Nature-inspired DMU Selection and Evaluation in Data Envelopment Analysis

Seyed Muhammad Hossein Mousavi[1]

1. Researcher at Pars AI Company, Tehran, Iran

Abstract: In order to have best efficient frontier, it is needed to have closer efficiency value for most of Decision Management Units (DMU) in Data Envelopment Analysis (DEA). So, selecting best DMU's for businesses in management before running it has high of importance. Also, having best DMU's pushes the business to the ideal point in the data space. Obviously, the closer features or DMU's to ideal point, the more efficient system is. Nature-inspired optimization algorithms gain the most optimized approaches based on natural selection which differentiates them from other un-intelligent mathematical selection models. Here Biogeography-Based optimization (BBO) algorithm is employed to select the best features or DMU's for business benchmark datasets which are contain a lot of samples. The system evaluates with four common DEA methods of Charles Cooper and Rhodes (CCR), Input-Oriented Banker, Chames and Cooper (IOBCC), Output-Oriented BCC (OOBCC) and Additive efficiently. Proposed approach compares with original DEA model, mathematical Lasso regularization and other nature-inspired methods such as GA and PSO feature selection for all four components of evaluation. Returned result depict more efficient values in most cases belongs to proposed nature-inspired DEA method. In order to present more comparisons, another optimization (Firefly) fuzzy regression algorithm is used for final linear regression part as they represent more correlation coefficient compare to traditional regression methods.

Keywords: Decision Management Units; Data Envelopment Analysis; Optimization; Nature-Inspired Feature Selection; Nature-Inspired Fuzzy Regression

1. Instruction

Businesses and companies are looking for best ways to increase their facilities productivity and efficiency in Data Envelopment Analysis (DEA) [1] by various of factors which in management, these factors call Decision Management Units or in short DMU's [1]. Considering the efficiency of hospitals (DMU's), and in that number of Doctors (nD), number of Nurses (nN) and number of Patients (nP) as input features and Treatment per day as output, total Efficiency could be calculated. For instance, nD=10, nN=20 and nP=50 is more efficient than nD=15, nN=27 and nP= 25, as with less resources more patients are cured. So, selecting these DMU's wisely are so crucial to have the most efficient business experience. To do so, number of techniques and algorithms are proposed by different researchers during years. Most of them used old traditional feature selection methods and some of them used these algorithms for other purposes of DEA. Here, DMU's or feature are selecting based on nature intelligence which is called nature-inspired or bio-inspired algorithm [2]. Two of the best nature-inspired algorithms for features selection are Particle Swarm Algorithm (PSO) and Genetic algorithm feature selection [3] but, none of them is as power full as Biogeography-Based optimization (BBO) algorithm [4] for this task. We come up to this fact by multiple experiment on these algorithms for feature selection purposes. BBO is faster and returns more optimized results in less iteration which, lead to select best DMU's (with higher value closer to 1) for the business. Basically, BBO feature selection, removes any weak DMU's which may result low value efficiency and DEA.

Furthermore, by increasing number of samples to for example 100 and number of DMU's or features to for example 10, selecting best DMU's is a laborious task. So, when dealing with big datasets, removing some of the weak samples is so rational. The main goal is to select those DMU's which are closer to efficient frontier.

This paper is consisted to 4 main sections. Section 1 is all about basics and fundamentals. Section 2 pays to some of the relevant researches conducted by other researcher in the field of feature selection. Section 3 firstly, describes the proposed method in details and secondly, pays to validations and result alongside with comparisons with other methods. Section 4 includes conclusion, future works and suggestions.

2. Literature

Feature selection or dimensionality reduction is a vital task in data mining [5] and big data [6]. By decreasing number of feature and selecting best ones, not only processing speed increases but also, outliers which are less desired, eliminate from the main process. Principle Component Analysis (PCA) [7] is one of the most effective but, traditional feature selection in data mining. Another more advanced feature selection algorithm is called Lasso regularization [8] which, is more time consuming but ends up with more effective features.

