International Journal of Engineering & Technology, 7 (1.3) (2018) 112-117

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Hybrid optimization for feature selection in opinion mining

Prasanna Moorthi N. [1] *, Mathivanan V [2]

[1] Research Scholar, Computer Science and Engineering Department AMET University, Chennai
[2] Research Supervisors, Computer Science and Engineering Department AMET University, Chennai
**Corresponding author E-mail: Prasannamoorthi.n@gmail.com*

Abstract

A sub-discipline of Information Retrieval (IR) is opinion mining and the lexicon of computers is not concerned of the subject of the document, but about the opinion expressed. It has caused a large impact in the arena of academics and industry as it has a wide area of research and the applications are widespread. Feature selection is a vital step in opinion mining, as its individual feature decides the opinions expressed by the customers. Feature selection reduces the dimensionality of data by avoiding non-relevant features; it can be considered as a necessary and excellent process for data mining applications. In this study, feature subset is optimized through Particle Swarm Optimization (PSO) algorithm, Cuckoo Search (CS) algorithm and hybridized PSO-CS algorithm. Classification is done through Naïve bayes and K-Nearest Neighbours (KNN) classifiers. Feature extraction has its basis on Term Frequency-Inverse Document Frequency (TF-IDF). The accuracy of classification precision is increased by the reduction in size of feature subset and computational complexity.

Keywords: *Opinion Mining; Feature Selection; Particle Swarm Optimization (PSO); Cuckoo Search (CS); Term Frequency-Inverse Document Frequency (TF-IDF); Naïve Bayes and K-Nearest Neighbours (KNN).*

1. Introduction

The process of opinion mining comes in handy when knowledge is extracted automatically from others' opinion about a specific topic or issue. For discovery of knowledge and decision support, human perception and opinion of the user is quite important. The main idea behind opinion mining is to make the computer identify and express emotions. Sentiment is a thought, view, or behaviour based on the emotion instead of reason; so opinion mining is also referred to as sentiment analysis. It is human tendency generally to seek opinion from friends at the time of making crucial decisions, which implies that it is natural to seek opinion of others in making a good decision [1].

The emotions of writers which are expressed as affirmative or negative comments through the evaluation of a hoard of documents is called Natural Language Processing (NLP) and Information Extraction (IE) which is a part of opinion mining. The techniques of IR and computational linguistics are combined in this. Sentiment analysis' work is nothing but classifying documents and determining its polarity. Polarity can be positive, negative or neutral. Sentiment analysis is takes place at three levels:

- Document level – Here whole documents are classified as positive, negative or neutral;
- Sentence level – As the name implies, the sentence is classified as positive, negative or neutral;
- Aspect and feature level – sentences/documents are classified as positive, negative or neutral on the basis of aspects of the sentences or documents generally termed as aspect level sentiment classification.

In feature extraction, data are excerpted from available subset of features applied in a learning algorithm. The least volume of dimensions are present in the best subset which contributes to accuracy and the rest of the unimportant dimensions are discarded. In pre-processing, this is a vital stage and can be considered as either of the methods available to avoid curse of dimensionality. This is an almost fundamental but an unavoidable task in sentiment analysis. The processed by converting a section of text into a feature vector. To make a precise learning task, effective feature selection is essential [3].

Feature is termed as an individual metric that is being observed. Classification can be performed with the help of a set of features in a machine learning algorithm. The problem of reduction of irrelevant or unwanted variables is done through several techniques which act as a load on challenging tasks. The advantage of feature selection/variable elimination is that it helps to understand data, reduces computation requirement, reduction of the effect of curse of dimensionality and to improve the performance of the predictor. The feature selection's focus is in selecting a subset of variables from input which describes input data efficiently by mitigating effects from noise or variables which are non-relevant and to provide good prediction results [4].

Feature selection techniques can be classified into three groups – filter, wrapper and embedded. Group of features are selected on the basis of a specific mathematical equation in filter category and can be used with any classifier; whereas, in wrapper and embedded techniques, the features that are selected are bound to a specific classifier. In spite of its rigidity as a classifier, the wrapper and embedded techniques require generally high allocation of resources and longer time of execution [5].

Techniques that are grouped under filter category include Document Frequency (DF), Chi-Squared (CHI) and Information Gain (IG). The words that are repeated frequently in a particular category are grouped under DF. Removal of features which come on the top and bottom of the list are removed. Particular feature that does not pertain to a particular class is calculated by CHI. On the basis of its relevance to a particular class, features are classified. The

Copyright © 2018 Prasanna K, Mathivanan V. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

relevance of feature is calculated by IG on the chance in which a word exists in a particular class.

