Optimizing Machine Learning Models using Multiobjective Grasshopper Optimization Algorithm

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

Maharaja Agrasen Institute of Technology,Delhi

Abstract. Multiobjective Grasshopper Optimization Algorithm is a recent meta-heuristic swarm intelligence algorithm developed by Mirjalili at. El. It is inspired from the movement of grasshopper swarms in nature. It can be applied in numerous domains due to its impressive characteristics like easy to use, scalable, flexible and better performance than classic methods in real problems. In the paper, MOGOA, which is a population based method has been used for feature selection. First, MOGOA has been used for feature extraction from six different datasets to form feature subsets from each dataset. Then three machine learning models - KNN, Logistic Regression and Random Forest have been implemented to predict the results before and after feature selection. Finally accuracy of results is obtained and comparison of results is performed. In the first section of this paper, theoretical foundation of multi-objective problems, feature selection and evolutionary algorithms is introduced. In second section, MOGOA ,its implementation and the three machine learning models are explained. Finally the accuracy results of the 6 different datasets implementing MOGOA along with the machine learning models has been reviewed and summarized. The paper is ended by mentioning the conclusion of the MOGOA application in the feature selection domain.

1 Introduction

Different machine learning models often implement the dataset with irrelevant input features that are of no use in the final prediction of result [1]. These type of datasets with a plethora of features increase the training and testing time of the ML algorithm [2].To solve this dimensional reduction, feature selection is performed to obtain a minimal subset of relevant features. Compactness is achieved by having low redundancy within the selected features of the data [3]. Recent reviews (from 2000 to 2014) on feature selection using multi-objective evolutionary algorithm (MOEA) has been presented in references [4] and [5]. For classification problems feature selection [6] helps in minimizing computational cost, over-fitting, comprehensibility and improving classification accuracy [7,8]. Feature selection is a NP-hard problem, for k-dimensional feature vector it finds $2\hat{k}$ subsets of features. Therefore, to extract sub-optimal feature subsets, numerous meta-heuristic algorithms [9] have been used with great global search ability. The common problems encountered in these meta-heuristics are high computational cost and stagnation in local optima. In 2017, Seyedali Mirjalili et al. [10] proposed Grasshopper Optimization Algorithm (GOA), a swarm intelligence optimization algorithm mimicking the interaction of grasshoppers. This algorithm solves the problem of stagnation in local optima. But in order to solve the feature selection problem task GOA is needed to be modified. With the help of experiments performed with multiple transfer function, S. Mirjalili [11] concluded that the performance of the v-shape family of transfer functions is better than the s-shape family of transfer functions. It has also helped in developing binary variants to other metaheuristics such as artificial bee cuckoo search algorithm [12], colony algorithm [13], firefly algorithm [14], etc. To solve the problem of feature selection, we need a multi-objective framework. In order to have a pareto front of non-dominated solutions, multi-objective GOA [15] had been proposed.

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, multiple mono-objective optimization algorithms were converted to multi-objective techniques.

An evolutionary algorithm comes under evolutionary computation and focuses on random sampling. In this all the possible solutions behave as the individuals in a population, and the quality of the solutions is determined by the fitness function. The population eventually evolve by repeating the above operators multiple time and the global best solution is obtained. Inspired from the above ideas we have come up with an application of MOGOA in the domain of feature selection. We have applied this algorithm along with other machine learning models on 5 different datasets of different fields and obtained an average accuracy of 72.072%, 77.856% and 77.908% with KNN, logistic regression and random forest respectively, against 73.64%, 75.63% and 70.796% without feature selection. The paper highlights how multiobjective grasshopper optimisation algorithm (MOGOA) is used as a population based technique to perform feature selection among the various relevant and irrelevant features of the problem set to filter out only the relevant features. Then three different classifiers have been used to evaluate the quality of selected features: (i) k-nearest neighbors(k-NN), (ii) Random Forest and (iii) Logistic Regression. In the following Section 2, introduction to MOGOA, its implementation for the purpose of feature selection, datasets description and the machine learning models have been presented; Section 3 discusses the results; Section 4 throws light on the accuracy comparison followed by the conclusion.

2 Methodology

2.1 MOGOA

In 2017, S. Mirjalili developed a population based meta-heuristic algorithm called Multi-objective grasshopper optimization algorithm(MOGOA). In this, the grasshoppers denote the solutions in the search space and their position is determined by three forces mainly the social interaction force among the grasshoppers, the gravitational force and the wind advection on the grasshoppers. Based on the grasshoppers position the position of grasshopper is shown by equation (1).

$$X_i^d = \left(\sum_{\substack{j=1 \\ j \neq i}}^{N} c \frac{ub_d - lb_d}{2} s\left(\left|x_j^d - x_i^d\right|\right) \frac{x_j - x_i}{d_{ij}} \right) + \overline{T}_d$$

Fig. 1. Position updating equation

Steps for finding the best global solution are as follows:

1. The algorithm is initiated by initialising the archive and setting the initial values of the parameters like max and min decreasing coefficient parameter (cmax, cmin), attraction intensity (l), attractive length scale (f) and max number of iterations (maxitr).
2. The grasshoppers are initialized randomly and the global best solution is assigned after evaluating each solution in the population with the help of objective function.
3. For every iteration, the decreasing coefficient parameter (c) is updated in order to shrink the comfort, repulsion and attraction zones.
4. The solutions are updated depending on the distance between it and the other solutions, the decreasing coefficient parameter, and the global best solution in the population (Td) as shown in fig 1.
5. The previous steps are repeated and the best global solution obtained for every iteration. In a mono-objective problem, the target can be easily chosen as the best global solution obtained so for but in multi-objective solutions the target is picked from Pareto optimal solutions which are added to the archive. The quality of distribution of solution in the archive depends upon the target value chosen. The target is chosen from the archive based on the roulette wheel and the probability of choosing the archive which is given by the inverse of the no. of solutions near the solution under consideration. This improves the distribution of solutions in the search space.

