Feature Selection using MultiObjective Grey Wolf Optimization Algorithm.

Dr. Deepak Gupta, Nimish Verma, Mayank Sehgal, and Nitesh

Maharaja Agrasen Institute of Technology,Delhi

Abstract. Multi Objective Grey wolf Optimization is one a meta-heuristic technique. The MOGWO has recently gained a huge research interest from numerous domains due to its impressive characteristics over other meta-heuristics optimization techniques: it has less parameters, derivation information is not required in the initial stage, scalable, flexible, easy to use. In this paper MOGWO, which is based on the leadership hunting technique of grey wolves is used for feature selection. The traditional GWO is useful for single objective optimization problems. Since, feature extraction is a multi-objective problem; this paper utilizes multi-objective GWO algorithm. In this paper, MOGWO is applied to 6 different datasets to understand its application in diverse set of problems. At first, MOGWO is used to obtain feature subsets from different datasets. Then machine learning models like KNN, random forest and logistic regression are used to obtain the accuracy results and comparison of the results is performed. The outputs from the 6 different datasets using MOGWO along with the machine learning models have been reviewed and summarized. The paper is concluded by mentioning the summary conclusion of MOGWO.

Keywords: Grey Wolf Optimization · Feature Selection · Evolutionary Algorithm · Machine Learning · Multi-objectivity

1 Introduction

Different high-dimensional machine learning modeling tasks often implement the data which require removal of irrelevant inputs[1]. The computation power of this high dimensional data is going to be large[2]. Thus, by reducing the number of features, the problem of dimensionality reduction is solved [2]. Compactness within the selected feature subset is aimed by having low redundancy[3]. Feature Selection using multi-objective evolutionary algorithms (MOEA) (2000-2014) have been discussed in [4] and [5]. Feature selection[6] in classification helps in reducing computational cost, over-fitting, comprehensibility and improving classification accuracy [7, 8]. To extract sub-optimal feature subsets, numerous meta-heuristics [9] have been used with great global search ability. Although, numerous metaheuristics have been used for extraction of features, but the problems encountered are high computational cost and stagnation in local optima. In 2014, S. Mirjalilia et al. [10] proposed a meta-heuristic algorithm, Grey Wolf

Optimizer derived from the hunting and the leadership hierarchy of grey wolves in which there exists a smooth transition between exploration and exploitation due to the adaptive values of GWO parameters.MOGWO was proposed which incorporated Pareto archive constituting non-dominated solutions.

Specific tools are required to solve different challenges in real world engineering problems. Multi-objectivity is one of the challenging characteristic of real world problems. When there are more than one objective to be optimized then it is called a multi-objective problem. In the initial years, numerous mono-objective optimizers were converted to multi-objective ones.

In our paper, inspired from the above ideas, we have come up with an improved MOGWO algorithm for feature selection. We have applied this algorithm with other machine learning models on 6 different datasets of different fields and obtained an average accuracy of 81.934%, 79.868% and 67.862% with KNN, random forest and logistic regression respectively, against 75.194%, 74.25%and 69.886% without feature selection. The paper highlights how multiobjective grey wolf optimisation (MOGWO) perform feature selection among the various relevant and irrelevant features of the problem set. Then the quality of the selected features have been evaluated using 3 ML models :
(i) k Nearest Neighbors(kNN)
(ii) Random Forest and
(iii) Logistic Regression.

In the following Section 2, the implementation of MOGWO for feature selection, introduction to the 6 datasets of different fields and the machine learning models used have been presented; Section 3 discusses comparison of accuracy results followed by the conclusion.

2 Methodology

In this section, the implementation of the algorithm, the datasets used and the machine learning models used have been discussed.

2.1 MOGWO and its implementation

The MOGWO Algorithm has been applied on the 6 different Datasets to perform feature selection among the various relevant and irrelevant features of the problem set to filter out only the relevant features. The algorithm of MOGWO is described below:
Input: Datasets with multiple features/attributes that require normalization and pre-processing before applying the algorithm.
Output: Subset of relevant features/attributes.

Load data CSV file into variable df
Remove any pattern in the data by shuffling
Normalize all attributes
Divide train and test datasets in 7:3 ratio

Initialize A, a, C and Xi X_a=SelectLeader(archive)
X_b= SelectLeader(archive)
X_d= SelectLeader(archive)
To the archive add back alpha and beta
TTT=1;
while (t < Max number of iterations)
for each agent: Update the position of agent by equations
end for
Update a, A, and C
Calculate the objective values of all search agents and find the non-dominated solutions
With respect to the obtained non-dominated solutions update the archive
X_a=SelectLeader(archive)
X_b= SelectLeader(archive)
X_d= SelectLeader(archive)
To the archive add back alpha and beta
TTT=TTT+1
end while
return archive

After running this algorithm over all the features of the data set, the subset of relevant features is obtained from the archive as the multi-objective solutions. Finally, the data set is trained with the machine learning models and accuracies with all the three models are computed before and after the feature selection is performed. This is done for all the six datasets.

