Multimedia Tools and Applications (2022) 81:38623–38641
https://doi.org/10.1007/s11042-022-12848-6

Detection of spam reviews using hybrid grey wolf optimizer clustering method

Sakshi Shringi[1] · Harish Sharma[1]

Received: 15 January 2021 / Revised: 4 February 2022 / Accepted: 9 March 2022 /
Published online: 25 April 2022

Abstract

Currently, online reviews play an essential role in the decision-making of customers. Various online websites such as Amazon, Yelp, Google Plus, BookMyShow, Facebook, Twitter, etc., allow its users to generate huge bulk of data. The data is generated in the form of feedback/reviews, comments, or tweets. This data is helpful for organizations to improve the quality of their products. Due to dependency on these online reviews, spam reviews are generated pretentiously by some organizations and people concerning promotion or demotion of the prominence of any product, organization, or person. Thus, identifying spam or non-spam review by the naked eye is nearly impossible. Classifying the reviews manually is also highly speculative. Hence, to overcome this issue, a hybrid Grey Wolf Optimizer (GWOK) based clustering method is proposed in this paper to identify spam reviews. In the proposed GWOK, the k-Means algorithm is used for initialization of the initial population for the basic GWO algorithm, and then the GWO algorithm is used for finding the optimal Cluster Heads. To prove that the proposed strategy is effective, three spam datasets, namely Synthetic Spam Reviews, Movie Reviews, and Yelp Hotel & Restaurant Reviews, have been used in our work. The reported results are compared with the existing state-of-art metaheuristic clustering methods like a genetic algorithm (GA), differential evolution (DE), particle swarm optimization (PSO), cuckoo search (CS), and k-Means. The results obtained by experimental and statistical analysis legitimize that the proposed GWOK algorithm surpasses contemporary techniques.

Keywords Grey Wolf Optimizer · k-Means · Data clustering · Spam detection · Metaheuristic methods

✉ Sakshi Shringi
sakshi.shringi@gmail.com

Harish Sharma
hsharma@rtu.ac.in

[1] Rajasthan Technical University, Kota, India

1 Introduction

The expansion of e-commerce and the Internet has raised excessively nowadays. The exponential increase in these sectors has given rise to online reviews and the reliance on such reviews has also increased tremendously. Some of the cases where online reviews play a vital role are:

1. When we want to buy something through an online retail website, product and seller both reviews are important.
2. When we want to buy software, online reviews are of great help.
3. Online reviews for movies help people in deciding whether to watch a movie or not.

Due to the rise of e-commerce, online customer reviews have become an important part of our decision when we are willing to purchase any product online. Since every customer who buys a similar product has his/her perspective of that product which is generated in the form of a review. It is quite difficult to judge whether a review is fake or genuine through the naked eye. Several e-commerce platforms, including Amazon, Flipkart, Myntra, eBay, etc provide the facility of writing user reviews of the product, the review may be of any type thus influencing the decision of the buyer. Hence, online reviews may enhance or degrade the reputation of any brand product. Thus, the prediction of spam reviews is a prerequisite.

Jindal et al. [13], classified spam reviews into three types.

1. Untruthful reviews: The reviews which purposely deceive readers or review mining systems by writing unworthy positive reviews for a specific target object for false promotions, also known as *hyper spam*, on the other handwriting negative reviews for some other specific objects to deteriorate their image, also known as *defaming spam*.
2. Non-reviews: Reviews that contain irrelevant content and commercials.
3. Review on brands: These reviews majorly focus on promoting a brand rather than focusing on the product.

Amongst all the three categories of reviews, detecting untruthful reviews is the hardest task. Some examples of untruthful reviews are as follows:

Review 1 “Fantastic Product. I haven’t seen a better HP LaserJet Pro with such cool features. It has wifi, lan connectivity. I can print from my phone laptop any device wireless to the machine. And it shows up instantly. Moreover, the whole setup process was a breeze. I would not have got a better HP Laserjet product with MFP capabilities at this price. Hats off!”

Review 2 “The copy quality of the printer is horrible, scan quality is just ok and print quality acceptable. It doesn’t have led light in the display so its gonna be more difficult and annoying for you to operate this printer”

The determination of fake and real reviews is highly crucial for the user. Therefore, some standard methods such as n-grams, bag of words, filtering, parts of speech, etc., are proposed. A continuous series of n features form a speech or text sample in n-gram[35]. The major drawback of n-grams is that the feature spaces represented by n-grams models are highly sparse. In bag of word-based method [2], to classify the spam reviews, individual words are considered as a feature, ignoring the meaning of words. Therefore, making such methods less efficient for spam identification. Many other methods such as unigrams-based methods [37, 38], lexical [4], and syntactical features [40] based methods were used by researchers for spam detection. Furthermore, supervised, unsupervised, and semi-supervised machine learning algorithms are commonly employed to detect spam reviews.

Saeed et al. [1] presented a study that compares several machine learning method predictive accuracy for predicting phishing emails. Tingmin et al. [51] presented a novel technique based on deep learning to identify Twitter Spam. Chang et al. [8] gave a comparative analysis of different methods which identify fake hotel reviews. Mukherjee et al. [24] introduced a method that identifies the group of fake reviewers affecting consumer products. Li et al. [15] presented a machine learning-based two-view co-training algorithms to identify review spam of products.

Petrescu et al. [31] presented an analysis of the impact of motivated review campaigns. His findings show that posting positive reviews of the product by the users is highly biased due to these motivating campaigns. Michael Luca [18] showed how online reviews affect the demand of restaurants using a dataset from Yelp.com and Washington State Department of Revenue's restaurant data. He gave three major conclusions through his work: firstly, 5% to 9% of increment in revenue occurs with just a one-star increase in Yelp rating; secondly, this rating does not affect the chain restaurants, only the independent restaurants are affected, thirdly, a declination in markets of chain restaurants due to increased Yelp penetration. Mesleh et al. [21] implemented a classifier for text classification based on Support Vector Machines (SVMs) which use Chi-square method for selecting relevant feature vector at the pre-processing phase for text classification. Catal and Guldan [7] presented a high-performance model that uses Multiple Classifier Problems (MCPs). To detect fraudulent negative reviews, this model combines the strengths of five high-performing individual classifiers and applies the majority vote aggregation technique. Xie et al. [53] proposed a method for detecting spam using significantly associated temporal emulations. Xu et al. [54] presented a technique called as Sparse Aspect Coding Model (SACM), to handle the problem of latent aspect mining. In SACM, the user intrinsic aspect-interest and item intrinsic aspect-quality, the latent variables are used to model the observed review text and complete rating. Hu et al. [10] proposed a novel multi-text summarization technique for hotel reviews. This method helps to identify the top- k most informative sentences. Sasaki and Shinnou [36] introduced a spam detection technique that uses a spherical k-Means algorithm to automatically identify disjoint clusters. This technique is used for text clustering and is based on the vector space model. Yang et al. [56] used a two-layered spam detection flow, showing the trade-off among accuracy and efficiency. McCord and Chuah [20] used four traditional classifiers Random Forest, Naïve Bayesian, Support Vector Machine, and KNN neighbor to identify spam in Twitter and showed that Random Forest classifier performed best with 95.7% precision and 95.7% F-measure values. Mateen et al. [19] proposed a combined method for disclosing of spammer's Twitter platform. This method uses content-based as well as graph-based features for identifying spam.

