ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BIOINSPIRED RELIABLE SUPPORT VECTOR MACHINE FOR ANALYZING SENTIMENTS IN BIG DATA

J. RATHIKA[1] , M. SORANAMAGESWARI[2]

[1]Assistant Professor, Department of Computing, Coimbatore Institute of Technology, Coimbatore, India

[2]Assistant Professor, Department of Information Technology, Government Arts College, Coimbatore, India

E-mail: [1] jradhika@cit.edu.in , [2] swarnakannappan@rediff.com

ABSTRACT

Real-time shopping and E-commerce benefit greatly from big data analytics and it is employed to increase sales of items and improve customer-retailer interaction. Shops increasingly employ internet marketing to identify top places to acquire quality items. The shopper's buying experience and thoughts about the retailer might be detected by observing social media activity on both sides. Sentimental Analysis is a great tool for identifying the emotional or emotional impact of the contents. SA investigates people's feelings and thoughts on various things. Weak sentimental analysis worsens the significance of gaining insight into customers' feelings when purchasing products and also it diminishes the impact of knowing the customers' perception about a product. This paper proposes a Bioinspired Reliable Support Vector Machine (BRSVM) for performing the sentiment analysis in big data. BRSVM is inspired from foraging behavior of ants and it is used to identify the sentiments in big review dataset. Lagrangian strategy utilized in BRSVM assist in achieving better optimization. Proposed classifier has been evaluated with fourproduct review big dataset using accuracy and f-measure performance metrics. Results make an indication that the proposed classifier attains better classification accuracy than existing classifiers.

Keywords: *Ant Colony, Classification, Sentiment Analysis, SVM, Amazon, Optimization*

1. INTRODUCTION

Popular services like podcasting, blogging, bookmarking and social networking are found all over the Internet because of the fast expansion of the Internet. This suite of services allows people to make and share information across open and private groups, as well as to boost the amount of data generation [1]–[3]. In an analysis of a study by IBM, "2.5 quintillion bytes of data" are being generated every day, and new data are rising annually. On Twitter, users make 400 million visits each month and generate 500 million tweets per day. Facebook now has over 1.845 billion daily active users and 1.797 billion monthly active users. The elements that are contributed to the growth of Big Data are (i) Data volume (ii) velocity (iii) authenticity (iv) diversity (v) value and (vii) volatility [4]–[7]. These things represent the common characteristics of big data. Big Data has emerged during the age of the smartphone and the internet browser, with the data generated from both sources, are recorded and stored in multiple ways. While the platform's storage and analytics functionality has been shown to be unable to deal with several distinct data sources and formats, there are no other alternatives. Since big data analytics has become a more and more common practice for processing and handling massive amounts of unstructured and structured data, it is now being referred to as big data analytics [8].

Sentiment Analysis (SA) is a big data strategy that seeks to analyze massive data sets to (i) discover patterns and correlations (ii) offer educated forecasts (iii) give actionable knowledge. The technique of employing text analytics to mine multiple sources of data in order to find out how people feel about particular issues is known as SA. The machine learning approach and lexical approach to the issue of SA fall into two broad groups [9].

When machine learning is used to analyze sentiments, the task of text classification is seen as a common issue, and solutions to it may be found by training a classifier on a large collection of labeled text. Supervised classification is the machine learning methodology relevant to SA [10]. Applying machine learning techniques to categorize the evaluations has shown to be beneficial. With the lexical technique, a list of emotional words or phrases is first compiled. Following this, a list of those words or phrases that include certain emotional terminology is compiled, and then it is used to determine the sentiment associated with the terms. The lexicon's dictionary was regulated by adding previously unassigned tags to the lexicon entries. The overall score determines the text's categorization [11].

Reviews are highly trusted. As Amazon customers do, they make carefully examine reviews of an online product before clicking the "purchase" button. It is clear that the product bought by a lot of customers is due to a big number of reviews that recommend it, while the abundance of good reviews points to the fact that it is of good quality [12]. More than 80% of Amazon customers base their purchasing decisions on product reviews, the same as they get personal suggestions. Taking a look at previous product reviews is a powerful tool for companies, and it may also be applied to make better use of the present product. Businesses should never disregard the value of studying product reviews. In amazon like an online shopping site, a popular product in each category could be found out. It involves personal reviewing/scanning of all reviews in a category, which is both time-consuming and expensive. Alternatively, users may group the reviews into one category to learn more about the product's global popularity. The analysis is based on sentiment, making this possible [13].