Experimental Setup CPU configuration used to test the algorithm is Intel Core TM i3-5005U CPU @ 2.00GHz 4 with 8 Gb of memory on Ubuntu 16.04 LTS platform. The algorithm is tested and implemented using Python 3.6.

Input parameters Table 1 comprises of all the parameters initialized for MOGWO.

Table 1. Input parameters

Parameter	Value
No. of Grey Wolves	80
No. of Iterations	20
Neural Network iterations	50
Size of archive	50
Upper Bound	1
Lower Bound	0
Grid Inflation Parameter	0.1
Pressure of selecting Leader	4
Extra members to be removed	2

2.2 Datasets Description

MOGWO is applied to 6 different datasets for feature selection. Following are the datasets discussed in detail.

2.2.1 Glass Identification Dataset This Dataset includes 10 features and 214 rows(data points). The goal is to determine if Glass is a 'float' type or not. To achieve the purpose the nearest neighbour algorithm, discriminant analysis and an algorithm (BEAGEL) is compared.

2.2.2 Wisconsin Diagnosis Breast Cancer (WDBC)
This dataset includes 32 features and 569 instances. The features of this daatse shows the properties of cell nuclei which is present in image of FNA(fine needle aspirate) of a breast mass.

2.2.3 Parkinsons Disease Detection Dataset (PDD)
This dataset includes 23 features and 197 instances. The goal is to find people suffering from Parkinson's disease. Those who are not suffering are labelled as healthy people. This is done on the basis of range of bio-medical voice measurements. Those measurements are taken from 31 people ,of which 23 have PD.

2.2.4 Statlog(heart)
This dataset includes 13 features and 270 instances. The goal is to predict if heart disease is present or not. Presence of disease is labelled with '2' and absence of disease is labelled with '1'. Features represent age ,sex,types of chest pain ,BP(blood pressure) etc.

2.2.5 Lung Cancer:
This dataset includes 56 features and 32 instances(with some missing values). The goal is to classify 3 kind of pathological lung cancer.

2.2.6 Letter Recognition
This is a multivariate dataset with 16 integer-valued features and 20,000 instances. The aim of this dataset is to identify the alphabets in the English language from the BW images. The images have 20 different fonts.

2.3 Machine Learning Models

In our implementation, 3 classic Machine Learning models are used. Namely - KNN, Random Forest and Logistic Regression. KNN is a supervised classification algorithm. The distance between two records is taken in an Euclidean space. The next model we have used is Random Forest, which is another supervised algorithm. It is based on classification using a multitude of decision trees used during

Table 2. Configuration of the ML model

Model	Parameters
K NN	No. of Neighbours= 5
Random Forest	Number of estimators = 10, min samples split = 2
Logistic Regression	Penalty = 12, Intercept scaling = 1

the training time. It is an advancement of decision trees where the overfitting occurs during training time. The last model we used is the Logistic Regression model which is a categorical technique based on Regression. It is a widely used statistical model that is used to describe a binary dependant variable using a logistic function. Table 2 lists the parameters used in these models.

3 Results and Discussion

The accuracy results by K Nearest Neighbour, Random Forest Logistic Regression before after feature selection of the datasets has been compared in this section.

'A' denotes after feature selection 'B' denotes before feature selection.

Table 3. Comparison of Accuracy

Dataset	KNN (B)	KNN (A)	RF (B)	RF (A)	LR (B)	LR (A)	Features (A/B)
PDD	0.7966	0.8305	0.9491	0.9322	0.8135	0.7796	11/22
Glass	0.7384	0.7384	0.723	0.723	0.6461	0.5534	5/9
Letter Recognition	0.947	0.947	0.938	0.9358	0.7076	0.77	15/16
Statlog	0.7777	0.7901	0.8024	0.8024	0.8271	0.7901	5/13
LungCancer	0.5	0.8	0.3	0.6	0.5	0.5	17/56
WDBC	0.75194	0.8212	0.7425	0.79868	0.69886	0.67862	14/30

MOGWO along with KNN, random forest and logistic regression were applied on same datasets namely PDD, Glass Identification, Letter Recognition, Statlog, Lung Cancer, WDBC and their accuracy was measured and analysed, as shown in Fig. 2. KNN, random forest and logistic regression showed an average accuracy of 81.934%, 79.868% and 67.862% respectively. The performance comparison depicts that feature selection using MOGWO along with KNN outperforms Random Forest and Logistic Regression models.

