generally have a multitude of hyperparameters that require tuning. Configuring these models can be difficult due to their sensitivity to hyperparameter selections; suboptimal selections may result in inadequate topic representations.

5) Overfitting: When using tiny datasets, neural networks are particularly susceptible to overfitting. Overfitting occurs when a model learns to discriminate noise in the training data, compromising its performance on new, unobserved data. Careful hyperparameter calibration and regularization techniques are required to prevent overfitting in NTMs.

6) Insufficient Topic Coherence: It is not easy to guarantee that the topics produced by an NTM are coherent and representative of significant themes. In contrast to conventional topic models, which generate coherent topics by nature, neural networks may produce more interpretable and coherent topics because of their complex non-linearities.

7) The necessity for pre-training on large-scale datasets: Pre-training on large-scale datasets is generally advantageous for many effective neural models, such as NTMs. The dependence on pre-training can pose difficulties when there is limited availability of broad and varied training data or where the specialization of the data in a particular domain is essential.

8) The integration of domain knowledge: The incorporation of specialized information from a particular field into neural topic models may present certain challenges. It is more difficult to incorporate prior knowledge into neural models as opposed to traditional models, which can do so readily via parameters such as Dirichlet priors.

9) Multimodal Difficulties: When dealing with data that contains many modes of information, such as text, graphics, and audio, incorporating these disparate modalities into a unified neural topic model presents extra difficulties. Achieving a comprehensive representation of the connections between various forms of data can be a challenging task.

When dealing with data that contains many modes of information, such as text, graphics, and audio, incorporating these disparate modalities into a unified neural topic model presents extra difficulties. Achieving a comprehensive representation of the connections between various forms of data can be a challenging task.

VI. Summary

The natural language processing community particularly widely uses topic modeling with LDA and its versions to manage vast amounts of unstructured data and classify them into specific topics. In this study, we conducted a detailed survey of various topic models such as LSA, PLSA, NMF, CTM, and LDA. Even though topic modeling has come a long way in the last two decades, there are still some unanswered concerns. These include:

1) Identifying the ideal quantity of themes or topics is crucial. Perplexity, harmonic mean, and cross-validation are among the methods used; however, the results still need clarification.
2) The selection of priors directly influences the quality of inferred topics and the probability of held-out documents. Which of these is the best, symmetric or asymmetric, and which is the most excellent fit for the problem? On the subject of priors, there is no substantial research available.
3) Which inference technique is best suited for the problem domain, such as Gibbs sampling and variational inference, has driven topic modeling for the past two decades, but which is the best is still unknown.

As topic modeling advances, many expanded versions of LDA based on related theories emerge, as well as neural topic models (NTMs), the most prevalent topic modeling research trend in the deep learning era. NTMs have a wide range of applications owing to their appealing flexibility, efficiency, and effectiveness. More flexibly than BPTMs, NTMs can offer topic distributions for documents and word occurrences for topics.

Since creating topic models, particularly NTMs, people have sought ways to incorporate external knowledge to enhance learning. Pre-trained language models (e.g., BERT) are a new approach to improving NTM performance by offering finer-grained, more advanced, and higher-level semantic information representations as contextual word embeddings.

This review provides readers with a comprehensive understanding of the essential elements of topic modeling and neural topic models, allowing them to better understand recent breakthroughs and gain insight into future research.

References

[1] Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, and Gao J, "Deep learning-based text classification: a comprehensive review," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp1-40, 2021. doi: 10.1145/3439726.

[2] Nouri H, and Sabri K, “Machine Learning Applications for Consumer Behavior Prediction,” *International Conference on Smart City Applications*, vol. 666, no. 675, pp666-675, 2022. doi: 10.1007/978-3-031-26852-6_62.

[3] Srivastava A, and Sahami M, "Text Mining: Classification, Clustering and Applications," *CRC Press*, vol.12, no.53, pp71-93, 2009. doi: 10.1201/9781420059458.