Another interning feature selection based on DEA research is belonged to Zhang, Yishi, et al [9]. They did something in revers of this research's purpose. Actually, they used DEA of each DMU as their final feature selection factor; in which those DMU's with higher DEA values got selected. Among nature inspired algorithms, PSO feature selection [10] and Genetic feature selection [11] are two best to mention. Proposed BBO feature selection for DMU would compare with all these four mentioned algorithms in validation section.

3. The Proposed Method and Evaluation

Clearly, DMU's in DEA, defines system's performance and in a small business, DMU's values could be determine easily. However, in a big business or in dealing with real big data, things are different. Nature has best approach for selection which is called natural selection. Any mathematical implementation of nature could handle certain number of mathematical problems. One of this nature inspire algorithms is BBO algorithm which, which we intend to used it for features (DMU) selection in DEA for the first time. Based on multiple experiment on different algorithm, BBO selected, as it returned more optimized result in faster runtime. Here, each DMU is considered as a feature in the feature space which makes our final feature matrix for the main process. After feature selection by BBO algorithm, Efficiency and DEA calculation starts and the process ends up to non-linear fuzzy [15] Firefly [12] regression algorithm which is another bio-inspire algorithm. Based on multiple experiment, nature inspired regression algorithms returned higher correlation coefficient than other traditional methods which, best performance among them that was Firefly algorithm is considered for this part. Figure 1 represents the whole process of our system.

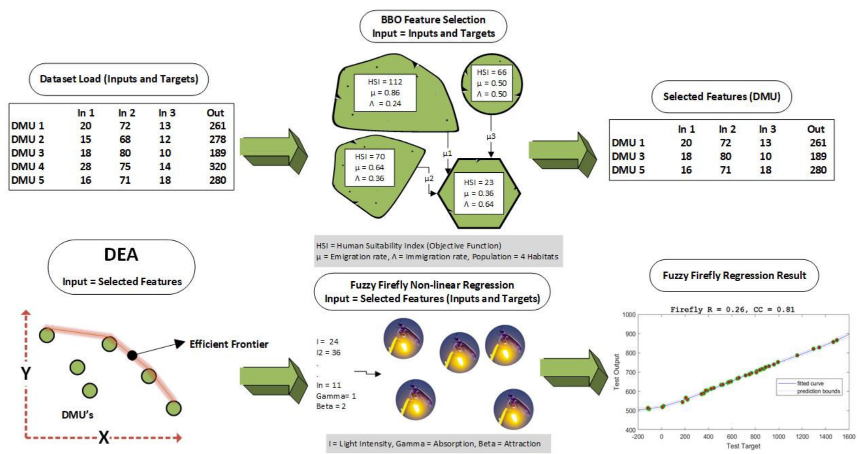

Fig. 1. Workflow of Proposed Method

- **BBO Feature Selection**

BBO algorithm [4] is consisted of important parameters of number of Habitants or "H", Human Suitability Index or "HIS", Emigrations Rate or μ, Immigration rate or λ and Suitability Index Variable (SIV). This algorithm is all about moving living creatures from one habitant to another with better life condition and room to grow. In feature selection, we are dealing with Number of Features of "NF", weight of feature or "w" and Mean Square Error (MSE) which should be minimized to select thee feature. Also, if x_i are values of NF then, $\hat{x_i}$ would be selected features out of NF. So, considering number of features entering the system, "y' would be the output and "t" would be the target. In order to calculate final error, e_i need to be calculated which is t_i - y_i . So final error is $\min MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$+w*NF. This goes for all features and finally those features with lowest MSE will be selected. In combination of BBO and feature selection, each feature vector or DMU is considered as a habitant with different HIS. Those habitants which could fit into final iteration would be selected alongside with their related features with lowers error as it mentioned. Pseudo code of BBO feature selection is presented in Table 1.

