November-December 2019
ISSN: 0193-4120 Page No. 4529 - 4535

BOVW Classification Method with Particle Swarm Optimization in Big Data

R.Vanitha
Assistant Professor, Department of Software Engineering
Periyar Maniammai Institute of Science and Technology
Thanjavur, India
itvanitha@pmu.edu.

Article Info
Volume 81
Page Number: 4529 - 4535
Publication Issue:
November-December 2019

Article History
Article Received: *5 March 2019*
Revised: *18 May 2019*
Accepted: *24 September 2019*
Publication: *23 December 2019*

Abstract
Big data technologies applied in many areas, including physical, biological and biomedical sciences and engineering. They were also used in government agencies, financial companies, large enterprises, etc. The key goal of this paper is to identify the problems and investigate possible solutions to the process of huge data handling. In this paper Introduce Bag of visual with PSO algorithms, PSO algorithm allows the time reduction expenditures for the developed Classifiers that are very important to the problem of Big Data classification. While comparing existing methods with the proposed PSO optimization with BOVW Classification method provide 98 % accuracy results.

Keywords: *Big data, BOVW, Partial swarm optimization*

1 INTRODUCTION

Big Data is a term for such large and/or complicated data sets that conventional data processing techniques are inadequate. They need software that can be used to generate and incorporate exponentially increasing data sets containing structured, semi-structured and unstructured data. The size, range, and speed of Big Data are three defining characteristics. Volume refers to the vast quantity of data, variety refers to the number of individual types, and speed refers to the velocity of data processing. Big Data management issues are the result of expanding all three functions. The storage of Big Data increases data sets transmission rate relative to the speed of simple data exchanges. The Big Data mining tools are very useful when solving their own real problems for end users. Some innovative methods actually need to be introduced when dealing with Big Data. In general, when it is necessary to retrieve, scan or classify information using the Big Data sets, the collection, clustering and classification of features plays a key role in analysis of big data [11]. Such methods are useful in fields such as pattern recognition, machine learning, bio-informatics, data mining etc.Since there are many algorithms available for selecting, clustering and classifying features, it is necessary to find the right algorithms to be selected correctly for the Big Data analysis problem.

Support Vector Machines (SVM) has emerged in the past as a properly standard for category models based on BOVW. In general, kernel SVMs based on Radial Basis Function (RBF) are commonly used. While the results obtained are very often satisfactory, very few researches specifically aims at comparing different methods of machine learning for models of categories of practice. As a result, in this paper introduce

November-December 2019
ISSN: 0193-4120 Page No. 4529 - 4535

classification based on novel BOVW method for achieve a better accuracy result.

2 RELATED WORKS

Susana M. Vieira et al [1] …have proposed MBPSO algorithm was tested showing better performance than state-of - the-art PSO methods and outcomes that are comparable or better than GA. Some drawbacks to the proposed MBPSO algorithm can occur if the parameters are improperly modified, as it has two more variable parameters than the default binary PSO version. Nonetheless, it has only one more variable than an overall performance close to that of the genetic algorithm.

Yi Yang et al [2] ...have proposed testing BOVW and spatial extensions in high-resolution overhead imagery for land-use identification. While the BOVW-based methods do not ultimately work better than the best standard approach, they are a reliable option that is more successful for some groups.

Sheng Xu et al [3] ... have proposed BOVW representation to solve the question of how to accurately depict complex and composite objects in high-resolution imagery. In VHR aerial photos, the experimental results suggest a good quality of such representation compared to the low-level features in the classification data.

S. L. Ho et al.[4] ... have proposed and tested a periodic mathematical function PSO-based vector algorithm and an inverse electromagnetic problem identified by its author as a very challenging MOP problem for a vector optimizer, the numerical results on the mathematical function show the robustness of the technique recommended in finding Pareto solutions to extremely difficult MOP problems, even though those results show the robustness of the method proposed.

YahyaRahmat et al [5] ... have proposed PSO over the GA is its algorithmic effortlessness. Normally 3 major operators are required by the GA: selection, crossover, and mutation. For each of these operators, there are several implementation choices. For example, you can pick a tournament or a choice that is proportionate. But there is one basic operator in the PSO: the estimation of velocity. The benefit of working with fewer operators is to reduce the equation and remove the process in order to select the best operator for optimization.

