



Sentiment Analysis Using Shuffled Frog Leaping Algorithm

¹S. P. Rajamohana, ²Dr. K. Umamaheshwari, ³Karthiga R

¹Assistant professor, ²Professor, ³PG Scholar,

^{1,2,3}Department of Information Technology, PSG College of Technology,
Coimbatore, Tamilnadu. India

Abstract: *Opinion Mining has become an indispensable part of online reviews in present scenario. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people can actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, attitudes and subjectivity in text that has occurred at least in part as a direct response to the surge of interest in new systems that directly deal with opinions. In the existing system, evolutionary methods such as Particle Swarm Optimization, Ant Colony Optimization, Ant Bee Colony Optimization algorithms were used for sentiment analysis. Evolutionary Algorithms possess a number of features that can help to position them within in the family of generate and test methods. The proposed system is implemented using Shuffled Frog Leaping Algorithm (SFLA). The SFLA is a real coded population based meta-heuristic optimization. It combines the benefits of the local search tool of the PSO and mixing information from Parallel Local Searches to move toward a global solution. SFLA identifies the best features compared to BPSO. The proposed method can be applied to analyze the reviews in social networks, forums and blogs.*

Key Words– *Opinion Mining, SFLA, Naive Bayes Classifier.*

I. INTRODUCTION

Web 2.0 nowadays provides a great means to share knowledge, including opinions which may be useful to all kinds of people: for example product reviews public reactions to political events, scientific blogs, and music trends and so on. With the web, one can express their views on almost anything on the Internet forums, blogs, etc. If one wants to buy a product, it is no longer necessary to ask opinions from their friends and families because there are plentiful of reviews present on the web which gives the opinions of existing users of the any topic.

Feature extraction is a process commonly wherein subsets of features available from the data are extracted for application of a learning algorithm. The best subset contains the least number of dimensions that most contribute to accuracy and discard the remaining, unimportant dimensions. This is an important stage of preprocessing and is one of the two ways of avoiding the curse of dimensionality. It is an extremely basic and essential task for Sentiment Analysis. In order to process the text, it is important to convert a piece of text into a feature vector. Effective feature selection is a must in order to make the learning task accurate. In text classification, the bag of words model, the documents are converted into vectors based on features weighting mechanism, which is critical to classification accuracy. The major feature types contain unigrams, bigrams and the mixtures of them, etc. Most of the existing work focused only the traditional approach. Evolutionary Algorithms (EAs) are inspired by the biological model of evolution and natural selection the natural world, evolution helps species adapt to their environments. Hence in this paper Shuffled frog leaping algorithm is applied to select the best features.

The rest of the paper is organized as follows: the literature survey is provided in the section II, section III describes the proposed system, section IV contains the result and analysis, the concluding remarks are provided in the section V.

II. LITERATURE SURVEY

Eusuff, M et al, [1] described the SFLA, a memetic meta-heuristic to solve combinatorial optimization problems. A memetic algorithm is a population-based approach for heuristic search. The SFLA starts with a sample virtual population of frogs, leaping in a swamp, searching for the optimum location of food. Frogs act as hosts of memes. A meme is considered as a unit of ideas or cultural evolution. Each meme consists of memotype(s). During the memetic evolution the memes of the frogs are changed, resulting in a change in their position towards the goal. The change in or evolution of memes occurs when the frogs are infected by other better ideas. The SFLA performance was compared with a GA for a series of test problems. The results for 11 theoretical test problems (functions) and two applications show that the SFLA performed better than or at least comparable with the GA for almost all problem domains and was more robust in determining the global solution. Four realistic engineering problems were also solved and results compared with literature results from several optimization algorithms. Again the results were extremely promising in terms of robustness of finding solutions and solution speed.

Jiang, J., et al, [2] proposed a new algorithm called Improved Shuffled Frog Leaping Algorithm. The population is separated into two parts which are searching population and competing population. In the searching population, the optimal solution is saved and the second best solution is updated. In the competing population, the mutation operator is imported to add the genotype. In the step updating process, the number of iterations is introduced to update step formula to adapt step changing. The algorithm proposed in this paper has improved the original algorithm in the aspects such as Initialization population with uniform ways, updating the second best solution to improve local search ability, Adding new genotype with Introducing mutation operator, Optimizing step updating formula. The results show that the new algorithm is better than the original algorithm in global and local searching ability. It not only can avoid falling into local optimal solution but also can improve the convergence speed and accuracy.

Farahani, M., et al, [3] presented an evolutionary algorithm based on the Shuffled Frog Leaping Algorithm (SFLA). Similar to the SFLA, our method partitions particles into different groups called memplexes; however, the best particle in each memplex thereafter determines its movement through the search space in each iteration of the algorithm toward the global best particle and the worst particle in each memplex keeps track of its coordinates in the solution space by moving toward the local best particle (the best particle in the same memplex). Not only does this method lessen computation costs and offer speedier solutions in comparison to the Particle Swarm Optimization, but it also has a distinct advantage over the SFLA in that it reduces the probability of the particles being trapped in the local minima by directing the best local particle toward the global best particle.