Singh and Singh [43] merged the efficiency of correlation-based feature selection technique (CFS) and Particle Swarm Based Optimization (PSO) to detect web spam. Singh and Batra [41] introduced a technique based on ensemble learning for spam detection. This method uses a quotient filter for efficient searching and locality-sensitive hashing for similarity searching. Bindu et al. [5], used the Twitter dataset for detecting spam by implementing unsupervised techniques. This technique was implemented on graphs, URLs, and community-based features. Li et al. [16] classified web spam with the help of de-noising auto-encoder and synthetic minority over-sampling techniques in Deep Belief Network (DBN) of machine learning. Liu et al. [17] identified spam tweets by introducing a hybrid method based on fuzzy-redistribution and asymmetric sampling. Miller et al. [22] used clustering methods like stream clustering methods and other methods such as streamKM++, and denStream to classify tweets that were spam. Narayan et al. [27] suggested a semi-supervised PU-learning-based technique to detect spam reviews. Inuwa-Dutse et al. [12]

discovered the accounts that post spam on Twitter by using the feature of account information. Singh et al. [42] gave a model for detection and blocking of counterfeit reviews and spam. Wu et al. [50] presented a hybrid method for spam filtering. This method uses rule-based processing and back-propagation neural networks to filter spam. Asghar et al. [3] achieved an accuracy of 96% by adding a revised scheme that gives weight to features. The experiment was performed based on a rule-based feature weighting scheme that tags the review sentence as spam or ham. Wu et al. [52] proposed a hybrid PU-learning-based Spammer Detection (hPSD) model to identify the hidden spam users based on the reviews of products. This model also detects multi-type spammers by identifying only a small portion of positive samples, which meets particularly real-world application scenarios. A hybrid method for feature selection was proposed by Rjamohana et al. [33], which uses cuckoo search along with harmony search to increase the processing rate and prediction accuracy. The Naive Bayes classifier was employed to classify the review into spam and non-spam.

To classify spam, recently some algorithms have also been employed, which are metaheuristics. Salehi et al. [34] detected email spam by using genetic algorithms. Idris et al. detected email spam by using two algorithms, differential evolution [44] and negative selection algorithms. Idris et al. [11] used a combined method based on particle swarm optimization (PSO) [14] and negative selection for detecting email spam. Shekhawat et al. [39] proposed a model that uses Spider Monkey Optimization (SMO) with k-Means to classify twitter sentiments. Rajamohana et al. [32] proposed a model that uses Adaptive Binary Flower Pollination Algorithm (BFPA), a global optimization technique for feature extraction and Naive Bayes (NB) classifier as the objective function to increase the classification accuracy. Pereira et al. [30] proposed a hybrid technique using the efficiency of local search and evolutionary algorithms to have a good search mechanism and to balance the difference in the population. As the metaheuristic strategies are stochastic in nature and often traps in the local optima, an effective solution is always required which can identify the optimal cluster-heads as well as can properly explore the search region without stagnating at some local optimal point.

Therefore, in this paper, to classify the spam and non-spam reviews, a novel metaheuristic method based on clustering has been proposed. The paper is sectioned in the following manner: The Grey Wolf Optimizer (GWO) and k-Means are described in Section 2. The proposed method, Grey Wolf Optimizer using k-Means (GWOK) is described in Section 3. Section 4 illustrates the experimental results and we have concluded our work in Section 5.

2 Preparatory

2.1 k-Means clustering

To organize the items of a set among segregated groups or clusters, partitioning is the most fundamental method of cluster analysis. One of the most well-known methods is k-Means [9]. Given a data-set D, which include n-items in the Euclidean space, the k-Means algorithm distributes the item n among k-clusters, C_i, ..., C_k, i.e., $C_i \subset$ D and $C_i \cap C_j = \emptyset$ for (1 $\leq$ i, j $\leq$ k). To estimate the quality of the partitioning method, an objective function is used, aspiring higher intercluster resemblance and lower intercluster resemblance.

To represent a cluster, k-Means uses the center of that cluster known as centroid C_i. The difference between an item q $\in C_i$ and $\mathbf{c_i}$, the centroid, is given by the Euclidean distance between any two points given by, dist(q,$\mathbf{c_i}$). To identify the quality of any cluster, the sum of *squared error* is calculated amongst all items in the cluster and the centroid, given as follows:

$$SE = \sum_{i=1}^{k} \sum_{q \in C_i} dist(q, c_i)^2 \tag{1}$$

where SE represents the squared error in the given dataset for all objects,
q represents a point in the given object,
c_i represents cluster centroid of cluster c_i

The pseudo-code of k-Means is given in Algorithm 1:

Algorithm 1 k-Means Algorithm.

```
Input:
  • D: n objects data set
  • k: the number of clusters
Output: k-clusters set
Mechanism:
1 Randomly select k-objects from the data set D as the initial cluster-centers.
2 Depending upon the mean value of the item in the cluster, re-designate each item to the cluster having maximum similarity to the item.
3 The data point with the minimum distance to all centroids is assigned to the cluster centroid.
4 Calculate the newly formed cluster centroid by (2):
      $C_i = \frac{1}{n_i} \sum_{j=1}^{n_i} t_i$      (2)
  where '$n_i$' indicates the number of datapoints in $i^{th}$ cluster.
5 The difference between distinctive datapoint and newly attained centroids is recomputed.
6 Go to Step-3, until all the data points are covered.
```