[4] George L E, and Birla L, “A Study of Topic Modeling Methods,” in 2018 *Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, vol. 109, no. 113, pp109-113, Jun. 2018. doi: 10.1109/ICCONS.2018.8663152.

[5] Vayansky I, and Kumar S A P, “A review of topic modeling methods,” *Information Systems*, vol. 94, pp101-582, 2020. doi: 10.1016/j.is.2020.101582.

[6] Xia Y, Luo D, Zhang C, & Wu Z, "A Survey of Topic Models in Text Classification," *the 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, vol. 51, pp244-250, 2019. doi: 10.1109/ICAIBD.2019.8836970.

[7] Kherwa P, and Bansal P, "Topic Modeling: A Comprehensive Review," *ICST Transactions on Scalable Information Systems*, vol. 23, pp159-623, 2018. doi: 10.4108/eai.13-7-2018.159623.

[8] Chen T H, Thomas S W, and Hassan A E,"A Survey on the Use of Topic Models When Mining Software Repositories. Empirical Software Engineering," *WIREs Data Mining and Knowledge Discovery*, vol. 21, no. 5, pp1843–1919, 2016. doi: 10.1007/s10664-015-9402-8.

[9] Abdouli A E, Hassouni L, and Anoun H," Mining Tweets of Moroccan Users Using the Framework Hadoop, NLP, K-means and Basemap," *Intelligent Systems and Computer Vision (ISCV)*, pp1–7, 2017. doi: 10.1109/ISACV.2017.8054907.

[10] Chaffai A, Hassouni L, and Anoun H," (2018). Informal Learning in Twitter: Architecture of Data Analysis Workflow and Extraction of Top Group of Connected Hashtags. In Big Data, Cloud and Applications," *Third International Conference BDCA*, vol. 4, no. 5, pp3-15, 2018 doi: 10.1007/978-3-319-96292-4_1.

[11] Habbat N, Anoun H, and Hassouni L," Topic Modeling and Sentiment Analysis with LDA and NMF on Moroccan Tweets," *Innovations in Smart Cities Applications*, vol. 4, no. 9, pp147–161, 2021. doi: 10.1007/978-3-030-66840-2_12.

[12] Sarumi O A, Adetunmbi A O, and Adetoye F A," Discovering Computer Networks Intrusion Using Data Analytics and Machine Intelligence," *Scientific African*, vol. 9, no. 500, pp20-36. doi: 10.1016/j.sciaf.2020.e00500.

[13] Habbat N, Anoun H, and Hassouni L,"Exploration, Sentiment Analysis, Topic Modeling, and Visualization of Moroccan Twitter Data," *International Conference on Advanced Intelligent Systems for Sustainable Development*, pp1067-1083, 2020. doi: 10.1007/978-3-030-90639-9_87.

[14] Habbat N, Anoun H, and Hassouni L," Sentiment Analysis and Topic Modeling on Arabic Twitter Data During Covid-19 Pandemic," *IJIAS*, vol. 2, no. 1, pp60–67, 2022. doi: 10.47540/ijias.v2i1.432.

[15] Qiang J, Qian Z, Li Y, Yuan Y, and Wu X,' Short Text Topic,' Modeling Techniques, Applications, and Performance: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 3, pp1427–1445, 2023. doi: 10.1109/TKDE.2020.2992485.

[16] Pal R, Sekh A A, Dogra D P, Kar S, Roy P P, and Prasad D K," Topic-based Video Analysis: A Survey," *ACM Computing Surveys*, vol. 54, no. 6,pp1-34, 2021. doi: 10.1145/3459089.

[17] Syed A A, Gaol F L, Suparta W, Abdurachman E, Trisetyarso A, and Matsuo T,"Prediction of the Impact of Covid-19 Vaccine on Public Health Using Twitter," I*AENG International Journal of Computer Science*, vol. 49, no. 1, pp19-29, 2022.

[18] Jelodar H," Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey," *Multimedia Tools and Applications*, vol. 78, no. 11, pp15169-15211, 2021. doi: 10.1007/s11042-018-6894-4.