1.1 Problem Statement

The fundamental function of sentiment analysis is the investigation of customer evaluations and feedback. Among different types of reviews text-based reviews occupies large space. Each assessment is viewed as a notable characteristic for exploration. It is noted that all collections of words often turn into huge warehouses of words, however these words may have little impact on evaluations and also it might be difficult to comprehend and analyze. To develop a better data mining algorithm capable of extracting meaningful comments and views from massive amounts of unstructured material. Rigorous optimization strategy is needed to overcome issues discussed above.

1.2 Research Objective

The main objective of this research work is to analyze, design, and develop a bio-inspired optimization-based classifier for big product review dataset that predicts the opinions with increased accuracy.

2. LITERATURE REVIEW

Influence Modelling [14] is proposed for aggregating the aspects to handle the calculation of global sentiments. Elements based on diffusion and prevalence like relevance, ambiguity, reach and influence was considered for evaluation. The weighted assessment of these aspects results in the prediction of future needs along with the estimation of current sentiments which were portrayed in the experimental results. Hybrid Framework [15] is proposed to analyze the sentiments in the Persian Language. It aims to integrate the rules of linguistics and deep learning for optimizing the detection of polarity. Sentiments are triggered to define the dependency relations and a neural network is used to perform classification. The performance outcome is evaluated by making use of the average margin based on the benchmark hotel reviews. Hybrid Topic-Based Sentiment Analysis (HTBSA) [16] is developed to find the relation election between words and their co-occurrences. The latent topics were learned through the lexical resources and the score for sentiments was calculated with sentiment orientation along with weighing for each topic. Results extracted from the experiment were demonstrated to illustrate the improved performance of the predicted technique through which the method can efficiently extend for monitoring the election in real-time and also for predicting the future polls. Semi-Supervised Classification [17] is developed for determining the feature weight in opinion mining. A selection approach was employed for increasing the classification performance in wherein the features from the speech are studied extensively. Benchmarking datasets were used for comparing the performance of the proposed work and it proves to work better than other techniques of analyzing the sentiments. Sentiment Classification using Lexicons [18] is proposed for detecting the sentiment in a specific type of sentence and it makes use of emoticon score learning. The algorithm proposed was validated with 1000 tweets wherein the experimental results prove that the defined algorithm was very efficient in classifying the sentiments and also able to detect the positive and negative sentiments in the sentences.

Fuzzy theory-based sentiment classification [19] is proposed for deriving the twitter sentiments in opinion mining. A valuable growth rate was obtained with energy-based keywords from the results and ranking was given based on its highest priority. Also, the different types of energy were used for the evaluation process. Unified Indexing [20] is proposed to analyze the opinions from different reviews and inconsistencies to detect to extract polarities. Geometric mean and the aspects were extracted for review of the aggregated polarities. The sentiment of the context was matched with extracted sentiment and the score obtained was able to solve the inconsistency with which it can apply for evolving aspects from both specific and general contexts. Fuzzy formal concept [21] is proposed as a descriptive analytics technique to define grievance analysis and to review the tedious and lengthy complaints in form of rules. To compare the sentimental analysis at the concept level hybrid formal fuzzy-based concept analysis and concept-based sentiment analysis are defined. A brief association rule list is generated along with related documents and interactive visualization was provided. The evaluation was performed by the annotator which yields a maximum score of matching the opinions for the aspects. Multi-class Sentiment Classification [22] is designed to fetch the essential features and to train the classifier with the algorithm of machine learning. An extensive study was performed for achieving the performance of algorithms using sentiments for classification. 10-fold cross-validation was performed for obtaining the accuracy for each algorithm to validate the sentiment classification and the results were portrayed for deriving the execution time wherein the existing technique proves to work better for developing classifiers. Experiments were carried out for fetching the service and product aspect words with the sentiment from the reviews. Appraisal Expression Pattern [23] is proposed to extract and express the opinions by using the Latent Dirichlet Allocation strategy. Each review is extracted with mutual sentiment topics and the information is incorporated with sentiment words and mining aspects respectively. The experimental result indicates better performance in defining the approaches to detect sentiment and aspect words.

