International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V68I12P223

Volume 68 Issue 12, 140-145, December 2020

Sentiment Detection Using Fish Optimization Genetic Algorithm

Sukhlal Sangule[#1], Dr. Sunil Phulre[*2]

[1#]HOD, Information Technology, S. V. Polytechnic College, Bhopal, MP, India

[#2]Associate Professor, Computer Science and Engineering Dept., LNCT University, Bhopal, MP, India

[1]ssangule@gmail.com,[2]sunilp@lnct.ac.in

Abstract - *Digital platforms are growing day by day, as people invest a large time in them. So human thoughts for any product, service, organization are easily available on this media platform. Analysis of user comments was done by text mining for understanding the status of any service or product. The sentiment of digital comments was extracted in the form of positive or negative class. This paper has proposed a Fish Optimization Genetic Algorithm for Sentiment Detection (FOGASD) in digital content. Collective Volitive and Feeding operator has increased the sentiment performance of work as well. Patterns are extracted from the input content, and as per the genetic algorithm, the output class was assigned to the patterns. The experiment was performed on a real amazon dataset having two sentiment classes. Results show that the proposed work has increased the evaluation parameter values as compared to another existing algorithm.*

Keywords — *Classification, Sentiment analysis, Ontology, Text Mining, Un-directed Classification.*

I. INTRODUCTION

The world is full of subjective as humans have its nature, thus concluding as per the significance. There causes the framework of the sentiments over the state of mind about any specific is expressed in the variety of books as per the survey. Mining work is the client undergo work is testing its helpful essentiality. There is an estimation investigation followed in three levels: record level, sentence level, or trait level. There will be decreasing of the all archive to the solitary conclusion at the estimation level in the report level. As the report estimation archive fails to speak, a similar report needs to contain the more diverse repudiating assessment about the same. While in the sentence level, opinion investigation is used to arrange the assessment for a similar element. For characterizing the sentence to be emotional or objective principal assessment is utilized.

Sentiment mining (commonly as estimation examination) is a venture to exploit client-created content's immense action. This mining uses the PC's full capacity to have the client's opinion and avoid further reuse. There are some earlier works of having some abstracts that begin in the mid-80s and 90s with the rise of web 2.0. the new technology boosted the new research with better techniques and data, having new issues, shortcomings, and challenges. Thus, there would be an obvious colossal increase in the research and work in this advanced decade.

Sentiment analysis is a natural language processing and information extraction method that denotes getting people's emotions in a two remark: a positive or negative remark, questions, and desires, by plugging or breaking many reports. In the sentiment analysis, there is an examination of the person's mentality or speaker and decide the general tone of that individual. The impetus behind the sentiment analysis is the increase in the exponential increment in the internet usage and trade of conclusions. As the web is collecting enormous sorted and unstructured data, investigation of such dormant general assessment is a tough task.

II. RELATED WORK

In [7], The Author proposes a dictionary-based technique for domain-specific sentiment analysis on the movie review dataset. The author makes use of a lexicon known as SentiWordNet (SWN-publically available dictionary), including adjectives, adverbs, and verbs. Report level examination includes utilizing phonetic highlights running from adverb + adjective to adverb + adjective + verb combination.

In [8] paper proposes a propel system for supposition mining that relates all the benefits of semantic web-guided answers to improve traditional NLP's general results (Natural Language Processing). The proposed framework makes use of domain ontology at the feature extraction stage. This enhancement makes huge changes in the feature-based sentiment classification. Existing machine learning techniques classify the words into a limited category such as positive/negative. The existing system also performs sentiment classification at the document level (i.e.) if it includes a huge no of positive than negative terms. It will be considered to be a positive document, otherwise a negative document. Dataset of Movie Reviews is used to check the performance of the proposed model.

In [9], Basha et al. presented that as the presence of E-business item reviews for an item were likewise developing quickly with an exponential factor. To settle on a choice among numerous choices where time and cash were valuable, other individuals' emotions would play a significant job. Presently, most of the associations had emotion/sentiment mining and slant investigation as a piece of their examination. Additionally, every business was affected by online networking sites and web journals that drove these organizations to do the nostalgic investigation.

