Novel Approach of Features selection by grey wolf optimization with SVM

Pallavi Dixit[1], Dr. Santosh Kumar[2]
Research Scholar[1], Associate Professor[2]
Maharishi University of information Technology, Lucknow

Abstract: Feature selection for classification in high-dimensional spaces can improve generalization, reduce classifier complexity, and identify important, discriminating feature "markers". For support vector machine (SVM) classification, a widely used technique is grey wolf optimization (GWO). We demonstrate GWO is not consistent with margin maximization, central to the SVM learning approach. We thus propose explicit margin-based feature elimination (MFE) for SVMs and show both improved margin and improved generalization, compared with RFE. Moreover, for the case of a nonlinear kernel, we show RFE assumes the squared weight vector 2-norm is strictly decreasing as features are eliminated. We demonstrate this is not true for the Gaussian kernel and, consequently, RFE may give poor results in this case. We show that MFE for nonlinear kernels gives better margin and generalization. We also present an extension which achieves further margin gains, by optimizing only two degrees of freedom – the hyperplane's intercept and its squared 2-norm – with the weight vector orientation fixed. In experiment use two dataset one is image segmented data and second is iris dataset. In experiment use different iteration of GWO and calculate precision, recall, accuracy.

Keywords - *GWO, feature selection, SVM*

I. INTRODUCTION

There are various issues of classification on a dataset which contains large number of attributes and redundant attributes, oftentimes containing relevant or irrelevant. Redundant and irrelevant attributes are not profitable for classification and they also reduce the precision due to its large search space. Here the main goal of the attributes reduction is to select subset of relevant attributes from a huge number of available attributes to acquired best classification precession using all attributes. Reduction can be reduced by eliminating redundant and irrelevant attributes. It is quite difficult task to reduce attributes. There are various search methods used to reduce attribute. In optimization, inspiration from nature is a main class optimization algorithm. The majority of the nature-inspired algorithms are based on some successful characteristics of the biological system. Therefore, the largest fractions of the nature inspired algorithms are biology-inspired or bio-inspired [1]. The main inspiration of proposed algorithm is heuristic optimization algorithm which imitates the way wolves by ignoring enemies and then a search on food and survives. WSA is able to rapid search of feature space for the reduction of optimal and near-optimal feature subsets by using fitness function. This function helps in reduction of feature size and classifies the accuracy.

1.1 Feature Selection

Feature selection is a process where it selects the relevant data or the best data from the large number of data. Feature selection is used to extract the data from large data instantly. Need of feature selection [2]:

1) Simplification of models to make model easier to interpret by users.
2) Due to short training time
3) To ignore the curse of dimensionality.
4) Enhancement of generalization by reducing over fitting.

There are two methods of feature selection:

1) Forward feature selection: In this feature selection process starts from empty features set having no features and used to search for feature subset(s) with the help of one features by selecting that feature which has the highest classification performances. Then after algorithm selects another features from the candidate features to add to S.

2) Backward feature selection: In this feature selection starts with features having available at that time, then after candidate features are simultaneously removed from the feature subset unless the further removal of any features does not increase the classification performance.

1.2 Feature Extraction

Reduction of attribution can be done by extracting data from the original one. Feature extraction is one of the methods used for extraction purpose. It transfers the data from high level to low level. There is linear and non-linear extraction techniques is exists.

There are techniques used for the reduction of attributes:

1) **Filter approach:** Filter approaches are not dependent of a learning technique and they are debated to be computationally inexpensive and more public than wrappers. It is used as a proxy measure rather than error rate to score features. They are less computational as compare to wrapper approach.[1]
   - Information Gain
   - Chi-Square

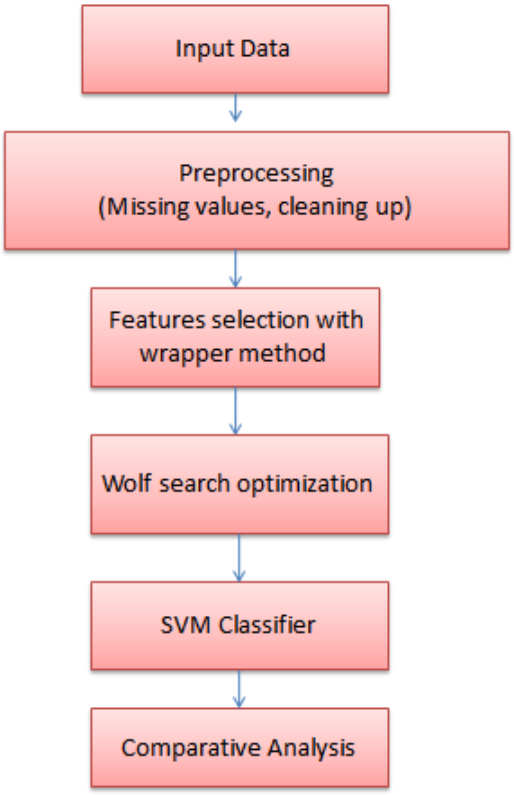
2) Wrapper approach: It contains a learning technique as part of the evaluation function. Thus, wrappers can usually obtain best results from filter approaches. It is a predictive model to score subsets. Every individual subset is used to train a model which is then tested on a hold-out-test. They are very computational intensive but usually provide the best performing features set for particular type of model. [1]