Consensus Vote Model [24] is proposed to fetch neutral features by fixing the boundary limit for positive and negative reviews. Different semantic analyses were defined for various corpora to fetch the sentiments. Neutral reviews were identified to filter the different models on the aggregation method. Finally, the classification performance was compared with aggregate and single models. The method performs better than the individual models which leads to positive and negative reviews for increasing the classification performance. Burst Detection [25] is proposed as joint sentiment analysis strategy by defining the Time-User Sentiment with Topic Latent Dirichlet Allocation (TUS-LDA) to determine the posts. The time slices of the aggregated post were defined to alleviate the issue which leads to designing sentiment-aware topics. The performance was measured to portray the improvement of the existing model and the sentiment-aware topics are visualized clearly with the proposed work. One-vs-One Multi-class Sentiment Classification [26] is proposed to classify sentiments with different classes. Binary class weight values are analyzed using k-nearest neighbor in which the text is trained from feature vectors and selected using an information-gain algorithm. The binary SVM classifier is used for training the feature vectors and to detect sentiment classes. The testing of sentiments was performed and the results of the proposed technique prove its increased efficiency for handling multi-class sentiments. Hybrid Classification Scheme [27] is proposed to predict the polarity of words in sentiments and perform classification by segregating the positive and negative feelings present in Twitter feeds. It shows the better performance of feeding the text classifier and it was experimented by handling the previous constraints for achieving the highest accuracy than the baseline techniques. Memory Networks [28] is proposed to deeply learn the words and opinions in interaction. The identification of aspects was performed by the process of categorization and this led to the development of end-to-end multi-task memory networks to extract the tasks present in particular categories. The entire defined task was learned by exploring the relationships between the commonalities. This was demonstrated with the state-of-art technique comparison on different benchmarked datasets. Optimization [29]–[35],[36] can be applied in different domains to increase the expected results.

Hybrid Ensemble Scheme (HES) [37] is proposed to prune the clusters the text sentiment classification. The ensemble classifiers are applied in different clusters for the prediction of features. This scheme is tested using balanced datasets and compared with Bagging, Random Subspace and AdaBoost algorithms. The results proved to

Journal of Theoretical and Applied Information Technology
15th September 2021. Vol.99. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

increase the efficiency and validity of the proposed scheme. Entropy based classifier (EBC) [38] is proposed to perform the feature and opinion classification. It was performed with source domain for the prediction. Comparison with different product reviews present in different domains is presented where it makes use of modified maximum entropy and biprate graph clustering. To ensure the effectiveness, the EBC is evaluated with domain-specific and independent words with the help of the SentiWordNet dataset.

3. BIOINSPIRED RELIABLE SUPPORT-VECTOR-MACHINE

3.1 Reliable Support Vector Machine

Support Vector Machine (SVM) is a special type of methodology for performing classification and it falls under supervised machine learning (SML). It expects the essential input and preferred output from the user. The user-provided data are labeled to perform classification to offer the most possible processing of data. Generally, SML provides better algorithms to identify and measure the support of taking future decisions. Shortly, all SMLs have X number of input variables, Y number of output variables, and the user utilizes an algorithm to learn the functions used for performing classification from the received input and expected output.

$$Y = f(X) \qquad (1)$$

SVM is used for classification and regression. Multiple fields have started utilizing SVM for classifying and predicting the future. However, in a real-time scenario, SVM is applied to classify disease, text, decision, spam, image, and sentiment. It works with the procedure of searching the hyperplane that partitions the input into two different classes as shown in Figure 1. Data pints that fall close to the hyperplane are support vectors. Suppose, data points near to hyperplane are removed, then there might have a difference in the hyperplane's position. Shortly, a hyperplane is a line that linearly classifies the available inputs.

Merits of SVM are efficiency and accuracy. Demerits include non-support towards large datasets (i.e., Big Data).

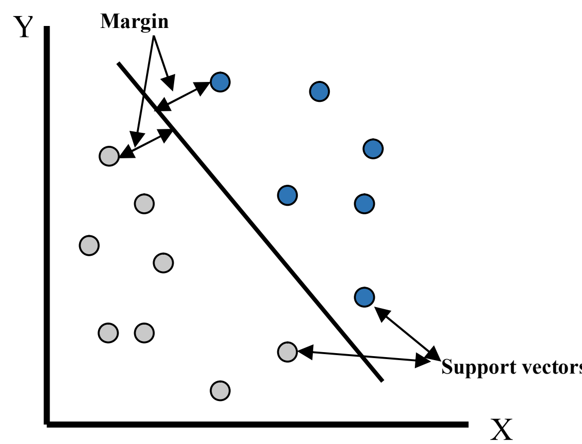

Figure 1. Support Vector Machine

Reliable SVM is an enhanced version of SVM which includes utilization equality constraints and functions of the least square loss method. It attempts to find a solution linearly rather than utilizing quadratic programming. LS-SVM is used to map the data into a High-Dimensional Feature Space (HDFS) where a linear separation hyperplane is then constructed.

