



Article LDA-Based Topic Modeling Sentiment Analysis Using Topic/Document/Sentence (TDS) Model

Akhmedov Farkhod¹, Akmalbek Abdusalomov¹, Fazliddin Makhmudov¹ and Young Im Cho^{2,*}

- ¹ Department of IT Convergence Engineering, Gachon University, Sujeong-Gu, Seongnam-Si 461-701, Gyeonggi-Do, Korea; farhodsam92@gmail.com (A.F.); akmaljon@gachon.ac.kr (A.A.); fazliddin12@gachon.ac.kr (F.M.)
- ² Department of Computer Engineering, Gachon University, Sujeong-Gu, Seongnam-Si 461-701, Gyeonggi-Do, Korea
- Correspondence: yicho@gachon.ac.kr

Abstract: Customer reviews on the Internet reflect users' sentiments about the product, service, and social events. As sentiments can be divided into positive, negative, and neutral forms, sentiment analysis processes identify the polarity of information in the source materials toward an entity. Most studies have focused on document-level sentiment classification. In this study, we apply an unsupervised machine learning approach to discover sentiment polarity not only at the document level but also at the word level. The proposed topic document sentence (TDS) model is based on joint sentiment topic (JST) and latent Dirichlet allocation (LDA) topic modeling techniques. The IMDB dataset, comprising user reviews, was used for data analysis. First, we applied the LDA model to discover topics from the reviews; then, the TDS model was implemented to identify the polarity of the sentiment from topic to document, and from document to word levels. The LDAvis tool was used for data visualization. The experimental results show that the analysis not only obtained good topic partitioning results, but also achieved high sentiment analysis accuracy in document- and word-level sentiment classifications.

Keywords: topic modeling; sentiment analysis; sentiment classification; polarity detection

1. Introduction

Since 2000, a great deal of research has been conducted on consumers' opinions and sentiments due to the increase in online commercial applications. Business sectors and organizations have put substantial efforts in determining consumers' opinions about the product or service they offer. This is because customers' decision-making process is significantly affected by the opinion of people around them (friends, family, etc.). As Bing et al. explained in [1], in research literature it is possible to see various names, e.g., "sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, and review mining"; however, all of them have a similar purpose and belong to the subject of sentiment analysis or opinion mining. Due to this, the importance of sentiment analysis (SA) is being realized in a range of different domains, such as consumer products, services, healthcare, political science, social sciences, and financial services [1–3]. A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence or feature level whether the expressed opinion in a document, a sentence or an entity feature is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as enjoyment, anger, disgust, sadness, fear, and surprise [4]. Meanwhile, with the rapid growth in the volumes of text data produced in recent years, the need for text mining has increased to discover hidden knowledge and insights from text data. Texts comprise meaningful words that describe people's opinions. Opinions are usually subjective expressions that describe people's feelings and sentiments toward entities,



Citation: Farkhod, A.; Abdusalomov, A.; Makhmudov, F.; Cho, Y.I. LDA-Based Topic Modeling Sentiment Analysis Using Topic/Document/Sentence (TDS) Model. *Appl. Sci.* **2021**, *11*, 11091. https://doi.org/10.3390/app112311091

Academic Editor: Antonio Fernández-Caballero

Received: 27 September 2021 Accepted: 21 November 2021 Published: 23 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). events, and properties [5]. Sentiment analysis, also known as opinion mining, is a task that uses natural language processing (NLP), text analysis, and computational techniques to automate the extraction or classification of general public's positive, negative, and neutral emotions about the products and services they have used. However, finding the sentiment polarity or opinion mining from a large amount of textual data is an overwhelming task. For example, there is a lack of capability in dealing with complex sentences; and the existing SA techniques do not perform well in certain domains and have inadequate polarity detection and accuracy. A detailed survey in [6,7] explained the challenges of SA and possible techniques that can be used to solve each problem. ASA task is considered a sentiment classification problem [8]. Sentiment classification is one of the most commonly studied areas for most researchers. Sentiment classification techniques can be divided into machine learning (ML), lexicon-based, and hybrid approaches [9]. ML approaches use various linguistic features for sentiment classification. The lexicon-based approach implements a sentiment lexicon. The hybrid approach, as the name suggests, combines both approaches. Thus, sentiment lexicons play an important role in most sentiment classification methods. In terms of different levels, the classification can be divided into document-level, sentencelevel, and aspect-based classification [10]. Sentence-level classification classifies sentiment expressions into every sentence. For this step, the subjectivity of the sentence is identified, and the classification process determines whether the sentence is positive or negative. An aspect-based level classifies the sentiments based on certain aspects of entities, such as identifying the entities and their aspects. To perform this analysis, supervised and unsupervised ML algorithms can be chosen.

To address the aforementioned problems, we present a robust and reliable, real-time sentiment analysis framework based on unsupervised machine learning approach. A motivation for this work was developed by reading JST (Joint Sentiment Topic model to be explained subsequently) model implementation in sentiment classification. JST will detect sentiment values of the topics without considering any external sentiment labels. This leads JST to slight drawbacks in sentiment classification accuracy, because without any external sentiment labels, topics discovery will suffer. To better topic quality, we used LDA (Latent Dirichlet Allocation) model which performs significantly well in inter- and intra-mixing distributional documents from multiple topics.