[19] Zhao H, Phung D, Huynh V, Jin Y, Du L, and Buntine W," Topic Modelling Meets Deep Neural Networks: A Survey" *the Thirtieth International Joint Conference on Artificial Intelligence* ,vol. 43, no. 5, pp4713-4720, 2021. doi: 10.24963/ijcai.2021/638.

[20] Mikolov T, Chen K, Corrado G, and Dean J," Efficient Estimation of Word Representations in Vector Space" *arXiv*, vol. 3, no.1301, pp37-81, 2013. doi: 10.48550/arXiv.1301.3781.

[21] Devlin J, Chang M W, Lee K, and Toutanova K," BERT: Pre-training of Deep Bidirectional Transformers for Language

Understanding," *arXiv*, vol. 18, no. 10, pp48-55, 2019. doi: 10.48550/arXiv.1810.04805.

[22] Radford A, Narasimhan K, Salimans T, and Sutskever I," Improving Language Understanding by Generative Pre-Training", 2018.

[23] Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, and Le Q V," XLNet: Generalized Autoregressive Pretraining for Language Understanding," *Advances in Neural Information Processing Systems*, vol. 32,no. 12, pp125-142, 2019. doi: 10.48550/arXiv.1906.08237.

[24] Hicham N, Nassera H, Karim S," A Thorough Analysis of E-commerce Customer Reviews in Arabic Language Using Deep Learning Techniques for Successful Marketing Decisions". *IAENG International Journal of Applied Mathematics*, vol. 53, no. 4, pp1540-1547, 2023.

[25] Pennington J, Socher R, and Manning C," GloVe: Global Vectors for Word Representation," *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp1532-1543, 2014. doi: 10.3115/v1/D14-1162.

[26] Joulin A, Grave E, Bojanowski P, Douze M, Jégou H, and Mikolov T," FastText.zip: Compressing Text Classification Models," *arXiv*, 2016. arXiv:1612.03651.

[27] Le Q V, and Mikolov T," Distributed Representations of Sentences and Documents, *arXiv*, 2014. arXiv:1405.4053.

[28] Peters M E," Deep Contextualized Word Representations," *arXiv*, 2018. doi: 10.48550/arXiv.1802.05365.

[29] Mikolov T, Sutskever I, Chen K, Corrado G S, and Dean J," Distributed Representations of Words and Phrases and Their Compositionality," *Advances in Neural Information Processing Systems*, vol. 26, no. 25, 2013. doi: 10.48550/arXiv.1310.4546.

[30] Naili M, Chaibi A H, and Ben Ghezala H," Comparative Study of Word Embedding Methods Topic Segmentation," *Procedia Computer Science*, vol. 112, no. 87, pp340-349, 2017. doi: 10.1016/j.procs.2017.08.009.

[31] Nandakumar N, Salehi B, and Baldwin T,"A Comparative Study of Embedding Models in Predicting the Compositionality of Multiword Expressions," *the Australasian Language Technology Association Workshop* , vol. 15, no. 51, pp71–76, 2018.

[32] Saurav K, Saunack K, Kanojia D, and Bhattacharyya P,"A Passage to India': Pre-trained Word Embeddings for Indian Languages," *1st Joint Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*,2020. doi: 10.48550/arXiv.2112.13800.

[33] George L, and Sumathy P," An Integrated Clustering and BERT Framework for Improved Topic Modeling," *International Journal of Information Technology*, vol. 15, no. 4, pp2187–2195, 2023. doi: 10.1007/s41870-023-01268-w.

[34] Sebbaq H, and Faddouli N E," MTBERT-Attention: An Explainable BERT Model Based on Multi-Task Learning for Cognitive Text Classification," *Scientific African*, vol. 21, no. 13, pp87-99, 2023. doi: 10.1016/j.sciaf.2023.e01799.