Zhang Jian-qi et al [6] ...have proposed the novel method of face recognition focused on supporting vector machine and optimizing particle swarm. Particle swarm optimization Is influenced by individuals ' social behavior such as blocking birds or grouping fish, which in broad search spaces can effectively seek optimal or near-optimal solutions. Therefore, PSO is used to modify SVM parameters in PSO-SVM at the same PSO-SVM, FERET human face database face recognition output is adopted and the proposed PSO-SVM system is compared to SVM, BPNN.

The SVM parameters suggested by Yukun Bao et al [7] are of great importance for SVM's results. This research proposed a memetic algorithm based on PSO and pattern search (PS) to better configure the parameters.

E.G. Karakasis et al [8] … have proposed a novel technique for the revival of images. The key development is the use of moment invariants as descriptors of local object areas taken from the SURF detector. In the present experimental study, three separate configuration models are considered. In the first, invariants are determined for one color affine moment. In the second, the invariants used are determined over the original image's chromites, while in the third, a standardization scheme occurs to expand the range of values of the invariants.

S Sergyan et al [9] …have proposed an image classification method. Using K nearest algorithm we perform the classification. The

November-December 2019
ISSN: 0193-4120 Page No. 4529 - 4535

proposed method provides 87% of accuracy result in 200 images and 4 classes of images.

Jing Yu et al [10] …have proposed an image recovery feature bag-based integration system and weighted clustering of K-means. This propose bug-based and image-based implementation of the features of SIFT, LBP, and HOG. The image-based integration provides the best performance centered on detailed experimental studies on benchmark object retrieval problems when implementing codebook size N1⁄4200 and weight K-means w1⁄40.6 compared to existing versions.

3 Proposed BOVW Classification with PSO

Bag of visual words bag creates the vocabulary for that collection of data. This classification technique gives the best result in accuracy when comparing the previous PSO (partial swarm optimization) algorithm used to get better the operation. It must be said that the swarm optimization algorithms inspired by the PSO algorithm and other design are very appropriate for the discrete engineering and treatment of unstructured high-volume information in Big Data examination [11].

3.1. BOVW Classification

The BOVW Classification method meant for classifies the data from the large data set. Comparing to existing methods Bag of visual classifier provide better accuracy result and classify the data from the huge data set. Second, take the data set from the cloud that the datasets train and check for future use. Second, eliminate selectively the redundancy present in captured images without losing any information which plays a key role in the overall process. Finally, functions are extracted using the extraction method of the SURF.

(The SIFT approach uses cascaded filters to detect scale-invariant characteristic points where the Gaussian difference (GD) is slowly calculated on rescaled images.) Filtering the square picture is much simpler when the integral I use. Fourth, the K-means algorithm in data mining starts with the first group of randomly selected centroids that are used as starting points for each cluster, and then iterative (repetitive) calculations are performed to optimize the centroid positions. Visual device bag is used for objects that are derived from the feature.

3.2 PSO OPTIMIZATION

As a rule, PSO is a more straightforward calculation than GA. It depends on the rule that every arrangement can be deciphered in a swarm as a particle (specialist).Each agent has a vector of position and velocity [11]. Every coordinate of the location represents a value of the parameter. PSO often requires a friction of fitness evaluation which takes the position of the agent and assigns a fitness value to it. The spot with the highest overall fitness value is called the global best (gkr,). The highest fitness value is also monitored by each agent [5]. This value's position is called its personal best (gbest). With a random location and random velocity, each agent is initialized. The velocity is accelerated in each of the n dimensions towards the own personal data based on the following equation:

$$y_n = V_n + C_1 rand() * \left(g_{best,n} - x_n\right) + C_2 rand() * \left(p_{best,n} - x_n\right) \dots\dots\dots (1)$$

Here rand () restores a number somewhere in the range of 0 and 1. The speeding up constants CI and CZ decide the relative "pull" of gbest and Pbest. The higher the steady, the more noteworthy the speeding up toward the position it is increasing. For this case 2.0 is utilized for the two constants. The Positions are changed as pursues based on their movement over a constant time interim (Af), typically set to 1:

$$X_n = X_n + V_n * \Delta t \dots\dots\dots (2)$$

November-December 2019
ISSN: 0193-4120 Page No. 4529 - 4535

Then the fitness was re-evaluated at each position. If any fitness is greater than g_{best}, then g_{est} becomes the new position and the velocities are increased to that level. If the fitness value of the agent is lower than p_{ko},the current position will replace P_{br} and the agent will be accelerated to that position [5].Occasionally one of the parameters considered the provincial best is used. This is the location from a small group of agents with the highest value.