Lu, Kong, et al [4] proposed an improved shuffled frog-leaping algorithm to solve the flexible job shop scheduling problem. The objective of the research is to minimize the makespan in FJSP. The algorithm possesses an adjustment sequence to design the strategy of local searching and an extremal optimization in information exchange. The shuffled frog-leaping algorithms with local search can guarantee the feasibility of the updated solution; however, the step size is still selected randomly. To attain a reasonable step size, the adjustment factor and adjustment order are adopted. Moreover, an extremal optimization is used to exchange the information among individuals. The effectiveness of the improved shuffled frog leaping algorithm is evaluated using a set of well-known instances. The results gained by SFLA are often the same or slightly better than those gained by the other algorithms for the FJSP. Furthermore, the improved strategies effectively improve the performance of the algorithm.

III. PROPOSED METHODOLOGY

In the proposed approach, Feature selection using Shuffled frog leaping algorithms with naive bayes methods are presented. The proposed system design is given in fig 1:

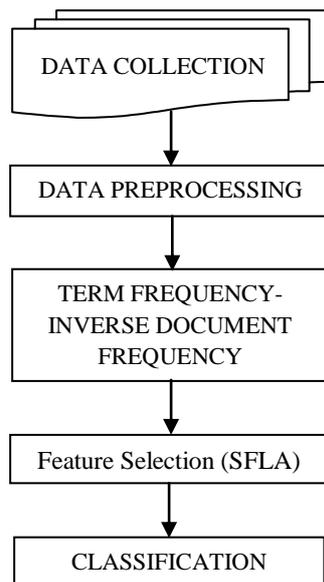


Fig.1. Framework of data processing

The stages of proposed system are as follows:

A. Preprocessing

All the irrelevant information in the review files are removed which are not necessary for analysis. Data preprocessing consists of tokenization, stop words removal and stemming. Text document has a collection of sentences which is split up into terms or tokens. Stop words such as articles like a, an, the, etc are removed. Stemming is the process for reducing words to their root word[27].

B. Shuffled Frog Leaping Algorithm

Initial Population

For an S-dimensional problem an initial population of P frogs is created randomly. Then the number of features, population size and number of iterations are given as input for the feature selection. Each frog represents a candidate

solution to the problem and group of frogs are called as initial population. The frogs are represented as binary string ranges from 0 to 1.

Sorting and distribution:

The Frogs are sorted in descending order according to the fitness value. The entire population is divided into m memplexes. Each memplex contains number of frogs (i.e., $P = m * n$). The first memplex goes to the first frog, the second memplex will go to the second frog, the m frog goes to the mth memplex and the process continues.

Memplex Evolution:

The frogs with the best fitness as X_b and worst fitness X_w is identified within each memplex. The global best fitness is considered as X_g . In order to improve the worst frog solution, the following equations are used:

Change in frog position

$$(D'i) = rand () * (X'b.X'w) \quad (13)$$

New position

$$X'w = \text{current position } X'w + D'i \quad (14)$$

$$D_{\max} \geq D_i \geq -D_{\max}$$

where $rand ()$ is a random number between 0 and 1 and D_{\max} is the maximum allowed change in a frog's position. It replaces the worst frog, otherwise the above calculations are repeated but If no improvement, then a new solution is randomly generated to replace the worst frog.

C. Performance Evaluation

On implementing sentiment classification system, an analysis was made over the performance of both LDA and SVM classifier. Performance analysis was made based on the evaluation measures such as precision and recall.

Evaluation Measure:

Precision: It is the number of relevant documents identified; precision is given in eqn (5)

$$\frac{\text{Number of true positives}}{\text{Number of true positives} + \text{False positives}} \quad (5)$$

Recall: It is the percentage of all topics that are correctly classified. It evaluates the performance of the particular system which is used in classification. Recall is given in eqn (6),

$$\frac{\text{Number of true positives}}{\text{Number of true positives} + \text{False negatives}} \quad (6)$$

F-Measure: F-score is used to measure a particular test's accuracy. It considers both the precision and recall to compute the score: p is the number of correct results divided by the number of result that should have been returned. Fscore is given by the eqn(7)

$$F\text{Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

IV. RESULT AND ANALYSIS

The movie review dataset is collected from the website cs.cornell.edu. It contains 2000 reviews consists of 1000 positive and 1000 negative reviews.

A. Data Preprocessing

Data preprocessing consists of several steps such as tokenization, stop-word removal and stemming. The dataset consists of 13 lakhs of words. After preprocessing, the number of words reduced to 5 lakhs.