[35] Haghighian R A, Afshar J, Lee W, and Lee S," PatentNet: Multi-label Classification of Patent Documents Using Deep Learning-Based Language Understanding," *Scientometrics*, vol. 127, no. 1, pp207–231, 2022. doi: 10.1007/s11192-021-04179-4.

[36] Gautam A V, and Masud S," Fake News Detection System Using XLNet Model with Topic Distributions: CONSTRAINT@AAAI2021 Shared Task," *International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situations* , vol. 17, no. 5, pp1-18, 2021. doi: 10.1007/978-3-030-73696-5_18.

[37] Chiorrini A, Mircoli A, Diamantini C, and Potena D," Emotion and Sentiment Analysis of Tweets Using BERT," *EDBT/ICDT Workshops* , vol. 3, 2021.

[38] Wang Y, Zheng J, Li Q, Wang C, Zhang H, and Gong J," XLNet-Caps: Personality Classification from Textual Posts," *Electronics*, vol. 10, no. 11,pp43-60, 2021. doi: 10.3390/electronics10111360.

[39] Adoma A F, Henry N M, and Chen W," Comparative Analyses of BERT, RoBERTa, DistilBERT, and XLNet for Text-Based Emotion Recognition," *17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, vol. 55, no. 23, pp117-121, 2020 doi: 10.1109/ICCWAMTIP51612.2020.9317379.

[40] Wang B, and Kuo C J," SBERT-WK: A Sentence Embedding Method by Dissecting BERT-Based Word Models," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 2, no. 8, pp2146-2157, 2020. doi: 10.1109/TASLP.2020.3008390.

[41] Rahman W," Integrating Multimodal Information in Large Pretrained Transformers," *58th Annual Meeting of the Association for Computational Linguistics*, vol. 56, no. 25, pp2359-2369, 2020. doi: 10.18653/v1/2020.acl-main.214.

[42] Aßenmacher M, and Heumann C," On the Comparability of Pre-trained Language Models," *5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS)*, 2020 doi: https://doi.org/10.48550/arXiv.2001.00781.

[43] dos S, da Silva A, da Silva N F F, & da Silva S A," Deep Learning Brasil - NLP at SemEval-2020 Task 9: Overview of Sentiment Analysis of Code-Mixed Tweets," *arXiv*, 2020. doi: https://doi.org/10.48550/arXiv.2008.01544.

[44] Landauer T K, Foltz P W, and Laham D," An Introduction to Latent Semantic Analysis," *Discourse Processes*, vol. 25, no. 3, pp259-284, 1998. doi: 10.1080/01638539809545028.

[45] Deerwester S, Dumais S T, Furnas G W, Landauer T K, and Harshman R," Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp391-407, 1990. doi: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9.

[46] Hofmann T," Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 42, no. 14, pp177-196, 2001 doi: 10.1023/A:1007617005950.

[47] Blei D, and Jordan M I," Latent Dirichlet Allocation," *Advances in Neural Information Processing Systems* vol. 14, no. 13, pp154-166, 2001.

[48] Griffiths T L, Jordan M I, Tenenbaum J B, and Blei D M," Hierarchical Topic Models and the Nested Chinese Restaurant Process," *Advances in Neural Information Processing Systems*, vol. 16, 2003.

[49] Blei D M, and Lafferty J D," A Correlated Topic Model of Science," *Annals of Applied Statistics*, vol. 1, no. 1,pp17-35, 2007. doi: 10.1214/07-AOAS114.

[50] Arora S, Ge R, Kannan R, and Moitra A," Computing a Nonnegative Matrix Factorization - Provably," *Forty-Fourth Annual ACM Symposium on Theory of Computing*, 2012. doi: 10.1137/130913869.

[51] Kingma D P, and Welling M," Auto-Encoding Variational Bayes," *arXiv*, 2014. doi: 10.61603/ceas.v2i1.33.

[52] Srivastava A, and Sutton C," Autoencoding Variational Inference for Topic Models," *arXiv*, 2017. doi: https://doi.org/10.48550/arXiv.1703.01488.