The group size is typically around 15 percent of the population size. Agents from their respective group are accelerated to l. However, its method was not used in this case. An additional factor Vmax is the most extreme speed of a specialist in some random measurement. For this situation; Vmax is set to a large portion of the measurements extend this is done to help deal with the swarm. Another factor known as the inertial weight (w) drops directly over the run term from 0.9 to 0.4. This has the effect of slowing down the same as the particle Running is moving. It makes the particles converge into gk instead of spinning around it. The new position formula is given by the following:

$$V_n = W * V_n + C_1 rand() * \left(g_{best,n} - x_n\right) + C_2 rand() * \left(p_{best,n} - x_n\right) \dots\dots\dots (3)$$

4 RESULTS AND DISCUSSIONS

The experiment of our proposed BOVW with PSO optimization is directed utilizing a dataset are gathered from the cloud. The experimental results are obtained using python installed in Intel core i5 processor.

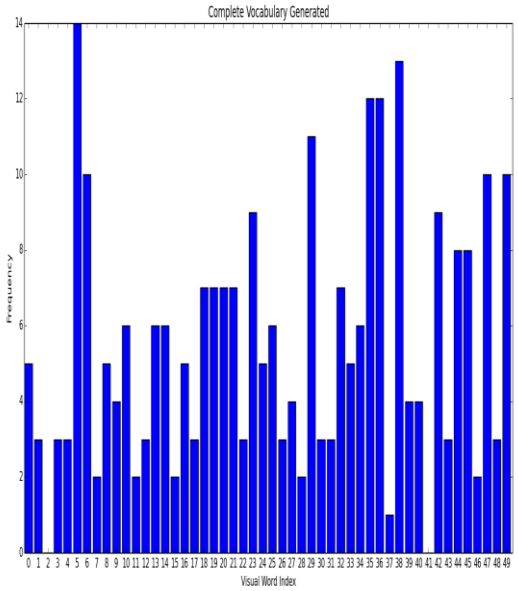

Fig 1: Normalized Vocabulary of BOVW

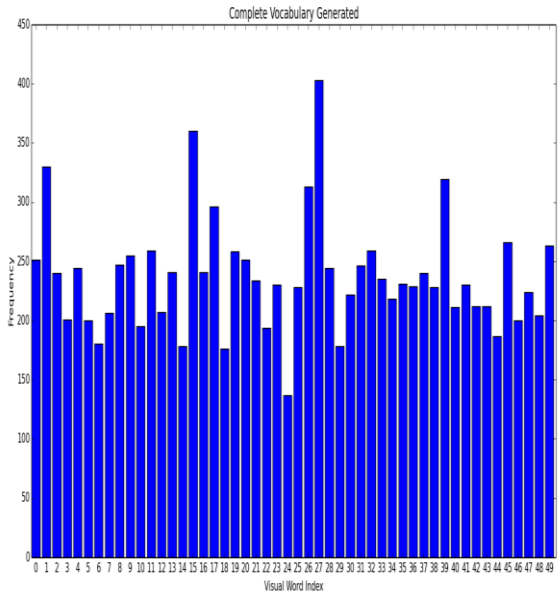

Fig 2: Visual Word Vocabulary

November-December 2019
ISSN: 0193-4120 Page No. 4529 - 4535

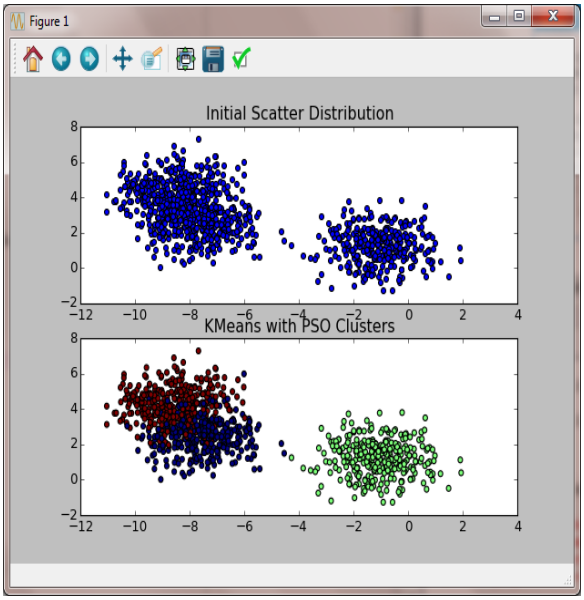

Fig 3: Bag of visual using K-mean with PSO

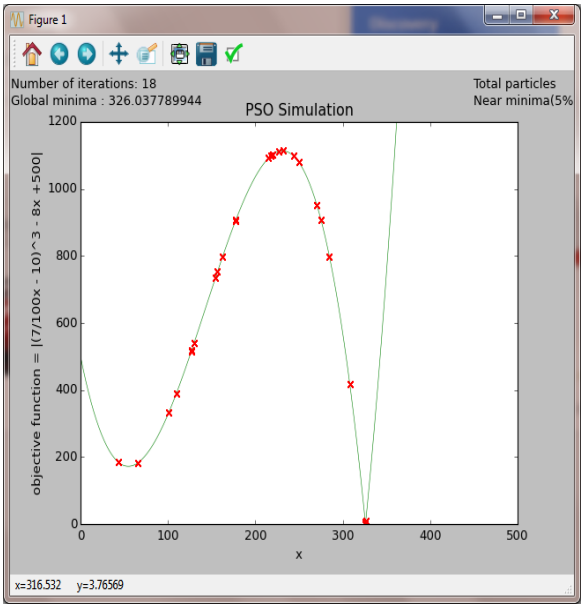

Fig 4: PSO Optimization

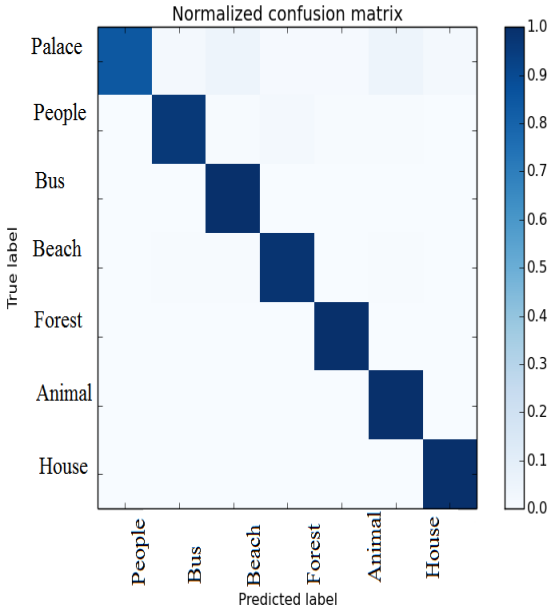

Fig 5: BOVW with PSO optimization

5 COMPARISION

Comparison graph for Existing and a proposed method shows that the proposed method achieves better classification of big data with 98% rather than other algorithms.