Through feature selection, original feature subsets are selected and through evaluation criterion, the optimality of feature subsets is measured. When there is an expansion in the domain dimensionality to N features, the number increases. An optimal feature subset is difficult to identify. The problems that are related to feature selection are NP-hard. Machine learning feature selection decreases the number of features, clears data that are irrelevant, noisy or redundant, which leads to accuracy in recognition and is a global optimization problem. This step affects the performance of pattern recognition system. Generally the problems involving single objective optimization techniques are solved through Genetic Algorithm (GA). Only a single quality measure is optimized, for instance, recall precision or F-measure at a time. At certain times, a single measure cannot capture the quality of a good classifier reliably. The qualities of a good classifier are recall, precision and F-measure values that are optimized simultaneously instead of the high value of any one parameter alone [6].

The accuracy of classification is affected through feature selection as a large number of features are present in the dataset, there will be large dimensional space, and precision of classification is degraded. Feature selection affects various aspects including classification pattern and accuracy, the duration in learning the functions of classification, the quantity of sample needed for learning and the cost that is associated with the features. A huge set of original features are reduced during feature selection which is an optimization process to comparatively small subset which can enhance the classification accuracy significantly, in a fast and effective manner [7].

Machine learning algorithms [8] enhance performance with experience automatically and the main result of performance is prediction. When an algorithm can predict a task's key elements if presented with a proper data, then it is said to be learned. Machine algorithms represent knowledge, which are characterized by languages. Studies have showed that almost all algorithms produced same results and so no single approach is superior. Success of the learning algorithm can be judged through the data's nature that characterizes the task to be learned. In machine algorithms, when data does not exhibit statistical regularity, learning fails. A fully automated technique is difficult when a new data is constructed in order to exhibit statistical regularity and facilitate learning.

Sentiment analysis can be classified as supervised and unsupervised learning – the machine learning task which infers function from labeled training data is supervised learning; whereas, in unsupervised learning, inferences are drawn from datasets which are present in the unlabeled data. There are defined rules in supervised learning and the results are known, whereas in unsupervised learning certain rules are followed to learn by itself and comes up with the result [9]. Some popular classification algorithms are present in many sentiment analysis tasks including Naive bayes, Support Vector Machines (SVM), KNN, Maximum Entropy (ME) etc.

This work proposes hybrid optimization algorithm for PSO-CS in opinion mining. The remainder of the work is structured in the following manner: in section 2, related works in literature are discussed. In section 3, the materials and methods used in the presented work is explained. Section 4 discusses the results, concluded in section 5.

2. Related works

A new technique was put forth by Hai et al., [10] located online review features which exploit the variation in the statistics beyond two corpora, i.e. domain-specific and domain independent corpus. Through Domain Relevance (DR) variation is captured by authors, which features the applicability of a term to a collection of text. Intrinsic Domain Relevance (IDR), Extrinsic Domain Relevance (EDR) and domain independent corpora is estimated for every candidate feature that is extracted. Less generic candidate features whose EDR score is less than threshold and domain-specific whose IDR is greater than another threshold is selected as opinion features.

Point-wise Mutual Information (PMI)–TFIDF which is a point similarity measure is introduced by Quan& Ren [11] in evaluating the relevance of candidate features and domain entities. Outcome of the study shows that feature extraction approach performs better than other modern methods and comparative domain corpora remains the lone external resources used. Thus it is generic and supervised. In contrast to conventional PMI, better distinction ability was shown by PMI-TFIDF. Feature oriented opinion determination was proposed by the author which was based on feature-opinion pair extraction and feature-oriented opinion lexicon generation. The effectiveness of the proposed technique was demonstrated through the results and showed that feature-oriented opinion lexicons proved to be better compared to other general opinion lexicons.

A new technique involving Ant Colony Optimization (ACO)-2OPT was proposed by Saraswathi & Tamilarasi. There is characterization of facets of sentiments in the medical sphere and identification of potential use cases. Feature sets are extracted by a classification framework from reviews with the help of TF-IDF and with the aid of the new technique features. Classification of the chosen features was done using Naïve Bayes and SVM. Results of the research demonstrated that is improved efficiency of feature selection of classifiers in the classification of opinions.

A novel GA was proposed by Keshavarz&Abadeh [13] in solving optimization issues and to find lexicon to classify text. Adaptive sentiment lexicons were generated through this algorithm and on its basis, which was used along with Bing Liu's lexicon and n-gram features. The experiments took place on six datasets. With regards to precision, the result of the study performed better compared to the modern techniques that were proposed in literature in two of the datasets. The proposed technique outperformed with regards to F-measure in four of the datasets. When the proposed technique was applied to six datasets, its accuracy was found to be greater than 80% in all the six datasets and also the F-measure was higher than 80% in four of these datasets.

