Novel Preprocessing for Diverse LDA Topic Modeling

Anonymous ACL submission

Abstract

This study introduces a novel preprocessing approach that applies dependency parsing to extract noun and verb heads, which are then used to generate unigram and n-gram representations. We investigate the trade-off between topic coherence and diversity in topic modeling, demonstrating how increased diversity enhances text pattern discovery. Using three preprocessing methods to train LDA models [3], we find that while coherence decreases slightly, topic diversity increases significantly, leading to the identification of novel patterns. By prioritizing topics with multi-word complements, our approach improves result granularity and highlights the role of diversity in uncovering deeper textual structures. To further validate these findings, we recommend additional diversity metrics.

1 Introduction

Topic modeling is an unsupervised machine learning techniques that aims to uncover subjects and classify large text corpora automatically. Modern topic modeling methods use either statistical and/or probabilistic approaches or leverage existing word embeddings and language models in order to extract insights from unstructured text.

Topic modeling has made significant progress in the past several years, with teams exploring new possibilities such as embedding-based models [7], [15], [26], [22] and n-gram-based preprocessing methods [30], [17], [16], [20] for both embedding-based models as well as the Latent Dirichlet Allocation (LDA) topic modeling algorithm [3].

1.1 Challenges

The LDA algorithm is classic in the domain of topic modeling and has been used with various preprocessing techniques, but all these techniques have their limitations. Embedding-based approaches and n-gram-based approaches both have the same goal, increasing context for better topic quality and more comprehensible results. In this study, we focus on optimizing preprocessing techniques for LDA topic modeling in the hopes of addressing concerns about the limitations of the algorithms as voiced by teams such as [28] and [8].

We show that by combining n-gram-based methods with syntactic dependency information, we can create n-gram document representations that are both smaller than standard n-gram approaches, and contain more relevant information.

This paper is organized as follows: an introduction to LDA and a brief overview of n-gram and multi-word isolation preprocessing techniques, a detailed explanation of syntactic n-grams and how we integrated them into our approach, a comparison of topic quality between standard approaches and our novel approach, and future perspectives to further optimize the approach.

2 LDA

The algortithm chosen to demonstrate this preprocessing technique is the classic LDA topic-modeling algorithm[3], which was chosen due to its prevalence in the topic-modeling field and will serve as a way to establish a baseline performance for the custom topic-modeling preprocessing approach.

This algorithm represents the text corpus as a Bag-of-Words, i.e. the vocabulary found in the corpus is processed without taking into account word order. As this is a limitation of the approach, many others have developed new methods in order to optimize the algorithm by adding neighboring words to the representations of the tokens [30], extracting expressions [20], and even combining domain-specific ontologies to filter corpora [18].

2.1 N-Grams

Accounting for n-grams and multi-word expressions is not new in topic modeling. Several approaches have used them for different purposes, including using n-gram recognition to find new terms and repeated terms. In [1], a novel approach was developed to find what can be characterized as complex stopwords. This approach find n-grams that are repeated throughout the corpus and systematically removes them. This approach has a similar goal of rendering the topic model's output readable and relevant.

A second interesting approach comes from [20]. With their novel approach, domain-specific n-grams and multi-word expressions are extracted in the preprocessing stages in order to increase the relevance of the extracted topics.

3 SN-Grams

SN-grams are a concept first discussed by [23] in 2013. Two separate applications were discussed by the team, namely authorship attribution [24] and English as a second language grammar correction [23]. The concept differs from traditional n-gram preprocessing techniques such as bigram & trigram approaches and skipgram approaches [5].

The main difference between n-gram techniques and sn-gram techniques is which elements are considered to be neighbors. In a traditional n-gram approach, the neighbors are simply the next and/or previous tokens in a sequence following a sliding window. However, in a sn-gram approach, the neighbors are found using a syntactic dependency tree.

As an example, take the sentence, "The quick brown fox jumped over the large lazy dog". This sentence is the shortest sentence using all letters of the English language, to which I have added an extra adjective to demonstrate the approach.