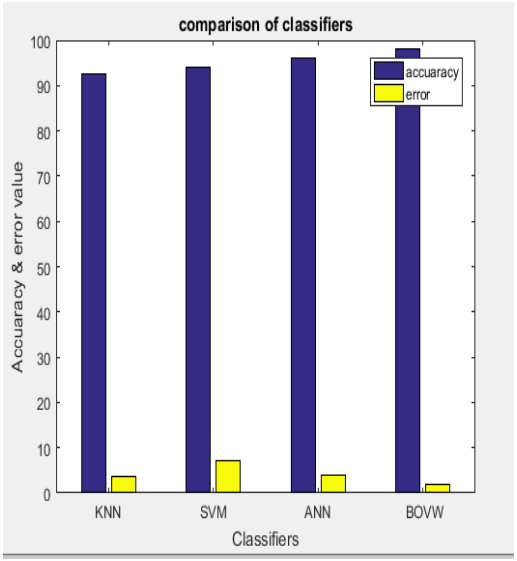

Fig 6: SVM and Proposed BOVW accuracy and error value

November-December 2019
ISSN: 0193-4120 Page No. 4529 - 4535

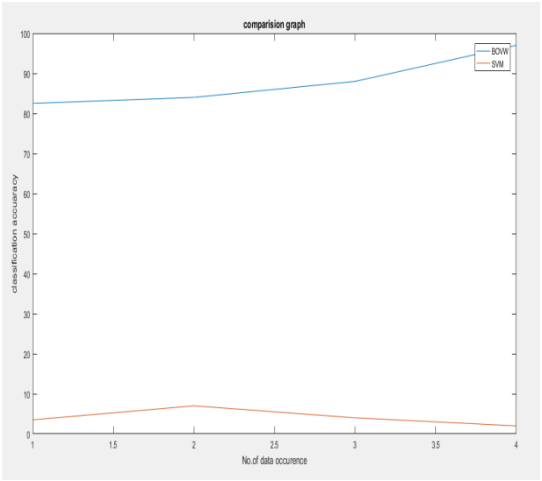

Fig 7: No of data occurrence graph for SVM and BOVW

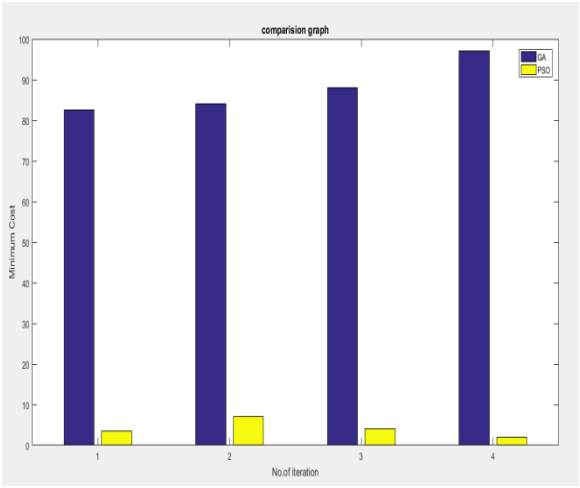

Fig 8: Number of iteration and Minimum Cost

6 CONCLUSION

The effectiveness of the recommended methodologies has been affirmed by results of proposed studies. The BOVW classifiers dependent on the updated PSO algorithm allow highly classified precision data to be categorized. The Key function of the modified version PSO algorithm is use of the particle "regeneration". The BOVW with PSO algorithm is used in the case of various strategies of image classification, decision making and helps to get better accuracy of classification. Due to integration of their performance, the deficiencies of some classifiers are offset by the strengths of other classifiers. The proposed algorithm obtained accuracy 98% and while comparing to existing method it provide best accuracy results.

REFERENCES

1. Vieira, Susana M., Luís F. Mendonça, Goncalo J. Farinha, and João MC Sousa. “Modified binary PSO for selection of features using SVM used for septic patient mortality prediction." *Applied Soft Computing* 13, no. 8 (2013): 3494-3504.
2. Yang, Yi, and Newsam Shawn. "Bag-of-visual-words and spatial extensions for identification of land-use." At the 18th International SIG SPATIAL Conference on Advances in Geographical Information Systems, pp. 270-279. ACM 2010. ACM 2010.
3. Xu, Sheng, Tao Fang, Deren Li, and Shiwei Wang. "Object classification of aerial images with BOVW." *IEEE Geoscience and Remote Sensing Letters* 7, no. 2 (2009): 366-370.
4. Ho, S. L., Shiyou Yang, Guangzheng Ni, Edward WC Lo, and Ho-ching Chris Wong. "A PSO - based method for multiobjective design optimizations." *IEEE Transactions on Magnetics* 41, no. 5 (2005): 1756-1759.
5. Rahmat, Yahya, Samii. "Genetic algorithm (GA) and (PSO) in magnetic engineering. “In 2003 at the 17th International Conference on Electromagnetic Applications and Telecommunications.
6. Wei, Jin, Zhang Jian-Qi, and Zhang Xiang. "Face recognition method based on SVM and PSO." *Expert Systems with Applications* 38, no. 4 (2011): 4390-4393
7. Bao, Yukun, Zhongyi Hu, and Tao Xiong. "A memetic algorithm based on PSO and pattern search for optimizing SVM parameters" *Neurocomputing* 117 (2013): 98-106.
8. Karakasis, Evangelos G., AngelosAmanatiadis, AntoniosGasteratos, and SavvasA.“Image instant invariants as local features for content-based image recovery using the visual-word model." *Pattern Recognition Letters* 55 (2015): 22-27.
9. S Sergyan. "Color The histogram is based on the image classification in CBIR systems." Sixth Symposium International on Applied Machine

Intelligence and Information Technology in 2008, pp. 221-224.
10. Yu Jing, Qin Zengchang, Wan Tao, and Zhang Xi. "Bag-of-function model feature integration analysis for object recovery." Neurocomputing 120 (2013): 355-364.
11. Demidova, Liliya, YuliaSokolova, EvgenyNikulchev. "Classification of large data using modified PSO and SVM classifiers. "AdvancedInformatics and applications (IJACSA) International Journal 7, No. 5 (2016): 294-312.