Machine learning techniques have found its place in many researches already so as to analyze the sentiments that are present in a particular document. Because of the increased feature set of data, the time of execution has been increased. The sentiment of a given document is determined through the participation of irrelevant features thus varying the accuracy of the algorithm. A technique was proposed by Shahid et al., [14] which is Biography Based Optimization (BBO) algorithm in which optimal features are selected from a given data. With the help of Naïve Bayes and SVM techniques, sentiment classification was performed of product reviews. For other classification problems also the proposed technique can be applied to other classification issues where there is a huge feature set.

Through statistical feature selection, formation of suboptimal feature subsets is done such as thresholding of the document frequency because of the NP-hard nature of the problem. In optimization problems, particularly swarm intelligence, swarm optimization renders feature selection through improvement of classification accuracy and reducing computational complexity and size of the feature. Firefly Algorithm (FA) was proposed by Kumar & Khorwal [15] for optimization of feature. For the task of classification, four varying datasets are used of which two in Hindi and two in English. Comparison of the proposed technique is done with feature selection using GA. This technique can optimize successfully the feature set and improve the system's performance with regards to accuracy.

Two unsupervised PSO-based opinion mining technique was presented by Souza et al [16]. Evaluation of these techniques was done through 18 experiments having different corpus types including domain, class balancing, language and preprocessing techniques. Better accuracy was reached on 12 experiments. Results were comparable in the case of corpora with less number of dimensions and specific domains. Best accuracy (0.79) was obtained by Discrete Improved Self-Adaptive PSO (IDPSO) on the OBCC

corpus, which outperformed supervised machine learning and lexicon-based approaches for this corpus.

Fitness proportionate selection Binary PSO (F-BPSO) was a feature selection technique proposed by Shang et al., [17]. The technique which can find its application in feature selection domain is the binary version of PSO called binary PSO (BPSO). A modification of BPSO is F-BPSO and the issues regarding conventional BPSO including unreasonable update formula of velocity lack of evaluation in every single feature is overcome by this technique. Then certain extensive variations are made on original F-BPSO, using fitness sum instead of average fitness in the fitness proportionate selection step. The modified technique is called Fitness Sum proportionate selection BPSO (FS-BPSO). The FS-BPSO is further modified making it more suitable for sentiment classification oriented feature selection domain, which is called Sentiment Classification-Oriented (SCO)-FS-BPSO.

3. Methodology

After sorting of features, the users select the specific features on the basis of the sorted value. Selection of features relies on the user and it is personalized. Generally, a novice who is unaware of this process takes more time for processing and also uses more resources. In this section, cell phones and accessories dataset is used. The classifier discussed here includes TF-IDF, PSO, CS, hybrid PSO-CS algorithm, KNN classifier and naïve bayes classifier.

a) Cell Phones and Accessories dataset

A real-world data must have certain features while computing the robustness of approach in seizing the fashion dynamics:

1) The dataset must be quite wide to understand the general preference of the public;
2) As the visual decision reasons are performed at different times [18] there should be a long period span, at least temporarily.

In this study, datasets have been taken from Amazon.com. Blogs, microblogs, and review sites provide the receptivity of products and services.

Blogs: Universal terminology of all blog sites put together is called blogosphere. A blog would contain the opinion of people on topics which they feel strongly about and would want to share. Opinions are collected from many investigations in a blog and sentiment analysis is performed.

Review sites: For an online purchaser, as he does not have the feel of the product he buys personally, he has to depend on the review on that particular product to make a decision. A huge database is provided by Internet for reviews that are generated by users for various services and products.

Micro-blogging: This again is another tool which is widely used for communication among the internet users. Websites such as Tumblr, Facebook, and Twitter host a tons of messages which seek opinion from a number of users and opinions are manifested and utilized as a source of data in case of sentiment classification.

b) Term Frequency - Inverse Document Frequency (TF - IDF)

The frequency of a term in a given document or text is said to be Term Frequency (TF) [20]. For instance, in a given text document consisting of 100 words, the term cat occurs four times, then the TF of cat is 4/100=0.04. Considering a document'd' the number of times a term occurs in it is called TF. In a lengthier document, there is a possibility of occurrence of a given term more number of times. Also, the probability of a word occurring more frequently in a longer document is more than a short document. So, as a method of normalization, TF is generally divided by the length of the document (1).

$$Normalized\ TF = \frac{tf\ (t,d)}{n_d} \tag{1}$$

Where,

$tf\ (t,d)$ = Term frequency

n_d = The total number of terms in document d.