2.2 Grey wolf optimizer(GWO)

The GWO algorithm developed by Mirjalili [23] is a population-based metaheuristic algorithm, designed to explore and construct a heuristic (partial search algorithm), to find an optimal solution for any optimization problem. All the algorithms with randomization and local search capacity are known as metaheuristic algorithms [55]. Metaheuristic algorithms can relatively handle problems with a huge population [46]. Unlike other optimization algorithms, metaheuristic algorithms do not assure to obtain the optimal solution, but they are capable of computing sub-optimal, good-quality solutions and take feasible execution time [28]. GWO is one such type of metaheuristic algorithm that mimes the attacking behavior and management hierarchy of grey wolves. In GWO, to fabricate the management hierarchy, the colony of wolves is divided into primarily four main classes, alpha (α), beta (β), omega (ω) and delta (δ). The alpha wolf is the leader and is considered the best ones. They are responsible for making decisions for hunting, time to sleep, waking up time, and so on. Beta is the second-level wolves. They are the auxiliary wolves that help alpha wolves in decision-making and other activities. Delta is the third-best, and it is responsible for sacrifice. These wolves are responsible for dominating other wolves. Omega is the lowest-ranked grey wolves. They are considered weaklings and are ready to sacrifice. All other wolves are delta wolves.

The working of GWO can be described in the following four steps:

1. Encircling prey
2. Hunting
3. Attacking prey
4. Search prey

2.2.1 Mathematical model

In this subsection, the mathematical model for encircling the prey, hunting, attacking, and searching the prey is illustrated.

2.2.2 Encircling the prey

At first, the grey wolves encircle the prey which can mathematically be given as:

$$D = \mid C.G_p(t) - G(t) \mid \tag{3}$$

$$G(t+1) = G_p(t) - A * D \tag{4}$$

$$A = 2 * a * r_1 - a \tag{5}$$

$$a = 2 - 2(\frac{t}{I}) \tag{6}$$

$$C = 2 * r_2 \tag{7}$$

where current iteration is indicated by t, vector A and C indicates coefficient vectors. G_p and G indicates the prey position and grey wolf position, respectively. a is a vector that linearly decreases from 2 to 0 over the course of iterations. r_1, r_2 are random vectors in the range [0,1]. Initially, vector A has a maximum value, which decreases gradually as the iterations increase which can be calculated by (6). I indicate the maximum number of iterations.

2.2.3 Hunting

To mathematically simulate the hunting behavior of grey wolves, the α, β, and δ are considered as the best solution which possesses the optimal information about where the prey is. Based on this information, other grey wolves update their positions using the following equations [23] :

$$G_1 = G_\alpha(t) - A_1 * D_\alpha \tag{8}$$

$$G_2 = G_\beta(t) - A_2 * D_\beta \tag{9}$$

$$G_3 = G_\delta(t) - A_3 * D_\delta \tag{10}$$

where,

$$D_\alpha = \mid C_1.G_\alpha(t) - G \mid D_\beta = \mid C_2.G_\beta(t) - G \mid D_\delta = \mid C_3.G_\delta(t) - G \mid$$

$$G(t+1) = \frac{G_1 + G_2 + G_3}{3} \tag{11}$$

2.2.4 Search prey (Exploration)

The exploitation and exploration behavior in the GWO algorithm, depends upon the A and C parameters. A lies randomly in the range of [-a,a]. The wolf shows exploration behavior when $\mid A \mid >1$ and $C >1$.

2.2.5 Attack on prey (Exploitation)

The attack on prey leads to exploitation. Exploitation occurs if $\mid A \mid <1$ and $C <1$.

The steps of GWO Algorithm are described in Algorithm 2.

Algorithm 2 Grey Wolf Optimization(GWO).

```
Input:
  • Number of search agents
  • Number of MaxofGeneration
  • fitness function
Output: Best Solution $G_\alpha$(t)
Initialize the population $G_p$ (p = 1, 2, ..., n);
Initialize $\alpha$, A, and C;
Calculate the fitness of each wolf:
$G_\alpha$(t) =the first-best known as the leader
$G_\beta$(t) =the second-best who assist $\alpha$
$G_\delta$(t) =the third-best is the subordinate
while $t < MaxofGeneration$ do
    for $each$ wolf do
        Modify the position of the current wolf by (11)
    end for
    Modify $\alpha$, $A$, and $C$
    Compute the fitness of individual wolf
    Modify $G_\alpha$(t),$G_\beta(t)$, $G_\delta$(t)
    t=t+1

end while
return $G_\alpha$(t)
```

3 Proposed clustering method for spam review detection

In this paper, we introduce a hybrid Grey Wolf Optimizer using k-Means (GWOK) clustering mechanism to identify the spam reviews. The GWOK helps in identifying spam reviews by the following phases:

1. Preprocessing of data
2. Feature Extraction Phase
3. Feature Selection using Chi-square

4. GWO clustering method
5. Testing the Proposed Clustering Method.

Figure 1 illustrates the flowchart of the proposed method.

3.1 Preprocessing of data

The online reviews collected from social media have noise in the form of unwanted and vague words, stop words, URLs, etc., which are not desired in feature extraction. The removal of such uncertain words is executed in two different phases, performed through python natural language toolkit (NLTK) [6]:

3.1.1 Phase 1

Eliminates all the noise and word with uncertainty by the following steps:

1. Conversion of all reviews in lowercase.
2. Remove special symbols from the online reviews such as , @, #, etc.
3. Remove stop words from reviews that do not consist of any relevant information such as with, at, of, to, etc. with the help of the NLTK library.
4. Replace multiple white spaces and add single white spaces in their place.
5. Remove all numbers from the reviews.
6. Remove few punctuations from the reviews such as hyphen, braces, slash, and quotation marks.

3.1.2 Phase 2

In this phase, the paragraphs are divided into sentences by applying lexical analysis or tokenization. After tokenization, the words are reduced to their root form by using lemmatization, like “changing” is altered to “change”.