Using a traditional approach, we would find that the bigrams generated for this sentence would be: *the quick, quick brown, brown fox, fox jumped, jumped over, over the, the large, large lazy, lazy dog*.

However, this approach does not take into account words that are syntactic neighbors and not direct neighbors in the sentence. Using the sn-gram approach, we can link nouns and their complements by parsing the syntactic tree and finding the matching dependencies. For example, bigrams can be formed from individual adjectives and the phrasal noun head. Bigrams can also be formed using the verb and its complement, in order to take actions into account as well.

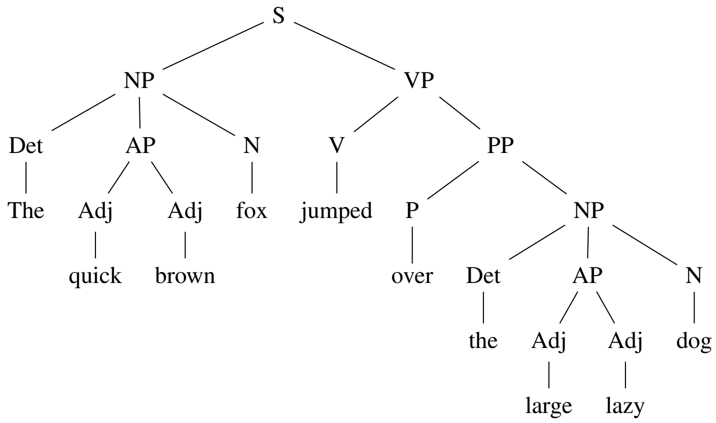

Figure 1: Syntactic Tree

Using the syntactic tree, we can create new, more relevant bigrams, such as *quick fox* and *large dog*, which is not possible using a traditional approach.

Since the original article describing sn-grams, this approach has been more frequently used in projects such as Semantic N-Gram Topic Modeling using point wise mutual information (PMI) and log frequency based mutual dependency (LGMD) [17] to find likely sn-grams. Another approach, Dependency-Based Open Information Extraction [11] seeks to enhance machine text understanding by introducing flexible, syntax-based sn-gram data structures to an unsupervised text-understanding algorithm that seeks to improve semantic information given in the preprocessing stage.

4 Novel Canonical NP & VP Processing

4.1 Context

Our novel approach is inspired by previous work involving syntactic dependency parsed language data for use on downstream tasks such as [11],[6], and [29]. While our novel approach is similar to the SN-gram approach detailed by [23], we focus solely on nouns associated with their complement(s) and verbs associated with their complement(s).

In a traditional bigram approach, we consider that a bigram is formed by a word and the word immediately following it. This representation allows for a greater context window to be explored as well as two-word terms, such as "renewable energy" to be extracted. However, there are several downsides to this approach. It produces a large vocabulary, can be costly to compute for large corpora, and cannot represent nouns and verbs with multiple complements and/or complements that are found at a distance of several tokens.

As of the writing of this paper, there is no consensus as to the best method to extract bigrams, trigrams, and n-grams from corpora while preserving the maximum amount of context and informative words. Several teams have proposed their solutions, including: [20], [30], [9], [1], and [16]. These approaches highlight the need for a standardized n-gram topic modeling approach while also demonstrating significant progress being made when n-grams are considered for preprocessing LDA input.

4.2 Methodology

The downsides discussed above are directly addressed by our new method. By creating unigrams (single-token terms) from single nouns and verbs, as well as bigrams & trigrams from individual nouns and/or verbs attached to their complements, we are able to reduce the size of the corpus by more than 40% compared to a traditional bigram approach. Using the syntactic dependency tree to extract syntactic heads (nouns and verbs) as well as all of their complements, the only limit at this stage is the accuracy of the dependency parse.

In line with similar approaches, we automatically filter out stopwords and do not keep determinant+noun pairs in our training corpus. We further lemmatize all nouns and verbs in order to reduce the vocabulary and reduce the amount of topics that are produced containing similar or inflected versions of terms found in other topics. This results in an increase in the number of verbs being found across topic-words, as forms like "be use" are created automatically following the syntactic bigram extraction and further lemmatization.