Consider B data points as the training set $\{p_j, q_j\}_{j=1}^{B}$ where $q_j \in E^s$ and it represents the input pattern of j and p_j represents a label holding a value that lies in $\{-1,1\}$. The mathematical expression of RSVM is expressed in Eq. (2).

$$\min_{v,f,c}((0.5 \times v^U \times v) + (0.5 \times \delta \times f^U \times f)) \qquad (2)$$
$$\text{s.t } A^U v + pc = 1_B - f$$

where $f \in R^B$ indicates erroneous variables where misclassification is adjusted using overlapping distributions $p = [p_1, p_2, p_3, \ldots, p_{B-1}, p_B]$ which represents the objective vector, where c and δ represent the bias term and a constant value greater than 0. $A^U \in E^{B \times s_g}$ is described as $A^U = [g(q_1)^U q_1, \ldots, g(q_B)^U q_B]$ where $g: E^s \to E^{s_g}$ indicates a feature map that illustrates the high-dimensional feature space which is used for mapping the input data. Function g often do not have an explicit definition, but are inferred by a kernel function of the form $KF: E^s \times E^s \to E$. The kernel function $KF(q_a, q_b) = g(q_a)^U g(q_b)$ is based on Mercer's condition, thus this research

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

work uses it to operate the HDFS without explicitly defining it.

It is unrealistic to work with this formulation when the HDFS is implicitly defined. To obtain the dual problem, we need to take the primal problem's Lagrangian strategy to derive the optimality conditions, and primal variables v and f elimination are expressed in Eq. (3)

$$\begin{pmatrix} 0 & z^U \\ z & \lambda + R_{B/\delta} \end{pmatrix} \begin{bmatrix} c \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 1_B \end{bmatrix} \quad (3)$$

where 1_B represents a single column vector having dimension B, R_B represents an identity matrix having dimension $B \times B$, λ represents a kernel matrix $(A^U A)$ which is labeled, kernel strategy is utilized inside the kernel matrix using Eq. (4).

$$\lambda = (z_a z_b) \times KF(z_a, z_b) a, b = 1,2,3, \dots (N-1), N \quad (4)$$

A classifier is constructed in dual space by using Eq. (5)

$$p(q) = sign\left(\sum_{j=1}^{B} \mu_j p_j KF(q, q_j) + j\right) \quad (5)$$

The error value for a data point is inversely proportional to the support vectors q_j for the associated support vector classifier in RSVM.

3.2 Ant Colony Optimization (ACO)

ACO is inspired by the natural characteristics of ants towards searching for food. When an ant looks for food, it first makes a random analysis regarding the space from its home. If an ant finds the food source, its checks the nature and capacity of food, then it carries a small quantity of food to its home. During the travel to the food source and home, ant spreads the chemical substance (termed as pheromone) in the path. The pheromone level that got spread indicates the measuring metric for other ants to check the quality of the path, where it is used to make a control on other ants to the food source. Hence, indirect communication gets starts between ants by using the trails of pheromone and it continues till it reaches food source and home. This unique character of ants utilized in classifying the sentiments in big product review dataset. In this paper, pheromone trails are made to incorporate in a probabilistic manner.

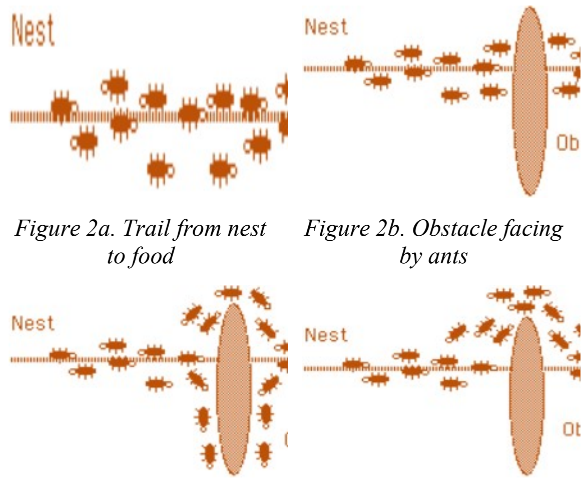

Figure 2a. Trail from nest to food

Figure 2b. Obstacle facing by ants

Figure 2c. Finding an alternate solution

Figure 2d. Alternate solution

Figure 2. Natural Behavior of Ant

Figure 2a to 2d represents the path from the source of food to home. All ants follow the trails of pheromone. If an obstacle is faced in the path of travel, then the first and immediately randomly seeks the next alternate path. In Figure 2, a better alternate path is found in the upper portion than in the lower portion to reach the destination in a shorter duration and distance. The shortest path tends to have a high level of pheromone than the longer path. It is considered that the path having stronger pheromone is the best path and can be used in finding better solutions neatly.

