Original Article

Analyzing the opinions and emotions of Internet customers using deep ensemble learning based on rbm

Sara Hajighorbani [1], Changiz Valmohammadi [2*], Kiamars Fathi Hafshejani [3]

[1] Department of Information Technology Management, South Tehran Branch, Islamic Azad University, Tehran, Iran.

[2] Department of Industrial Management, South Tehran Branch, Islamic Azad University, Tehran, Iran.

[3] Department of Management, South Tehran Branch, Islamic Azad University, Tehran, Iran.

Corresponding author and reprints: Changiz Valmohammadi, Associate Professor, Department of Industrial Management, South Tehran Branch, Islamic Azad University, Tehran, Iran.

Email: ch_valmohammadi@azad.ac.ir

Received: 29 Apr 2023 **Accepted:** 17 Aug 2023 **Published:** 23 Aug 2023

Abstract

Background: The emotions and opinions of Internet users are critical, as they directly influence the provision of proper services. The aim of this study was analyzing the opinions and emotions of internet customers using deep ensemble learning based on rbm.

Methods: Method of this study was based on the deep ensemble learning technique which uses a deep ensemble neural network based on Gaussian restricted Boltzmann machine and cost-sensitive tree the opinions and emotions of Internet customers were analyzed in terms of semantics and linguistics in virtual shops. To analyze behavior or emotions, existing algorithms were divided into groups of semantic approach, language approach and machine learning. The semantic, linguistic and group learning aspects (machine learning) were considered together. The opinions, feelings, and behaviors of Internet customers were analyzed. The proposed method was implemented in MATLAB software. To evaluate this method, conventional criteria that were /applied in data mining applications have been used including accuracy, recall, and F score.

Results: Based on the experiments performed and by evaluating this method against individual and ensemble methods plus the approaches presented in data mining so far, it was revealed that the proposed model outperforms other methods regarding data mining assessment criteria.

Conclusion: Based on social engineering, the suggested model is provided to forecast consumer behavior. In addition to analyzing customers' behavior which examined their emotions and feelings based on their opinions. The results of this study can be used by planners in the field of competitive internet markets.

Keywords: Data Mining; Deep Learning; Emotions; Internet; Sentiment Analysis; Social Networking.

Cite this article as: Hajighorbani S, Valmohammadi C, Fathi Hafshejani K. Analyzing the opinions and emotions of Internet customers using deep ensemble learning based on rbm. *Soc Determinants Health*. 2023;9(1):1-17. DOI: http://dx.doi.org/10.22037/sdh.v9i1.41964

Introduction

The emotions and opinions of Internet users are critical, as they directly influence the provision of proper services. In order to analyze this massive and unstructured information, the wide scope of data mining comes into play whose main duty is to analyze the opinions, beliefs, ideas, attitudes, emotions, behaviors, and tendency of individuals about a subject, entity, event, phenomenon, as well as their features for summarizing and extracting hidden and valuable information (1). For this purpose, there are various methods such as lexicon-based methods, which only emphasize on textual and linguistic features. In the lexicon-based approach, there is the problem of shortage of lexical resources. Other methods include t techniques based on data mining and machine learning (2).

Decision-making is greatly influenced by the word meanings in these texts. Lexical analysis is thus linguistically and semantically significant in order to be able to draw a suitable conclusion (3). Bertola and Patti focused on investigating users' information based on the resources related to online artistic sets as a source of information and labels as textual works users use to demonstrate the artistic designs on social media. They used a framework in which various methods and tools, such as a set of fields, semantic web technologies, social web, and natural language processing, were employed in which the artworks would be organized based on the users' op. The users of that system participated in developing a semantic space for identifying viewpoints (4).

According to studies that have been done on analyzing the opinions and emotions of Internet users, the current research gap is the lack of an integrated model that can deal with doing so while also lowering the detection error and doing so with a high degree of classification accuracy. The simultaneous investigation of each of these issues requires the creation of a powerful model that can do the analysis and provide appropriate and palatable evaluation criteria. In order to solve the above challenges in the model proposed in this paper, in this research, we tried to following question: what extent linguistic and semantic concepts as well as use of deep ensemble learning affect the prediction of opinions and emotions of Internet customers?"

Methods

Study design

In this research, first proper data sets were collected from social networks. Then, their contents and subjects were investigated. The data included the words and labels that characterize the customers' preferences. Using the deep ensemble learning method which involved different categories including neural network and decision tree, we categorized this information into suitable classes. Also, the cost of each classification has been calculated on different data, whereby the one with minimum cost for the relevant data was chosen. For such classification, part of the data was presented to the training section, and we performed the assessment with the remaining part of data. In the proposed method, the datasets utilized for training the learning models were a set of training samples, with each sample characterized by a vector of features. Indeed, using the features of the samples in a learning algorithm, they learn the required training.

The proposed method

In order to analyze the opinions and emotions, the algorithms available were categorized into semantic, linguistic, and machine learning approaches. In the mentioned model, the semantic, linguistic, and deep ensemble learning (machine learning) aspects were considered alongside each other. Indeed, all of the three semantic, linguistic, and machine learning models together create a comprehensive model which resolves the weaknesses of the others and investigates opinions plus emotions from all aspects, where the Internet customers' opinions and emotions are analyzed. This suggested technique has an advantage over others since it simultaneously examines the consumers' emotions and opinions, in addition to their viewpoints. Using deep ensemble learning and Boltzmann machine, the detection accuracy increased and the classification error was minimized. Also, using a cost-sensitive decision tree, the best classification function for each type of data was identified and chosen as the classifier. The proposed model was comprehensive which can offer a thorough analysis on Internet customers, and employs both linguistic and semantic approaches alongside the machine learning approach. Also, through deep ensemble learning and cost sensitive approach, it provides us with the ability of managing the strong and weak points of each model and directs us towards the best possible solution.

In this research, in order to analyze the opinions and emotions of Internet customers, deep ensemble learning approach was used. Also, by considering the linguistic and semantic features together, it was possible to develop a model which can examine the emotions and opinions from different dimensions, and eventually would be completed using the deep ensemble learning approach and ensemble classifications. In the suggested technique, the available data were preprocessed and their essential characteristics were retrieved because the information pertaining to consumers is unstructured and cannot be utilized directly in constructing learning models. The next step was to separate each sentence into its component words and give each one the appropriate weight. Last but not least, an ensemble model was employed to classify the behaviors.