Table 1. Pseudo Code of BBO Feature Selection

```
Start
    Load dataset (DMUs)
    Generate a random set of habitats (H1, H2, …, Hn (DMU's or features)
    Define FN (number of features) and w (weights for DMUs)
    Compute HSI value (Fitness function and sort best to worst)
        While termination criterion is not satisfied
            Keep the best individuals (elites (best DMUs))
            Calculate immigration rate λ and emigration rate μ for each habitat based on HSI
            Start Migration
                Select Hi with probability by λ
                Select Hj with probability by μ
                Randomly select a SIV from Hj
                Replace random SIV Hj with Hi
                End of migration
            Start Mutation
                Select a SIV in Hi with probability of mutation rate
                If Hi (SIV) is selected
                    Replace Hi (SIV) with a randomly generated SIV
                End if
            End of Mutation
                Recalculate the HIS value of new habitats
                Calculate MSE of DMU's
                Sort population (best to worst (cost))
                Replace worst with preview generation's elites (DMUs with best cost)
                Sort population (best to worst (cost))
        End of While
        Select NF first ones
End
```

- **Data Envelopment Analysis (DEA)**

Originally introduced by Charnas and Cooper in 1978 [13] based on Farrell ideal in 1957 [14], DEA faces multiple forms of it by different researchers around the world but, still one of the most flexible tools in various areas specifically in management. DEA employs to evaluate the performance of different kinds of entities in various range of activities and applications [1]. Obviously, DEA main target is to estimate the system's efficiency by different input and output factors which leads to select best DMU's for having the most efficient system. Evaluates takes place with four common DEA methods of Charles Cooper and Rhodes (CCR), Input-Oriented Banker, Chames and Cooper (IOBCC), Output-Oriented BCC (OOBCC) and Additive efficiently [1]. The CCR model considered that constant return to scale exists at the efficient frontiers whereas BCC says variable retunes to scale frontiers. CCR model measures the Overall Technical Efficiency (OTE), while BCC model assesses the Pure Technical Efficiency (PTE). Actually, CCR model has a straight-line efficiency frontier, whereas the BCC model has a convex line efficiency frontier as Figure 2 represents.

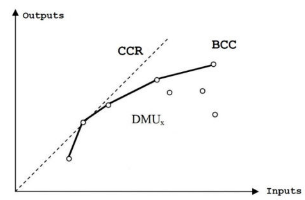

Fig. 2 CCR and BCC models

- **Fuzzy Firefly Regression**

Firefly Algorithm (FA) [12] is one the most robust and famous bio-inspired optimizations algorithms which has proper application in learning for regression task. Now, this regression could be aided with other clustering techniques as input data, which one of them is Fuzzy [15] C-means clustering (FCM) [16]. Fuzzy C-means clustering is fuzzy model of K-means or Lloyd's clustering algorithm [17]. By clustering the data in the initial step, data will be organized in an optimized manner for training step. Clearly, more clusters, means more accuracy but brings more computational time too. The goal here is to adjusting base fuzzy parameters according to modeling error by FA and returning best fuzzy parameters values as the final result. Considering p_i^* as final regression optimized value, two parameters of it which are x_i and p_i^o determines by FA and Fuzzy logic, accordingly. Firstly, data (inputs and targets) passing through fuzzy system divides into train and test parts by 70% and 30%, respectively. Second step is to define linguistic variables, constructing membership functions, sets and rules and finally converting crisp feature (inputs and target) matrix to fuzzy model (fuzzification) which end up to initial fuzzy model ready for training by FA algorithm. Fuzzy part uses "Sugeno" inference system as it performs better than "Mamdani". Each input represents one feature (each three membership functions) and three rules ("and" operator) followed by an output which contains targets in this step.

The fuzzy model od DMU's sends to FA as input for adjusting basic fuzzy parameters by nature inspired behavior of FA under x_i value as it mentioned above. The change goes over membership functions and changes gaussian curve as range and variance to fittest form by FA. Just like any other bio-inspired algorithm, number of population and iterations, plays important rule in the algorithm. Here, population and iteration considered 15 and 1000, respectively. Also, number of decision variable and its lower and upper bounds are considered to be 10, -10 and 10, respectively.