II. LITERATURE REVIEW

Artificial bee colony algorithm is used for feature selection and it is integrated with genetic operators and sorting procedure. In this ABC algorithm is used in two ways that are ABC with continuous and ABC with binary representation. The performance of the proposed system is compared with stepwise backward selection approach **[1]**.Feature selection method is used to identify the spam in Facebook social media. In this method, spam is detected by reviewing the comments and post. It identifies the spam content and protects to publish that content on social media. This method provides the secure environment for users on social media **[2]**. Feature selection is also done by using rule-based knowledge bases. It is also a part of mining knowledge from the bases. It improves the quality of knowledge by selecting the essential features according to the user need. In this clusters are made on the basis of rules and their selection of best data **[3]**.Particle swarm based feature selection approach is proposed by the author in this article. In this method, search is improved by using extra learning approach with subset cardinality. Learning on cardinality helps in position updates of swarms. Performance evaluation of the proposed method is done with the help of naïve Bayes and K-nearest neighbor algorithms **[4]**.In this paper, the author proposed deep feature extraction and selection method which combines the weighted feature and stack feature selection together. The stack auto encoding method is helpful in the representation of information from the inputs. These features are combined with modified weighted features selection which is inspired by shallow structured machine learner. Set of features is used to reduce the bias of proposed learner model and also computational complexity **[5**].Genetic programming approach is proposed by the author in this paper for feature selection. This technique performs well on balanced and unbalanced data. It combines the feature sets which are selected by the distinct feature selection process. The proposed method is evaluated on the biological and textual datasets. It also increases the efficiency of learning process and enhances the effectiveness of the system **[6]**. In this paper the author proposed binary optimization approach for feature selection. This method is based on the opinion formation process. This process is simulated by artificial agent by encode the subset of selected features. The problem of feature selection is solved by the proposed method. Agents are interacting by using interaction network for share their opinions to find the better solution of the problem **[7]**. In this paper, the modified cuckoo algorithm is proposed for feature selection from the rough sets of the high dimensional data. This algorithm works on the behavior of cuckoo. It uses rough set theory to make fitness function which contains useful features only. In this work results are also compared with existing algorithms and evaluate the performance of the system. This method also improves the classification process **[8]**. Support Vector machine is an effective learning algorithm which is used in many applications for learning. The performance of the SVM is depends upon its kernel type. In this work Grasshopper optimization Algorithm is proposed which is inspired from the biological behavior of grasshoppers. This algorithm used to optimize the parameters of SVM. It selects the best features subset and enhances the accuracy of the proposed approach. The results of the approach is compared with grid search approach and found that proposed approach performs well **[9]**.In this paper the feature selection process is done on the image by using the Genetic algorithm and Tabu search. In this work premature convergence of the genetic algorithm is used by using Tabu search. Convergence level is defined by using prematurity index defined during search. The proposed feature selection method provides effective classification of features **[10]**.

III. PROPOSED METHODOLOGY

Step1 : Firstly input the data.
Step2 : In this step preprocessing of data is done. In this process missing values, noisy data is removed. It is also called as cleaning of data.
Step3 : Features are selected by using wrapper methods in this step.
Step4 : Apply wolf search optimization algorithm for optimized result.
Step5 : Apply SVM classifier and the compare the results.



IV RESULTS

In this section, results on the segmented data set and IRIS dataset is shown in table and also with the help of graphical method. These graphs explain the changes in output in various classifiers.

Table 1.1 Output on Segmented dataset

Classifiers	Precision	Recall	Accuracy
SVM	89.23	86.23	88.14
SVM+GWO	92.23	90.13	93.33
Random Forest	87.23	88	91
Random Forest+ GWO	91.23	90	1.55
Max Entropy	90.20	88.23	87.13
Max Entropy + GWO	91.16	90.13	90.15
Naïve Bayes	89.13	86.13	89.13
Naïve Bayes+ GWO	90.45	87.23	91.23

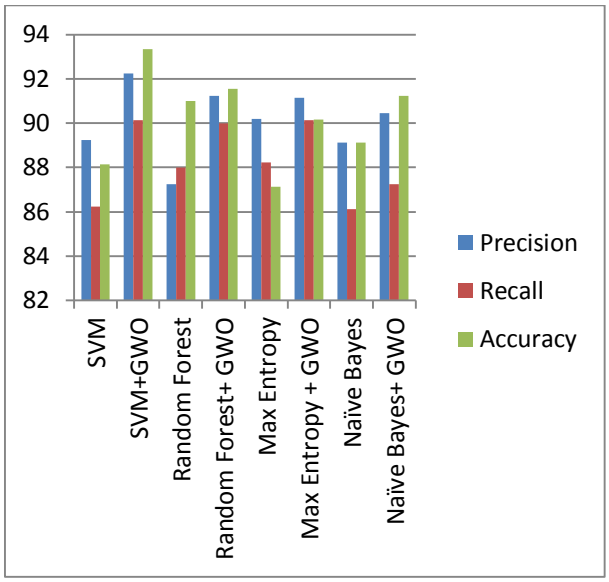


Figure 1.1 Comparison of Precision, Recall and Accuracy of classifiers in segmented dataset

Figure 1.1 shows a combined graph of Precision, Recall and Accuracy of classifiers. The overall better result is given by SVM+GWO classifier which shows the highest value among all classifiers results.