As a concrete example, we can take the following article on heat transfer characteristics, [21], present in our corpus extracted from Semantic Scholar[1]. The abstract is as follows:

> ’One of the long term renewable energy conversion methods is an ocean thermal energy conversion (OTEC) operating on a closed Rankine cycle that is typically composed of a boiler, condenser, pump and turbine. As well known, since the OTEC cycle efficiency is quite low, the improvement of boiler and condenser heat exchanger efficiency is very important. Over the past three decades, many new working fluids such as R1234yf have been suggested and are available in the market for the use of OTEC power generation. In this paper, boiling and condensation heat transfer characteristics of commercially available eight working fluids are predicted and compared for the design of high efficiency boiler and condensers of the future closed OTEC power plants. The results show that R32 has the best heat transfer and environmental properties among the fluids compared.’

[1] https://www.semanticscholar.org/

In a traditional bigram approach, the bigrams generated would be:

> one-of, of-the, the-long, long-term, term-renewable, renewable-energy, energy-conversion, conversion-methods, methods-is, is-an, an-ocean, ocean-thermal, thermal-energy, energy-conversion, conversion-(otec), (otec)-operating, operating-on, on-a, a-closed, closed-rankine, rankine-cycle...

This approach generates 135 bigrams, not all of which are semantically relevant to a topic-model approach. Bigrams such as of-the, is-an, for-the, in-this, etc only serve to add noise to the training corpus. It is possible to remove stopwords, and establish lists of n-grams to remove from the corpus, as was done by [30], but there is no current standardized preprocessing for this approach. As such, our demonstration in the next section will use this bigram approach as a comparison against our novel approach.

Using our novel preprocessing approach, the bigrams become:

> operate-on, be-compose, turbine, as-know, over-suggest, in-predict, compare, close-plant, show-compare, have-transfer, compare, long-method, term-method, renewable-method, energy-method, conversion-method, method, ocean-conversion, thermal-conversion, energy-conversion, conversion, otec, closed-cycle, rankine-cycle...

This approach, combining relevant noun and verb unigrams with bigrams & trigrams generated with base noun/verb and complement pairs contains significantly less n-grams than a traditional bi-gram approach. In this specific example (chosen at random from the corpus), we decrease the number of n-grams generated from 135 to 76. This is a 43.7% decrease in the quantity of n-grams to be given to the LDA algorithm, which allows for faster processing times at scale when compared to a classic bigram approach.

5 LDA Topic Modeling

All LDA models in this study were trained on a uniform corpus of 300,000 abstracts sourced from articles retrieved through a search for "renewable energy" using the Semantic Scholar API. All models are set to generate 40 topics, ensuring a broad and representative sample of the field. To maintain a fair comparison, all models were trained with the same configuration. The key difference across models was the preprocessing techniques applied to the data. The three models are as follows:

1. Classic Bag-of-Words approach: all words are lemmatized and stopwords are removed.

2. Classic bigram approach: the corpus is divided into bigrams, no further preprocessing occurs.

3. Novel syntactic dependency approach: based on the dependency parse, noun and verb heads become unigrams, and bigrams & trigrams are formed by nouns and verbs plus their complements.

The results unveil some striking findings, particularly in terms of topic diversity. Despite being trained on identical data, these different preprocessing approaches led to substantial differences in the range and distinctiveness of the topics generated. This indicates that preprocessing choices can significantly influence the granularity and breadth of topic modeling results. Moreover, the findings suggest a potential trade-off between topic diversity and other evaluation metrics, such as coherence and perplexity. Models with greater topic diversity often showed lower coherence scores, implying less tightly clustered terms within each topic. Similarly, an increase in topic diversity was sometimes associated with higher perplexity values, reflecting potential challenges in predicting unseen data due to the added complexity of a broader set of topics. These insights highlight the possibility of prioritizing topic diversity when preprocessing techniques are tailored to specific research objectives, even if it means accepting a trade-off in coherence and perplexity.

6 Comparison

Our novel approach is shown to be particularly useful in creating unique topic clusters. Using the PyLDAvis library [25], we can visualize topic similarity on a two-dimensional plane providing an intuitive understanding of the models' topic diversity and coherence.