B. TF-IDF:

TF-IDF values are calculated by extracting the words from all documents that are given as input. The results of TF-IDF calculation is given in table I

Table I. TF-IDF values

Doc_no	Word	TF*IDF
cv000_29416.txt	good	0.071921
cv000_29590.txt	good	0.115073
cv001_18431.txt	good	0.028768
cv001_18431.txt	person	0.069315
cv000_29416.txt	person	0.866431
cv000_29590.txt	dark	0.138629
cv001_18431.txt	dark	0.069315
cv000_29590.txt	success	0.138629
cv001_18431.txt	success	0.069315

These TF-IDF values are given as input to the SFLA. SFLA take the given features; calculate the fitness value. Sort the features according to its fitness values calculated and divide it into memplexes as specified. After calculating the fitness of each feature, GBest values will be calculated for each memplex. For each iteration, overall best features will be calculated from those values. By comparing the result of BPSO, SFLA produces better results. The result of BPSO and SFLA are given in table II

Table II. Comparison of BPSO and SFLA based on features

ALGORITHM	TOP 100	TOP 200	TOP 300	TOP 400
BPSO	0.6034	0.6069	0.6076	0.6060
SFLA	0.7452	0.9312	0.9366	0.9969

Naive Bayesian classifier is used to classify the documents according to the best features calculated by the evolutionary algorithm. Table III gives the classification results of BPSO and SFLA.

Table III Classification results

ALGORITHM	NAIVEBAYES (Acc%)
BPSO	90%
SFLA	93%

SFLA gains better results when compared to BPSO. The performance of BPSO and SFLA based on precision, recall and accuracy is given in the following figures 2, 3 & 4.

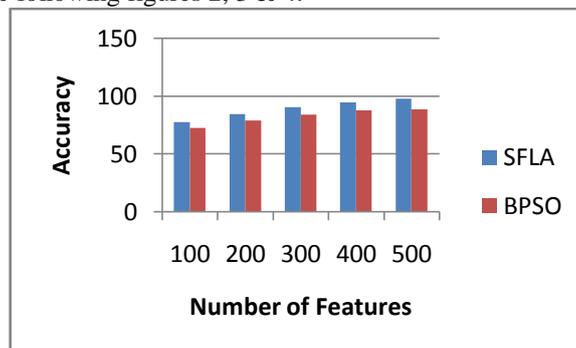


Fig 4. Performance Analysis of BPSO and SFLA based on accuracy, precision and recall.

V. CONCLUSION

Managing the explosion of digital content on the internet requires new tools for automatically mining, searching, indexing and browsing large collections of text data. Recent research in data mining and statistics has developed a new brand of techniques called evolutionary approach for text. The system provides the mining results of different movies based on data mining and natural language processing methods. The proposed system presents SFLA with Naive Bayesian classifier. The objective of this system is to maximize the accuracy of sentiment classification and to improve the performance of the system. The particle swarm optimization with local search can guarantee the feasibility of the updated solution; however it achieves minimum accuracy as like other evolutionary approaches. The effectiveness of the shuffled frog leaping algorithm is evaluated using a set of well-known instances. This indicates that the results gained by SFLA are better than those gained by the other algorithms for the Sentiment Classification. In future, topic modeling technique will be used to identify aspects along with the opinions. The result of topic modeling will be compared with the existing evolutionary methods used in opinion mining.

REFERENCES

- [1] Eusuff, M, Lansey, K, & Pasha, F, "Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization", *Engineering Optimization*, Vol.38 (2), pp.129-154, 2006.
- [2] Jiang, J, Su, Q, Li, M, Liu, M, & Zhang, L, "An Improved Shuffled Frog Leaping Algorithm", *J. Inf. Comput. Sci*, Vol.10, pp.4619-4626, 2013
- [3] Farahani, M, Movahhed, S. B, & Ghaderi, S. F, "A hybrid meta-heuristic optimization algorithm based on SFLA", In 2nd Int. Conf. on Engineering Optimization, September, 2010
- [4] Lu, K, Ting, L, Keming, W, Hanbing, Z, Makoto, T, & Bin, Y, "An Improved Shuffled Frog-Leaping Algorithm for Flexible Job Shop Scheduling Problem. *Algorithms*, Vol.8 (1), pp.19-31, 2015
- [5] Taghikhani, M. A, "DG Allocation and Sizing in Distribution Network Using Modified Shuffled Frog Leaping Algorithm", *International Journal of Automation and Power Engineering*, Vol.1, pp.10-19, September, 2010.
- [6] Liping, Z, Weiwei, W, Yi, H, Yefeng, X, & Yixian, C, "Application of shuffled frog leaping algorithm to an uncapacitated SLLS problem", *AASRI Procedia*, Vol.1, pp. 226-231, 2012