The main contributions of this paper are as follows:

- Increasing automatic discovery of topics from data or corpus by a joining proposed method with LDA approach.
- Providing more precise sentiment representation over topics, documents and words by integrating accurate topic and document discovery.

The remainder of the paper is organized as follows: Section 2 reviews existing conventional studies for the sentiment analysis. Section 3 presents the proposed sentiment analysis approach in detail. The experimental results based on IMDB databases are discussed in Section 4. Section 5 highlights certain limitations of the proposed method. Finally, Section 6 concludes the paper by summarizing our findings and future research directions.

2. Related Works

Sentiment analysis is one of the hardest tasks in natural language processing because even humans struggle to analyze sentiments accurately. Data scientists are getting better at creating more accurate sentiment classifiers, but challenges of machine-based sentiment analysis remaining exist. In general, existing systems related to sentiment analysis technologies can be divided into two categories: traditional sentiment analysis approaches based on computer vision and AI-based sentiment analysis systems using machine learning (ML) and deep learning (DL). In this section, we focus mainly on discussing the aforementioned AI based approach, which are appropriate only under certain conditions. However, these features are not sufficient to establish accurate classification from text analysis. To overcome these limitations, additional sentiment attributes are required, such as the feedback data, multilingual efficacy, speed and scale, social media and multimedia [11].

2.1. Computer Vision and Image Processing Approaches for Sentiment Analysis

Ortis et al. [12] introduced the research field of image sentiment analysis, reviewed the related problems, provides an in-depth overview of current research progress, discusses the major issues and outlines the new opportunities and challenges in this area. The first paper on Visual Sentiment Analysis aims to classify images as "positive" or "negative" and dates back to 2010 [13]. In this work, the authors studied the correlations between the sentiment of images and their visual content. They assigned numerical sentiment scores to each picture based on their accompanying text (i.e., meta-data). Udit et al. [14] developed an improved sentiment analysis method using image processing techniques based on visual data. Other research related to aspect-based sentiment analysis (ABSA) classify the sentiment of a specific aspect in a text presented in [15–17]. Earlier research has shown that these approaches are appropriate only under certain conditions.

2.2. Artificaial Intelligence Approaches for Sentiment Analysis

In recent years, DL approaches have been significantly and effectively implemented in sentiment analysis research areas in different ways. In contrast to the techniques reviewed earlier that rely on handcrafted characteristics, ML and DL approaches can automatically identify, extract, quantify, and analyze complicated point features. Another benefit is that deep neural networks can be implemented flexibly and successfully in automatic feature extraction using learned data or analyzed customer's feedback data; instead of spending time extracting functions, they can be modified to create a robust database and an appropriate network structure.

There has been a great deal of research focus on the problem of sentiment classification. Supervised and unsupervised ML approaches are separately used by most researchers to classify SA. In addition, supervised and unsupervised approaches can also be combined to analyze the sentiment. In [18], authors used supervised and unsupervised methods together in the case of proposing meta-classifiers to develop a polarity classification system. Furthermore, in [19], unsupervised learning algorithm was applied to automatically categorize text, which resulted in training sets by using keyword lists. They classified documents into a certain number of predefined categories. The purpose of their study was to overcome the problems associated with the creation of labeled training documents and manually categorizing them. To evaluate the proposed methods, they embodied a traditional system by supervised learning using the same naïve Bayes classifier, and then tested and compared the performance check.

Similar research was conducted by Turney et al. [20], applies an unsupervised learning algorithm to classify the semantic orientation regarding mutual information between document phrases and a small set of positive/negative paradigm words. Lin et al. [21] proposed a framework based on the LDA approach called JST to detect sentiments and topics simultaneously. Their JST model is also an unsupervised ML algorithm. In another interesting study, Adnan et al. used a statistical approach in the feature selection process as detailed in [22]. The hidden Markov model (HMM) and LDA method were used to separate the entities in a review document from the subjective expressions according to the polarity of those entities. The proposed scheme achieved competitive results for document polarity classification. Although some approaches have applied unsupervised and semi-supervised learning methods [23,24], using supervised learning techniques for aspect-based sentiment analysis is also a very popular concept in machine learning [25]. The LDA model allows documents to be explained by latent topics. There is also a very applicable supervised ML approach called latent semantic indexing (LSI) [26]. LSI is a well-known feature selection method that attempts to reduce the dimensionality of the data by transforming the text space into a new axis system. However, compared with LSI, LDA has a better statistical foundation to define the topic-document distribution θ , by allowing inferences in new documents based on previously estimated models, and avoids the overfitting problem.

3. Proposed Method

Pang et al. [6] mentioned that the sentiment classification problem is comparatively more challenging than the traditional topic-based classification because sentiments can be expressed in a more subtle manner while topics can be identified more easily with respect to the co-occurrence of keywords. According to the Appraisal group, the improvement of sentiment polarity detection accuracy is related to incorporating prior information or the subjectivity lexicon.