In comparing three different preprocessing approaches, distinct patterns emerged. The first approach, shown in Figure 2, resulted in all topics being tightly clustered together, suggesting low variation in the subjects covered. This outcome indicates that the topics were relatively homogeneous, with significant overlap in content, potentially limiting the model's ability to uncover nuanced differences within the corpus.

The second approach, shown in Figure 3, which used classic bigrams without further preprocessing, showed higher variation in the topic distribution, with clusters more spread out across the visualization. However, this increased diversity came at the cost of interpretability, as the model generated a higher number of "nonsense" topics dominated by stopwords and irrelevant terms. This suggests that while the model captured more varied themes, the inclusion of common and semantically weak terms reduced the overall quality of the topics.

In contrast, the third approach in Figure 4 produced the highest variation, with well-spaced clusters indicating distinct topic groups. This preprocessing technique balanced stopword removal and text normalization methods, such as stemming or lemmatization, to refine the content while preserving key thematic elements. The result was a more diverse set of topics with minimal overlap and better-defined boundaries, highlighting the value of our novel preprocessing strategy for creating clear and meaningful topic clusters in large text corpora. Finally, when analyzing the topics themselves, we find that the terms are more semantically relevant and allow us to distinguish more specific energy categories such as wind power, hydroelectric power, battery technology, and electric vehicles.

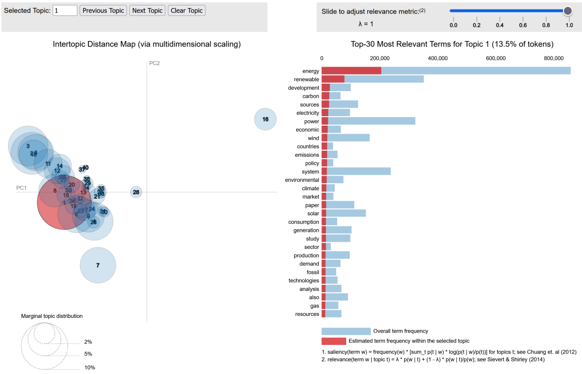

Figure 2: Classic Unigram Approach

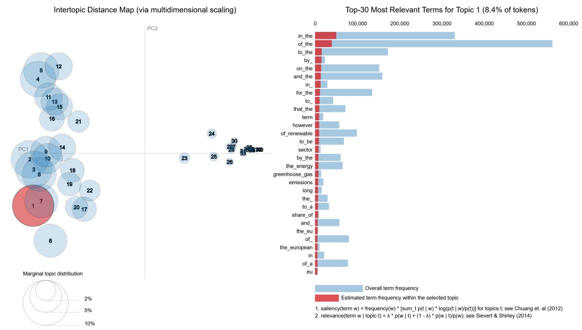

Figure 3: Classic Bigram Approach

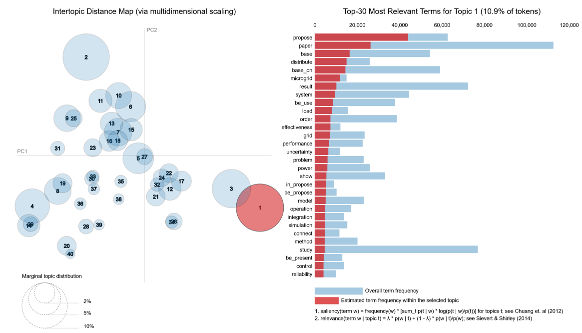

Figure 4: Novel N-gram Approach

6.1 Evaluation

Here we argue that while the classic coherence scores are slightly lower for our novel approach than with for the unigram and classic bigram approaches, the novel approach manages to show its usefulness in terms of topic diversity. We start by comparing the proportion of unique words/terms generated in the 40 topics analyzed. If a term is present in the top 10 words of more than one topic, or multiple times in the top 10 words of a single topic, the measurement decreases. A perfect 100% would mean that all top 10 words × 40 topic slots are filled with unique terms, with no repetition across or within topics. The results of this analysis are in Table 1.