3.3 Bioinspired Reliable Support Vector Machine

BRSVM is an ensemble of 2 algorithms namely ant colony optimization and reliable support vector machine. BRSVM is discussed in this section for classifying the sentiments in big product review dataset. SVMs parameter estimation is optimized in a progressive manner by analyzing the procedure of meta-heuristic. SVM algorithm performs the classification task in an ACO manner for possible classifications. BRSVM intakes the input as $S = \{(xy_i, yz_i) | x_i \epsilon H, yz_i \epsilon \{\pm 1\}\}$ where i holds the value from $(1,2,\dots,l)$, then finally produces the output. It attempts to develop a function for making decision function f by mapping input vectors xy and $yz \epsilon \chi$ onto labels $yz \epsilon \{-1, 1\}$ for effectively classifying the positive, negative and neutral comments in product review big dataset. BRSVM is expected to run only after the inputs are received. To set the optimization for BRSVM, below fundamental steps are to be followed necessarily.

Fundamentally, it's essential to check the features of input and factors of algorithm. The framework intervals of every parameter are to be necessary to be calculated using Eq. (6)

$$h_j = \left(Obs_{xy}^{upper} - Obs_{xy}^{lower}\right) / N, (xy = 1, \dots, m) \quad (6)$$

where $Obst_{xy}^{up}$ and $Obst_{xy}^{low}$ represent the breaking points of the upper and lower value of the SVM parameter. Further calculation of pheromone trails provides $Obst_{xy}^{up} = 2^{x>0}, Obst_{xy}^{low} = 2^{x} < 0$, where xy falls in $1, 2, \dots, m$.

Firstly, pheromone levels of ant blend the data that are equivalent in range. It is maintained by a distribution function consistently. Hence, most of the ants possibly select the first record in the available dataset. Then, ants tend to select the parameter, which got blend with data. BRSVM checks individual parameters by analyzing the objective function. This step is repeated till it attains the maximum number of iterations. Finding the subscript of the data with higher pheromone quality will lower the scope of parameter:

$$\boldsymbol{Obs_{xy}^{low}} \leftarrow \boldsymbol{Obs_{xy}^{low}} + (\boldsymbol{m_j} - \Delta) \times \boldsymbol{h_j}, \quad (7)$$

$$\boldsymbol{Obs_{xy}^{upr}} \leftarrow \boldsymbol{Obs_{xy}^{up}} + (\boldsymbol{m_j} + \Delta) \times \boldsymbol{h_j}, \quad (8)$$

where Δ represents the class coefficient and ants tend to be available in the space with higher pheromone till from the first step of the algorithm.

4. DATASET AND PERFORMANCE METRICS

4.1 Dataset

This research work makes of Amazon product review datasets to evaluate the performance of proposed classifier against HES [37] and EBC [38]. In Amazon, different review dataset are available different products. Among the different review datasets, this research work has chosen Book, DVD, Electronics and Kitchen Appliances dataset. Count of instances available in the chosen datasets are provided in Table 1.

Table 1. Count of Instances in Dataset

Amazon Product Review Big Dataset	Total
Book	146294
DVD	142885
Electronics	88127
Kitchen appliances	72839

4.2 Performance Metrics

This research work makes use Accuracy and F-Measure to measure the performance of proposed classifier against HES [37] and EBC [38]. Accuracy and F-measure are calculated using 4 variables which are true positive ($TrPs$), true negative ($TrNg$), false positive ($FlPs$) and false negative ($FlNg$).

- $TrPs$: Output of correctly predicted positive class.
- $TrNg$: Output of correctly predicted Negative class.
- $FlPs$: Output of incorrectly predicted positive class
- $FlNg$: Output of incorrectly predicted negative class

4.2.1 Accuracy

It is the count of correctly predicted instances against the count of predictions made. It is mathematically expressed as Eq. (9)

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TrPs + TrNg}}{\boldsymbol{TrPs + TrNg + FlPs + FlNg}} \quad (9)$$

4.2.2 F-Measure

It is the measure of classification accuracy. It is mathematically expressed as Eq. (10)

$$\boldsymbol{F - Measure} = \frac{\boldsymbol{2TrPs}}{\boldsymbol{2TrPs + FlPs + FlNg}} \quad (10)$$

4.3 Performance Metric Variables Analysis

To make better understanding about the proposed classifier, results obtained for the performance metric variables namely $TrPs, TrNg, FlPs$ and $FlNg$ are provided in Table 2.