In [10], a novel method for extracting Web video groups' hierarchical structure based on sentiment-aware signed network analysis is presented to realize Web video retrieval. First, the proposed method estimates latent links between Web videos using multimodal features of contents and sentiment features obtained from texts attached to Web videos. Thus, our method enables the construction of a signed network that reflects not only similarities but also positive and negative relations between topics of Web videos. Moreover, an algorithm to optimize a modularity-based measure, which can adaptively adjust the balance between positive and negative edges, was newly developed. This algorithm detects Web video groups with similar topics at multiple abstraction levels; thus, successfully extracting the hierarchical structure becomes feasible. By providing the hierarchical structure, users can obtain an overview of many Web videos, and it becomes feasible to retrieve the desired Web videos successfully.

In [11], a new method for calculating the polarities and strengths of Chinese sentiment phrases is proposed in this study, which could be used to analyze the semantic fuzziness of Chinese. It uses a probability value, rather than a fixed value for sentiment phrases' polarity strengths, compared with the conventional methods. According to those phrases' polarities and strengths, this paper proposes two multi-strategy sentiment analysis methods based on SVM and NB. Particularly, in the method based on NB, this paper considers adversative conjunctions. The two methods could be used for the sentiment analysis of documents.

In [12], paper this paper focuses on fusing textual information of Twitter messages and sentiment diffusion patterns to obtain better performance of sentiment analysis on Twitter data. This paper first analyzes sentiment diffusion by investigating a phenomenon called sentiment reversal and finding some interesting properties of sentiment reversals. This paper then considers the inter-relationships between textual information of Twitter messages and sentiment diffusion patterns. It proposes an iterative algorithm called SentiDiff to predict sentiment polarities expressed in Twitter messages.

III. PROPOSED METHODOLOGY

The proposed FOGASDmodel performed social platform review content sentiment mining. The block diagram did the proposed model from raw data processing to ontology creation for sentiment identification. 1. Keyword extraction or phrase based data was prepared to find the sentiment ontology by using a fish schooling genetic algorithm.

A. Preprocessing

Rating of any product or service was done by two types first was an objective questionnaire, while the second was a user text review. This work pre-processes text review for finding the user sentiment towards the product or service.

Stop-word Filtration: Reviews have few words which were used to frame the sentence. English words that help in sentence building are term as stop-words. Example of those like: {a, be, the, in, to, for, an, can, etc.}. In this step of sentiment mining, input review was arranged in a set of words and each word was compared with the stop-word S dictionary. If a word is found in the dictionary, remove it; otherwise, keep the original word [13]. So if input data have r number of reviews and S is a stop-word list, then pre-processed data PD is obtained by:

Input D, S

Output: PD

1. Loop 1:R
2. Loop 1:W // W: number of words in R
3. Loop S
4. If D[R] Do not intersect S
5. PD←D[R]
6. EndIf
7. endloop
8. EndLoop
9. EndLoop

B. Generate Phrase

Two kinds of features were extracted from the review content obtained from PD. The First was single term based, and the other was more than one term based. This paper performs sentiment mining by more then one term-based approach, also known as phrase-based. Consecutive word sets are found in reviews for declaring a phrase. So if any consecutive words were present in more than one review paper, put that phrase in P [14]. The phrase generation was done by the below function, which generates a successive pattern and traces those in PD.