In this study, we propose an unsupervised ML TDS approach to determine sentiment polarity first at the topic level, and then at document and word levels. We applied the LDA feature selection model to discover the topics in the IMDB dataset [27]. To visualize topic distribution, we used the LDAvis data visualization tool, which was developed by Carson et al. [28], which revealed aspects of the topic–term relationships, including topical distance calculation, number of clusters, terms, and value of lambda. Lambda (λ) determines the weight given to the probability of term under the topic relative to its lift (measuring both on the log scale) and the value of lambda used to compute the most relevant terms for each topic. The overall workflow of this study is shown in Figure 1.



Figure 1. Workflow of proposed method.

3.1. Introduction to LDA

LDA is a well-known method for topic modeling. First introduced by David et al. in [29], LDA shows topics using word probabilities. LDA is an unsupervised generative probabilistic model of a corpus. The main task of LDA is that documents are represented in a random mixture over latent topics, where a topic is characterized by a distribution over words [30]. LDA assumes that every document can be represented as a probabilistic distribution over latent topics, as shown in Figure 2. In this process, the topic distribution in all documents shares a common Dirichlet prior. Each latent topic represented in the LDA model is also represented as a probabilistic distribution over words, and the word distributions of topics share a common Dirichlet prior. Given a corpus *D* consisting of *M* documents, with document *d* having N_d words ($d \in 1, ..., M$), LDA models *D* according to the following generative process [31]:

- (a). Choose a multinomial distribution φ_t for topic t (t $\in \{1, ..., T\}$) from a Dirichlet distribution with parameter β .
- (b). Choose a multinomial distribution θ_d for document d ($d \in \{1, \dots, M\}$) from a Dirichlet distribution with parameter α .
- (c). For a word w_n ($n \in \{1, ..., N_d\}$) in document d,
 - i selection of a topic z_n from θ_d ,
 - ii selection of a word w_n from φ_{zn} .



Figure 2. Latent Dirichlet Allocation model.

Regarding the generative process above, words in documents are only observed variables, while others are latent variables (φ and θ) and hyper parameters (α and β). To infer the latent variable and hyper parameters, the probability of the observed data *D* is computed and maximized as follows:

$$p(D|\alpha,\beta) = \prod_{d=1}^{M} \int p(\theta_d|a) (\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn},\beta)) d\theta_d$$
(1)

The α parameters of the topic Dirichlet prior and the distribution of words over topics, which are drawn from the Dirichlet distribution, are given by β . T is the number of topics, M is the number of documents, and N is the size of the vocabulary. The Dirichlet multinomial pair for corpus-level topic distributions (α , β) was considered. The Dirichlet multinomial pair for topic-word distributions is given by β and φ . Variables θ_d are the document-level variables. z_{dn} and w_{dn} are word-level variables that are sampled for each word in each text document.

There have been many studies related to topic models using LDA in different fields, such as topic modeling in linguistic science, political science, medical and biomedical fields, and other research areas [32–34].

3.2. Topic Document Sentence (TDS) Model

The existing LDA framework represents three hierarchical layers, where topics are associated with documents and words are associated with topics. However, many researchers [21] have implemented a joint sentiment topic (JST) unsupervised ML model by adding an additional layer to the LDA model. More specifically, sentiment labels of JST are associated with documents, topics are associated with sentiment labels, and words are associated with both sentiment labels and topics of the entire corpus. The difference between TDS and LDA is that LDA supports three layers for processing, while TDS has one additional sentiment label layer for giving higher classified sentiment. It is important

to note that, other than the one additional layer, the TDS model is similar to the JST model. However, what distinguishes the TDS model from the JST model is the implementation of sentiment label analysis in three parts of sentiment analysis sections, such as topic sentiment analysis, document sentiment analysis, and word sentiment analysis.

Let us assume that we have a corpus with a collection of **D** documents denoted by $\mathbf{D} = \{d_1d_2d_3, \ldots d_M\}$, and the corresponding vocabulary of the document collection is denoted by $\mathbf{V} = \{w_1w_2w_3, \ldots w_V\}$. Each document in the corpus is a sequence with N_d words denoted by $\mathbf{d} = \{w_{d1}w_{d2}w_{d3}, \ldots w_{N_d}\}$, where *d* is the document number of the collection d_D , *V* is the number of words, and w_{N_d} is the **n**th word in document **d**, $w_{N_d} \in V$. Assume that *S* is the number of distinct sentiment labels, and *T* is the total number of topics.

The generative process shows in Figure 3a representing of the original JST model [35]. Like the JST model, the proposed TDS model development parameters are as follows:



(a) JST graphical model

(b) TDS graphical model

Figure 3. Graphical models of JST and TDS.

For every document d, choose a distribution $\pi_d \sim Dir(\gamma)$. For each sentiment label l under document d, choose a distribution $\theta_{d,l} \sim Dir(\alpha)$. For each word ω_i in document d

- choose a sentiment label $l_i \sim \pi_d$,
- choose a topic $z_i \sim \theta_{d,l_i}$,
- choose a word ω_i from the distribution over words defined by the topic z_i and sentiment label l_i , $\varphi_{z_i}^{l_i}$.