4 Conclusion

In this work, a bio-inspired multi-objective optimization algorithm for feature selection named multi-objective grey wolf optimisation algorithm along with three machine learning models namely KNN, random forest and logistic regression has

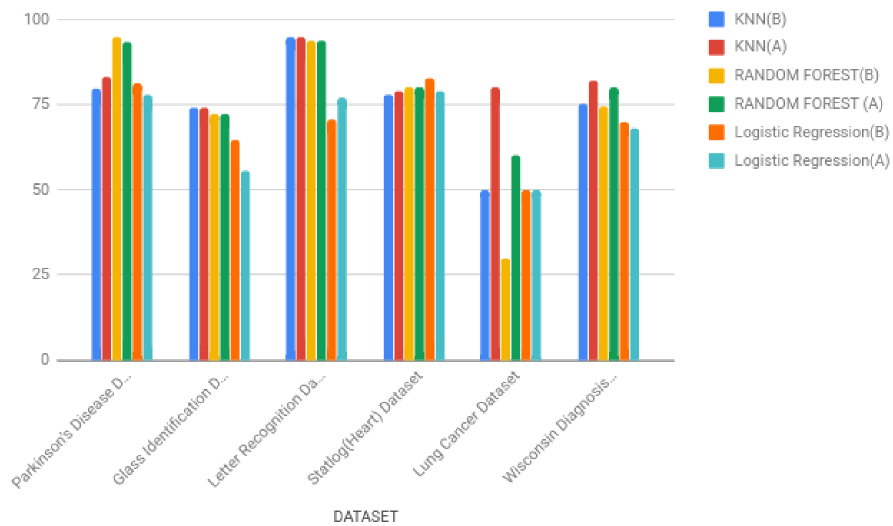

Fig. 1. Overall Comparison of Accuracy

been proposed to get a subset of features by rejecting all the irrelevant features of a dataset and measure comparative accuracy results without compromising the performances. The algorithm along with KNN has selected reduced set of relevant features with highest accuracy of 81.934% in contrast to the accuracy of 79.868% and 67.862% of other machine learning models namely- random forest and logistic regression. The results imply that MOGWO algorithm along with KNN outperforms the other two models in the case of datasets we have considered. Research practitioners can implement this algorithm along with KNN to perform feature selection among the various relevant and irrelevant features of the problem set to filter out only the relevant features.

References

1. Martn-Smith, Pedro, et al. "A Label-Aided Filter Method for
2. Multi-objective Feature Selection in EEG Classification for BCI." International Work-Conference on Artificial Neural Networks. Springer International Publishing, 2015. DOI: 10.1007/978-3-319-19258-1_12 Das, Ayan, and Swagatam Das. "Feature Weighting and Selection with a Pareto-optimal Trade-off between Relevancy and Redundancy." Pattern Recognition Letters (2017). http://dx.doi.org/10.1016/j.patrec.2017.01.004
3. De la Hoz, Emiro, et al. "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps." Knowledge-Based Systems 71 (2014): 322-338. http://dx.doi.org/10.1016/j.knosys.2014.08.013

4. Mukhopadhyay, Anirban, et al. "A survey of multiobjective evolutionary algorithms for data mining: Part I." IEEE Transactions on Evolutionary Computation 18.1 (2014): 4-19.
5. B. Xue, M. Zhang, W. N. Browne and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," in IEEE Transactions on Evolutionary Computation, vol. 20, no. 4, pp. 606-626, Aug.2016. Doi: 10.1109/TEVC.2015.2504420
6. A. Jain, D. Zongker, Feature selection: Evaluation, application, and small sample performance, IEEE Trans. Pattern Anal. Mach. Intell., 19(1997), 153158. doi: 10.1109/34.574797
7. I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3(2003), 11571182.
8. H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. On Knowl. and Data Eng. 17, 4 (April 2005), 491-502. Doi: 10.1109/TKDE.2005.66 http://dx.doi.org/10.1109/TKDE.2005.66.
9. C. Blum, A. Roli. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. ACM Computing Surveys. 35(2003). pp. 268308.
10. S. Mirjalili, S. M. Mirjalili, A. Lewis, Grey Wolf Optimizer, Adv. Eng. Softw. 69 (March 2014), 46-61. Doi: 10.1016/j.advengsoft.2013.12.007 http://dx.doi.org/10.1016/j.advengsoft.2013.12.007
11. S. Mirjalili, S.Z. Mohd Hashim, G. Taherzadeh, S.Z. Mirjalili, S. Salehi. A Study of Different Transfer Functions for Binary Version of Particle Swarm Optimization. Int'l Conf. Genetic and Evolutionary Methods (2011), 169-174.
12. K. Chandrasekaran, S. P. Simon. Multi-objective unit commitment problem using Cuckoo search Lagrangian method. International Journal of Engineering, Science and Technology 4, 2 (2012), 89-105.
13. K. Chandrasekaran, S. Hemamalini, S. P. Simona, N. P. Padhy. Thermal unit commitment using binary/real coded artificial bee colony algorithm. Electric Power Systems Research 84(2012), 109119.
14. B. Crawford, R. Soto, M. O. Suarez, F. Paredes, F. Johnson. Binary Firefly algorithm for the set covering problem. 2014 9th Iberian Conference on Information Systems and Technologies (CISTI June 2014), 1- 5. Doi: 10.1109/CISTI.2014.6877090
15. S. Mirjalili, S. Saremi, S. M. Mirjalili, L. Coelho, Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization, Expert Systems with Applications, in press, DOI: http://dx.doi.org/10.1016/j.eswa.2015.10.039