P←Successive_Words(IPD)----------Eq. 1

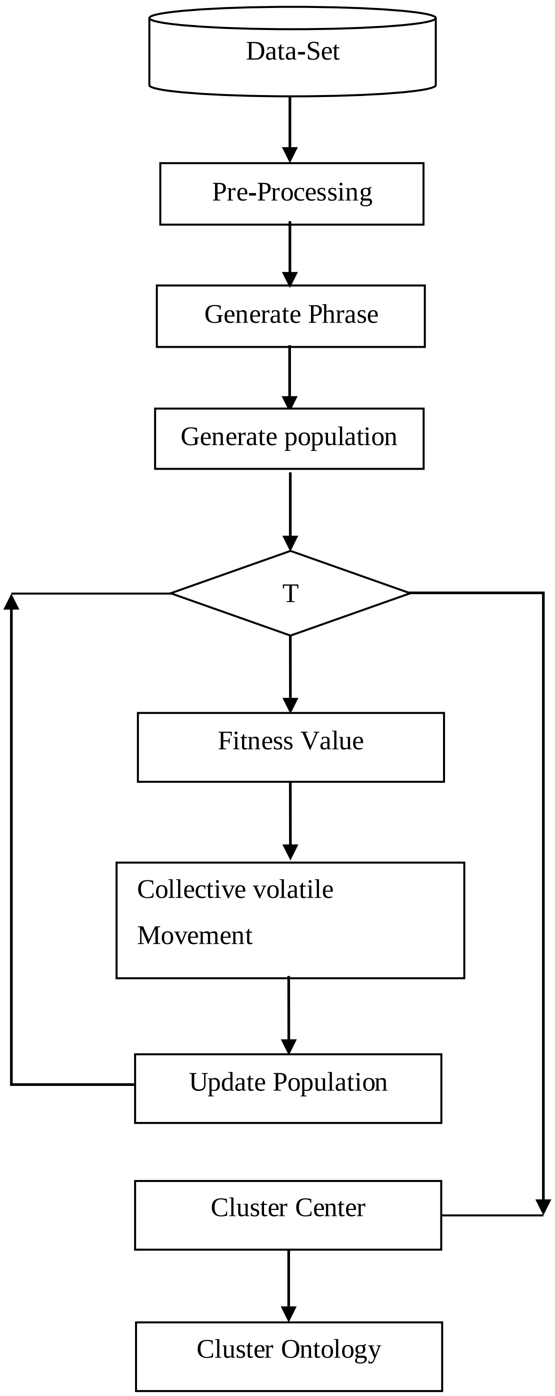

Fig. 1 Proposed work Block diagram.

C. Phrase Graph

Phrases obtained from reviews arranged in the graph data structure. Graph G_{PxP} has nodes in the form of phrases and weight in the similarity between phrase terms present in phrase for a common sentiment class review [15]. Relation between phrase and uses as per sentiment was developed in the graph. The similarity word count of two phrases for G_{PxP}in a review was obtained by:

```
Input: P, PD
Output: G
Loop 1:Pr  // Phrase in Row
        Loop 1:Pc// Phrase in Column
                Loop 1:PD
                        RC←Similar Word (PD, Pr)
                        CC←Similar Word (PD, Pc)
                        // Weight update condition
                        If RC and CC is greater than 0
                        G[Pr, Pc, 1]← G[Pr, Pc]+1
                        St←Sentiment_Type(PD)
                        G[Pr,St,2]← G[Pr,St,2]+1
                        EndIf
                endloop
        EndLoop
EndLoop
```

RC (Row Counter) is common terms in Pd and P_r counter. Similarly, RC (Row Counter) is common terms in Pd and P_r. Graph weight value between a phrase increases if review PD has common terms from any of phrase terms. Higher the weight value stronger the relationship between the phrases.

D. Fish Schooling Genetic Algorithm:

This FSGA did the selection of sentiment representative phrases. The algorithm performed fish Schooling Genetic algorithm search food for dispersion and assembly operation for higher area search. The phrase set selection was done by rotating and fish (phrase) in a deep area [16]. This genetic algorithm's major steps involve fish population generation, Collective Volitive Movement, more than one fish movement after that Feeding operator and Crossover for fish group shuffling was performed.

Generate Fish Population: The primary unit of genetic algorithm work is a chromosome. In the FSGA algorithm group of fish, each fish was phrase P. Number of fish in chromosome depends on sentiment type. So a random FP set was developed by the FSGA. FP is a matrix of having m number of rows representing a row and a column representing a cluster center. Based on highly connected phrases in the G cluster center, the phrase was select by

FP←Generate_Fish(P, St, m, G)------Eq. 2

Fitness Function

Random generation of chromosome in a population need fitness estimation. This step of FSGA finds fitness value by evaluating the difference between fitness sentiment value as per review G[Pr, St,2].