Document Frequency (DF) of a given term t measures of out all the documents available, how many of them contains the given term t. Inverse Document Frequency (IDF) is the exact opposite to DF. IDF measures the importance of a term. All terms are considered to be important while computing ``TF", [21]. However it is known that certain terms, such as ``is", ``of", and ``that", may appear a lot of times but have little importance. So, the frequent terms should be weighed down while the rare ones scaled up. . IDF is calculated as (2):

$$IDF(t) = \log \frac{n_d}{n_d(t)} \tag{2}$$

Where,

n_d = The total number of documents.

$n_d(t)$ = The number of documents that contain the term (t).

Text documents can be characterized by TF-IDF [22]. It can be understood as a weighted TF, which is especially useful if stop words have not been removed from the text corpus. The TF-IDF approach assumes that the importance of a word is inversely proportional to how often it occurs across all documents in (3).

$$TF - IDF = tf_n(t,d) * idf\ (t) \tag{3}$$

Though TF-IDF is used to rank documents through relevance in various text mining tasks such as page ranking by search engines, it can be applied also to text classification through Naïve Bayes. For instance, a document of 100 words is considered and the word cat appears 4 times; then its TF is 4/100=0.04. Now it is assumed there are 10 million documents and the same word appears in 1000s of these. Then TDF is log (10,000,000/1000) =4. The product of these quantities is the TF-IDF weight that is 0.04*4=0.16.

c) Particle swarm Optimization (PSO) Algorithm

An intelligent evolutionary computation learning algorithm is PSO. Social behavior is simulated through PSO on the basis of information exchange and holds good for practical applications. When considered in a problem space, every potential solution is taken into consideration as a particle in a swarm. The direction of path can be adjusted according to the velocity of each particle based on its own flight experience and the experiences of its companions. The optimal regions of complex search spaces are mined based on this superior strategy through the interaction of individuals in a set of particles [23].

There are four steps within a process period in PSO. First step is the initialization step where the particles are initialized in a population of random solutions. In the second step, the $pbest_j$, of each particle is obtained, where its current fitness is compared with the fitness of its previous position; $pbest_j$ is the position of the j-th particle with the highest fitness value at a given iteration. This can be considered as the best solution with regards to the j-th particle. In the third step, gbest for all particles in a given population is determined. gbest is the best position or global best for all pbest particles and is considered as the best. By updating generations, a search for optimal solutions is executed in every algorithm. In each generation, the position and velocity of the j-th particle are updated with the $pbest_j$ and gbest of the swarm population. The update equations can be formulated as (4 & 5):

$$v_j(t+1) = \omega v_j(t) + d_1(pbest_j - x_j(t)) + d_2(gbest - x_j(t)) \tag{4}$$

$$x_j(t+1) = x_j(t) + v_j(t+1) \tag{5}$$

Where ω is the inertia weight, v_j is the velocity, x_j is the particle position, and d1 and d2 are learning factors.

In a feature space, every feature subset is considered as a point. The subset which has the least length and highest classification precision is the optimal point. There is random distribution of initial swarm over search space and the particles take a position. The particles' goal is to fly to the best position. As time goes by, the position is changed through communication with each other and the search around local and global best position. Ultimately convergence should take place in good and possibly optimal positions as there is exploration ability which equip them in order to perform feature selection and to discover optimal subsets. In order to deal with feature selection, PSO should be extended. The position of the particle is considered to be binary it strings. A feature is represented by every bit; the selected feature is represented through bit value '; while the bit value 0 represents a non-selected feature. Every position is considered as a feature subset.

d) Cuckoo Search (CS) Algorithm

The latest addition to the group of meta-heuristic algorithms is Cuckoo Search (CS). It takes its inspiration from the brood parasitism of a specific species of cuckoo, in combination with Lévy flight behavior of certain birds and fruit flies. This is again, a population-based optimization technique and it starts with random initial population as with many other metaheuristic algorithms. CS can be clearly described through the breeding behavior of certain cuckoo species [24].

A unique and aggressive reproduction strategy is followed by the cuckoo species. Cuckoos belonging to the species of Ani and Guira show an aggressive attitude towards reproduction, that is, they lay their eggs in communal nests which is known as brood parasitism and nest take over. A particular species 'Tapera' can even imitate the color and pattern of eggs of a few host species. Generally, the eggs of cuckoo hatch slightly earlier compared to that of the host eggs thus increasing their productivity. Similar to their parents, the first instinct of the hatched cuckoo chick is to push out the host bird's egg by propelling it out of the nest.