Data preprocessing

The data related to Internet customers are textual and unstructured data which cannot be normally used for opinion mining or analysis. In other words, machine learning methods cannot apply textual data that have no primary structure or concept. In all data mining tasks using machine learning algorithms, the datasets utilized for training the learning models are a set of training samples, with each sample characterized by a vector of features. Indeed, using the features of the samples in a learning algorithm, it receives the necessary training. Therefore, a primary phase is required for preparing textual data (5).

Data preparation is a method of processing that involves extracting and initializing the essential characteristics from the text, which is in fact the data pertaining to Internet users (5). All words in the given data are initially converted to tiny letters for this purpose. Then, if the text contains any Internet addresses, they are deleted. The data is also cleared of any punctuation and additional characters like spaces and numerals. The aforementioned activities represent the initial step of preprocessing. Every phrase includes a label that reflects the client's desire, and we are aware that this label contains information about the consumer. Each phrase is then separated into its component words. Tokenization is the term for this. Sentences that have been transformed into a vector of words are the result of this step. In the preprocessing stage known as tokenization, this is the second step. After that, inconsequential words are eliminated (the step of deleting words), including very common terms like prepositions and auxiliary verbs that are regularly used in everyday phrases. The data set's whole content has been retrieved to date, and all of its essential terms have been accounted for. Now, in order to perform the main task in the data preprocessing, which is extraction of meaningful features for learning methods, we have considered two approaches. The implementation of these two approaches is explained hereunder.

Linguistic model in feature extraction

In the first approach, every n-gram is considered a feature, where the possible values for n can be 1, 2, and 3. n-gram means every n consecutive words in the sentence constitute a separate feature. For example, in the phrase "in the name of God", after tokenization and eliminating the preposition "in", and assuming that n=2, the extracted features are: {"name of God", "forgiving God", and "the kind forgiving"}. As a result, with this method, all grammatical possibilities are retrieved from the dataset's data. The term frequency - inverse document frequency (TF.IDF) approach is then employed in accordance with Relation 1 to initialize these characteristics. This approach assigns each term (statement) a weight based on how frequently it appears in the document.

Relation 1 $TF.IDF_{t,i} = tf_{t,i} * \log [N / df_t]$

In Relation 1, t represents a term (statement) and i indicates a document. Further, TF.IDFt,i displays the weight assigned to the word t in document i, whereas tft,i displays how prevalent t is there; dft stands for the number of sentences where t has been noticed. The total number of texts is also shown by the letter N. The importance of a phrase for a document is indicated by this weighting scheme. This has various applications in retrieving information. They weight of the term increases as its frequency grows in the text, however the weight is controlled by the number of terms in the text. This is because we know that if the length of a text is large, some terms will naturally emerge more than others, though they may not be that important for the meaning. Thus, highly frequent terms such as prepositions will be assigned a lower weight. The largest weight calculated occurs when t has been repeated more frequently in fewer documents. Thus, most probably term t is an important term across the available documents and is related to their subject. On the other hand, the minimum weight occurs when that term emerges in all or most documents, or has been repeated very few times in only a few documents (6). We call this method "linguistic approach in feature extraction". After extracting all possible n-gram in the entire information available, this weight is calculated and initialized for each of them. We consider each text as a vector, where each of its elements is a possible n-gram in the feature space, and its corresponding value is the TF-IDF weight calculated for that element in that text. Hence, zero weight is assigned to the items that do not exist in a text.

Semantic model in feature extraction

In the second approach, every separate term is considered a feature. Indeed, every 1-gram is a feature. In this approach, in order to initialize the features, instead of using TF-IDF, lexicon-based method is used. In this method, we have used a well-known tool known as machine usable dictionary (MRC). This instrument is a web-based dictionary database which contains 150837 words, with more than 26 linguistic features for each word. It is used in artificial intelligence applications especially natural language processing. Examples of features include the numbers of letters of words, the number of synonyms of words, the level of familiarity of the word, the level of concreteness of the word, the degree of imagery of the word, and the degree of concreteness of the word, etc. (7). In this research, we have used three features of level of familiarity, level of concreteness, and imagery level of the words. The values that are extracted for these features for every word from MRC range between 100 and 700. These features have been chosen arbitrarily. After extracting all features from the available texts, every term is applied as an input for MRC which has been presented online, and the values of the three features of interest are returned as the output. Then, the mean values of the output are kept as the final value for the intended term. In this method as with the previous method, every text is created as a vector, each of the elements are one of the extracted features, and the value calculated via MRC is considered as the value of that feature. We call this method "semantic approach in feature extraction". Finally, after implementing each of the above approaches, the output resulting from the preprocessing stage is a matrix, each of its rows indicates information about the customer, and each of its colors indicates a feature in the feature space. The elements of this matrix are created from implementing the above approaches. This output matrix is used as the final dataset which can be employed in the training the learning model. The available information is mapped from a textual and unstructured space to a structured feature space, which can be used for any machine learning algorithm.

The learning method in the proposed model is homogeneous in which deep learning of decision has been used. Also, the learning method is dynamic in which different features are employed in combining the classes. Eventually, the final output of classification is obtained as a linear combination of the output of learning factors. In the model learning, begging algorithm which employs cost sensitive decision trees and neural network based on Boltzmann machine has been used. In this method, several classifiers together are used for each decision-making, which are the branches utilized for the decision-making trees. Different decision-making branches with various features chosen randomly are arranged alongside each other and eventually their prediction is collated, and the output develops as a random forest for a special input which involves the feature vectors related to the customers. Any class mentioned by more trees is considered as the output.

Based on the vector of linguistic and semantic features, a model is developed for the feature learning. This model finds the feature of the shared representative from among different features. A deep neural network based on a Gaussian-limited Boltzmann machine is trained beforehand before the training procedure in order to function as a preprocessor for transforming the actual observed values into feature vectors.

After finding the representations of the hidden feature or the trained shared feature, by combining several classifiers in a cost-sensitive hierarchical method, which is based on the cost sensitive decision tree, we create a classifier for the classification. We considered the following trend for choosing the patches.

- The proposed patch should not overlap any of the selected patches by more than 50%.
- The patches whose mean p value is lower than the mean p value across all of the candidate patches are chosen from among those candidate patches that have satisfied the first criteria.