In addition to this simple measurement, we will use the average Jaccard similarity, discussed by [27], and defined as:

$$JS(T_i, T_j) = \frac{|T_i \cap T_j|}{|T_i \cup T_j|}$$

where:

- $T_i \cap T_j$ represents the intersection of the two sets (common elements), and $|T_i \cap T_j|$ is its size.
- $T_i \cup T_j$ represents the union of the two sets (all unique elements), and $|T_i \cup T_j|$ is its size.

The numerator $|T_i \cap T_j|$ measures the shared elements, while the denominator $|T_i \cup T_j|$ normalizes the similarity by the total number of unique elements. The resulting value lies between 0 (no overlap) and 1 (identical sets). This measurement will be computed as an average of the Jaccard similarity between all topics of each topic model.

In order to have a more robust analysis of topic diversity within our three models, we will use a third topic diversity metric. Explained in [2], the topic diversity measurement:

> Inversed Rank-Biased Overlap (p) evaluates how diverse the topics generated by a single model are. We define p as the reciprocal of the standard RBO (Webber et al., 2010; Terragni et al., 2021b). RBO compares the 10-top words of two topics. It allows disjointedness between the lists of topics (i.e., two topics can have different words in them) and uses weighted ranking. I.e., two lists that share some of the same words, albeit at different rankings, are penalized less than two lists that share the same words at the highest ranks. p is 0 for identical topics and 1 for completely different topics.

We have also questioned the relevance of automatic evaluation methods such as topic coherence, whose downsides have been highlighted by [10]:

> Traditional topic quality metrics are not robust to stopwords. We next show that two standard measures of topic quality—coherence and PMI—perform counterintuitively in situations in which the

corpus contains many common but irrelevant words. This situation is common in many real corpora, where there is standard vocabulary that is often repeated in the text but is generally uninformative. That said, our analysis does not invalidate the use of these measures in cases where the vocabulary has been carefully curated for relevance. After a discussion of coherence and PMI, we introduce another metric, log lift, that alleviates these found concerns in the case of the stopword problem

We will therefore analyze the traditional coherence measurements against the topic diversity measurement for the models. We use the c_uci coherence measurement first proposed by [19] and implemented with the Gensim library [31]. The results of this comparison are in Table 1. We will then compare coherence topic by topic in order to fully comprehend the potential gains in diversity alongside the potential losses in coherence of this new approach.

As shown by Table 1, we see that coherence and topic diversity are inversely related when examining the models. When comparing our approach against the classic unigram approach, we see an astounding 195.54% increase in topic diversity. Inversely, we see a decrease in coherence that may undermine the advantages of this new approach.

As for the classic bigram model, we do see a higher level of diversity, but as a large portion of the topic words are semantically empty, we do not see a high coherence measure. This words include *development_of, in_the, and_the, of_power*. This further highlights the need for a more robust framework for LDA evaluation as well as highlights the need for a unified stopword-removal approach for bigrams automatically extracted from corpora.

When evaluating the coherence seen in the model topic by topic in A, we find that although many topics have very low coherence values when compared to the standard unigram model, some have similar values to the more coherent model. This suggests that coherence can be improved with further research into preprocessing techniques, hyperparameter setting, and evaluation techniques.

7 Conclusions & Future Perspectives

This project presents a novel approach to unigram and bigram creation for topic model preprocessing, but does have certain limitations. As demonstrated, the new model's coherence has greater variability, which means that extracted topics need to be manually evaluated in order to determine their value. It would be interesting to find ways to improve topic coherence while also preserving the significant increase to topic diversity that this approach has demonstrated.

As this project relied on [13], and the en_core_web_sm model, it would be interesting to judge performance of other, more recent models, as well as the well-known Stanford dependency parser [4] to find the optimal configuration for large-scale implementation of the discussed preprocessing techniques.

A potential future project would be to compare embedding-based approaches using the proposed preprocessing method in order to see if this approach can improve existing embedding-based algorithms such as BERTopic[12].

Finally, as highlighted by our results as well as sources such as [10] and [14], coherence as a measure needs to robustly evaluated and novel evaluation measures need to be created and tested.