Table 2 Results of performance metrics variables

Amazon Product Review Big Dataset	Classification Algorithms	TrP	$TrNg$	$FlPs$	$FlNg$
Book	EBC [38]	44125	43008	29496	29665
	HES [37]	48147	47369	25926	24852
	BRSVM	63111	60148	12547	10488
DVD	EBC [38]	40124	38012	32458	32291
	HES [37]	45698	44123	26249	26815
	BRSVM	63473	61457	9810	8145
Electronics	EBC [38]	23143	22365	21458	21161
	HES [37]	28129	27145	16477	16376
	BRSVM	38986	38498	5127	5516
Kitchen Appliances	EBC [38]	21369	20981	14143	16346
	HES [37]	24658	23899	12381	11901
	BRSVM	31476	31478	4145	5740

4.4 Accuracy Analysis

In Fig 3, the x-axis is marked with amazon product review big dataset while the y-axis is marked with the percentage of correctness (i.e., accuracy). From Fig 3, It is easy to make an understanding that BRSVM has better performance than HES [37] and EBC [38]. Enhanced accurate classification of instances assist BRSVM to achieve better f-measure than HES [37] and EBC [38]. Due to performing general classification by HES [37] and EBC [38] it yields to poor results. Numerical values of Figure 3 are provided in Table 3.

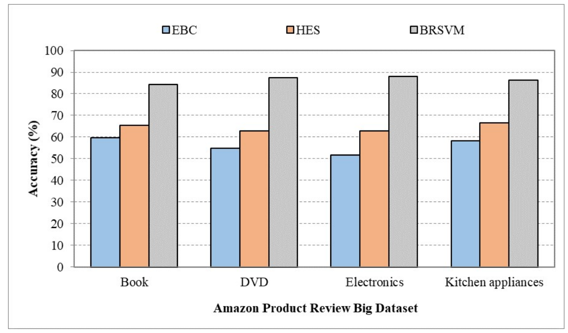

Fig 3 BRSVM Vs Accuracy

Table 3 Accuracy

Amazon Product Review Big Dataset	EBC	HES	BRSVM
Book	59.5602	65.2904	84.2543
DVD	54.6845	62.8624	87.4340
Electronics	51.6391	62.7208	87.9231
Kitchen appliances	58.1419	66.6635	86.4290

4.5 F- Measure Analysis

In Fig 4, the x-axis is marked with amazon product review big dataset while the y-axis is marked with the F-Measure in percentage. From Fig 4, it is evident that BRSVM achieves better f-measure than HES [37] and EBC [38]. BRSVM identifies the sentiments using ant colony optimization and then it performs classification, but HES [37] and EBC [38] concentrates on classification alone. Neural network present in IGWO-ELM assists in improved accuracy. Lagrangian strategy utilization assist BRSVM to achieve better optimization that leads to improved accuracy. Numerical values of Figure 4 are provided in Table 4.

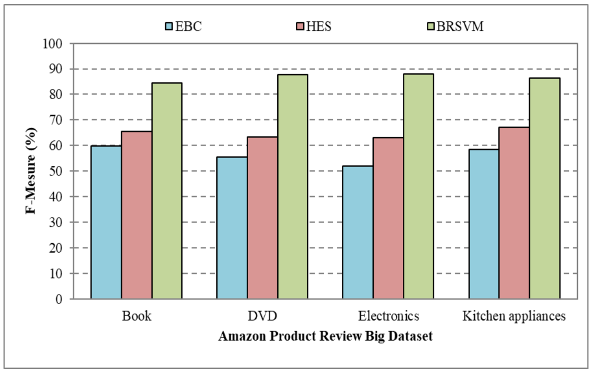

Fig 4 F-Measure Vs Accuracy

Table 4 F-Measure

Amazon Product Review Big Dataset	EBC	HES	BRSVM
Book	59.8666	65.4741	84.5669
DVD	55.3446	63.2673	87.6088
Electronics	52.0623	63.1325	87.9896
Kitchen appliances	58.3637	67.0073	86.4286

5. CONCLUSION

This paper has proposed a Bioinspired Reliable Support Vector Machine (BRSVM) to perform classification in big product review dataset. BRSVM is inspired from the biological characteristics of ants towards searching for its food. In BRSVM, sentiments are detected using ant colony optimization and the same is provided as input to RSVM to perform classification. Utilization of kernel function to handle high dimensional feature space yield to better classification. BRSVM is evaluated in MATLAB R2018b with three big product review data set using the performance metrics accuracy and f-measure. BRSVM has attained the average accuracy of 86.5101%, where Entropy Based Classifier and Hybrid Ensemble Scheme have achieved 56.0064% and 64.3843% respectively. Future enhancement of this research work can be focused with enhancing the classification accuracy even more by adopting enhanced machine learning techniques.

REFERENCES:

[1] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data," *Inf. Process. Manag.*, vol. 58, no. 1, p. 102435, 2021, doi: https://doi.org/10.1016/j.ipm.2020.102435.

[2] G. Zhai, Y. Yang, H. Wang, and S. Du, "Multi-attention fusion modeling for sentiment analysis of educational big data," *Big Data Min. Anal.*, vol. 3, no. 4, pp. 311–319, Dec. 2020, doi: 10.26599/BDMA.2020.9020024.