Restricted Boltzmann machine (RBM)

A deep neural network called an RBM contains visible and hidden units (for variables) in each layer and is capable of automatically detecting the intrinsic patterns in data through input regeneration. Eventually, these interchangeable units and variables are used in the present study. The assumption stating that there is a symmetric connection of W between the hidden layers and visible layers, and the assumption arguing that there is no connection between the layers and each layer has a bias have been shown with a and b respectively.

The visible layer units have been shown as $v = [v_i], i = \{1, ..., D\}$. In the same vein, the hidden layer units have been indicated as $h = [h_j], j = \{1, \cdots, F\}$. Structural models are those that are dependent on the visible variables; here, F and D represent the number of visible units and number of hidden units.

An RBM has been considered as an auto-encoder. This is a desirable feature which has been used by RBM parameters and in order to run the learning process in RBM. A shared probability between (v,h) is calculated using Relation 2 .

Relation 2

Where,

$\Theta = \{W = [W_{ij}] \in R^{D \times F}, a = [a_i] \in R^D, b = [b_j] \in R^F\}, E(v, h; \Theta)$ is an energy function and $Z(\Theta)$ is a separator function which is obtained through summing up all possible elements of v and h.

For further simplicity, through summing up the hidden and visible variables, the energy function $E(v,h;\Theta)$ is defined according to Relation 3.

$$E(\mathbf{v},\mathbf{h};\Theta) = -\mathbf{h}^{\top}\mathbf{W}\mathbf{v} - \mathbf{a}^{\top}\mathbf{v} - \mathbf{b}^{\top}\mathbf{h} = -\sum_{i=1}^{D}\sum_{j=1}^{D} v_i W_{ij} h_j - \sum_{i=1}^{D} a_i v_i - \sum_{j=1}^{F} b_j h_j.$$

Both the conditional distribution of visible variables based on the hidden variables and the conditional distribution of visible variables based on the hidden variables may be determined using Relations 4 and 5, respectively.

Relation 4:

$$P(h_j = 1|\mathbf{v};\Theta) = sigm\left(b_j + \sum_{i=1}^{D} W_{ij} v_i\right)$$

Relation 5:

$$P(v_i = 1|\mathbf{h};\Theta) = sigm\left(a_i + \sum_{j=1}^{F} W_{ij} h_j\right)$$

Where, $\mathrm{sigm}(.) = \frac{\exp(.)}{1+\exp(.)}$ is a logistic sigmoid function. According to the invisible hidden variables, the objective function as marginal distribution of the visible variables is defined according to Relation 6.

$$P(\mathbf{v};\Theta) = \frac{1}{Z(\Theta)}\sum_{\mathbf{h}} exp(-E(\mathbf{v},\mathbf{h};\Theta)).$$

The values of the visible patches are the real values of $v \in R^D$. In this case, a Gaussian RBM has been used, whose energy function is defined according to Relation 7.

Relation 7:

$$E(\mathbf{v},\mathbf{h};\Theta) = \sum_{i=1}^{D}\frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{i=1}^{D}\sum_{j=1}^{F}\frac{v_i}{\sigma_i} W_{ij} h_j - \sum_{j=1}^{F} b_j h_j$$

Where, σ_i represents a standard deviation of the ith visible variable. Also, for Θ we have $\Theta = \{W, a, b, \sigma = [\sigma_i] \in R^D\}$.:

Deep Boltzmann machine is a model which has been created from aggregation of several RBMs in a hierarchical method. An RBM includes a visible layer called v and a series of hidden layers as $h^1 \in \{0, 1\}^{F_1}, \ldots, h^i \in \{0, 1\}^{F_i}, \ldots, h^L \in \{0, 1\}^{F_L}$, where Fi represents the number of units available in the ith heating layer, while L denotes the total number of hidden layers (3).

Given the hierarchical structure of cost sensitivity in the suggested model, one of its key aspects is its ability to capture extremely intricate and nonlinear statistical patterns, such as the correlations between the input values. The ability to immediately complete the learning process without the need for human interaction by representing the hierarchical hidden characteristics is another crucial aspect.

In other words, the duty of expressing the features is given to a DBM, and this DBM will locate the desired features independently of the training samples, in contrast to earlier approaches that primarily examined preset features or the output of specified functions. A hidden representation of the characteristics that are available in a patch may be found by taking into account the representational features and self-taught learning of DBM.

Different layers of the network reflect varying degrees of this information when an input patch is submitted to a DBM. In this approach, the network's lowest layer contains the most basic patterns, while the top layer contains more abstract or complicated patterns that are inherent to the input data.

The relations DBM use for representational features are as follows:

can learn the internal hidden representations. These hidden representations record complex nonlinear patterns or statistical patterns using a hierarchical method. However, following the initial bottom-to-top pass, the inferential technique also often incorporates top-to-bottom feedback, in contrast to other deep network models like the deep belief network and stacking auto-encoder. As a result, the DBM is able to leverage high-level information to address the ambiguity in the mid-level characteristics. Thus, development of representations and other statistics required for learning is dependent on better data. In DBM, the (v,h1,h2) state energy is calculated according to Relation 8.

Relation 8:

$$E\left(\mathbf{v},\mathbf{h}^1,\mathbf{h}^2;\Theta\right) = -\mathbf{v}^{\top}\mathbf{W}^1\mathbf{h}^1 - \left(\mathbf{h}^1\right)^{\top}\mathbf{W}^2\mathbf{h}^2$$

Where, $\mathrm{W}^1 = \left[w_{ij}^{1}\right] \in R^{D\times F_1}$ and $\mathrm{W}^2 = \left[w_{jk}^{2}\right] \in R^{F_1\times F_2}$ represent the symmetric collections of (v,h1) and (h1,h2), and we have $\Theta = \left\{W^1, W^2\right\}$.. Then, the probability of allocation of the model to visible vector v is calculated according to Relation 9.

Relation 9

$$P(\mathbf{v};\Theta) = \frac{1}{Z(\Theta)}\sum_{\mathbf{h}^1,\mathbf{h}^2} exp\left(-E\left(\mathbf{v},\mathbf{h}^1,\mathbf{h}^2;\Theta\right)\right)$$

Where, $Z\left(\Theta\right)$ is a normalization factor. Considering the values of units in the neighboring layer(s), the probability of the units being visible or hidden units has been set to zero, and calculated according to Relations 10-12.