The natural phenomenon o foraging of animals take place in a random or quasi random basis and its direction depends on mathematically designed model called Lévy flights. Lévy flight is generally described by the mathematicians as random walk trajectories which are composed of self-similar jumps which are distributed according to power law that is (6):

$$y = x^{-\alpha} \tag{6}$$

Where $1< \alpha < 3$ and therefore has an infinite variance.

To simplify CS algorithm certain rules are used [25]:

- A cuckoo lays one egg at a time and drops it in the host nest;
- The highest quality egg is selected to carry on the progeny can be termed as a kind of eclecticism;
- There is a fixed number of host nests. The probability of discovery by the host bird on the egg laid by a cuckoo $P_a \in [0, 1]$.

To simplify things, the fraction of n nests is replaced by new random solutions or new nests. It can be assumed that the egg in a nest represents new solution and the objective is to locate the worst solutions and replace them with potentially better solutions (cuckoos). Mathematically speaking, the first process is the randomization process so that a new solution can be generated randomly by random walk or levy flights [26]. When generating new solutions say for a cuckoo i, a Levy flight random walk is performed as mentioned in (7):

$$x_i^{t+1} = x_i^t + \alpha \oplus Levy(\lambda) \tag{7}$$

Where $\alpha > 0$ is known as the step size and it is related to the scales of the problem of interests. The sign $\oplus$ stands for entry wise multiplication. In this algorithm the significant advantage is the use of Levy flight, which is more efficient in exploring the search space as its step length is much longer in long run for both local and global searching. The random step length is drawn from a Levy distribution in (8):

$$Levy \sim u = t^{-\lambda},\ (1<\lambda \leq 3) \tag{8}$$

This equation has an infinite variance and an infinite mean. Here the steps form a random walk process which follows the power law step length distribution and it is heavy tailed.

The solutions are updated in search space in standard CS towards continuous valued positions. Binary string is the way that is used in encoding the feature selection. Here the feature position is generated by random values. Additionally, as the problem is to select or not a particular feature, there is employment of a solution binary vector, where 1 represents the selection of a feature to compose a new dataset and 0 otherwise.

The advantages of PSO are easy understanding, quick searching and operational simplicity. However, when a large problem is solved, PSO gets trapped in the local minima. The practicability of PSO should be overcome through this weakness. The advantages of CS are certain control parameters such as high efficiency but there are some defects also which includes slow convergence speed and low accuracy. The search process is made to jump quickly from one area to another through high randomness of Levy flight. Thus, the algorithm has a strong global search ability. However, the high randomness of Levy flight initiates a blind search process, slow convergence speed and there is reduction in the searching efficiency close to optimal solution [27].

The performance of CS is enhanced through the hybridization of PSO with an update process of CS and so PSOCS hybrid technique is developed. Levy flight is used to search the search space in PSO-CS and then the particles are accelerated through the PSO update mode to optimal solution convergence. Simultaneously, local optima can be escaped through the random elimination mechanism of CS thus improving the searching performance for optimal solution. Each particle is located as a binary string. Selection of feature is represented through a value of 1 and value of 0 implies that the feature is not selected.

The terms of algorithm are defined as follows [28]:

1) Population and Population Size (sizepop). Population is made up of particular number of individuals; the total number of individuals is the population size or sizepop.
2) Fitness. Fitness implies the individual quality. Generally, a large fitness value corresponds to good result and vice versa.
3) Search Space Upper Bound (Ub) and Search Space Lower Bound (Lb). The search space for the optimization problem is depicted through Ub and Lb are the upper bound and lower bound, respectively.
4) Maximum Search Velocity (Vmax) and Minimum Search Velocity (Vmin). Speed is limited as the algorithm performs a search. Consider Vmax = a∗Ub, where a is the adjustment coefficient in the range of (0, 1).Consider Vmin = b∗Lb, where b is also the adjustment coefficient in the range of (0, 1).
5) PSO search Mode. Through the process of PSO, an individual updates its position and velocity in this mode.
6) CS Mode. The process of CS is used in updating the position. There is no speed and velocity in updating the formula and an individual in PSO search mode has both position and velocity. There is no updating of the individual velocity in CS mode and the current velocity of the individual and the velocity updated by PSO search mode are same.
7) Discovery Probability. Through the random elimination mechanism in CS mode, the host has probability P_a of finding foreign eggs.