3.2 Feature extraction phase

Once the preprocessing phase is complete, relevant features are extricated with the help of Linguistic Inquiry and Word Count (LIWC 2015) [29]. The LIWC program has the main

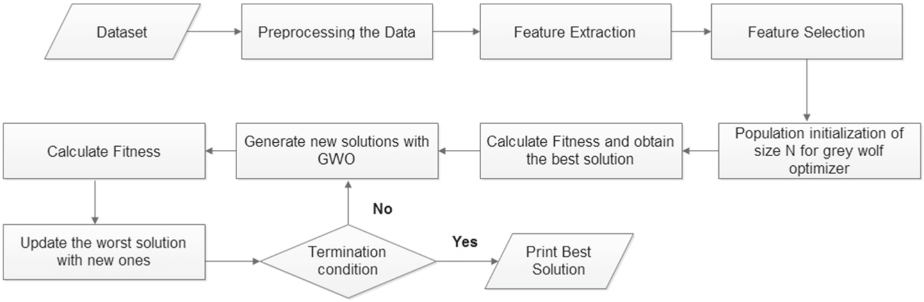

Dataset
Preprocessing the Data
Feature Extraction
Feature Selection
Population initialization of size N for grey wolf optimizer
Calculate Fitness and obtain the best solution
Generate new solutions with GWO
Calculate Fitness
Update the worst solution with new ones
Termination condition
No
Yes
Print Best Solution

Fig. 1 Proposed clustering method flowchart

text analysis module in addition to a group of built-in dictionaries. The work of LIWC is to read the text given and count the percentage of words reflecting various emotions, styles of thinking, concerns of society, and even parts of speech. The features provided by LIWC are generally 93.

3.3 Feature selection using chi-square

In this phase, the relevant attributes are selected from the 93 extracted features given by LIWC. Feature selection plays an important role due to the following reasons:

1. Eliminates the repetitive data.
2. Picks relevant attributes/variables.
3. Minimizes the occurrence of overfitting.
4. Minimizes the training time.

93 features were selected using the LIWC tool. The features may be irrelevant and may cause the problem of overfitting [48]. As the number of features increases, the training time [37] also increases [47]. To eliminate the redundant and irrelevant features, the Chi-square feature selection method used for text classification is implemented [57]. It is one of the simplest tools for univariate feature selection by performing the univariate statistical analysis for classification. The feature interactions are neglected in this process. It is mainly used for categorical data and is extensively used for text data. The Chi-square method is majorly used for categorical features in a dataset. The chi-square is calculated amongst each feature and the target and the feature with the highest Chi-square value is selected.

The Chi-square score is given by the following equation:

$$Ch = \frac{(Observedfrequency - Expectedfrequency)^2}{Expectedfrequency} \tag{12}$$

where, Observed frequency = Total number of class observations,
Expected frequency = Number of expected class observations without any relation between feature and the target. The relevant features are selected using the Chi-square method for all three datasets.

3.4 k-Means GWO (the proposed clustering mechanism)

To classify the reviews into spam and non-spam, the proposed method finds out optimal cluster heads with the help of Grey Wolf Optimizer(GWO). Spam reviews are classified using the following steps:

1. Within the search space, randomly initialize the population of Grey Wolf Optimizer.
2. Position of the search agents represents the cluster head coordinates of spam and non-spam reviews.
3. For each search agent, the fitness is calculated by using accuracy as the objective function.
4. Use Grey Wolf Optimizer to optimize the clusters.

Algorithm 3 shows the steps of the proposed mechanism

Algorithm 3 Proposed clustering method.

```
Input:
  • Number of clusters
  • Dataset
Output: Optimum Cluster Head (CH)
Define number of features (f)
$-d$ Dimension of search agents (d) = 2 * f
$-N$ Number of Search Agents
$-R$ Number of Runs
$-MaxIter$ Number of max. iterations
Apply Chi-square on the dataset to get best 'f' features.
for $i = 1$ $to$ $R$ do
    Initialize Alpha, Beta, and Delta position to zero.
    Initialize Alpha, Beta, and Delta score to minus infinity (-$\infty$).
    for $i = 1$ $to$ $N$ do
        Generate a random Cluster Head (CH) within the search space and assign it to $i^{\text{th}}$ search agent.
    end for
    for $i = 1$ $to$ $MaxIter$ do
        for $i = 1$ $to$ $N$ do
            Calculate the fitness of $i^{\text{th}}$ search agent.
            Update the position of Alpha, Beta, and Delta.
        end for
    end for
    Update the value of $a$
    for $i = 1$ $to$ $N$ do
        for $i = 1$ $to$ $d$ do
            Calculate the vectors A1, A2, and A3 form equation (5)
            Calculate the vectors C1, C2, and C3 form equation (7)
            Evaluate $G_1$, $G_2$ and $G_3$ from (8), (9) and (10) respectively.
            Calculate the new position of $i^{th}$ search agent along $d^{\text{th}}$ dimension from equation (11)
        end for
    end for
end for
Return Optimum Cluster Head (CH)
```

4 Experimental results

The effectiveness of the proposed Grey Wolf Optimizer using the k-Means (GWOK) clustering method is tested on the Synthetic Spam Review, Movie Review, and Yelp review datasets. We have used MATLAB 2016a to simulate all the experiments. The experiments were performed on a system with the configuration of 20.21 GHz Intel(R) Core(TM) i7 processor and size of RAM 16 GB.

4.1 Performance analysis of proposed k-Means grey wolf optimizer (GWOK)

To train a model, using k-Means, as the number of samples of the training data increases, the computational time and space also increase, due to the Mean computed in each iteration. In the proposed GWOK, to reduce this computational complexity of taking mean in each iteration, we randomly shift a Cluster Head(CH), to get a new combination of Cluster Heads(CH). This helps to find the optimal set of Cluster Heads(CH) within a given time or iteration.

The proposed GWOK is compared with the existing nature-inspired algorithms, like cuckoo search (CS) algorithm, Differential Evolution (DE) algorithm, Particle Swarm Optimization (PSO) algorithm, and Genetic Algorithm (GA) for validating the accuracy and computational time of the proposed method. The population size (N) is 50 and the maximum iteration (MaxIter) is 1000 for all the algorithms. The average value of accuracy and computational time for 30 runs, with dimension, equals twice the number of features (f), on all three datasets is presented in Table 1. It can be seen that the accuracy of the Grey Wolf Optimizer using k-Means is better than the other methods.

4.2 Experimental analysis of proposed k-Means GWO (GWOK)

Synthetic spam reviews, Yelp hotel and restaurant reviews, and Movie reviews datasets are used to test the proposed method. A brief description of all three datasets is given in Table 2.

4.2.1 Synthetic spam reviews dataset

The synthetic spam dataset was unlabelled [49] and was retrieved from the *Database and Information System Laboratory, University of Illinois (TripAdvisor Dataset)*. The spamming methods known as, synthetic review method for spamming have been used to generate spam reviews [45]. This method generated 479 reviews, out of which 316 reviews were spam and 163 reviews generated were non-spam.

4.2.2 Movie review

The movie review dataset is generated from the popular site IMDB. This dataset has a total of 8544 reviews, out of which 3998 were spam reviews and 4546 reviews were non-spam.