Relation 10

$$P\left(h_j^1 = 1\middle|\mathbf{v},\mathbf{h}^2\right) = sigm\left(\sum_{i=1}^{D} W_{ij}^1 v_i + \sum_{k=1}^{F_2} W_{jk}^2 h_k^2\right)$$

Relation 11

$$P\left(h_k^2 = 1\middle|\mathbf{h}^1\right) = sigm\left(\sum_{j=1}^{F_1} W_{jk}^2 h_j^1\right)$$

Relation 12

$$P(v_i = 1|\mathbf{h}^1) = sigm\left(\sum_{j=1}^{F_1} W_{ij}^1 h_j^1\right).$$

DBM differs from other deep learning models in that it combines the topmost hidden layer (layer h2) and the bottommost visible layer (layer v) when computing the probability of hidden units h1.

Results

In the previous section about the model proposed in this research for analyzing internet customers, we discussed the concepts related to the proposed model and its simulation. We have created and put into practice certain experiments in this part to explore the impact of various model parameters on detection accuracy. Acceptable values for the free parameters of the model are established through these tests. Next, we have compared the suggested model with existing approaches in order to show its strength and effectiveness, and the results have been given.

The tested datasets

Users of the Amazon social network made up the sample used in this study. Txt and XML files that have been added to the programming environment are among this data. More than 10,000 samples that were evaluated were records connected to Amazon users.

Examining different parameters in the model

Generally speaking, any algorithm has certain free parameters that can determine how effective the method is. This is not an exception for the suggested method, which contains various parameters whose ideal modification would increase the program's classification accuracy and power. We've used a few studies in this section to illustrate the effects of various settings. Then, various criteria for evaluating classification algorithms in data mining have been utilized, including recall (sensitivity), precision, and F-measure, which are important in most data mining applications, in order to compare and evaluate this approach against other methods.

The number of trees

The number of trees in the model is one of the key factors determining classification accuracy. In the model, each branch operates at a relatively low efficiency, and the detection power of the classifier is in fact dependent on the combined efficiency of all group branches. This is a characteristic of ensemble techniques where the ultimate choice is made by a form of member voting mechanism. The confidence interval for the final detection based on various branches would increase as these branches were included in the model more frequently (3).

The following experiment has been carried out in order to examine the impact of this parameter on the algorithm's effectiveness and to choose the appropriate number of branches in the model:

Experiment 1: number of branches

The following conditions were used to run the suggested model in this experiment:

1. The criterion for choosing the breakpoint: Gini Index

2. The parameter m: n variable denotes that the total number of features extracted using that method is the same as the parameter m. The various values for n in the extraction of n-grams are 1, 2, and 3, depending on whether a linguistic approach is of interest. The resulting feature space varies for each distinct n. If n = 1, feature space has a size of 2520; if n = 2, it has a dimension of 5989; and if n = 3, it has a dimension of 5308. Thus, this experiment is repeated for every feature space. Also, if we have adopted the semantic approach in feature extraction, the feature space includes the separate terms available. An example is a state in statistical approach n=1 where every 1-gram is considered a feature. Thus in this state, the feature space has 2520 dimensions.

3. The number of decision-making branches: it is variable, and is initialized from 10 layers, with incremental steps of 10 up to a maximum of 150.

Also, in all experiments, according to the main algorithm of creating the proposed model, one-third of the training data have been used for estimating OOB error, and calculation of other assessment criteria has been used as experimental data.

After running the above experiment, different assessments have been performed and the results have been mentioned further.

The OOB error is shown in Figures 1 and 2 as the number of decision-making branches in the model for each feature extraction method space.

Figures 1 and 2 show that the OOB error has discovered a falling trend with an increase in the number of layers. In reality, the model's detection capability has grown with an increase in the number of decision-making branches. In some regions in the diagrams presented in Figure 1, some fluctuations are observed within the [-5,5] range. The random behavior of this method during the sampling of training data to construct the model is what causes this increase and decrease in inaccuracy to exist.

Also, according to Figure 1, with increase in n, in the feature extraction linguistic approach, OOB error maintains its descending trend, but generally its level has increased. When n=1, every feature is considered an individual term, and it has no relationship or interdependence with other terms. When n grows to 2 and 3, every feature is a 2- and 3-member possible combination of all available terms. With this combination, a kind of interdependence between terms develops; in most cases, these terms have no relationship with each other and this interdependence is wrong. Due to the misleading and inaccurate training of the learning algorithm caused by the establishment of an unnecessary dependency between the features, the effectiveness of the classifier algorithm is reduced. Also, compared to the case when every independent term is considered as a feature, the OOB error increases.

Figure 2 compares the results of implementing the experiments using the semantic approach in feature extraction against the results of implementation when 1-grams of the statistical approach constitutes the feature space. This comparison has been made since in the semantic approach every independent term indicates a feature. It is observed that training the model with the data created in the semantic approach has had worse results and the efficiency of the algorithm has diminished considerably. This is

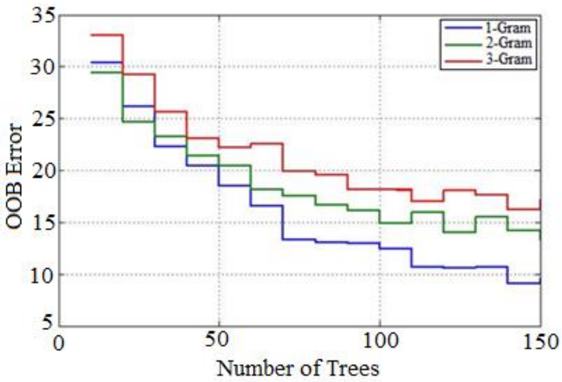

Figure 1. OOB error estimation in the experiment of examining the effect of number of decision-making branches parameter based on the linguistic approach of feature extraction

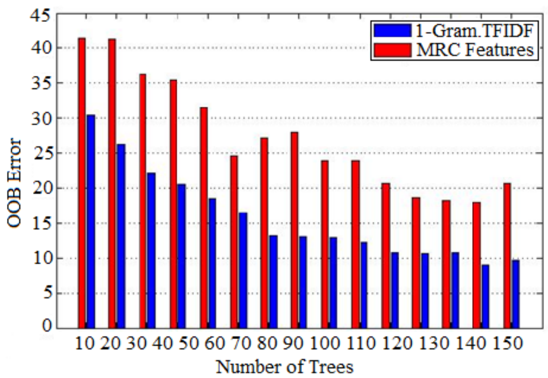

Figure 2. OOB error estimation in the experiment of examining the effect of number of decision-making branches parameter based on the semantic approach of feature extraction

because initialization of these features in the semantic approach is heavily dependent on the lexicons employed in the MRC tool. If these dictionaries or lexicons do not have a wide range, the values extracted from MRC for each feature would have low accuracy, thereby reducing the detection power of the learning model.