- [7] Jadidoleslam, M, Bijami, E, Amiri, N, Ebrahimi, A, & Askari, J, "Application of shuffled frog leaping algorithm to long term generation expansion planning", *International Journal of Computer and Electrical Engineering*, Vol.4, No.2, April 2012
- [8] Lee, D., Jeong, O. R., & Lee, S. G, "Opinion mining of customer feedback data on the web", In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pp. 230-235, January, 2008
- [9] Jandail, R. R. S, & Ratan, R, "A proposed Novel Approach for Sentiment Analysis and Opinion Mining", *International Journal of UbiComp*, Vol. 5, 2014
- [10] Tsutsumi, K, Shimada, K, & Endo, T, "Movie review classification based on a multiple classifier", In the 21th Pacific Asia Conference on Language, Information and Computation (PACLIC), 2007
- [11] Haydari, M., Banejad, M., & Hahizadeh, A, "Using the Modified Shuffled Frog Leaping Algorithm for Optimal Sizing and location of Distributed Generation Resources for Reliability Improvement", *Journal of AI and Data Mining*, Vol.1(2), pp.103-110, 2013
- [12] Darvishi, A, Alimardani, A, Vahidi, B, & Hosseinian, S. H, "Shuffled Frog-Leaping Algorithm for Control of Selective and Total Harmonic Distortion", *Journal of Applied Research and Technology*, Vol.12(1), 2014
- [13] Tripathi, A, Singh, V, & Sharma, T. K, "Application of Shuffled Frog Leaping Algorithm in Determining Optimal Manpower Size", *Matrix Academic International Online Journal Of Engineering And Technology*, ISSN: 2348-3326, October, 2015
- [14] Afzalan, E, Taghikhani, M. A, & Sedighzadeh, M, "Optimal placement and sizing of dg in radial distribution networks using sfla", *International Journal of Energy Engineering*, Vol.2(3), pp.73-77, 2012
- [15] Karimzadeh, M. K, "Improved Shuffled Frog Leaping Algorithm for the Combined Heat and Power Economic Dispatch", *HCTL Open Int. J. of Technology Innovations and Research*, HCTL Open IJTIR, Volume 2, e-ISSN: 2321-1814, ISBN: 978-1-62776-111-6, March 2013.
- [16] Basari, A. S. H, Hussin, B, Ananta, I. G. P, & Zeniarja, J., "Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization", *Procedia Engineering*, Vol.53, pp.453-462, 2013
- [17] Suin Kim, Jianwen Zhang, "A Hierarchical Aspect-Sentiment Model for Online Reviews", *Association for the Advancement of Artificial Intelligence*, Oct, 2013
- [18] Ravendra Ratan Singh Jandail, "A proposed Novel Approach for Sentiment Analysis and Opinion Mining", *International Journal of UbiComp (IJU)*, Vol.5, Apr, 2014
- [19] Richa Sharma, Shweta Nigam and Rekha Jain, "Mining Of Product Reviews At Aspect Level", *International Journal in Foundations of Computer Science & Technology (IJFCST)*, Vol.4, May, 2014
- [20] Oh J H, Torisawa K, Hashimoto C, et al. "Why question answering using sentiment analysis and word classes". *Proceedings of EMNLPCNLL*, pages.368-378, Apr, 2012
- [21] Poongodi s, radha n, "A Survey based on Opinion Classification Techniques" *International Journal of Engineering Sciences & Research Technology*, October, 2013
- [22] Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar, "Opinion Mining and Analysis: A Survey", *International Journal on Natural Language Computing (IJNLC)* Vol. 2, No.3, June 2013
- [23] Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar, "Opinion Mining and Analysis: A Survey", *International Journal on Natural Language Computing (IJNLC)* Vol. 2, No.3, June 2013
- [24] Mita K. Dalal, Mukesh A. Zaveri, "Automatic text classification: a technical review", *International Journal of Computer Applications (0975 – 8887)* Volume 28– No.2, August 2011.
- [25] Y. H. Li and A. K. Jain, "Classification of Text Documents", *the computer journal*, vol. 41, no. 8, 1998
- [26] Pravesh Kumar Singh, Mohd Shahid Husain, "Analytical Study Of Feature Extraction Techniques In Opinion Mining", *Department of Computer Science and Engineering*, 2013.
- [27] SP.Rajamohana Dr.K.Umamaheswari and Karthiga R "Sentiment Classification based on LDA using SMO Classifier", *International Journal of Applied Engineering Research*, Vol 10, 2016.
- [28] SP.Rajamohana Dr.K.Umamaheswari and Karthiga R, "Sentiment Classification based on Latent Dirichlet Allocation", *International Journal of Computer Applications*, 2015.