Table 1 Parameters for k-Means GWO

S. No.	Parameter	PSO	DE	GA	CS	GWO
1.	Population (N)	50	50	50	50	50
2.	Maximum Iteration	1000	1000	1000	1000	1000
3.	Dimension	2*f	2*f	2*f	2*f	2*f
4.	Inertia weight (w)	0.8	-	-	-	-
5.	Crossover Rate (CR)	-	0.5	-	0.8	-
6.	Cognitive constant (c_1)	2	-	-	-	-
7.	Social constant (c_2)	2	-	-	-	-
8.	Mutation rate (F)	-	0.8	0.3	-	-
9.	Probability ($Prob_i$)	-	-	-	0.25	-
10.	Step scaling factor (α)	-	-	-	0.01	-

Table 2 Dataset Description

S. NO.	Dataset used	Total Reviews	Spam Reviews	Non-Spam Reviews
1.	Synthetic Spam Reviews	478	163	315
2.	Movie Reviews	8544	3998	4546
3.	Yelp Hotel & Restaurant Reviews	4952	3709	1243

4.2.3 Yelp hotels & restaurant reviews dataset

Yelp hotels & restaurant review dataset is extracted through Yelp.com, consisting of 85 hotels and 130 restaurant reviews, in the areas of Chicago [25, 26]. The reviews of both popular and disliked hotels and restaurants are considered in the dataset. There is a total of 4952 reviews, out of which 3709 are spam and 1243 are non-spam.

These datasets are noisy. Therefore, we prepreprocess the datasets to remove the noise as described in Section 3.1. After preprocessing, 93 features are extracted from these datasets using the LIWC 2015 tool. as all the features may not be of relevance, the Chi-square feature selection method, as described in Section 3.3 is applied to these 93 features to obtain the best features. The optimum number of features selected from all three datasets is given in Table 3.

For measuring the efficacy of the proposed clustering method, we have evaluated the classification accuracy. The values of precision and recall are also evaluated to check the performance of GWOK and to compare it with other existing techniques, as classification accuracy alone can be deceptive if the number of instances is not equal in each class.

4.3 Results

For computing the values of accuracy, precision, and recall, a confusion matrix sized $n \times n$ is generated. The confusion matrix has n number of classes with C_{ji} representing the number of patterns predicted in i by j.

The values used in Table 4 have the following significance:

- TP represents precisely predicted spam reviews.
- TN represents precisely predicted ham review predicted correctly.
- FP represents imprecisely predicted ham reviews.
- FN represents imprecisely predicted spam review.

Table 5 illustrates the value of precision and recall for the three datasets over original and optimal features.

Using confusion matrix, the value of accuracy, precision, and recall can be calculated by (13), (14), and (15) respectively.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

Table 3 Feature Selected

S. No.	Dataset Used	Optimum Features Selected
1.	Synthetic Spam Reviews	20
2.	Yelp Reviews	21
3.	Movie Reviews	11

Table 4 Confusion Matrix

	Spam	Non-spam
Spam	TP	TN
Non-spam	FP	FN

Table 5 Comparative analysis of measure of Mean values of precision and recall of proposed method with other methods for original and optimum features dataset

Dataset	Method	Original Dataset		Optimal Dataset	
		Precision (%)	Recall (%)	Precision (%)	Recall (%)
Synthetic Spam Reviews	k-Means	55.19	57.95	60.61	62.69
	PSO	59.57	60.01	62.32	60.61
	DE	59.33	61.06	61.26	62.16
	GA	57.09	59.85	60.09	61.89
	CS	59.57	60.85	60.17	62.99
	GWOK	**63.75**	**65.40**	**66.68**	**68.20**
Movie Review	k-Means	52.03	53.02	54.51	55.64
	PSO	53.02	54.32	55.22	56.10
	DE	51.12	53.24	54.08	54.89
	GA	52.67	54.04	54.97	56.21
	CS	55.15	56.43	56.84	57.82
	GWOK	**58.47**	**60.82**	**61.94**	**62.43**
Yelp Hotel & Restaurant Review	k-Means	70.19	84.72	74.27	88.97
	PSO	69.87	87.39	70.78	89.66
	DE	67.90	80.42	70.05	87.83
	GA	64.72	88.24	85.13	90.95
	CS	70.42	89.95	71.06	92.98
	GWOK	**72.99**	**92.94**	**75.99**	**99.64**

Table 6 Comparative analysis to measure mean and standard deviation values of Accuracy of the proposed method with other methods over datasets with original features

Dataset	Method	Accuracy (%)		Computational Time (s)	
		Mean	STD	Mean	STD
Synthetic Spam Reviews	k-Means	70.28	1.04	300.71	1.45
	PSO	68.47	1.05	317.87	1.89
	DE	70.57	1.02	290.82	1.58
	GA	68.67	0.951	295.64	1.57
	CS	70.63	0.84	289.49	1.86
	GWOK	**75.43**	**0.524**	**281.62**	**1.56**
Movie Review	k-Means	59.56	1.26	4844.23	2.28
	PSO	58.24	1.18	4891.05	1.96

Table 6 (continued)