Parameter m

The model has another variable called m that may be changed. Keep in mind that this parameter refers to the quantity of features drawn at random from the entire collection of the issue features and placed in each tree layer. The best characteristic for the breakdown is determined using this narrowed-down subset. This parameter is often set to Sqrt (nVariable) or Log (nVariable) and is used as a constant for all trees while building the model (Karta et al., 2020). We have carried out the following experiment to see the impact of this parameter and to determine a potential value for it:

Experiment 2: parameter m

The model was used in this experiment under the following circumstances:

1. Number of trees: 150

2. The criterion for determining the fracture feature: Gini Index

3. Parameter m: It can take on three different values and is variable: 1. the total number of extracted features; 2. the square root of the total features in each dataset; and 3. the logarithm of the total features in each dataset.

The desired characteristics are retrieved and initialized in this experiment, just like in experiment 1. These characteristics are used to create the training model. The size of the space for different values of n resembles experiment 1 in this regard.

The outcomes of the model evaluation in these three states are shown in Table 1. Figure 3 contrasts the outcomes of applying various values of the parameter m depending on the OOB error.

Due to the fact that OOB error has grown while other evaluation criteria have decreased, it is evident from the findings above that reducing the number of features to sqrt (nVariable) and log (nVariable) does not give promising outcomes for the current situation. In the training data utilized for developing the learning model, we know that every sample is a vector data whose elements are the very features extracted, and the values corresponding to the elements in each feature extraction approach are calculated separately.

Table 1. Model assessment per different values of parameter m

Parameter m	F-score	Accuracy	Recall	Feature space
nVariable	0.89	0.87	0.91	---
	0.86	0.85	0.90	2-gram
	0.83	0.82	0.87	3-gram
	0.77	0.76	0.8	MRC Fea.
Sqrt(nVariable)	0.86	0.83	0.89	1-gram
	0.85	0.82	0.87	2-gram
	0.78	0.79	0.78	3-gram
	0.74	0.76	0.76	MRC Fea.
Log(nVariable)	0.85	0.81	0.9	1-gram
	0.84	0.81	0.87	2-gram
	0.77	0.79	0.77	3-gram
	0.76	0.75	0.81	MRC Fea.

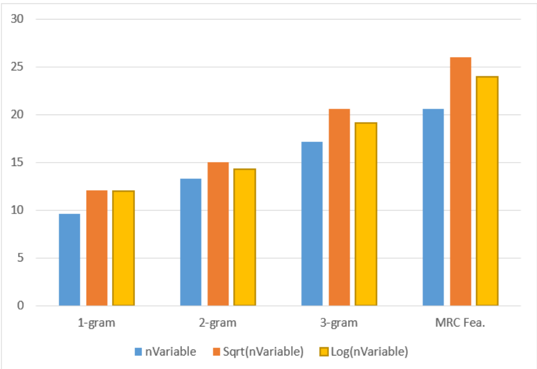

Figure 3. Illustrates the OOB error per different values of parameter m.

The important thing to note is that this feature space is enormous, and every vector that is transformed into it starts off with zero initialization for any features that are not present in it. As a result, there are a lot of zeros in the samples. As the value of m decreases, it's possible that valuable features with various nonzero values are lost when choosing the fracture feature in each layer's nodes, leading to the creation of branches with less significant features that are initialized to zero in most samples. This lowers the model's accuracy and efficiency.

The criterion of determining the fracture feature

The choice of an appropriate criterion for detecting the fracture at the time of generating the decision model is one of the most crucial factors in the proposed model and all techniques that rely on cost-sensitive ensemble learning. Since decision trees and deep neural networks make up the majority of the proposed ensemble model, choosing the optimum criterion for finding the point, or the feature for data segmentation, is crucial (8). There are several criteria for this aim, all of which are fully discussed.

In order to determine the optimal criterion for choosing the fracture and see how this parameter affects it,

Experiment 3: the criterion of determining the fracture

The model was used in this experiment under the following circumstances:

1. Number of branches: 150

2. nVariable: m. We think of m's value as the number of variables, or the characteristics in the data, because of the prior experiment and the possibility that a decrease in m's value may negatively affect the algorithm's performance. Similar to experiment 1, the intended characteristics

are retrieved and initialized in this experiment. The model is then created based on these. The size of the feature space for different values of n resembles experiment 1 in size.

3. A different criterion, such as the information ratio, gain ratio, or Gini index, is chosen and examined each time to determine the fracture.

Table 2 lists the findings from the model evaluation in these three distinct states. Figure 4 also visualizes the results. In Figure 4, the classification criteria have been compared based on OOB error.

Based on the findings, it can be shown that Gini index selection of the breakdown feature has greater efficiency than the two other criteria. The effectiveness of these criteria relies on the data utilized and the issue being studied; hence, it has not yet been shown which criterion would work better than the others without a doubt. On the other hand, the Gini index and the information ratio often function similarly (9). However, it was found that using the Gini index would produce superior findings for the issue this research looked at.

The limitations of the proposed method

One of the limitations of the proposed model was the long time required for model learning, as numerous branches and layers should be individually trained in the forest. Another limitation is the case when the magnitude of huge and unstructured information increases. In this case, a large volume of information should be trained, and thus the training model in its branches should perform long processes; some of these processes may be repeated and there is absolutely no need to perform them.

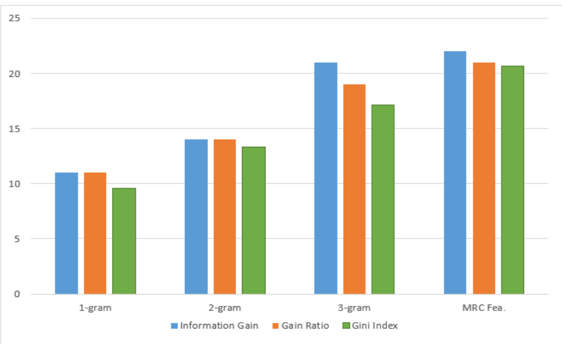

Figure 4. OOB error per different values of the criterion of selecting fracture

Table 2. Model assessment based on different parameters of determining the fracture

Data classification	F score	Accuracy	Recall	Feature space
Information Gain	0.87	0.87	0.89	1-gram
	0.83	0.82	0.86	2-gram
	0.8	0.82	0.8	3-gram
	0.77	0.77	0.77	MRC Fea.
Gain Ratio	0.87	0.87	0.89	1-gram
	0.83	0.82	0.85	2-gram
	0.82	0.82	0.84	3-gram
	0.77	0.77	0.77	MRC Fea.
Gini Index	0.88	0.87	0.91	1-gram
	0.86	0.85	0.89	2-gram
	0.84	0.82	0.85	3-gram
	0.78	0.76	0.8	MRC Fea.