A flowchart of hybrid PSO-CS is shown in figure 1

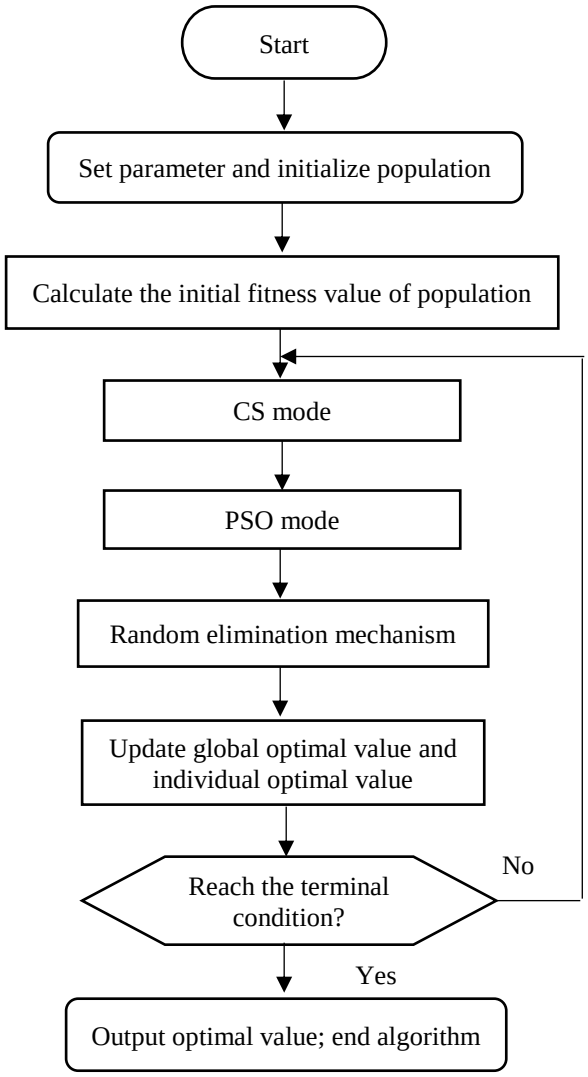


Fig. 2: Flowchart for Hybrid PSO-CS Algorithm.

e) K-Nearest Neighbours (KNN) Classifier

A specific type of instance-based or lazy learning is KNN, where there is approximation of function only locally and computation takes place only after classification. This can be deemed as a non-parametric technique for classification or regression. Output is the class membership with regards to classification, there is classification of objects through majority of neighbors, with assigning of object to the class which is most common among KNN. The entire training set is retained during learning and a query is assigned to each class which is represented by majority label of KNN in the training set [29].

The simplest form of KNN is Nearest Neighbor rule, where K=1. Given an unknown sample and a training set, all the distances between the unknown sample and all the samples in the training set can be computed. The training set closest to the unknown sample corresponds to the distance with smallest value. Thus, classification of unknown sample can take place on the basis of classification of its nearest neighbor. It is quite easy to understand and implement KNN n=and is a powerful tool to dispose sentiment analysis. KNN can be powerful, as there is no assumption about the data other than consistent calculation of distance measure between two instances. As such, it is called non-parametric or non-linear as it does not assume a functional form.

The weighted sum in KNN classification [30] can be represented as (9):

$$score(d,t_i) = \sum_{d_j=KNN(d)} sim(d,d_j)\delta(d_j,c_i) \quad (9)$$

Where, KNN (d) denotes the set of KNN for the review (d). If belongs to ci, then could either be equal to 1 or 0. In the case of the test review, d, it should fit into the class with the highest resulting weighted sum.

4. Results and discussion

In this section, the CS-KNN, CS-naïve bayes, PSO-CS-KNN and PSO-CS-naïve bayes methods are used. The summary of results is shown in table 1. The true positive rate for positive, neutral and negative and precision for positive, neutral and negative as shown in figures 2 & 3.

Table 1: Summary of Results

	CS - KNN	CS - Naïve Bayes	PSO-CS-KNN	PSO-CS-Naïve Bayes
True Positive Rate - Positive	0.906	0.921	0.9615	0.9635
True Positive Rate-Neutral	0.9091	0.9265	0.9696	0.9743
True Positive Rate - Negative	0.9341	0.9489	0.9711	0.9711
Precision - Positive	0.8913	0.9151	0.9605	0.9645
Precision- Neutral	0.9277	0.9413	0.9725	0.9727
Precision - Negative	0.9296	0.9409	0.9693	0.9718

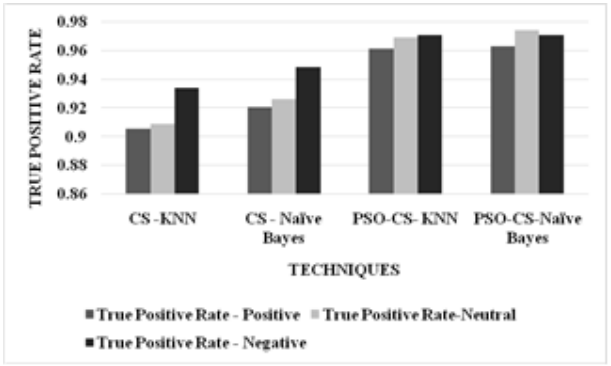


Fig. 2: True Positive Rate.