The JST model has two determinant parameters, α and β , for the document collection layer. In this example, α is a prior parameter of topic distribution in the document collection, and β is the word distribution of every latent topic. Sentiment label 1 is associated with the number of topics, and parameter γ can be integrated into 1 in corpus observations. The JST consists of three sets of latent variables:

- JST with document distribution— θ ,
- JST with word distribution— φ ,
- Sentiment document distribution— π .

As mentioned in [18], π plays an important role in identifying the document polarity. In our model implementation, π is the main change. It is worth noting that sentiment document distribution is applied to the total number of topics, T stage.

Formula implementation of TDS model with joint sentiment.

$$P(w,z,l) = P(w|z,l)P(z,l) = P(w|z,l) * P(z|l,d)P(l|d)$$
(2)

For the first term, by integrating out φ , we obtain:

$$P(w|z,l) = \left(\frac{\Gamma(V\beta)}{\Gamma(\beta)^{V}}\right)^{T*S} \prod_{j} \prod_{k} \frac{\prod_{i} \Gamma\left(N_{i,j,k} + \beta\right)}{\Gamma\left(N_{j,k} + V\beta\right)}$$
(3)

where *V* is the size of the vocabulary, *T* is the total number of topics, *S* is the total number of sentiment labels, and $N_{i,j,k}$ subscripts are used to loop for the number of times the word i appears in topic j and for sentiment label k. $N_{j,k}$, is the number of times words are assigned to topic j and sentiment k, and Γ is the gamma function.

The remaining terms of Equation (4) are obtained by integrating out the term— θ :

$$P(w|z,l) = \left(\frac{\Gamma(Ta)}{\Gamma(a)^T}\right)^{S*D} \prod_k \prod_d \frac{\prod_i \Gamma(N_{j,d,k} + a)}{\Gamma(N_{k,d} + Ta)}$$
(4)

In Equation (4), S is the total number of sentiment labels, D is the total number of documents in the collection, $N_{j,d,k}$ is the number of times a word from document d has been associated with topic j and sentiment label k. $N_{k,d}$ is the number of times sentiment label k has been assigned to some word tokens in document d.

For the fifth and sixth terms, integrating out π :

$$P(l|d) = \left(\frac{\Gamma(S\gamma)}{\Gamma(\gamma)^{S}}\right)^{D} \prod_{d} \frac{\prod_{k} \Gamma(N_{k,d} + \gamma)}{\Gamma(N_{d} + S_{\gamma})}$$
(5)

$$P(s|z, y) = \left(\frac{\Gamma(\sum_{k} y_{k})}{\prod_{k} \Gamma(\gamma k)'}\right)^{T} \prod_{j} \frac{\prod_{j} \Gamma(n_{j,k} + y_{k})}{\Gamma(n_{j} + \sum_{k} y_{k})}$$
(6)

where D is the total number of documents in the collection, $N_{k,d}$ is the number of times sentiment label k has been assigned to some word tokens in document d. N_d is the total number of words in the document collection.

4. Experimental Results and Analysis

This section presents the topic modeling and experimental setup of sentiment polarity classification based on the IMDB dataset. We implemented and tested the proposed method in Visual Studio 2019 C++ on a PC with a 3.20-GHz CPU, 32 GB of RAM, and two Nvidia GeForce 1080Ti GPUs.

4.1. Preprocessing the Dataset

Data cleaning is one of the most important processes for obtaining accurate experimental results. To test our model, we used the IMDB movie review dataset [27]. The dataset included 50,000 reviews that were evenly divided into positive and negative reviews. The dataset was divided into a 80% (4000) training set and a 20% (1000) testing set. First, unnecessary columns were removed. The second process included some spelling corrections, removing weird spaces in the text, html tags, square brackets, and special characters represented in text and contraction. They were then handled with emoji by converting them to the appropriate meaning of their occurrence in the document. Thereafter, we put all the text to lowercase, and removed text in square brackets, links, punctuation, and words containing numbers. Next, we removed stop words because having these makes our analysis less effective and confuses our algorithm. Subsequently, to reduce the vocabulary size and overcome the issue of data sparseness, stemming, lemmatization, and tokenization processes were applied. We normalized the text in the dataset to transform the text into a single canonical form. With the aim of achieving better document classification, we also performed count vectorization for the bag-of-words (BOW) model. The BOW model can be used to calculate various measures to characterize the text. For this calculation process, the

term frequency-inverse document frequency (TF-IDF) is the best method. Basically, TF-IDF reflects the importance of a word [36]. We applied the N-gram model to avoid the short-comings of BOW when dealing with several sentences with words of the same meaning. The N-gram model parses the text into units, including TF-IDF values. The N-gram model is an effective model representation used in sentiment analysis. Each N-gram related to parsed text becomes an entry in the feature vector with the corresponding feature value of TF-IDF [37].