$F_m = \sum_{i=1}^{P} Min(\text{Euclidian}(G_{m,St}, G_{P,St}))_1^{St}$--------Eq. 3

The searching group of fish has a fitness value $F_m$where m is the number of chromosomes in the population. $G_{m,St}$is chromosome phrase value of sentiment count for St type of class. Similarly $G_{P,St}$ is P phrase value of sentiment count for St type of class.

Collective Volitive Movement

Searching for a portion of food is done by a collective operation where each chromosome undergoes assembly or dispersion. In case a fish found a barry center, then assembly of other fish done by reducing their distance with other fish. This assembly operation was performed by eq. 3. If food is not found, then the fish's distance increases as per eq. 4 *[16]*. So if the fitness value of tth iteration is higher than t+1, then apply case 1, otherwise case2.

Case1

$$\boldsymbol{x(t+1) = x(t) - M_{vol} \times R \times \left(\frac{x(t) - B(t)}{Distance(x(t), B(t))}\right)}$$

Case2

$$x(t+1) = x(t) + M_{vol} \times R \times \left(\frac{x(t) - B(t)}{\boldsymbol{Distance}(x(t), B(t))}\right)$$

$$B(t) = \frac{\sum_{i=1}^{N} x(t)W(t)}{\sum_{i=1}^{N} W(t)}$$

Feeding Operator

$$W(t+1) = W(t) + \left(\frac{\Delta f}{max(|\Delta f|)}\right)$$

$$\Delta f = F(t+1) - F(t)$$

$$M_{vol} = M_{vol} - \frac{M}{Iteration_{max}}$$

Where

M_{vol}: Maximum displacement performs in an operator.
X(t): Random position of a phrase in the tth iteration.
W(t): Summation of fitness value.

Crossover

AS per new position of fish x(t) in tth iteration values range between from null to St. This value modifies the fish in the chromosome. Other chromosome fish values were modified as per the best chromosome in a current population of tth iteration. So the result of this step was new chromosome sets. Population updation was done by evaluating the child chromosome's fitness value if the child chromosome value is better than keeping the child and parent chromosome removed from the population. Otherwise, if the parent chromosome fitness value is better, then the child chromosome was removed from the population.

E. Cluster Phrase

After the maximum T umber of iteration was completed, the obtained population's fitness value was further estimated by Eq. 3. Best chromosomes act as the final cluster center for sentiment identification. Other phrases that are not part of the cluster center were the sentiment class group as per weighted graph G.

IV. EVALUATION PARAMETER

The experimental part was done on a real dataset of Amazon review content obtained from [17]. Implementation of the proposed algorithm was done on MATLAB software. Comparison of FOGASD was done with the method proposed in [18]. In [18], term based sentiment analysis was done where the genetic algorithm reduces the feature set of work. The experimental dataset contains sentences labeled with positive or negative sentiment, extracted from reviews of products, movies, and restaurants [17].

$$\text{Precision} = \frac{\text{True\_Positive}}{\text{True\_Positive} + \text{False\_Positive}}$$

$$\text{Recall} = \frac{\text{True\_Positive}}{\text{True\_Positive} + \text{False\_Negative}}$$

$$F\_Score = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$Accuracy = \frac{Correct\_Classification}{Correct\_Classification + Incorrect\_Classification}$$

The above true positive value is obtained by the system when the ranked review/comment favors the user query, and the system also says that review/comment favors the user query. While in the case of false-positive value, the system obtains when the input review/comment favors the user query, and the system does not rank that review/comment in their list.

Results:

Table 1 shows the review sentiment class detection accuracy value obtained from previous work in [18] and proposed FOGASD work. The use of the Fish Optimization algorithm for sentiment phrase detection has increased the accuracy of work. The proposed model does not use Terms as done in [18], as a single word results in false clustering. Clustering, by finding a pair of words, has enhanced the correct sentiment class detection.

Table 1. Accuracy based comparison of sentiment detection techniques.

Dataset Size	Previous Work [18]	FOGASD
100	40.404	67.6768
200	36.1809	68.8442
300	25.4181	50.8361
400	17.2932	51.6291

Table 2. Precision based comparison of sentiment detection techniques.