FA is consisted of five main parts of number of population (fireflies) or x, light intensity of each firefly or I, light absorption coefficient or γ, attraction coefficient or β and mutation rate. It simply works as moving lower light intensity fireflies toward higher ones, effecting mutation and updating old and new solutions. Now, fuzzy input model is transformed to a better fuzzy model after taking effect by FA on its membership functions and parameters. By evaluating fuzzy FA model using fuzzy inference engine, final trained data (train and test) is available. In order to calculate error fuzzy data should return to its original crisp mode which this action is called defuzzification. Clearly, inputs are train and test inputs; and evaluated version of them are train and test outputs. The difference between train, test outputs and train, test targets, provide system error which here are MSE, RMSE, Mean Error and STD Error. Pseudo code of fuzzy Firefly regression is available via Table 2. Also, Table 3 presents, BBO and Firefly algorithms' parameters which is used in the experiment to evaluate.

Table 2. Pseudo Code of Fuzzy Firefly Regression

Start
Loading Data (Extracted features or DMUs)
Diving Train and Test data (Both into Inputs and Targets)
Generating Basic Fuzzy C-Means Model
Define Linguistics Variables
Construct Membership Functions, Sets and Rules
Initial Training Using Fuzzy Logic
Fuzzification (Crisp Inputs to Fuzzy)
Training Using Firefly Algorithm (Input: Fuzzy Sets and Rules)
Goal: Adjusting Base Fuzzy Parameters According to Modeling Error by Firefly
Objective function f (x), $x=(x1,x2,\ldots,xd)^T$
Generating population of fireflies xi (i=1,2,…,n)
Define light intensity Ii in xi by f(xi)
Define light absorption coefficient γ or gamma
    While *maximum generation is not satisfied*
        For *i=1 to n fireflies*
            For *j=1 to n fireflies (inner loop)*
                If *(Ii < Ij), firefly i goes toward firefly j*
                End if
                    Change attractiveness by distance of r via exp [- γr]
                    Evaluate new solutions and update light intensity
            End
        End
        *Sort and rank fireflies and find the current global best or g**
    End of while
Inference (Evaluating Fuzzy Rules and Combining Results Based Firefly Model Error)
Defuzzification (Fuzzy Outputs to Crisp)
Getting Optimized Value of $p_i^* = x_i p_i^o$ $(x_i \text{ by } Firefly\ Algorithm \text{ and } p_i^o \text{ by } Fuzzy\ Logic)$
Calculating Polynomial Nonlinear Regression Between Targets (Inputs Labels) and Outputs (p_i^*)
Returning MSE, RMSE, Error Mean and Error STD
End

Table 3. Nature-Inspired Algorithms' Parameters

-	BBO Algorithm	Firefly Algorithm
Iteration	1000	1000
Population	20 Habitats	15 Fireflies
Human Suitability Index (HIS)	Objective Function	*
Light Intensity of I	*	Objective Function
Variables	10	10
Lower Bound (var min)	-10	-10
Upper Bound (var max)	10	10
Keep Rate	0.2	*
No of Kept Habitats	Keep Rate * Habitats	*
No of New Habitats	Habitats – Kept Habitats	*
Emigration Rate	μ = 0.2	*
Immigration Rate	λ = 0.3	*
Alpha	α = 0.9	α = 0.2 (mutation coefficient)

Mutation Probability	0.1	*
Sigma	ς = 0.02*(var max -var min)	*
Light Absorption Coefficient	*	$\gamma = 0.1$
Attraction Coefficient	*	$\beta = 2$
Delta (Mutation Range)	*	δ = 0.05*(var max-var min)

Table 4 shows the details of datasets which are used in the experiment. Figure 3 illustrates DEA calculation for first 15 samples of daily demand forecasting order dataset before BBO feature selection and after features selection by BBO algorithm in 7, 8 and 10 features (DMUs) out of 13, respectively. Also, figure 4 represents, BBO algorithm training stage over 1000 iterations. Table 5 is retuned results from proposed method plus four other methods including original DEA and three feature selection methods before DEA. These results are average of CCR, IOBCC, OOBCC and Additive methods on four mentioned datasets in three 25 %, 50% and 75 % ranks. Proposed BBO DMU selection method compared with traditional Lasso regularization [8], Genetic Algorithm (GA) feature selection [11] and Particle Swarm Optimization (PSO) feature selection [10]. Obviously, higher DEA value to 1 means more efficiency for the system. Figure 5 represents acquired results from the experiment in the Table 5 as box plots.