[53] Duan X," Cleaning of Transient Fault Data in Distribution Network Based on Clustering by Fast Search and Find of Density Peaks," *Engineering Letters*, vol. 31, no. 4, pp763-775, 2023.

[54] Hicham N, Karim S, Habbat N," Enhancing Arabic Sentiment Analysis in E-Commerce Reviews on Social Media Through a Stacked Ensemble Deep Learning Approach," *Mathematical Modelling of Engineering Problems*, vol. 10, no. 3, pp790-798, 2023. doi: 10.18280/mmep.100308.

[55] Anupriya P, and Karpagavalli S," LDA-Based Topic Modeling of Journal Abstracts," *International Conference on Advanced Computing and Communication Systems* ,pp1-5, 2015. doi: 10.1109/ICACCS.2015.7324058.

[56] Blei D M, Griffiths T L, and Jordan M I," The Nested Chinese Restaurant Process and Bayesian Nonparametric Inference of Topic Hierarchies," *Journal of the ACM*, vol. 57, no. 2, pp1-30, 2010. doi: 10.1145/1667053.1667056

[57] Krishnan K, and McLachlan G J, "The EM Algorithm," *Handbook of computational statistics: concepts and methods*, pp139-172, 2012. doi: 10.1002/9780470191613.

[58] Rezende D J, Mohamed S, and Wierstra D," Stochastic Backpropagation and Approximate Inference in Deep Generative Models," *arXiv*, 2014. doi: https://doi.org/10.48550/arXiv.1401.4082

[59] Miao Y, Yu L, and Blunsom P," Neural Variational Inference for Text Processing," *arXiv*, 2016. doi: https://doi.org/10.48550/arXiv.1511.06038

[60] Burkhardt S, and Kramer S," Decoupling Sparsity and Smoothness in the Dirichlet Variational Autoencoder Topic Model," *Journal of Machine Learning Research*, vol. 20, no. 131,pp1-27, 2019.

[61] Joo W, Lee W, Park S, and Moon I C," Dirichlet Variational Autoencoder," *arXiv* , 2019. doi: 10.1016/j.patcog.2020.107514

[62] Ning X, Zheng Y, Jiang Z, Wang Y, Yang H, and Huang J," Nonparametric Topic Modeling with Neural Inference," *arXiv*, 2018. doi: 10.1016/j.neucom.2019.12.128

[63] Miao Y, Grefenstette E, and Blunsom P," Discovering Discrete Latent Topics with Neural Variational Inference," *arXiv*, 2018.

[64] Hirchoua B, Ouhbi B, and Frikh B," Topic Modeling for Short Texts: A Novel Modeling Method," *AI and IoT for Sustainable Development in Emerging Countries: Challenges and Opportunities*, pp573–595, 2022. doi: 10.1007/978-3-030-90618-4_29

[65] Cui D," A BTM-Based Adaptive Objectionable Short Text Filtering Framework," *Wireless Communications and Mobile Computing*, vol. 12, no. 135, pp1–12, 2022. doi: 10.1155/2022/6668344

[66] Shah N, Li J, and Mackey T K," An Unsupervised Machine Learning Approach for the Detection and Characterization of Illicit Drug-Dealing Comments and Interactions on Instagram," *Substance Abuse*, vol. 43, no. 1, pp273–277, 2022. doi: 10.1080/08897077.2021.1941508

[67] Albalawi R, Yeap T H, and Benyoucef M," Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis," *Frontiers in Artificial Intelligence*, vol. 3, no. 42, pp122-135, 2020. doi: 10.3389/frai.2020.00042

[68] Luo L," Network Text Sentiment Analysis Method Combining LDA Text Representation and GRU-CNN," *Personal and Ubiquitous Computing*, vol. 23, no. 3, pp405–412, 2019. doi: 10.1007/s00779-018-1183-9

[69] Chang L, and RuiLin H," Hot Topic Discovery across Social Networks Based on Improved LDA Model," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 15, no. 11, 2021. doi: 10.3837/tiis.2021.11.004