After preprocessing, the LDA model was applied. LDA is a three-level hierarchical Bayesian model that creates probabilities at the word level, on the document level, and on the corpus level. Corpus level means that all documents exist in the dataset. Then, a model was developed for identifying unique words in the initial documents and the number of unique words after removing rare and common words. For the analysis of visualizing document relationships, the LDA developed model was applied at the corpus level, which was used for whole document visualization. The LDA application greatly reduced the dimensionality of the data. We used a web-based interactive visualization system, called LDAvis, developed by Carson et al. [28]. LDAvis helps to understand the meaning of each topic and measures the prevalence and relationship of topics to each topic. The removal of rare and common tokens from documents decreased the number of unique words. Initially documents covered 163,721 unique words, and after the process, we had 24,960 unique words. From 50,000 documents, the experiments showed 24,960 unique tokens as shown in Table 1.

Table 1. Processed data.

Number of unique words in initial documents	163,721
Number of unique words after removing rare and common words	24,960
Number of documents	50,000
Number of unique tokens	24,960

Experiments were benchmarked on the IMDB dataset [27], which contains an even number of positive and negative reviews. The IMDB dataset included movie reviews retrieved from the IMDB. In text mining, datasets probably contain existing data with type errors. Misspelled words decrease having high classification accuracy. Coming to the main purpose of this research, increasing sentiment classification accuracy, we used TextBlob python library to correct misspelled words. TextBlob library is useful in applying manifolds in Natural Language Processing (NLP), such as detecting sentiment orientation, every word intensity in the sentence and spelling correction of words.

4.2. Performance Measurement

For the evaluation check, we used a confusion matrix that is commonly used to describe the data classification performance in three metrics: accuracy, recall, and precision, as highlighted in Equations (7)–(9), and Table 2.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(7)

$$Recall = \frac{TP}{TP + FN}$$
(8)

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

	Predicted			
Actual	Positive Documents	Negative Documents		
Positive documents	# True Positive samples (TP)	# False Negative samples (FN)		
Negative documents	# False Positive samples (FP)	# True Negative samples (TN)		

True positive (TP)—outcome of correctly predicted positive class, False positive (FP)—outcome of incorrectly predicted positive class, True negative (TN)—outcome of correctly predicted negative class, False negative (FN)—outcome of incorrectly predicted negative class.

To evaluate our model, we first divided each document into two parts and analyzed whether the topics assigned were similar. To implement this analysis process, we used a corpus LDA model transformation. The LDA transformation in every document returns topics with non-zero weights. This function then creates a matrix transformation of documents in the topic space. To compute the LDA transformation, we chose the cosine similarity method, which is a simple and effective method. For two given vectors of attributes *A* and *B*, the cosine similarity is represented as:

$$\cos\theta = \frac{A * B}{\parallel A \parallel \parallel B \parallel} \tag{10}$$

where *A* represents the target sentence's word vector, and *B* represents the word vector of the compared sentence. A * B represent the number of shared words between word vectors *A* and *B*. $\parallel A \parallel$ and $\parallel B \parallel$ refer to the number of words in *A* and *B*.

Table 3 shows an evaluation check of document similarity at the corpus level. We can say that the intra-topic similarity was highly accurate. This can be explained by the fact that topics in the corpus are perfectly modeled. Moreover, inter-topic similarity was well separated with an accuracy of 99.92%. Intra and inter document evidence has proved effective in improving sentence sentiment classification [38]. We ran the LDAvis tool for four-topic visualization to determine the most relevant terms for every topic and percentage of tokens. Moreover, this tool was also helpful for visualizing topic correlations.

Table 3. Cosine similarity.

Intra similarity: Cosine similarity for corresponding parts of a document	0.9997066
Inter similarity: Cosine similarity between random parts of a document	0.9992671

Table 4 highlights the most frequent terms in the case of the four topics. Frequency rate of all topics are low, which indicate the analyzed data is big or very sparse. From the frequency of words, topics can be indicated. However, as the main purpose of this study is to find those topics, sentences and words are representing which sentiment class. The TDS model implementation identified sentiment classes of data, as can be seen in Figure 4.

Sentiment polarities are presented in Figure 4 with document level and in Figures 7 and 8 with word level. The Tokenized Word Length Distribution Table represents the polarity of documents. Length displays how the document includes sample tokenized words. As mentioned before, we have 50,000 documents and of them 24,960 are unique tokens. As can be seen from Figure 5, most of the text length is [0, 80], and the text between [0, 40] is the vast majority. The text with the length between [10, 40] is selected for the experiment as corpus sentiment analysis, and the selected experimental text length covers 14,000 texts.

 Table 2. Confusion matrix.

Тор	Topic 1 Topic 2		Тор	Topic 3		Topic 4	
Term	Frequency	Term	Frequency	Term	Frequency	Term	Frequency
rol	0.008	lif	0.005	look-lik	0.007	kil	0.006
perform	0.008	world	0.004	bet	0.006	girl	0.005
hi_wif	0.007	view	0.004	im	0.006	wom	0.004
cast	0.006	feel	0.003	didnt	0.005	old	0.004
best	0.006	cannot	0.003	laugh	0.005	sex	0.004
New_york	0.005	tru	0.003	funn	0.005	car	0.004
star	0.005	year	0.003	whi	0.004	guy	0.004
john	0.004	liv	0.003	spec_effect	0.004	back	0.004
comed	0.004	war	0.003	dont_know	0.004	little	0.004
wif	0.004	two	0.003	iv	0.004	find	0.004

Table 4. Term Frequency representation for four topics.