Table 4. Experiment's Datasets

Name	Area	Associated Task (s)	Instances	Features	Reference
Clickstream Data for Online Shopping	Business	Classification, Regression, Clustering	165474	14	[18]
Daily Demand Forecasting Orders	Business	Regression	60	13	[19]
Online News Popularity	Business	Classification, Regression	39797	61	[20]
Statlog (Australian Credit Approval)	Financial	Classification, Regression	690	14	[21]

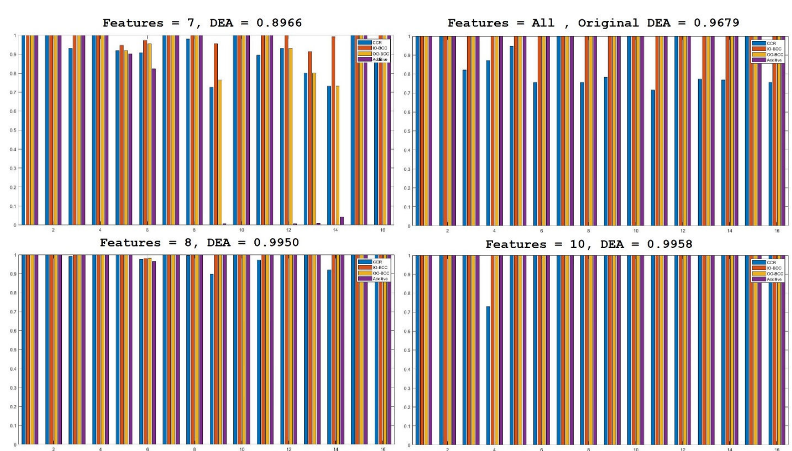

Fig. 3 Testing bar plot of BBO feature selection on samples of daily demand forecasting order dataset versus DEA on original data

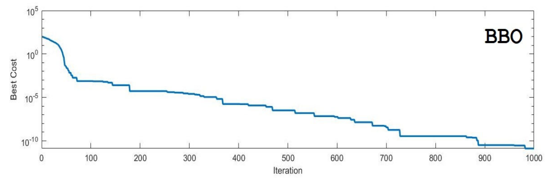

Fig. 4 BBO algorithm training stage over 1000 iterations based on Table 3 parameters

Table 5. Average of CCR, IOBCC, OOBCC and Additive methods on datasets and comparison with other methods in three ranks

Features = DMUs	**Clickstream**	**Daily Demand**	**Online News**	**Statlog**
Original DEA (all features)	0.837	0.926	0.844	0.883
Lasso DEA				
Rank 1 = 75 % of Features	0.691	0.839	0.799	0.817
Rank 2 = 50 % of Features	0.601	0.782	0.749	0.781
Rank 3 = 25 % of Features	0.588	0.754	0.694	0.758
GA Features DEA				
Rank 1 = 75 % of Features	0.825	0.937	0.905	0.899
Rank 2 = 50 % of Features	0.799	0.879	0.866	0.857
Rank 3 = 25 % of Features	0.767	0.867	0.805	0.796
PSO Features DEA				
Rank 1 =75 % of Features	0.812	0.951	0.857	**1.000**
Rank 2 = 50 % of Features	0.786	0.928	0.817	0.881
Rank 3 = 25 % of Features	0.772	0.876	0.800	0.856
BBO Features DEA				
Rank 1 = 75 % of Features	0.919	**1.000**	0.961	**1.000**
Rank 2 = 50 % of Features	0.884	0.969	0.863	0.957
Rank 3 = 25 % of Features	0.803	0.900	0.840	0.883