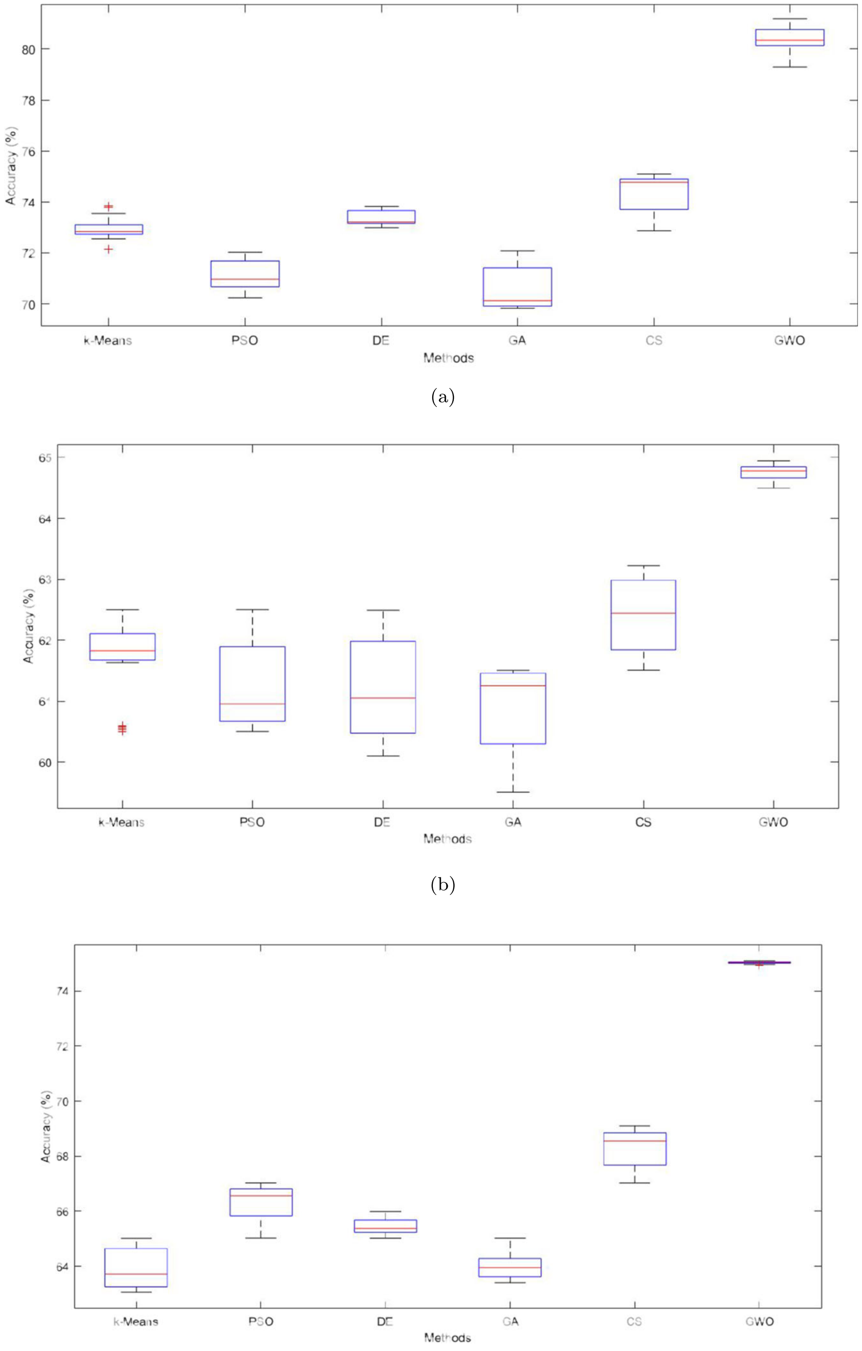
Dataset	Method	Accuracy (%)		Computational Time (s)	
		Mean	STD	Mean	STD
	DE	60.49	1.25	4894.49	1.88
	GA	58.67	1.12	4875.62	2.15
	CS	60.07	1.24	4929.78	1.97
	GWOK	**61.59**	**1.21**	**4496.17**	**1.51**
Yelp Hotel & Restaurant Review	k-Means	60.24	1.71	2539.12	1.92
	PSO	59.25	1.25	2584.49	2.37
	DE	59.74	1.40	2674.34	2.58
	GA	59.61	1.22	2785.12	2.05
	CS	60.12	1.34	2589.73	1.76
	GWOK	**70.09**	**1.09**	**2459.42**	**1.29**

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

Table 7 Comparative analysis to measure mean and standard deviation values of Accuracy of the proposed method with other methods over datasets with optimum features

Dataset	Method	Accuracy (%)		Computational-Time (s)	
		Mean	STD	Mean	STD
Synthetic Spam Reviews	k-Means	72.68	0.405	292.73	1.44
	PSO	70.24	0.563	311.07	1.69
	DE	73.85	0.410	287.67	1.58
	GA	71.67	0.852	293.66	1.65
	CS	74.63	0.712	288.02	1.68
	GWOK	**80.43**	**0.269**	**280.51**	**1.35**
Movie Review	k-Means	61.66	0.587	4823.63	2.11
	PSO	61.58	0.746	4852.95	1.95
	DE	61.22	0.802	4845.49	1.81
	GA	60.96	0.677	4863.35	2.05
	CS	62.04	0.591	4914.74	1.87
	GWOK	**64.75**	**0.12**	**4468.31**	**1.41**
Yelp Hotel & Restaurant Review	k-Means	64.52	0.715	2511.42	1.65
	PSO	60.39	0.609	2534.44	1.32
	DE	60.42	0.275	2474.30	1.22
	GA	60.25	0.448	2726.70	1.03
	CS	61.95	0.665	2534.82	1.76
	GWOK	**75.01**	**0.032**	**2409.71**	**0.984**