[3] E. Tromp, M. Pechenizkiy, and M. M. Gaber, "Expressive modeling for trusted big data analytics: techniques and applications in sentiment analysis," *Big Data Anal.*, vol. 2, no. 1, pp. 1–28, Dec. 2017, doi: 10.1186/s41044-016-0018-9.

[4] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Syst.*, vol. 226, p. 107134, Aug. 2021, doi: 10.1016/j.knosys.2021.107134.

[5] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[6] J. Xu, Z. Li, F. Huang, C. Li, and P. S. Yu, "Social Image Sentiment Analysis by Exploiting Multimodal Content and Heterogeneous Relations," *IEEE Trans. Ind. Informatics*, vol. 17, no. 4, pp. 2974–2982, Apr. 2021, doi: 10.1109/TII.2020.3005405.

[7] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis," *J. Big Data*, vol. 5, no. 1, pp. 1–25, Dec. 2018, doi: 10.1186/s40537-017-0111-6.

[8] A. R. Pathak, M. Pandey, and S. Rautaray, "Topic-level sentiment analysis of social media data using deep learning," *Appl. Soft Comput.*, vol. 108, p. 107440, Sep. 2021, doi: 10.1016/j.asoc.2021.107440.

[9] Y. Wang, Q. Chen, M. Ahmed, Z. Li, W. Pan, and H. Liu, "Joint Inference for Aspect-Level Sentiment Analysis by Deep Neural Networks and Linguistic Hints," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 2002–2014, May 2021, doi:

10.1109/TKDE.2019.2947587.

[10] S. Hu, A. Kumar, F. Al-Turjman, S. Gupta, S. Seth, and Shubham, "Reviewer Credibility and Sentiment Analysis Based User Profile Modelling for Online Product Recommendation," *IEEE Access*, vol. 8, pp. 26172–26189, 2020, doi: 10.1109/ACCESS.2020.2971087.

[11] C. Wu *et al.*, "Multiple-element joint detection for Aspect-Based Sentiment Analysis," *Knowledge-Based Syst.*, vol. 223, p. 107073, Jul. 2021, doi: 10.1016/j.knosys.2021.107073.

[12] S. Al-Dabet, S. Tedmori, and M. AL-Smadi, "Enhancing Arabic aspect-based sentiment analysis using deep learning models," *Comput. Speech Lang.*, vol. 69, p. 101224, Sep. 2021, doi: 10.1016/j.csl.2021.101224.

[13] V. Iosifidis and E. Ntoutsi, "Sentiment analysis on big sparse data streams with limited labels," *Knowl. Inf. Syst.*, vol. 62, no. 4, pp. 1393–1432, Apr. 2020, doi: 10.1007/s10115-019-01392-9.

[14] R. Neves-Silva, M. Gamito, P. Pina, and A. R. Campos, "Modelling Influence and Reach in Sentiment Analysis," in *Procedia CIRP*, Jan. 2016, vol. 47, pp. 48–53, doi: 10.1016/j.procir.2016.03.239.

[15] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, "A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks," *Neurocomputing*, vol. 380, pp. 1–10, 2020, doi: https://doi.org/10.1016/j.neucom.2019.10.009.

[16] B. Bansal and S. Srivastava, "On predicting elections with hybrid topic based sentiment analysis of tweets," *Procedia Comput. Sci.*, vol. 135, pp. 346–353, 2018, doi: https://doi.org/10.1016/j.procs.2018.08.183.

[17] F. H. Khan, U. Qamar, and S. Bashir, "SWIMS: Semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis," *Knowledge-Based Syst.*, vol. 100, pp. 97–111, May 2016, doi: 10.1016/j.knosys.2016.02.011.

[18] S. Bahri, P. Bahri, and S. Lal, "A Novel approach of Sentiment Classification using Emoticons," in *Procedia Computer Science*, Jan. 2018, vol. 132, pp. 669–678, doi: 10.1016/j.procs.2018.05.067.

[19] M. Abdar *et al.*, "Energy choices in Alaska: Mining people's perception and attitudes from geotagged tweets," *Renew. Sustain. Energy Rev.*, vol. 124, p. 109781, May 2020, doi: 10.1016/j.rser.2020.109781.

[20] A. Valdivia *et al.*, "Inconsistencies on TripAdvisor reviews: A unified index between users and Sentiment Analysis Methods," *Neurocomputing*, vol. 353, pp. 3–16, 2019, doi: https://doi.org/10.1016/j.neucom.2018.09.096.

[21] K. Ravi, V. Ravi, and P. S. R. K. Prasad, "Fuzzy formal concept analysis based opinion mining for CRM in financial services," *Appl. Soft Comput. J.*, vol. 60, pp. 786–807, Nov. 2017, doi: 10.1016/j.asoc.2017.05.028.