Table 5.2 Results on IRIS Dataset.

Classifiers	Precision	Recall	Accuracy
SVM	88.13	90.13	91.23
SVM+GWO	90.56	92.23	93.33
Random Forest	89.23	90.13	92.13
Random Forest+ GWO	90.13	91.23	92.56
Max Entropy	90	91	91.56
Max Entropy + GWO	92.23	93.33	92.99
Naïve Bayes	88.16	87.34	89.23
Naïve Bayes+ GWO	89.13	89	90.23

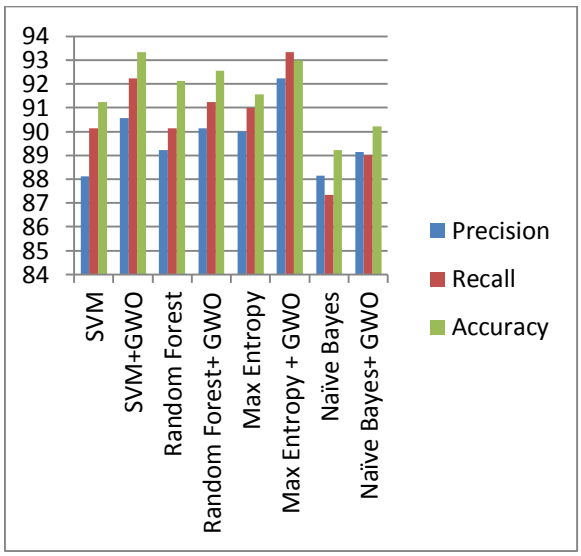


Figure 1.2 Precision, Recall, and Accuracy of classifiers in IRIS dataset

Figure 1.2 shows a combined graph of Precision, Recall and Accuracy of classifiers. The overall better result is given by Max Entropy +GWO classifier which shows the highest value among all classifiers results.

V. CONCLUSION

Reaching the final step of this research many lessons were learned not only regarding to the security point of view but also from machine learning topics. Initially, it was seen that there are absence of "goodware" and malware datasets that researches can use to compare results. However, great tools as Virus total can be used to download many behavioral and metadata information. Regarding to the objectives of this thesis, feature selection showed that it is possible to use few features to reach high level of accuracy for malware detection, which could implies to reduce resources in programs that look for detecting malware, and warn software developers to study further how these important features could be related to security breaches. Thus, the best classification accuracy can be gotten using the 9 features ranked by Random forest by decrease on accuracy. In addition, SVM algorithm showed a great performance compared to RF and NN; however combining algorithms can lead to better results at it happened in our case

VI. REFERENCES

[1]. Hancer, Emrah, et al. "Pareto front feature selection based on artificial bee colony optimization." *Information Sciences* 422 (2018): 462-479.

[2]. Sohrabi, Mohammad Karim, and Firoozeh Karimi. "A Feature Selection Approach to Detect Spam in the Facebook Social Network." *Arabian Journal for Science and Engineering* 43.2 (2018): 949-958.

[3]. Nowak-Brzezińska, Agnieszka. "Feature Selection Approach for Rule Based Knowledge Bases." *Advances in Feature Selection for Data and Pattern Recognition*. Springer, Cham, 2018. 163-182.

[4]. Hafiz, Faizal, et al. "A two-dimensional (2-D) learning framework for Particle Swarm based feature selection." *Pattern Recognition* 76 (2018): 416-433.

[5]. Aminanto, Muhamad Erza, et al. "Deep Abstraction and Weighted Feature Selection for Wi-Fi Impersonation Detection." *IEEE Transactions on Information Forensics and Security* 13.3 (2018): 621-636.

[6]. Viegas, Felipe, et al. "A Genetic Programming approach for feature selection in highly dimensional skewed data." *Neurocomputing* 273 (2018): 554-569.

[7]. Hamedmoghadam, Homayoun, Mahdi Jalili, and Xinghuo Yu. "An opinion formation based binary optimization approach for feature selection." *Physica A: Statistical Mechanics and its Applications* 491 (2018): 142-152.

[8]. El Aziz, Mohamed Abd, and Aboul Ella Hassanien. "Modified cuckoo search algorithm with rough sets for feature selection." *Neural Computing and Applications* 29.4 (2018): 925-934.

[9]. Aljarah, Ibrahim, et al. "Simultaneous Feature Selection and Support Vector Machine Optimization Using the Grasshopper Optimization Algorithm." *Cognitive Computation* (2018): 1-18.

[10]. Shi, Lei, et al. "Feature Selection for Object-Based Classification of High-Resolution Remote Sensing Images Based on the Combination of a Genetic Algorithm and Tabu Search." *Computational Intelligence and Neuroscience* 2018 (2018).