Dataset Size	Previous Work [18]	FOGASD
100	0.5882	0.6471
200	0.5446	0.6733
300	0.3861	0.4810
400	0.2891	0.5071

Table 2 shows the precision value of sentiment Identification for a testing review with different number dataset size. FOGASD has improved the precision value by 21.68%% as compared to the genetic algorithm used in [18]. This raise of precision was obtained by a collective operation done by the I fish genetic algorithm. It controls crossover operation value as per the best fitness value of a chromosome in an algorithm.

Table 3. Recall based comparison of sentiment detection techniques.

Dataset Size	Previous Work [18]	FOGASD
100	0.4412	0.7021
200	0.4044	0.701
300	0.3262	0.5390
400	0.2531	0.5459

Table 3 shows the recall value of sentiment Identification for testing reviews with a different number of dataset size. FOGASD has improved the precision value by 42.7% compared to the genetic algorithm used in [18]. This raise of precision was obtained by a collective operation done in the fish genetic algorithm. It controls crossover operation value as per the best fitness value of a chromosome in an algorithm.

Table 4. F-measure based comparison of sentiment detection techniques.

Dataset Size	Previous Work [18]	FOGASD
100	0.5042	0.6735
200	0.4641	0.6869
300	0.3536	0.5084
400	0.2699	0.5258

Table 4 shows the review sentiment class detection F-measure value obtained from previous work in [18] and proposed FOGASD work. The use of the Fish Optimization algorithm for sentiment phrase detection has increased the accuracy of work. The proposed model does not use Terms as done in [18], as a single word results in false clustering. Clustering by finding a pair of words has enhanced the correct sentiment class detection.

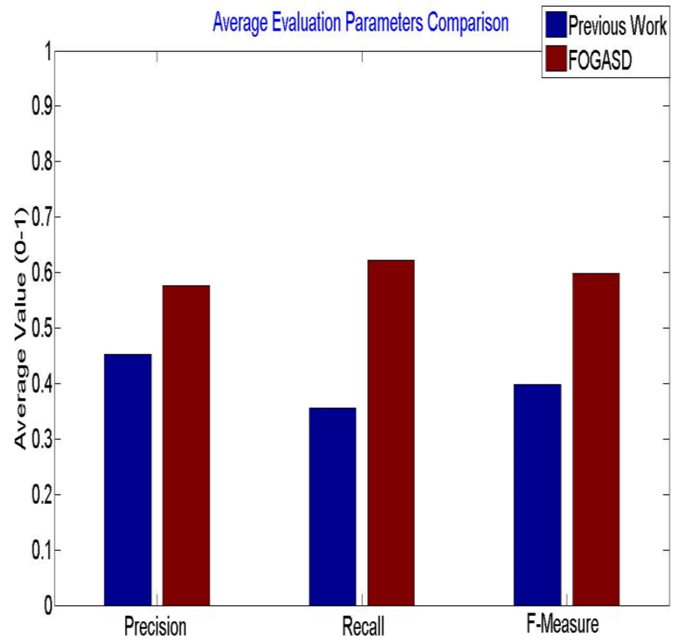

Fig. 2 Comparison of sentiment detection techniques average Evaluation parameters values.

Fig. 2 shows the review sentiment class detection accuracy value obtained from previous work in [18] and proposed FOGASD work. The use of the Fish Optimization algorithm for sentiment phrase detection has increased the accuracy of work. The proposed model does not use Terms as done in [18], as a single word results in false clustering. This raise of precision was obtained by a collective operation done in the fish genetic algorithm. It controls crossover operation value as per the best fitness value of a chromosome in an algorithm.

V. CONCLUSIONS

In a few decades, miners have many open dimensions of the research field. Out of those fields, this paper has worked on sentiment mining. Paper has proposed Fish Optimization Genetic Algorithm Sentiment Detection. For increasing the efficiency of the sentiment identification work, utilize input content in the form of phrases. So extracted feature was arranging into a graph data structure for enhancing the fitness function output. The experiment was done on the amazon review dataset, and results show that the proposed FOGASD has increased the precision value by 21.68%, while the accuracy value was improved by 50.08%. In the future, the researcher can enhance the sentiment detection work by involving another learning model.