We begin by choosing the ideal values for the parameters of the suggested model before comparing it to alternative approaches. For this, we compare the suggested technique to existing ensemble and single algorithms as well as the most effective ones that have been developed at this point.

Initially, in order to assess the proposed method against individual learning approaches, we have employed support vector machine (SVM), Naive Bayes (NB), and decision tree which are among well-known and practical methods and utilized in numerous studies. SVM which belongs to supervised classification maps training samples from the vector space model to a Table 3. Assessment of the proposed method against individual learning methods

Method	F-score	Accuracy	Recall
Proposed model	0.89	0.87	0.91
SVM	0.68	0.68	0.7
NB	0.71	0.69	0.74

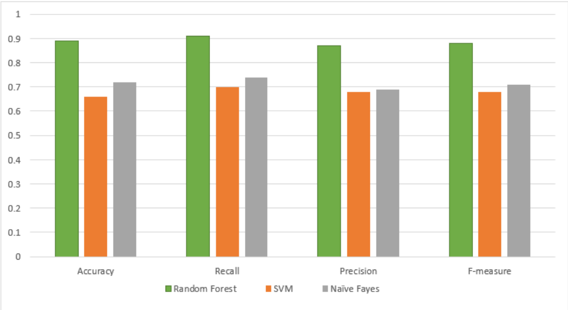

Figure 5. Comparing the proposed model against individual learning algorithms

space with higher dimensions. Once the samples are mapped to higher dimensional space, the samples can be classified linearly. In the tested SVM model, linear function has been used as the Kernel function. The results of this assessment are provided in Table 3. Also, Figure 5 presents the results obtained based on the efficiency criteria.

Table 3's findings show that the suggested model's performance has been much superior to that of individual methods. This

Table 4. Evaluation of the suggested approach using additional ensemble models

Method	F-score	Accuracy	Recall
Proposed model	0.89	0.87	0.91
SVM	0.8	0.8	0.81
NB	0.81	0.79	0.85

is because in a feature space with very high dimensions, individual models have less potential in learning the entire space of the problem. However, in an ensemble algorithm which is a combination of many individual learning models, complex feature space is well-managed.

Also, in order to assess the proposed model against other ensemble methods, we have trained a set of Naive Bayes and SVM models via Bagging method. In the Bagging method, learning models are developed through a subset of training data such as the ensemble learning. In each of these ensemble methods, 150 learning models have been trained. In the SVM ensemble method, linear function has been used as the kernel function in all models, with the results shown in Table 4. Also, Figure 6 presents the results obtained based on various efficiency criteria.

On the basis of the findings shown in Table 4 and Figure 6, it can be seen that the algorithm suggested in this study has also outperformed previous ensemble techniques and meets the evaluation requirements. Thus, it can be stated that in on some learning methods, selection of the type of learning model created as the members of this set is very important. For example, the decision tree has advantages over NB.

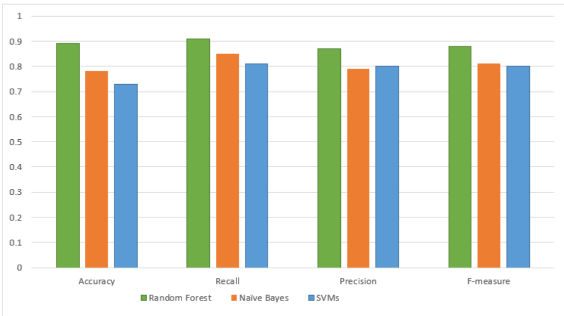

Figure 6. The suggested model is contrasted with previous ensemble learning techniques.

For instance, when creating the decision tree, there is a criterion which determines what the best feature is for placement in a node of the tree. However, there is no such ability in the NB and there is no guarantee for it either. Also, both decision tree and neural network can work with samples consisting of both ensemble features and continuous features. On the other hand, in NB and SVM, the data should be continuous. Also, decision tree has advantages over SVM. In SVM, determining the kernel function is not easy. Also, in this model, unlike decision tree, it is not possible to use ensemble data, and we are allowed to use only continuous data.

Table 5. Comparison of the proposed model with other models

Method	F-score	Accuracy	Recall
Proposed model	89%	87%	91%
Random forest and linguistic models	75%	67.3%	80.8%
Random forest and semantic models	82%	85.7%	82.2%

Next, based on different efficiency criteria, the suggested model has been contrasted with the models of applying random forest and linguistic models, as well as the models of applying random forest and semantic models. Table 5. For these evaluations, the breakpoint has been determined using a trained random forest method with 150 trees, m equal to the total number of features, and the Gini index. Based on the results presented in Table 5, it is observed that the model proposed in this research is of top priority in all three assessment criteria, and has had a better efficiency.

Discussion

The single-lean vector machine method is about 23%, compared to the simple Bayesian method is about 17%, compared to the mass method is about 16% and compared to the collective Bayesian method. About 11 per cent improved. The proposed model measures readability by about 21% compared to the single backup vector machine method, about 17% compared to the simple Bayesian method, about 10% compared to the mass vector machine method, and about 6% compared to the collective Bayesian method. It has improved relative to the single support vector machine technique by approximately 19%, the basic Bayesian approach by about 18%, the mass vector machine method by about 7%, and by about 8%. Compared to the Bayesian collective method. The proposed model in criterion F is about 20% better than the single vector support method, about 17% better than the simple Bayesian method, about 8% better than the collective vector method and about 7% better than the collective Bayesian method. Compared to individual algorithms, the performance of the suggested model is significantly superior. Additionally, it is noted that the suggested model outperforms the group techniques.