[70] Chandrasekaran K, Gerych W, Buquicchio L, Alajaji A, Agu E, and Rundensteiner E," CARTMAN: Complex Activity Recognition Using Topic Models for Feature Generation from Wearable Sensor Data," *IEEE International Conference on Smart Computing (SMARTCOMP)*, pp39–46, 2021. doi: 10.1109/SMARTCOMP52413.2021.00026

[71] Steuber F, Schneider S, and Schoenfeld M," Embedding Semantic Anchors to Guide Topic Models on Short Text Corpora," *Big Data Research*, vol. 27,no. 13, pp107-333, 2022. doi: 10.1016/j.bdr.2021.100293

[72] Yu G, Zhang L, Zhang Y, Zhou J, Zhang T, and Bi X," Prediction and Risk Stratification from Hospital Discharge Records Based on Hierarchical sLDA," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, pp14-32, 2022. doi: 10.1186/s12911-022-01747-3

[73] Wang W, Guo B, Shen Y, Yang H, Chen Y, and Suo X," Neural Labeled LDA: A Topic Model for Semi-Supervised Document Classification," *Soft Computing*, vol. 25, no. 16, pp14561-14571, 2021. doi: 10.1007/s00500-021-06310-2

[74] Yang Y, Pan B, Cai D, and Sun H," TopNet: Learning from Neural Topic Model to Generate Long Stories," *27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* , pp. 1997–2005, 2021. doi: 10.1145/3447548.3467410

[75] Qin X, Rao Y, Xie H, Wang J, and Wang F L," TACN: A Topical Adversarial Capsule Network for Textual Network Embedding," *Neural Networks*, vol.144, no. 45, pp766–777, 2021. doi: 10.1016/j.neunet.2021.09.026

[76] Pathak A R, Pandey M, and Rautaray S," Topic-Level Sentiment Analysis of Social Media Data Using Deep Learning," *Applied Soft Computing*, vol.108, no.10, pp74-90, 2021. doi: 10.1016/j.asoc.2021.107440

[77] Hicham N, Karim S, and Habbat N," Customer Sentiment Analysis for Arabic Social Media Using a Novel Ensemble Machine Learning Approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol.13, no.4, pp4504-4515. doi: 10.11591/ijece.v13i4.pp4504-4515

[78] Van L N, Bach T X, and Than K," A Graph Convolutional Topic Model for Short and Noisy Text Streams," *arXiv*, 2021. doi: https://doi.org/10.48550/arXiv.2003.06112

[79] Murakami R, and Chakraborty B," Investigating the Efficient Use of Word Embedding with Neural-Topic Models for Interpretable Topics from Short Texts," *Sensors*, vol.22, no.3, pp852-867, 2022. doi: 10.3390/s22030852

[80] Liu W, Pang J, Du Q, Li N, and Yang S," A Method of Short Text Representation Fusion with Weighted Word Embeddings and Extended Topic Information," *Sensors*, vol.22, no.3, pp1066-1075, 2022. doi: 10.3390/s22031066

[81] Truică C O, Apostol E S, Șerban M L, and Paschke A," Topic-Based Document-Level Sentiment Analysis Using Contextual Cues," *Mathematics*, vol.9, no.21, pp2722-27219. doi: 10.3390/math9212722

[82] Kim S, Park H, and Lee J," Word2vec-Based Latent Semantic Analysis (W2V-LSA) for Topic Modeling: A Study on Blockchain Technology Trend Analysis," *Expert Systems with Applications*, vol.152, no.145, pp113401-113435, 2020. doi: 10.1016/j.eswa.2020.113401

[83] Cheng Q," Bert-Based Latent Semantic Analysis (Bert-LSA): A Case Study on Geospatial Data Technology and Application Trend Analysis," *Applied Sciences*, vol.11, no.24, pp11897-11921, 2021. doi: 10.3390/app112411897