8 Limitations

While our syntactic n-gram preprocessing approach shows promise for increasing topic diversity, several limitations should be noted. First, the dependency parsing required for generating syntactic n-grams adds significant computational overhead compared to traditional n-gram approaches. This may limit scalability for very large corpora.

Second, the quality of the syntactic n-grams depends heavily on the accuracy of the dependency parser. Parsing errors can propagate through to the topic model and impact results. This is particularly relevant for technical or domain-specific texts where parsers may struggle with specialized terminology.

Third, while our approach increases topic diversity, it comes at some cost to topic coherence as measured by standard metrics. Further work is needed to better understand this trade-off and potentially develop new evaluation metrics that can better capture both coherence and diversity simultaneously.

Fourth, our current implementation only considers noun-complement and verb-complement relationships when generating syntactic n-grams. Other potentially valuable syntactic relationships are not

Table 1: Coherence vs. Diversity Among Topic Modeling Approaches

Model	Coherence	Unique Words/Topic (%)	Pairwise Jaccard Distance	IRBO
Classic Unigram	0.03	28.00	0.79	0.58
Classic Bigram	-1.65	79.25	0.97	0.96
Novel N-gram	-2.11	82.75	0.99	0.98

captured. Expanding to additional dependency types could yield further improvements but would increase complexity.

Finally, our evaluation focused on English language texts. The effectiveness of this approach for other languages, particularly those with substantially different syntactic structures, remains to be investigated. Additional language-specific adjustments may be needed for optimal performance across languages.

These limitations suggest several promising directions for future work while highlighting important considerations for practitioners applying this approach.

References

[1] Mohamad Almgerbi, Andrea De Mauro, Adham Kahlawi, and Valentina Poggioni. Improving Topic Modeling Performance through N-gram Removal. pages 162–169.

[2] Federico Bianchi, Silvia Terragni, and Dirk Hovy. *Pre-Training Is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence*.

[3] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. 3:993–1022.

[4] Danqi Chen and Christopher Manning. A Fast and Accurate Dependency Parser using Neural Networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750. Association for Computational Linguistics.

[5] Winnie Cheng, Chris Greaves, and Martin Warren. From n-gram to skipgram to concgram. 11(4):411–433.

[6] Elnaz Delpisheh and Aijun An. Topic Modeling Using Collapsed Typed Dependency Relations. In *Advances in Knowledge Discovery and Data Mining*, pages 146–161. Springer International Publishing.

[7] Adji B. Dieng, Francisco J. R. Ruiz, and David M. Blei. Topic Modeling in Embedding Spaces.

[8] Roman Egger and Joanne Yu. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. 7:886498.

[9] given-i=HARISHKUMAR family=S, given=HARISHKUMAR and Bhuvaneshwaran R. S. Enhanced DGA Detection in BotNet Traffic: Leveraging N-Gram, Topic Modeling and Attention BiLSTM.

[10] Angela Fan, Finale Doshi-Velez, and Luke Miratrix. Assessing topic model relevance: Evaluation and informative priors. 12(3):210–222.

[11] Pablo Gamallo, Marcos Garcia, and Santiago Fernández-Lanza. Dependency-Based Open Information Extraction. In *Proceedings of the Joint Workshop on Unsupervised and Semi-Supervised Learning in NLP*, pages 10–18. Association for Computational Linguistics.

[12] Maarten Grootendorst. BERTopic: Neural topic modeling with a class-based TF-IDF procedure.

[13] Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

[14] Alexander Hoyle, Pranav Goel, Andrew Hian-Cheong, Denis Peskov, Jordan Boyd-Graber, and Philip Resnik. Is Automated Topic Model Evaluation Broken? The Incoherence of Coherence. In *Advances in Neural Information Processing Systems*, volume 34, pages 2018–2033. Curran Associates, Inc.

[15] Zhuolin Jiang, Manaj Srivastava, Sanjay Krishna, David Akodes, and Richard Schwartz. Combining Word Embeddings and N-grams for Unsupervised Document Summarization. *Preprint*, arXiv:2004.14119.