Based on several performance criteria, the suggested model was compared with a random hybrid forest model, a linguistic model, and a random hybrid forest model, as well as a random forest hybrid model and a semantic model. Compared to the random forest technique and linguistic model combination and the random forest method and semantic model combination, respectively, the findings revealed that the suggested model had an accuracy criterion improvement of roughly 19% and 4%. In comparison to the random forest approach plus the linguistic model and the random forest method plus the semantic model, the suggested model has increased the reading criteria by roughly 10% and 9%, respectively. The suggested model provides a 20% to 2% accuracy criteria improvement over the random forest technique and linguistic model combined, and a 2% improvement over the random forest method and semantic model combined. The suggested model in criteria F is about 13% better than the combination of the random forest technique and the linguistic model and approximately 6% better than the combination of the random forest method and the semantic model. In the Amazon, Sadasiuam and Kalivarhan created a technique for assessing emotions. Yi & Liu, noted that a recommendation system is categorized into hybrid recommendation systems. The suggestion is based on the weighted average of the scores of users who have previously scored objects similarly to those of interest in approaches that employ the ratings provided to items to detect similarities. By evaluating the score of other product items, recommendations' products are projected. Both user-based and item-based forms of recommendation exist. On the other hand, content-based techniques examine a collection of item descriptions that have previously been rated by a user and then build an interest profile for that user based on the characteristics of those things. While this is going on, the hybrid recommendation system aims to integrate the two aforementioned approaches to reap the benefits of both (9). Khalid et al. also looked at machine learning techniques and proposed a technique based on gradient-boosted SVM. After preprocessing and disassembling sentences into their component words, the classification function is given the estimated weight of each word for decision-making. The gradient function serves as the support vector machine's kernel function in this approach. This strategy produces extremely good results in identifying positive samples and reduces error in judgment (10), which is congruent with the current study.

The suggested model was evaluated using the best techniques already proposed as well as individual and ensemble algorithms. According to the findings, the suggested model outperformed single SVM by approximately 21%, NB by about 17%, ensemble SVM by about 10%, and the ensemble Bayes method by about 6% in terms of recall criteria. Also, the proposed model showed improvement in accuracy over single SVM by around 19%, NB by about 18%, ensemble SVM by approximately 7%, and ensemble Bayes by nearly 8%. Concerning the F-criterion, the proposed model showed improvements as follows: over single SVM, NB, ensemble SVM, and ensemble Bayes by around 21, 18, 9, and 8%, respectively. In this method, the Bayesian ensemble algorithm and support vector machine (SVM) were used alongside each other. After preprocessing and extracting valuable information from the data, they were used for training the ensemble model. To determine the vote, the intersection of SVM and Bayesian votes were taken, whereby the final vote was specified based on the majority of votes. The results obtained from experiments indicated that the combination of these two algorithms and forming a categorizer using them could enhance the accuracy of detection and characterization. It showed a more favorable performance compared to the state when only one of these two algorithms was used in decision-making (11).

The outcome of this research was a model for classifying the customers' emotions and opinions. The information used in this research included textual data, which were preprocessed and then the training features of the model were extracted. In order to weigh the extracted features and to examine the semantic and linguistic load of features, the linguistic approach is based on n-grams and TF. IDF as well as the semantic approach based on MRC were used. The views and feelings were then identified using deep ensemble learning, which involved the use of a cost-sensitive decision tree and a neural network built using rbm. After that, the suggested model was evaluated. The MATLAB program was used to complete all of the calculations and phases described, compare the stated variables to other models, and compare the results. The comparisons' findings revealed that the model put out in this study performed better than expected across the board. In their research, Kang et al. proposed a unique classification approach based on the hidden Markov model. In this paper, the hidden Markov model was introduced and keywords were emphasized. First, the sentences were divided into their constituent words, and then a weight was assigned to each word. This weighting was based on the extent of the importance of the word in the sentence. The results indicated that the classification of texts in this way was right, and the detection accuracy increased using the proposed method. The results also suggested that proper selection of experimental and test data sets also influences the improvement of the method (12).

Support vector machines (SVM), artificial neural networks, and ensemble learning techniques are all components of the machine learning methodology. In this case, supervised classifications are the main learning technique used since they are more accurate. This is because each of the classes is trained using the known data in the dataset. However, the semantic approach to the exploration of information is unsupervised learning as it requires no previous training. Tokenization, stemming, stop-word removal, punctuation removal, and POS tagging are all components of the linguistic method, which encompasses natural language processing (13).

Afterward, the suggested model was contrasted with the random forest and linguistic hybrid model and the random forest and semantic hybrid model based on a number of efficiency factors. The obtained results showed that the proposed model revealed improvement in recall criterion over random forest and linguistic hybrid model by about 10%, and over random forest and semantic hybrid model as about 9%. Concerning the accuracy, the model offered the following improvements: over random forest and linguistic hybrid model as about 20% and over random forest plus semantic hybrid model as about 2%. Finally, the improvement in F criterion by the proposed model was as follows: by about 14% and 7% over random forest plus linguistic hybrid model and random forest plus semantic hybrid model, respectively. In the method designed for analysis in opinion mining and depth of data, George et al. utilized data heterogeneity. This method converts several data sources to a single source for processing. In this paper, first opinion mining has been examined. Then, the essential aspects have been explored, and eventually, the effect of data heterogeneity method was tested on opinion mining and evaluated further (14). Chen & Qi, examined the thematic aspect of indeed semantic modelling. In these methods, ensemble aspects have been employed and the features of each group have been extracted according to their semantic group. Theme-based methods include plea and LDA, with the disadvantage being the fact that it requires a huge volume of data about a specific theme in order to generate reliable results for the text and introduced 100 opinions into the model, though the results were not satisfactory. It also suffered poor performance in detecting repeated phrases (6). Rill et al., again used plsa and LDA for opinion analysis in the text. Using these two methods, the terms in the text are isolated and analyzed. It can also detect cluttered terms randomly. This method has wide applications in predicting political subjects on Twitter. However, in some assessment criteria, it underperforms other machine learning methods (15). Hu et al. (7) as well as Chen & Liu (16), focused on the thematic aspect. Indeed, these papers have dealt with semantic modelling of the theme. In these methods, ensemble aspects have been used, and the features of each group have been extracted based on their thematic group. The advantage of natural language processing is that it contributes to better semantics and simplicity of feature extraction. However, Nakov et al., considered no criterion for assessing the effective features (17). To look into every part of a text, Poria et al. employed a convolutional deep layer neural network. Additionally, in order to draw a more accurate conclusion, the neural network and linguistic pattern were integrated (18). Though combining linguistic patterns with deep neural networks to increase the detection pattern made the process more difficult and time-consuming, the findings showed that this technique was more successful than other well-known ways.