From the figure 2, it can be observed that the PSO-CS-naïve bayes has higher true positive rate by 6.15%, 6.92% & 3.88% for CS -KNN, by 4.51%, 5.02% & 2.31% for CS-naive bayes and by 0.2%, 0.48% & same value for PSO-CS-KNN when compared with positive, neutral and negative.

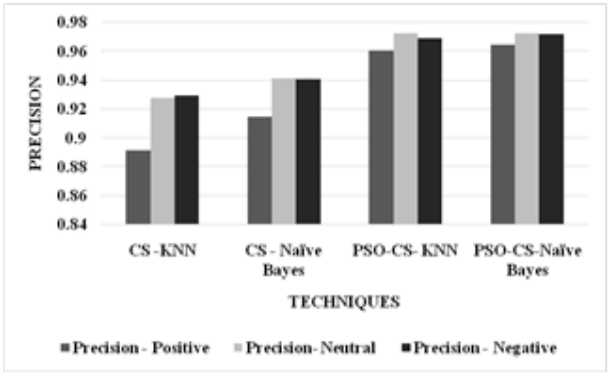


Fig. 3: Precision.

From the figure 3, it can be observed that the PSO-CS-naïve bayes has higher precision by 7.88%, 4.73% & 4.43% for CS -KNN, by 5.25%, 3.28% & 3.23% for CS-naive bayes and by 0.41%, 0.02% & 0.25% for PSO-CS-KNN when compared with positive, neutral and negative.

5. Conclusion

Opinion mining is the fastest developing area of research. Relevant features are selected through feature selection on the basis of specified measurement, the purpose of which is to simplify training and reduce the time taken to train. As PSO and CS are powerful optimization techniques, they are being used in this study in solving combinatorial optimization problems. This technique is incorporated in optimizing feature subset selection. The advantage of PSO is increased convergence speed; while its disadvantage is that of premature convergence, and falling into local optimal. Again the advantage of CS is the use of few control parameters and increased efficiency and its defects is slow convergence speed and low accuracy. When both are used as hybridized algorithm, the performance can be improved, both complementing each other. Results show that the PSO-CS-naïve bayes has higher precision by 7.88%, 4.73% & 4.43% for CS -KNN, by 5.25%, 3.28% & 3.23% for CS-naive bayes and by 0.41%, 0.02% & 0.25% for PSO-CS-KNN when compared with positive, neutral and negative

References

[1] Jandail, R. R. S. (2014). A proposed Novel Approach for Sentiment Analysis and Opinion Mining. International Journal of UbiComp, 5(1/2), 1. https://doi.org/10.5121/iju.2014.5201.

[2] Sharma, R., Nigam, S., & Jain, R. (2014). Opinion mining of movie reviews at document level. arXiv preprint arXiv:1408.3829.

[3] Jeong, H., Shin, D., & Choi, J. (2011). Ferom: Feature extraction and refinement for opinion mining. Etri Journal, 33(5), 720-730. https://doi.org/10.4218/etrij.11.0110.0627.

[4] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024.

[5] Samsudin, N., Puteh, M., Hamdan, A. R., &Nazri, M. Z. A. (2013). Immune based feature selection for opinion mining. In Proceedings of the World Congress on Engineering (Vol. 3, pp. 3-5).

[6] Sumathi, T., Karthik, S., &Marikkannan, M. (2014). Artificial Bee Colony Optimization for Feature Selection in Opinion Mining. Journal of Theoretical & Applied Information Technology, 66(1).

[7] Wahyudi, M., & Kristiyanti, D. A. (2016). Sentiment Analysis of Smartphone Product Review Using Support Vector Machine Algorithm-Based Particle Swarm Optimization. Journal of Theoretical & Applied Information Technology, 91(1).

[8] Isabella, J., & Suresh, R. (2012). Analysis and evaluation of Feature selectors in opinion mining. Indian Journal of Computer Science and Engineering (IJCSE), 3(6), pp. 757-762.

[9] Behera, R. N., Manan, R., & Dash, S. (2016). Ensemble based Hybrid Machine Learning Approach for Sentiment Classification-A Review. International Journal of Computer Applications, 146(6).

[10] Hai, Z., Chang, K., Kim, J. J., & Yang, C. C. (2014). Identifying features in opinion mining via intrinsic and extrinsic domain relevance. IEEE Transactions on Knowledge and Data Engineering, 26(3), 623-634. https://doi.org/10.1109/TKDE.2013.26.