[16] Noriaki Kawamae. Supervised N-gram topic model. pages 473–482.

[17] Pooja Kherwa and Poonam Bansal. Semantic N-Gram Topic Modeling. 0(0):163131.

[18] Hyon Hee Kim and Hey Rhee. An Ontology-Based Labeling of Influential Topics Using Topic Network Analysis. 15:1096–1107.

[19] David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 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.

[20] Michael Nokel and Natalia Loukachevitch. Accounting ngrams and multi-word terms can improve topic models. In *Proceedings of the 12th Workshop on Multiword Expressions*, pages 44–49. Association for Computational Linguistics.

[21] Myungkyu Park, Joo-Hyung Kim, and Dongsoo Jung. Heat Transfer Characteristics of Working Fluids for Closed OTEC Power Plants. OnePetro.

[22] Sattar Seifollahi, Massimo Piccardi, and Alireza Jolfaei. An Embedding-Based Topic Model for Document Classification. 20(3):52:1–52:13.

[23] Grigori Sidorov. Syntactic dependency based n-grams in rule based automatic English as second language grammar correction. 4(2):169–188.

[24] Grigori Sidorov, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, and Liliana Chanona-Hernández. Syntactic Dependency-Based N-grams as Classification Features. In *Advances in Computational Intelligence*, Lecture Notes in Computer Science, pages 1–11. Springer.

[25] Carson Sievert and Kenneth Shirley. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 63–70. Association for Computational Linguistics.

[26] Silvia Terragni, Elisabetta Fersini, and Vincenzina Messina. Word Embedding-Based Topic Similarity Measures. pages 33–45.

[27] Nam Khanh Tran, Sergej Zerr, Kerstin Bischoff, Claudia Niederee, and Ralf Krestel. *Topic Cropping: Leveraging Latent Topics for the Analysis of Small Corpora*, volume 8092.

[28] Ike Vayansky and Sathish A.P. Kumar. A review of topic modeling methods. 94:101582.

[29] Xuerui Wang, Andrew McCallum, and Xing Wei. Topical N-Grams: Phrase and Topic Discovery, with an Application to Information Retrieval. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 697–702.

[30] Linkai Zhu, Maoyi Huang, Maomao Chen, and Wennan Wang. An N-gram based approach to auto-extracting topics from research articles.

[31] Radim Řehůřek and Petr Sojka. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50. ELRA.

A Full Coherence

Topic	Classic Unigram	Classic Bigram	Novel N-gram
0	-1.84	-2.53	-1.85
1	-1.56	-7.95	-2.03
2	-1.96	0.58	-2.18
3	-1.27	-2.12	-5.42
4	-2.31	-0.20	-2.03
5	-1.46	0.89	-2.51
6	-1.37	0.57	-3.47
7	-1.58	-0.13	-2.84
8	-1.89	-7.25	-13.22
9	-1.45	0.19	-6.58
10	-1.29	2.19	-3.97
11	-1.93	-6.10	-10.48
12	-1.50	-0.20	-16.97
13	-1.60	-5.84	-18.45
14	-1.48	0.70	-5.15
15	-2.50	-5.55	-13.16
16	-1.99	0.71	-15.84
17	-2.20	-0.08	-1.88
18	-2.23	0.01	-2.12
19	-2.02	0.38	-2.38
20	-1.33	-5.85	-1.50
21	-1.50	0.17	-2.51
22	-1.25	0.10	-9.13
23	-1.43	-3.33	-12.25
24	-2.32	0.21	-2.04
25	-1.65	0.52	-3.59
26	-1.37	0.08	-10.82
27	-1.73	0.90	-2.07
28	-1.54	-4.77	-1.85
29	-1.49	-0.81	-18.44
30	-1.27	-5.61	-2.18
31	-1.27	0.36	-9.41
32	-1.67	0.18	-11.12
33	-1.58	-4.85	-2.56
34	-1.24	-0.05	-5.67
35	-1.38	-0.17	-2.60
36	-1.29	0.35	-3.35
37	-1.31	-4.99	-2.17
38	-1.65	0.42	-15.95
39	-1.34	-7.17	-13.61