[22] Y. Liu, J. W. Bi, and Z. P. Fan, "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms," *Expert Syst. Appl.*, vol. 80, pp. 323–339, Sep. 2017, doi: 10.1016/j.eswa.2017.03.042.

[23] X. Zheng, Z. Lin, X. Wang, K. J. Lin, and M. Song, "Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification," *Knowledge-Based Syst.*, vol. 61, pp. 29–47, May 2014, doi: 10.1016/j.knosys.2014.02.003.

[24] A. Valdivia, M. V. Luzón, E. Cambria, and F. Herrera, "Consensus vote models for detecting and filtering neutrality in sentiment analysis," *Inf. Fusion*, vol. 44, pp. 126–135, 2018, doi: https://doi.org/10.1016/j.inffus.2018.03.007.

[25] K. Xu, G. Qi, J. Huang, T. Wu, and X. Fu, "Detecting bursts in sentiment-aware topics from social media," *Knowledge-Based Syst.*, vol. 141, pp. 44–54, Feb. 2018, doi: 10.1016/j.knosys.2017.11.007.

[26] Y. Liu, J. W. Bi, and Z. P. Fan, "A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm," *Inf. Sci. (Ny).*, vol. 394–395, pp. 38–52, Jul. 2017, doi: 10.1016/j.ins.2017.02.016.

[27] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," *Decis. Support Syst.*, vol. 57, no. 1, pp. 245–257, Jan. 2014, doi: 10.1016/j.dss.2013.09.004.

[28] W. Wang, S. J. Pan, and D. Dahlmeier, "Memory networks for fine-grained opinion mining," *Artif. Intell.*, vol. 265, pp. 1–17, Dec. 2018, doi: 10.1016/j.artint.2018.09.002.

[29] J. Ramkumar and R. Vadivel, "Performance Modeling of Bio-Inspired Routing Protocols in Cognitive Radio Ad Hoc Network to Reduce End-to-End Delay," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 221–231, 2019, doi: 10.22266/ijies2019.0228.22.

[30] J. Ramkumar and R. Vadivel, "Improved Wolf prey inspired protocol for routing in cognitive radio Ad Hoc networks," *Int. J. Comput. Networks Appl.*, vol. 7, no. 5, pp. 126–136, 2020, doi: 10.22247/ijcna/2020/202977.

[31] J. Ramkumar and R. Vadivel, "Meticulous elephant herding optimization based protocol for detecting intrusions in cognitive radio ad hoc networks," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 8, pp. 4549–4554, 2020, doi: 10.30534/ijeter/2020/82882020.

[32] J. Ramkumar and R. Vadivel, "Intelligent Fish Swarm Inspired Protocol (IFSIP) For Dynamic Ideal Routing in Cognitive Radio Ad-Hoc Networks," *Int. J. Comput. Digit. Syst.*, vol. 10, pp. 2–11, 2020, Accessed: Dec. 02, 2020. [Online]. Available: http://journals.uob.edu.bh.

[33] J. Ramkumar and R. Vadivel, "Multi-Adaptive Routing Protocol for Internet of Things based Ad-hoc Networks," *Wirel. Pers. Commun.*, pp. 1–23, Apr. 2021, doi: 10.1007/s11277-021-08495-z.

[34] M. Lingaraj, T. N. Sugumar, C. Stanly Felix, and J. Ramkumar, "Query Aware Routing Protocol for Mobility Enabled Wireless Sensor Network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 3, p. 258, Jun. 2021, doi: 10.22247/IJCNA/2021/209192.

[35] R. Vadivel and J. Ramkumar, "QoS-Enabled Improved Cuckoo Search-Inspired Protocol (ICSIP) for IoT-Based Healthcare Applications," pp. 109–121, 2019, doi: 10.4018/978-1-7998-1090-2.ch006.

[36] S. Boopalan and S. Jayasankari, "Dolphin Swarm Inspired Protocol (DSIP) for routing in underwater wireless sensor networks," *Int. J. Comput. Networks Appl.*, vol. 8, no. 1, pp. 44–53, Feb. 2021, doi: 10.22247/IJCNA/2021/207981.

[37] A. Onan, S. Korukoğlu, and H. Bulut, "A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification," *Inf. Process. Manag.*, vol. 53, no. 4, pp. 814–833, Jul. 2017, doi: 10.1016/j.ipm.2017.02.008.

[38] J. S. Deshmukh and A. K. Tripathy, "Entropy based classifier for cross-domain opinion mining," *Appl. Comput. Informatics*, vol. 14, no. 1, pp. 55–64, Jan. 2018, doi: 10.1016/j.aci.2017.03.001.