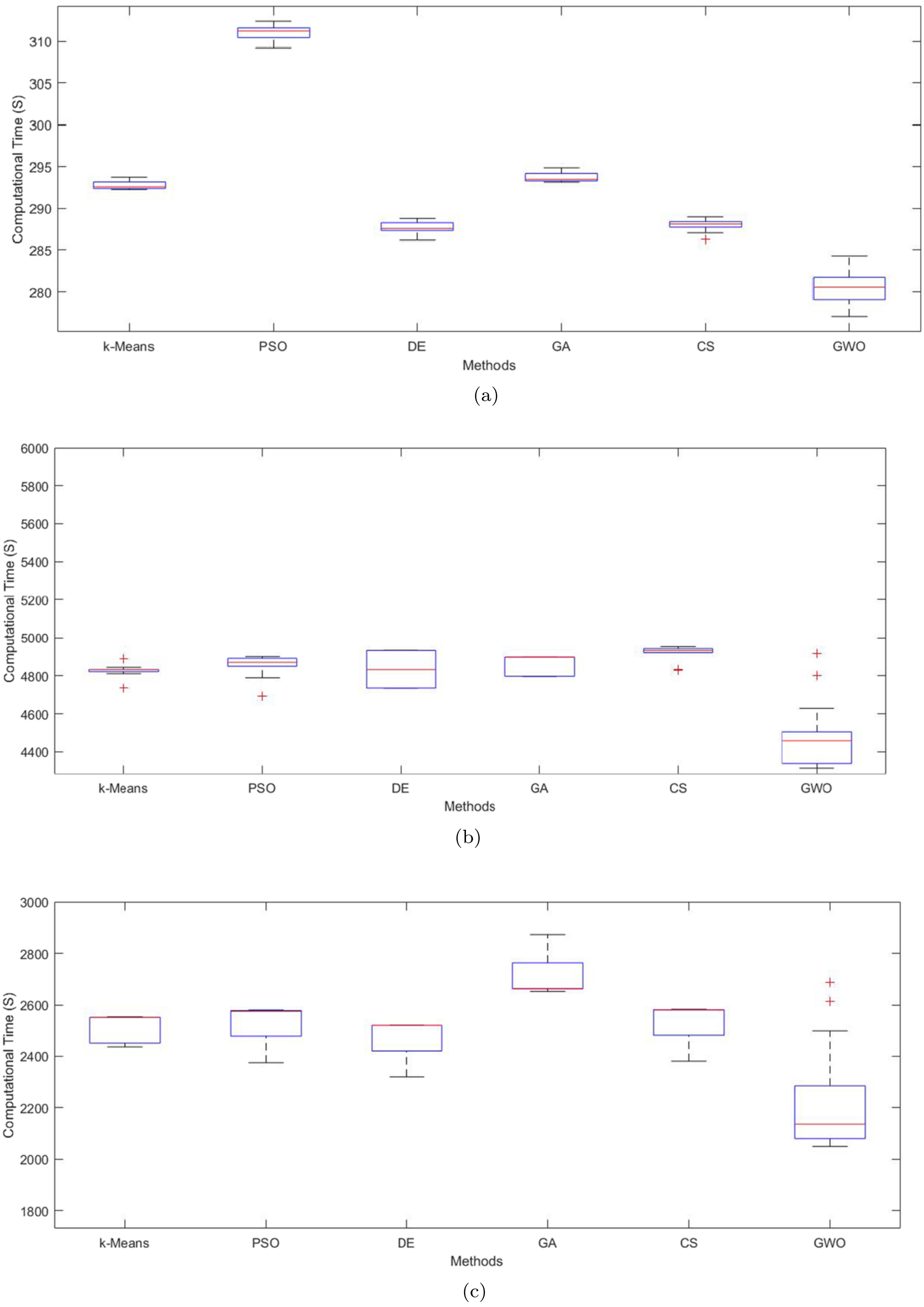


(a)

(b)

(c)

Fig. 2 Box plots of Accuracy of proposed Grey Wolf based clustering method and other nature-inspired algorithms of (a) Synthetic Spam Review Dataset, (b) Movie Review Dataset, and (c) Yelp Hotels & Restaurant Review Dataset



(a)

(b)

(c)

Fig. 3 Box plots of Computational Time of proposed Grey Wolf based clustering method and other nature-inspired algorithms of (a) Synthetic Spam Review Dataset, (b) Movie Review Dataset, and (c) Yelp Hotels & Restaurant Review Dataset

The mean and standard deviation values for optimum feature selected after feature selection, on all three datasets is described in Tables 6 and 7 and compared with other existing nature-inspired algorithms. From the Table 7, we can see that the proposed Grey Wolf Optimizer clustering method outperforms other existing methods in respect of accuracy and computational time.

The boxplots of accuracy and computational time (seconds) for all the three datasets over the optimal features are depicted in Figs. 2 and 3 respectively. From the box plots, we can see that the accuracy of the proposed grey wolf optimization clustering method is high as compared to other metaheuristic algorithms. The proposed method also takes less time for execution as compared to other methods, thus proving the efficacy of our algorithm.

So, through experiments, it is proved that the GWOK is an efficient and Robust algorithm to solve the spam review detection problem. So this algorithm can be considered as an prominent solution to solve the spam review detection problem.

5 Conclusion

In this paper, for the classification of spam reviews, a novel clustering approach, Grey Wolf Optimizer with k-Means is proposed. The proposed method is tested on three different datasets namely synthetic spam reviews, movie reviews and, yelp hotels & restaurant reviews. The proposed method is compared with k-Means, PSO, DE, GA, and CS. The experimental results demonstrate that the proposed clustering method for detecting spam outperforms the existing nature-inspired methods like PSO, DE, GA, and CS. Boxplots show the consistency of the proposed method. The Chi-square feature selection method is used for finding the optimum features. In the future, some new optimization strategies or a modified Grey Wolf Optimizer-based hybrid approach can be employed in combination with other feature selection methods such as wrapper-based, Pearson's correlation, etc. to attain higher accuracy for the given datasets.

Funding We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

1. Abu-Nimeh S, Nappa D, Wang X, Nair S (2007) A comparison of machine learning techniques for phishing detection. In: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, pp 60–69
2. Angeli A, Filliat D, Doncieux S, Meyer J-A (2008) Fast and incremental method for loop-closure detection using bags of visual words. IEEE Trans Robot 24(5):1027–1037
3. Asghar MZ, Ullah A, Ahmad S, Khan A (2020) Opinion spam detection framework using hybrid classification scheme. Soft Comput 24(5):3475–3498
4. Baeza-Yates R, Ribeiro-Neto B et al (1999) Modern information retrieval, vol 463. ACM Press, New York
5. Bindu PV, Mishra R, Santhi Thilagam P (2018) Discovering spammer communities in twitter. J Intell Inform Syst 51(3):503–527
6. Bird S, Klein E, Loper E (2009) Natural language processing with python: analyzing text with the natural language toolkit. "O'Reilly Media Inc."
7. Catal C, Guldan S (2017) Product review management software based on multiple classifiers. Iet Softw 11(3):89–92