2.2 Implementation of the algorithm

In this section, the experimental setup, input parameters and the datasets description have been presented.

Table 1. Input parameters

Parameter	Value
No. of Grasshoppers	80
No. of Iterations	20
Neural Network iterations	50
Size of archive	50
Upper Bound	1
Lower Bound	0
cMin	0.00004
cMax	1

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 and its related libraries such as SkLearn, Pandas, Numpy and Matplotlib.

Input parameters Table 1 comprises of all the parameters initialized at the beginning of the MOGOA algorithm.

Datasets MOGOA 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 daatset 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.

2.3 IMPLEMENTATION

The Multi-objective Grasshopper Optimisation Algorithm (MOGOA) 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 input to the algorithm is the datasets with multiple features/attributes that require normalization and preprocessing before applying to the algorithm. Using MOGOA, a subset of feature was obtained for each dataset. The corresponding subset of the dataset was then trained and tested on 3 different machine learning models- Logistic Regression, Random Forest and K-Nearest Neighbors. In the next step, the accuracy was compared for an individual dataset among the 3 different machine learning models before and after the feature selection is performed. The ratio of training to testing data considered in our implementation is 7:3 respectively.

2.4 Machine Learning Models

Table 2. Configuration of the algorithms adopted

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

In our implementation, 3 classic Machine Learning models are used. Namely - KNN, Random Forest and Logistic Regression. KNN is a supervised learning algorithm that works on the concept of maximum number of nearest neighbours belonging to the same class. The distance between two records is taken in an Euclidean space. The second model is Random Forest, which is also a supervised algorithm. It is based on classification using a multitude of decision trees used during 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 shows the initial parameters used in these models.

3 Results and Discussion

This section focuses on the results generated when the Multiobjective Grasshopper Optimisation Algorithm along with different machine learning models have been applied to 5 different datasets. The accuracy(True Positive + True Negative/Total number of testing samples) results by KNN, Random Forest and Logistic Regression before and after feature selection of the dataset has been compared.

'A' denotes the accuracy after feature selection and 'B' denotes accuracy before feature selection for each model

Table 3. Comparison of Accuracy

Dataset	KNN (A)	KNN (B)	LR (A)	LR (B)	RF (A)	RF (B)	Features (A/B)
PDD	0.7796	0.7966	0.9152	0.8983	0.9322	0.9491	13/22
Glass	0.7384	0.7384	0.6461	0.6461	0.67	0.5534	5/9
Statlog	0.7572	0.7612	0.7960	0.7943	0.7434	0.6921	11/13
LungCancer	0.4	0.5	0.6	0.5	0.6	0.4	28/56
WDBC	0.9356	0.9473	0.9415	0.9473	0.9532	0.9473	14/30

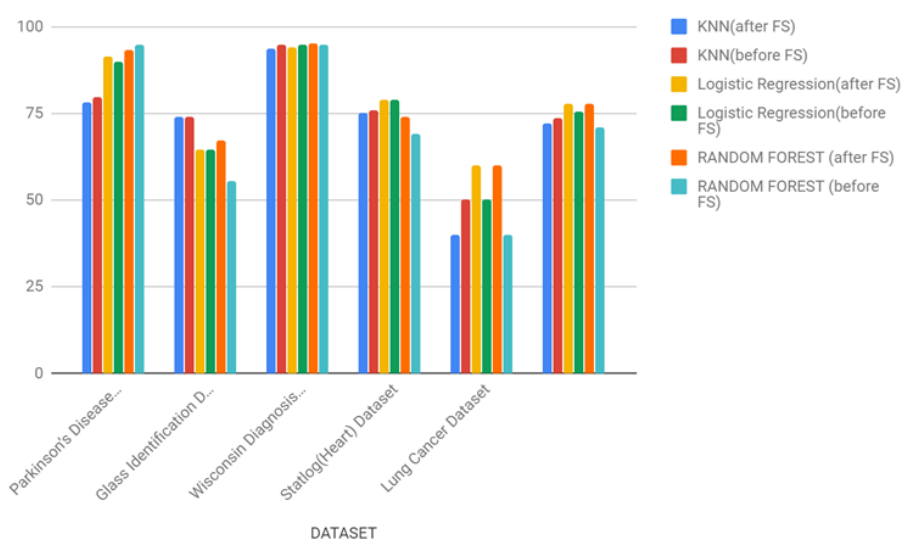

Fig. 2. Comparison of Accuracy on 3 different models for all datasets

3.1 Comparison

MOGOA 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, logistic regression and random forest showed an average accuracy of 72.072%, 77.856% and 77.908% respectively. The performance comparison depicts that feature selection using MOGWO along with Random Forest outperforms KNN and Logistic Regression models.

4 Conclusion

In this work, a bio-inspired multi-objective optimization algorithm for feature selection named multi-objective grasshopper optimisation algorithm along with three machine learning models namely KNN, random forest and logistic regression has 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 of the model. The algorithm and datasets description is given in Section 2. The algorithm along with Random Forest has selected reduced set of relevant features with higher accuracy of 77.9% as compared to the average accuracy of 72% and 77.8% of other machine learning models such as KNN and logistic regression respectively. The results imply that the MOGOA algorithm along with Random Forest outperforms the other two models. Research practitioners can implement this algorithm along with Random Forest 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