Recommendations

By increasing the range of lexicons and raising the values extracted from MRC per each feature, the detection power of the learning model can be enhanced. This model may be used in other related fields, such as data mining in social networks, which encompasses more individuals and resources, to evaluate the suggested approach. Also, in the feature extraction statistical approach, instead of using

TF.IDF criterion, other methods can be employed and via weighting methods, the features can be weighted such that it could improve the performance accuracy of the learning model. In addition to the above points, since in data mining tasks, the extracted feature space is very large, and in turn due to this large dimension, extensive training data are required for the learning model. The proposed method can be implemented based on big data techniques, which have recently received the attention of researchers. Finally, to reduce the model's learning time and resolve the processing problem on huge volume of data, big data techniques can be used.

Conclusion

The decision tree model's benefits over other models, along with the fact that machine learning methods often perform better than other methods, allow the suggested model to outperform other group approaches, according to the findings of the current study. The application of the suggested approach to any company or trustee that provides the good or service online is, in the end, pretty promising, and with additional research, better outcomes may be obtained. The ability to simultaneously analyze users' behavior and emotions based on applied information, textual data, and their preprocessing using linguistic and semantic approaches is the model of the current study. It can be used to analyze Internet users' behavior on social networks based on social engineering. It can be used by planners and managers of Internet businesses to use the existing capacities in an appropriate way to develop business and generate revenue and expand the range of customers.

Authors' contribution

Sara Hajighorbani and Changiz Valmohammadi developed the study concept and design. Kiamars Fathi Hafshejani acquired the data. Sara Hajighorbani and Changiz Valmohammadi analyzed and interpreted the data, and wrote the first draft of the manuscript. All authors contributed to the intellectual content, manuscript editing and read and approved the final manuscript.

Informed consent

Questionnaires were filled with the participants' satisfaction and written consent was obtained from the participants in this study.

Funding/financial support

There is no funding.

Conflict of interest

The authors declare that they have no conflict of interests.

References

1. Pang, B. and Lee, L. Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval. 2008;2(1):1-135. https://doi.org/10.1561/1500000011
2. Keshwani K, Agarwal P, Kumar D. Prediction of Market Movement of Gold, Silver and Crude Oil Using Sentiment Analysis. In: Bhatia, S., Mishra, K., Tiwari, S., Singh, V. (eds) Advances in Computer and Computational Sciences. Advances in Intelligent Systems and Computing, vol 554. Springer, Singapore;2018. https://doi.org/10.1007/978-981-10-3773-3_11
3. Luo X, Jiang C, Wang W, Xu Y, Wang J, Zhao W. User behavior prediction in social networks using weighted extreme learning machine with distribution optimization. Future Generation Computer Systems. 2019;93(1):1023-1035. https://doi.org/10.1016/j.future.2018.04.085
4. Bertola F, Patti V. Ontology-based affective models to organize artworks in the social semantic web. Information Processing & Management. 2016;52(1):139-162. https://doi.org/10.1016/j.ipm.2015.10.003
5. Jain Y, Tiwari N, Dubey S. Jain S. A comparative analysis of various credit card fraud detection techniques. International Journal of Recent Technology and Engineering. 2019;7(2):402-407. https://www.researchgate.net/publication/332264296_A_comparative_analysis_of_various_credit_card_fraud_detection_techniques
6. Chen L, Qi L. Social opinion mining for supporting buyers' complex decision making: exploratory user study and algorithm comparison. Social Network Analysis and

Mining. 2021;1(4):301-320. https://doi.org/10.1007/s13278-011-0023-y
7. Hu Y, Boyd-Graber JL, Satinoff B, Smith-Renner A. Interactive topic modeling. Machine Learning. 2014;95(3):423-469. https://doi.org/10.1007/s10994-013-5413-0
8. Pavlou PA, Fygenson M. Understanding and Predicting Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior. MIS Quarterly. 2006;30(1):115-143. https://doi.org/10.2307/25148720
9. Yi S, Liu X. Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. Complex & Intelligent Systems. 2020;6(5):621-634. https://doi.org/10.1007/s40747-020-00155-2
10. Khalid MA, Ashraf I, Mehmood A, Ullah S, Ahmad M, Choi GS. GBSVM: Sentiment Classification from Unstructured Reviews Using Ensemble Classifier. Applied Sciences. 2020;10(8):2788-2798. https://doi.org/10.3390/app10082788
11. Sadhasivam J, Babu R. Sentiment Analysis of Amazon Products Using Ensemble Machine Learning Algorithm. International Journal of Mathematical, Engineering and Management Sciences. 2019;4(2):508-520. https://doi.org/10.33889/IJMEMS.2019.4.2-041
12. Kang M, Ahn JC, Lee KS. Opinion mining using ensemble text hidden Markov models for text classification. Expert Systems. 2018;94(1):218-227. https://doi.org/10.1016/j.eswa.2017.07.019
13. Severyn A, Moschitti A, Uryupina O, Plank B, Filippova K. Multi-lingual Opinion Mining on YouTube. Information Processing & Management. 2015;52(1):46-60. https://doi.org/10.1016/j.ipm.2015.03.002
14. Balazs JA, Velásquez JD. Opinion Mining and Information Fusion: A survey. Information Fusion. 2016;27(1):95-110. https://doi.org/10.1016/j.inffus.2015.06.002
15. Rill S, Reinel D, Scheidt J, Zicari RV. PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. Knowledge-Based Systems. 2014;69(1):24-33. https://doi.org/10.1016/j.knosys.2014.05.008
16. Chen Z, Liu B. Mining topics in documents: standing on the shoulders of big data. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data. 2014;2(1):1116-1125. https://doi.org/10.1145/2623330.2623622
17. Nakov P, Rosenthal S, Kiritchenko S, Mohammad SM, Kozareva Z, Ritter A, Stoyanov V, Zhu X. Developing a successful SemEval task in sentiment analysis of twitter and other social media texts. Language Resources and Evaluation. 2016;50(1):35-65. https://doi.org/10.1007/s10579-015-9328-1
18. Poria S, Cambria E, Gelbukh A. Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems. 2016;108(1):42-49. https://doi.org/10.1016/j.knosys.2016.06.009