Figure 4. Tokenized documents with sentiment polarity.





Every word appearing in a sentence has the meaning of identifying document polarity. Every document has a role in identifying the entire topic, and the topic can be classified as positive or negative. To adopt this theory, word-level sentiment classification is important. For the word-level sentiment class, we identified the top seven most common positive and negative sentiments. Figure 6 represents the most frequent words in an entire dataset. Words are classified in sentiment classes with their frequency count numbers.



Figure 6. Most common sentiments.

Representative figures are examples of the seven most commonly used words as positive and negative in our entire dataset. Some words appear in all sentiment classes; however, this measurement is performed subjectively. When classification was done at the sentence level, all group-dependent sentiments were classified in the class as they presented. Nevertheless, comparatively in all sentiment classes, it can be predicted which sentences are more dominant in terms of sentiment class, and that sentiment class is more suitable in that identified sentiment class. From the experimental results, it can be conducted that a word represented comparatively more in the positive class than in the other classes, and then the word classified as a positive sentence. Emotional tendency score judgment rules as follows:

$$\{Positive \text{ score } \ge 0.05, Negative \text{ score } \le -0.05\}$$
(11)

The score indicates how negative or positive the analysis is of the overall text. Anything below a score of -0.05 we tag as negative and anything above 0.05 we tag as positive.

Table 5 describes performance of TDS model. Our data is well balanced, and showing relatively high results in positive precision, negative recall and negative F1 scores. To check the validity of the proposed model, the following models are compared.

Table 5. Model comparison with other models.

	Model Classification			
	Precision	Recall	F1-Score	Support
Positive	0.94	0.18	0.30	4993
Negative	0.55	0.99	0.70	5009
Accuracy			0.58	1000
Marco avg	0.74	0.58	0.50	1000
Weighted avg	0.74	0.58	0.50	1000

Table 6 and Figures 7 and 8 show the analysis results of the sentiment classification results in comparison with the SVM, CNN, and LSTM models. The precision, recall, and F1 score of the positive sentiment class are highlighted by P_{pos} , R_{pos} , and $F1_{pos}$. For the negative sentiment class, precision was highlighted as P_{neg} , recall as R_{neg} , and F1 score as $F1_{neg}$, respectively. The F1 score is the weighted average of precision and recall. Hence, this score considers both false positives and false negatives. Intuitively, it is not as easy to understand as the accuracy, but F1 is more commonly used than accuracy. It can be seen that the deep learning models performed better in sentiment classification than SVM, which is a traditional method. The F1 score calculated using following Equation (12).

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$
(12)

Table 6. TDS model classification.

Model _		Positive			Negative		Marco F1
	P _{pos}	R _{pos}	F1 _{pos}	P _{neg}	R _{neg}	F1 _{neg}	- 101u100_11
SVM	84.80	80.70	82.22	82.10	83.50	82.79	82.51
CNN	86.00	85.60	85.80	85.17	85.82	85.49	85.65
LSTM	85.10	82.50	83.78	84.30	86.10	85.19	84.49
Our model	94.0	18.0	30.0	55.0	99.0	70.0	58.0



Figure 7. Positive sentiment class.



Figure 8. Negative sentiment class.

Precision is defined as the number of true positives over the number of true positives plus the number of false positives. Recall is defined as the number of true positives over the number of true positives plus the number of false negatives.

It can be seen from Figure 7 that the TDS model performs better only in the positive precision sentiment class compared with the other three models, reaching 94.0% in the TDS model and 84.8%, 86.0%, and 85.1% in the SVM, CNN, and LSTM models, respectively. Nevertheless, the TDS model reaches the lowest classification results in the recall and F1 score classes. The range of the text sentiment score was between [-1, 1] for all distributed samples, which selected experimental data length was between [10, 40].

Figure 8 shows the negative sentiment class classification performances of the three models compared with our model TDS. Negative recall sentiment class (R_{neg}) results indicated that classification of our model performed relatively high classification reach than other models, to around 99.0%, while SVM, CNN and LSTM models reached around 83.5%, 85.82% and 86.1%, respectively.

5. Limitations

Some classification results of TDS model were comparatively higher than other models. However, the model is achieving dominance in all classifications' parts. Moreover, the F1 score is consistently showing low results in both positive and negative sentiment classifications. Furthermore, we will improve the data analysis by separating score intervals into several parts in order to keep balance of the learning rate.

6. Conclusions

In this study, a topic-based SA was analyzed. The LDA, which is an unsupervised ML technique, was successfully used for topic modeling. Furthermore, we implemented the TDS model in this study, with the main idea for which was to model topics with the aim of increasing sentiment classification. The topic similarity checks yielded accurate results. Thereafter, from highly similar topics, we classified sentiments at the topic, document, and word levels. Our experimental results confirm that the TDS model is an excellent ML technique for modeling the topic and classifying sentiment polarity. However, the results showed that the TDS model accurate results only in the positive and negative recall scores. According to our assessment, one of the main reasons for having low classification performance in R_{pos} and F1_{pos} positive sentiment classes, P_{neg} and F1_{neg} negative sentiment classes, is probably because of having high fluctuation representation of score accuracy in underscore intervals.