[11] Quan, C., & Ren, F. (2014). Unsupervised product feature extraction for feature-oriented opinion determination. Information Sciences, 272, 16-28. https://doi.org/10.1016/j.ins.2014.02.063.

[12] Saraswathi, K., & Tamilarasi, A. (2016). Ant Colony Optimization Based Feature Selection for Opinion Mining Classification. Journal of Medical Imaging and Health Informatics, 6(7), 1594-1599. https://doi.org/10.1166/jmihi.2016.1856.

[13] Keshavarz, H., & Abadeh, M. S. (2017). ALGA: Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs. Knowledge-Based Systems, 122, 1-16. https://doi.org/10.1016/j.knosys.2017.01.028.

[14] Shahid, R., Javed, S. T., & Zafar, K. (2017, April). Feature selection based classification of sentiment analysis using Biogeography optimization algorithm. In Innovations in Electrical Engineering and Computational Technologies (ICIEECT), 2017 International Conference on (pp. 1-5). IEEE. https://doi.org/10.1109/ICIEECT.2017.7916549.

[15] Kumar, A., & Khorwal, R. (2017). Firefly Algorithm for Feature Selection in Sentiment Analysis. In Computational Intelligence in Data Mining (pp. 693-703). Springer, Singapore. https://doi.org/10.1007/978-981-10-3874-7_66.

[16] Souza, E., Oliveira, A. L., Oliveira, G., Silva, A., & Santos, D. (2016, October). An Unsupervised Particle Swarm Optimization Approach for Opinion Clustering. In Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on (pp. 307-312). IEEE. https://doi.org/10.1109/BRACIS.2016.063.

[17] Shang, L., Zhou, Z., & Liu, X. (2016). Particle swarm optimization-based feature selection in sentiment classification. Soft Computing, 20(10), 3821-3834. https://doi.org/10.1007/s00500-016-2093-2.

[18] He, R., &McAuley, J. (2016, April). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In Proceedings of the 25th International Conference on World Wide Web (pp. 507-517). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872427.2883037.

[19] Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. International Journal, 2(6), 282-292.

[20] Kumar, J. A., & Abirami, S. (2015). An Experimental Study Of Feature Extraction Techniques In Opinion Mining. International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), 4(1).

[21] Basari, A. S. H., Hussin, B., Ananta, I. G. P., &Zeniarja, J. (2013). Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. Procedia Engineering, 53, 453-462. https://doi.org/10.1016/j.proeng.2013.02.059.

[22] Patil, P. K., &Adhiya, K. P. (2015). Automatic Sentiment Analysis of Twitter Messages Using Lexicon Based Approach and Naive Bayes Classifier with Interpretation of Sentiment Variation. International Journal of Innovative Research in Science Engineering and Technology, 4(9).

[23] Li, X., Li, J., & Wu, Y. (2015). A global optimization approach to multi-polarity sentiment analysis. PloS one, 10(4), e0124672. https://doi.org/10.1371/journal.pone.0124672.

[24] Adnan, M. A., &Razzaque, M. A. (2013, March). A comparative study of particle swarm optimization and Cuckoo search techniques through problem-specific distance function. In Information and Communication Technology (ICoICT), 2013 International Conference of (pp. 88-92). IEEE. https://doi.org/10.1109/ICoICT.2013.6574619.

[25] Gandomi, A. H., Yang, X. S., &Alavi, A. H. (2013). Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. Engineering with computers, 29(1), 17-35. https://doi.org/10.1007/s00366-011-0241-y.

[26] Pandey, A. C., Rajpoot, D. S., &Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. Information Processing & Management, 53(4), 764-779. https://doi.org/10.1016/j.ipm.2017.02.004.

[27] Ghodrati, A., & Lotfi, S. (2012). A hybrid CS/PSO algorithm for global optimization. Intelligent Information and Database Sstems, 89-98. https://doi.org/10.1007/978-3-642-28493-9_11.

[28] Guo, J., Sun, Z., Tang, H., Jia, X., Wang, S., Yan, X., & Wu, G. (2016). Hybrid Optimization Algorithm of Particle Swarm Optimization and Cuckoo Search for Preventive Maintenance Period Optimization. Discrete Dynamics in Nature and Society, 2016. https://doi.org/10.1155/2016/1516271.

[29] Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment Analysis of Review Datasets Using Naive Bayes and K-NN Classifier. arXiv preprint arXiv:1610.09982.

[30] Alsaffar, A., & Omar, N. (2015). Integrating a Lexicon based approach and K nearest neighbour for Malay sentiment analysis. Journal of Computer Science, 11(4), 639. https://doi.org/10.3844/jcssp.2015.639.644.