[84] Liu W, Pang J, Li N, Zhou X, and Yue F," Research on Multi-Label Text Classification Method Based on tALBERT-CNN," *International Journal of Computational Intelligence Systems*, vol.14, no.1, pp201-233, 2021. doi: 10.1007/s44196-021-00055-4

[85] Bianchi F, Terragni S, and Hovy D," Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence," *59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* , vol.2, no.14, pp759–766, 2021 doi: 10.18653/v1/2021.acl-short.96

[86] Habbat N, Anoun H, and Hassouni L," Extracting Topics from a TV Channel's Facebook Page Using Contextualized Document Embedding. International Archives of the Photogrammetry," *Remote Sensing and Spatial Information Sciences*, vol.4, no.5, pp245–249, 2021. doi: 10.5194/isprs-archives-XLVI-4-W5-2021-245-2021

[87] Habbat N, Anoun H, Hassouni L, and Nouri H," Using Neural Topic Model and CamemBERT to Extract Topics from Moroccan News in the French Language," *AIP Conference Proceedings*, vol.2814, no.1, pp20004-20017, 2023. doi: 10.1063/5.0148733

[88] Wang P, Li S, Liu S, Tang J, Wang T," Plan and Generate: Explicit and Implicit Variational Augmentation for Multi-Document Summarization of Scientific Articles," *Information Processing & Management*, vol.60, no.4, pp103409-103437, 2023. doi: 10.1016/j.ipm.2023.103409

[89] Ebeling R, Nobre J, and Becker K," A Multi-Dimensional Framework to Analyze Group Behavior Based on Political Polarization," *Expert Systems with Applications*, vol.233, no.120, pp768-789, 2023. doi: 10.1016/j.eswa.2023.120768

[90] Zhu L, and Cunningham S W," Unveiling the Knowledge Structure of Technological Forecasting and Social Change (1969–2020) through an NMF-Based Hierarchical Topic Model," *Technological Forecasting and Social Change*, vol.174, no.12, pp1277-1293, 2022. doi: 10.1016/j.techfore.2021.121277

[91] Narasimhulu K, Abarna K T M, Kumar B S, and Suresh T," A Novel Sampling-Based Visual Topic Models with Computational Intelligence for Big Social Health Data Clustering. *The Journal of Supercomputing*, vol.15, no.73, pp1654-1673, 2022. doi: 10.1007/s11227-021-04300-7

[92] Harrando I, and Troncy R," Discovering Interpretable Topics by Leveraging Common Sense Knowledge," *11th on Knowledge Capture Conference*, pp265–268, 2021. doi: 10.1145/3460210.3493586

[93] Zhang H, Chen B, Cong Y, Guo D, Liu H, and Zhou M,"Deep Autoencoding Topic Model with Scalable Hybrid Bayesian Inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.43, no.12, pp4306–4322, 2021. doi: 10.1109/TPAMI.2020.3003660

[94] Yang Y, Hsu J H, Löfgren K, and Cho W," Cross-Platform Comparison of Framed Topics in Twitter and Weibo: Machine Learning Approaches to Social Media Text Mining," *Social Network Analysis and Mining*, vol.11, no.1, pp75-94, 2021. doi: 10.1007/s13278-021-00772-w

[95] Wu Q," BATS: A Spectral Biclustering Approach to Single Document Topic Modeling and Segmentation," *arXiv*, 2021. doi: 10.1145/3468268

[96] Mohammed S H, and Al-augby S," LSA & LDA Topic Modeling Classification: Comparison Study on E-Books," *Indonesian Journal of Electrical Engineering and Computer Science*, vol.19, no.1, pp353–362, 2020. doi: 10.11591/ijeecs.v19.i1.pp353-362

[97] Wu X, Li C, Zhu Y, and Miao Y," Short Text Topic Modeling with Topic Distribution Quantization and Negative Sampling Decoder," *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp1772–1782, 2020. doi: 10.18653/v1/2020.emnlp-main.138