Author Contributions: This manuscript was designed and written by A.F., A.A. wrote the program in C++ and conducted all the experiments. F.M. and Y.I.C. supervised the study and contributed to the analysis and discussion of the algorithm and experimental results. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2021-2017-0-01630) supervised by the Institute for Information & Communications Technology Promotion (IITP), and by the Gachon University research fund of GCU-2019-0722.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The first author A.F., would like to express his sincere gratitude and appreciation to the supervisor, Young Im Cho (Gachon University) for her support, comments, remarks, and engagement over the period in which this manuscript was written. Moreover, the authors would like to thank the editor and anonymous referees for the constructive comments in improving the contents and presentation of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bing, L. Sentiment Analysis and Subjectivity. In *Handbook of Natural Language Processing*, 2nd ed.; Indurkhya, N., Damerau, F.J., Eds.; MIT Press: Cambridge, MA, USA, 2010; pp. 4–10.
- Martínez Cámara, E.; Almeida-Cruz, Y.; Díaz Galiano, M.C.; Estévez-Velarde, S.; García Cumbreras, M.Á.; García Vega, M.; Gutiérrez, Y.; Montejo Ráez, A.; Montoyo, A.; Munoz, R.; et al. Overview of TASS: Opinions, Health and Emotions. In TASS 2018 Workshop on Sentiment Analysis as SEPLN; CEUR Workshop: Sevilla, Spain, 2018.
- Martínez-Cámara, E.; Martín-Valdivia, M.T.; Urena-López, L.A.; Montejo-Ráez, A.R. Sentiment analysis in Twitter. *Nat. Lang. Eng.* 2014, 20, 1–28. [CrossRef]
- Ho, V.A.; Nguyen, D.H.-C.; Nguyen, D.H.; Pham, L.T.-V.; Nguyen, D.-V.; van Nguyen, K.; Nguyen, N.L.-T. Emotion Recognition for Vietnamese Social Media Text. In Proceedings of the 2019 International Conference of the Pacific Association for Computational Linguistics (PACLING 2019), Hanoi, Vietnam, 11–13 October 2019.
- 5. Liu, B.; Indurkhya, N.; Damerau, F.J. (Eds.) *Handbook of Natural Language Processing*, 2nd ed; Chapman & Hall: Cambridge, MA, USA, 2010.
- 6. Pang, B.; Lee, L. Opinion mining and sentiment analysis. Found Trends Inf. Retrieve 2008, 2, 1–135. [CrossRef]
- 7. Liu, B. Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. 2012, 5, 1–167. [CrossRef]
- 8. Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* **2014**, *5*, 1093–1113. [CrossRef]
- 9. Maynard, D.; Funk, A. Automatic detection of political opinions in tweets. In Proceedings of the 8th International Conference on the Semantic Web, ESWC'11, Heraklion, Greece, 29–30 May 2011; pp. 88–99.
- 10. Tsytsarau, M.; Palpanas, T. Survey on the mining subjective data on the web. Data Min. Knowl. Discov. 2012, 24, 478–514. [CrossRef]
- 11. Yousif, A.; Niu, Z.; Tarus, J.K.; Ahmad, A. A survey on sentiment analysis of scientific citations. *Artif. Intell. Rev.* 2019, 52, 1805–1838. [CrossRef]
- Ortis, A.; Farinella, G.M.; Battiato, S. An Overview on Image Sentiment Analysis: Methods, Datasets and Current Challenges. In Proceedings of the 16th International Joint Conference on e-Business and Telecommunications, ICETE 2019, Prague, Czech Republic, 26–28 July 2019.
- Siersdorfer, S.; Minack, E.; Deng, F.; Hare, J. Analyzing and predicting sentiment of images on the social web. In Proceedings of the 18th ACM International Conference on Multimedia, ACM, Firenze, Italy, 25–29 October 2010; pp. 715–718.
- 14. Doshi, U.; Barot, V.; Gavhane, S. Emotion Detection and Sentiment Analysis of Static Images. In Proceedings of the 2020 International Conference on Convergence to Digital World—Quo Vadis (ICCDW), Mumbai, India, 18–20 February 2020; pp. 1–5. [CrossRef]
- 15. Noh, Y.; Park, S.; Park, S.-B. Aspect-Based Sentiment Analysis Using Aspect Map. *Appl. Sci.* **2019**, *9*, 3239. [CrossRef]
- 16. Tao, J.; Fang, X. Toward multi-label sentiment analysis: A transfer learning based approach. J. Big Data 2020, 7, 1. [CrossRef]
- 17. Abdelgwad, M.M.; Soliman, T.H.A.; Taloba, A.I.; Farghaly, M.F. Arabic aspect based sentiment analysis using bidirectional gru based models. *arXiv* **2021**, arXiv:2101.10539.
- Martín-Valdivia, M.T.; Martínez-Cámara, E.; Perea-Ortega, J.M.; Ureña-López, L.A. Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Syst. Appl.* 2013, 40, 3934–3942. [CrossRef]
- 19. Jeong, H.; Ko, Y.; Seo, J. How to Improve Text Summarization and Classification by Mutual Cooperation on an Integrated Framework. *Expert Syst. Appl.* **2016**, *60*, 222–233. [CrossRef]
- 20. Turney, P.D.; Littman, M.L. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *arXiv* 2002, arXiv:cs/0212012.