REFERENCES

[1] P Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, Social Media Mining – An Introduction, Cambridge University Press, Publisher Location, 2014.

[2] Kavanaugh, A. L., Fox, E. A., Sheetz, S. D., Yang, S.Li, L. T., Shoemaker, D. J., …Xie, L. Social media use by the government: from the routine to the critical. Government Information Quarterly, 29(4), 2012.

[3] Twitter Usage Statistics – Live Internet Stats. (n.d.). Retrieved Oct. 22, from http://www.internetlivesstats.com/twitter-statistics. 2018.

[4] Liu, B., Sentiment analysis, and opinion mining. Synthesis lectures on human language technologies, 2012.

[5] Dandan Jiang1, Xiangfeng Luo1, Junyu Xuan, And Zheng Xu .“Sentiment Computing for the News Event-Based. on the Social Media Big Data”. Digital Object Identifier 10.1109/ACCESS.2016.2607218 IEEE Acess, 2017.

[6] Priyanka Patil, Pratibha Yalagi, Sentiment Analysis Levels and Techniques: A Survey in International Journal of Innovations in Engineering and Technology, April 2016.

[7] Isidro Peñalver-Martinez, Francisco Garcia-Sanchez, Rafael Valencia-Garcia,” Feature-based opinion mining through ontologies,” Expert Systems with Applications-2014.

[8] V.K. Singh, R. Piryani, A. Uddin, P. Waila," Sentiment Analysis of Movie Reviews," conference on IEEE-2013.

[9] Basha, S. M., Zhenning, Y., Rajput, D. S., Iyengar, N., &Caytiles, D. R.. Weighted Fuzzy Rule-Based Sentiment Prediction Analysis on Reviews. International Journal of Grid and Distributed Computing, 10 (6), (2017) 41-54.

[10] Ryosuke Harakawa, Takahiro Ogawa, And Miki Haseyama. Extracting Hierarchical Structure Of Web Video Groups Based On Sentiment-Aware Signed Network Analysis., Sept. 19, 2017.Digital Object Identifier 10.1109/Access.2017.2741098.

[11] Ying Fang, Hai Tan, And Jun Zhang. Multi-Strategy Sentiment Analysis Of Consumer Reviews Based On Semantic Fuzziness. Ieee Access May 2, 2018.Digital Object Identifier 10.1109/Access.2018.2820025.

[12] Lei Wang, JianweiNiu, And Shui Yu. Sentidiff: Combining Textual Information and Sentiment Diffusion Patterns For Twitter Sentiment Analysis.”Journal Of Latex Class Files, 14(8), August 2018.

[13] Dr. Nitesh M Sureja, Dr. Sanjay P Patel , Solving A Combinatorial Optimization Problem Using Artificial Fish Swarm Algorithm, International Journal of Engineering Trends and Technology 68(5) (2020) 27-32.

[14] Prof. Durgesh Wadbude, A.P. Jayshree Boaddh, Rajeev Kumar, "A Survey on Various Techniques and Characteristics of Text Document Fetching." ijsret.com 4(I) , Mar-apr-2018

[15] Lubomir Stanchev. Semantic Document Clustering Using Information from WordNet and DBPedia. IEEE 12th International Conference on Semantic Computing (ICSC), 2018.

[16] DANDAN JIANG, XIANGFENG LUO, JUNYU XUAN, AND ZHENG XU. Sentiment Computing for the News Event Based on the Social Media Big Data. IEEE Access, Mar. 15, 2017.

[17] X. Li, Z. Shao, J. Qian, Anoptimizing method base on autonomous animates: fish- swarm algorithm, Systems Engineering Theory and Practice 22 (2002) 32–38

[18] UCI ML Repository_Sentiment Analysis Dataset. Accessed: Jun. 8, 2018. Available: http://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences.

[19] Farkhund Iqbal, Jahanzeb Maqbool Hashmi, Benjamin C. M. Fung, Rabia Batool, Asad Masood Khattak, Saiqa Aleem, And Patrick C. K. Hung. A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction. IEEE Access Feb. 8, 2019.