8. Chang T, Hsu PY, Cheng MS, Chung CY, Yi LC (2015) Detecting fake review with rumor model—case study in hotel review. In: International conference on intelligent science and big data engineering. Springer, pp 181–192
9. Hartigan JA, Wong MA (1979) Algorithm as 136: A k-means clustering algorithm. J R Stat Soc Ser C (Appl Stat) 28(1):100–108
10. Hu Y-H, Chen Y-L, Chou H-L (2017) Opinion mining from online hotel reviews–a text summarization approach. Inform Process Manag 53(2):436–449
11. Idris I, Selamat A, Omatu S (2014) Hybrid email spam detection model with negative selection algorithm and differential evolution. Eng Appl Artif Intell 28:97–110
12. Inuwa-Dutse I, Liptrott M, Korkontzelos I (2018) Detection of spam-posting accounts on twitter. Neurocomputing 315:496–511
13. Jindal N, Liu B (2008) Opinion spam and analysis. In: Proceedings of the 2008 international conference on web search and data mining, pp 219–230
14. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks, vol 4. IEEE, pp 1942–1948
15. Li FH, Huang M, Yi Y, Zhu X (2011) Learning to identify review spam. In: Twenty-second international joint conference on artificial intelligence
16. Li Y, Nie X, Huang R (2018) Web spam classification method based on deep belief networks. Expert Syst Appl 96:261–270
17. Liu S, Zhang J, Xiang Y (2016) Statistical detection of online drifting twitter spam. In: Proceedings of the 11th ACM on Asia conference on computer and communications security, pp 1–10
18. Luca M (2016) Reviews: reputation, and revenue: The case of yelp. com. Com (March 15, 2016). Harvard Business School NOM Unit Working Paper (12-016)
19. Mateen M, Iqbal MA, Aleem M, Islam MA (2017) A hybrid approach for spam detection for twitter. In: 2017 14Th international bhurban conference on applied sciences and technology (IBCAST). IEEE, pp 466–471
20. Mccord M, Chuah M (2011) Spam detection on twitter using traditional classifiers. In: International conference on autonomic and trusted computing. Springer, pp 175–186
21. Mesleh AMoA (2007) Chi square feature extraction based svms arabic language text categorization system. J Comput Sci 3(6):430–435
22. Miller Z, Dickinson B, Deitrick W, Hu W, Wang AH (2014) Twitter spammer detection using data stream clustering. Inf Sci 260:64–73
23. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
24. Mukherjee A, Liu B, Glance N (2012) Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st international conference on World Wide Web, pp 191–200
25. Mukherjee A, Venkataraman V, Liu B, Glance N (2013) What yelp fake review filter might be doing? In: Proceedings of the international AAAI conference on web and social media, p 7
26. Mukherjee A, Venkataraman V, Liu B, Glance N et al (2013) Fake review detection: Classification and analysis of real and pseudo reviews. Technical Report UIC-CS-2013–03. University of Illinois at Chicago. Tech Rep
27. Narayan R, Rout JK, Jena SK (2018) Review spam detection using semi-supervised technique. In: Progress in intelligent computing techniques: theory, Practice, and Applications. Springer, pp 281–286
28. Nesmachnow S (2014) An overview of metaheuristics: accurate and efficient methods for optimisation. Int J Metaheuristics 3(4):320–347
29. Pennebaker JW, Boyd RL, Jordan K, Blackburn K (2015) The development and psychometric properties of liwc2015. Technical report
30. Pereira FB, Marques JMC (2009) A study on diversity for cluster geometry optimization. Evol Intel 2(3):121
31. Petrescu M, O'Leary K, Goldring D, Mrad SB (2018) Incentivized reviews: Promising the moon for a few stars. J Retail Consum Serv 41:288–295
32. Rajamohana SP, Umamaheswari K, Abirami B (2017) Adaptive binary flower pollination algorithm for feature selection in review spam detection. In: 2017 International conference on innovations in green energy and healthcare technologies (IGEHT). IEEE, pp 1–4
33. Rajamohana SP, Umamaheswari K, Vasantha Keerthana S (2017) An effective hybrid cuckoo search with harmony search for review spam detection. In: 2017 Third international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB). IEEE, pp 524–527
34. Salehi S, Selamat A, Bostanian M (2011) Enhanced genetic algorithm for spam detection in email. In: 2011 IEEE 2Nd international conference on software engineering and service science. IEEE, pp 594–597
35. Santos I, Penya YK, Devesa J, Bringas PG (2009) N-grams-based file signatures for malware detection. ICEIS (2) 9:317–320

36. Sasaki M, Shinnou H (2005) Spam detection using text clustering. In: 2005 International conference on cyberworlds (CW'05). IEEE, pp 4–pp
37. Sedhai S, Sun A (2017) Semi-supervised spam detection in twitter stream. IEEE Trans Comput Soc Syst 5(1):169–175
38. Shehnepoor S, Salehi M, Farahbakhsh R, Crespi N (2017) Netspam: A network-based spam detection framework for reviews in online social media. IEEE Trans Inform Forens Secur 12(7):1585–1595
39. Shekhawat SS, Shringi S, Sharma H (2020) Twitter sentiment analysis using hybrid spider monkey optimization method. Evol Intel, 1–10
40. Shojaee S, Murad MAA, Azman AB, Sharef NM, Nadali S (2013) Detecting deceptive reviews using lexical and syntactic features. In: 2013 13Th international conference on intellient systems design and applications. IEEE, pp 53–58
41. Singh A, Batra S (2018) Ensemble based spam detection in social iot using probabilistic data structures. Futur Gener Comput Syst 81:359–371
42. Singh M, Kumar L, Sinha S (2018) Model for detecting fake or spam reviews. In: Ict based innovations. Springer, pp 213–217
43. Singh S, Singh AK (2018) Web-spam features selection using cfs-pso. Procedia Computer Science 125:568–575
44. Storn R, Price K (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. J Global Optim 11(4):341–359
45. Sun H, Morales A, Yan X (2013) Synthetic review spamming and defense. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 1088–1096
46. Talbi E-G (2009) Metaheuristics: from design to implementation, vol 74. Wiley, New York
47. Tran CT, Zhang M, Andreae P, Xue B, Bui LT (2018) Improving performance of classification on incomplete data using feature selection and clustering. Appl Soft Comput 73:848–861
48. Van der Aalst WMP, Rubin V, Verbeek HMW, Van Dongen BF, Kindler E, Günther CW (2010) Process mining: a two-step approach to balance between underfitting and overfitting. Softw Syst Model 9(1):87
49. Wang H, Yue Lu, Zhai C (2010) Latent aspect rating analysis on review text data: a rating regression approach. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 783–792
50. Wu C-H (2009) Behavior-based spam detection using a hybrid method of rule-based techniques and neural networks. Expert Syst Appl 36(3):4321–4330
51. Wu T, Liu S, Zhang J, Xiang Y (2017) Twitter spam detection based on deep learning. In: Proceedings of the australasian computer science week multiconference, pp 1–8
52. Wu Z, Wang Y, Wang Y, Wu J, Cao J, Lu Z (2015) Spammers detection from product reviews: a hybrid model. In: 2015 IEEE International conference on data mining. IEEE, pp 1039–1044
53. Xie S, Wang G, Lin S, Yu PS (2012) Review spam detection via temporal pattern discovery. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 823–831
54. Xu Y, Lin T, Lam W, Zhou Z, Cheng H, So AM-C (2014) Latent aspect mining via exploring sparsity and intrinsic information. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management, pp 879–888
55. Yang X-S (2010) Nature-inspired metaheuristic algorithms. Luniver Press
56. Yang Z, Nie X, Xu W, Guo J (2006) An approach to spam detection by naive bayes ensemble based on decision induction. In: Sixth international conference on intelligent systems design and applications, vol 2. IEEE, pp 861–866
57. Zhai Y, Song W, Liu X, Liu L, Zhao X (2018) A chi-square statistics based feature selection method in text classification. In: 2018 IEEE 9Th international conference on software engineering and service science (ICSESS). IEEE, pp 160–163

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.