- Lin, C.; He, Y. Joint Sentiment/Topic Model for Sentiment Analysis. In Proceedings of the 18th ACM Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 375–384.
- Duric, A.; Song, F. Feature selection for sentiment analysis based on content and syntax models. In Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA '11), Portland, OR, USA, 24 June 2011; pp. 96–103.
- Hercig, T.; Brychcín, T.; Svoboda, L.; Konkol, M.; Steinberger, J. Unsupervised Methods to Improve Aspect-Based Sentiment Analysis in Czech. Comput. Sist. 2016, 20, 365–375. [CrossRef]
- 24. Matsuno, I.P.; Rossi, R.G.; Marcacini, R.M.; Rezende, S.O. Aspect-Based Sentiment Analysis Using Semi-Supervised Learning in Bipartite Heterogeneous Networks. *Inf. Data Manag.* **2016**, *7*, 141–154.
- Pannala, N.U.; Nawarathna, C.P.; Jayakody, J.T.K.; Rupasinghe, L.; Krishnadeva, K. Supervised Learning Based Approach to Aspect Based Sentiment Analysis. In Proceedings of the 2016 IEEE International Conference on Computer and Information Technology (CIT), Nadi, Fiji, 8–10 December 2016; pp. 662–666. [CrossRef]
- Chakraborti, S.; Lothian, R.; Wiratunga, N.; Watt, S. Sprinkling: Supervised Latent Semantic Indexing. In Proceedings of the 28th European Conference on Advances in Information Retrieval, London, UK, 10–12 April 2006; pp. 510–514. [CrossRef]
- 27. IMDB Dataset. Available online: https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews (accessed on 10 August 2021).
- 28. Sievert, C.; Shirley, K.E. LDAvis: A method for visualizing and interpreting topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MD, USA, 27 June 2014. [CrossRef]
- 29. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. J. Mach. Learn. Res. 2003, 3, 993–1022.
- 30. Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.; Jiang, X.; Li, Y.; Zhao, L. Latent Dirichlet Allocation (LDA) and Topic modeling: Models, application, a survey. *Multimed. Tools Appl.* **2019**, *78*, 15169–15211. [CrossRef]
- Agarwal, A.; Xie, B.; Vovsha, I.; Rambow, O.; Passonneau, R.J. Sentiment Analysis of Twitter Data. In Proceedings of the Workshop on Language in Social Media (LSM 2011), Portland, Oregon, 23 June 2011.
- Bauer, S.; Noulas, A.; Séaghdha, D.O.; Clark, S.; Mascolo, C. Talking places: Modelling and analyzing linguistic content in foursquare. In Proceedings of the 2012 International Conference on Social Computing (SocialCom), Amsterdam, The Netherlands, 3–5 September 2012.
- Eidelman, V.; Boyd-Graber, J.; Resnik, P. Topic models for dynamic translation model adaption. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Jeju Island, Korea, 8–14 July 2012.
- 34. Godin, F.; Slavkovikj, V.; De Neve, W.; Schrauwen, B.; Van de Walle, R. Using topic models for twitter hashtag recommendation. In Proceedings of the 22nd International Conference on World Wide Web, ACM, Rio de Janeiro, Brazil, 13–17 May 2013.
- 35. Mukherjee, S.; Basu, G.; Joshi, S. Joint Author Sentiment Topic Model. In Proceedings of the 2014 SIAM International Conference on Data Mining, Philadelphia, PA, USA, 24–26 April 2014. [CrossRef]
- 36. Rajaraman, A.; Ullman, J.D. Mining of Massive Datasets; Cambridge University Press: Cambridge, UK, 2011. [CrossRef]
- 37. Phan, H.T.; Tran, V.C.; Nguyen, N.T.; Hwang, D. Improving the Performance of Sentiment Analysis of tweets containing fuzzy sentiment using the feature ensemble model. *IEEE Access* 2020, *8*, 14630–14641. [CrossRef]
- 38. Yan-Yan, Z.; Bing, Q.; Ting, L. Integrating Intra-and Inter-document Evidences for Improving Sentence Sentiment Classification. *Acta Autom. Sin.* **2010**, *36*, 1417–1425.
- Sahlgren, M.; Cöster, R. Using bag-of-concepts to improve the performance of support vector machines in text categorization. In Proceedings of the 20th International Conference on Computational Linguistics (COLING '04), Geneva, Switzerland, 23–27 August 2004; Association for Computational Linguistics: USA; p. 487-es. [CrossRef]
- 40. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014. [CrossRef]
- 41. Zheng, W.; Liu, X.; Yin, L. Sentence Representation Method Based on Multi-Layer Semantic Network. Appl. Sci. 2021, 11, 1316. [CrossRef]