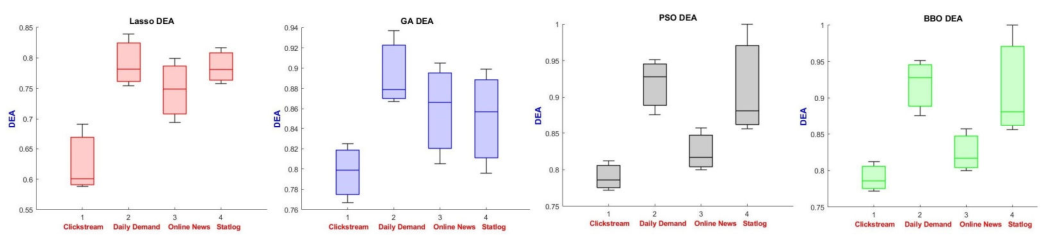

Fig. 5 Comparison of different methods as box plots

By looking at the results from Table 5 and Figure 5, performance of the Lasso regularizations DEA meets the modest. However, original DEA on all data have achieved medium performance as it is clear in the Table 5. GA DEA shows a little bit better performance than original DEA in all ranks. Second places go to PSO DEA and best results belongs to BBO DEA. Figure 6, illustrates fuzzy firefly regression result on 250 samples of online news popularity dataset, using 25% of the features (DMUs). Returned Correlation Coefficient (CC) [23] and errors, shows outstanding performance between target and out for train and test. Also, Figure 7 shows related error for the same experiment. Figure 8. Presents Fuzzy Firefly Regression performance in training stage.

Table 6, covers Correlation Coefficient (CC) and Means Square Error (MSE) [23] for the experiment on four datasets in three ranks of 25%, 50% and 75% which are achieved from fuzzy firefly regression and compared with fuzzy regression [22] on extracted DMUs from previews step by BBO features selection algorithm.

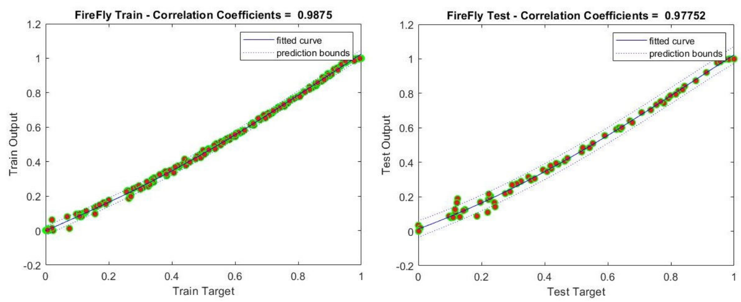

Fig. 6 Fuzzy Firefly non-linear regression test on samples of online news popularity dataset after BBO feature selection by 25 % of DMUs

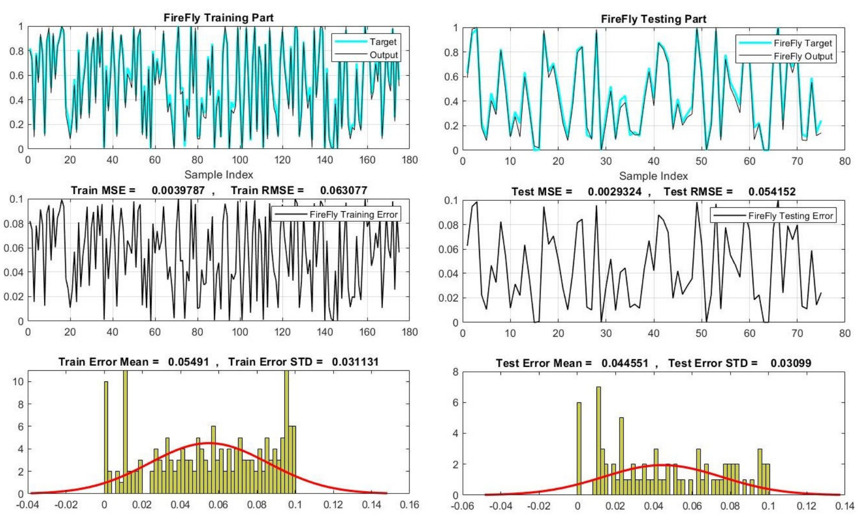

Fig. 7 MSE, RMSE, Error Mean, Error STD for regression experiment of Figure 6

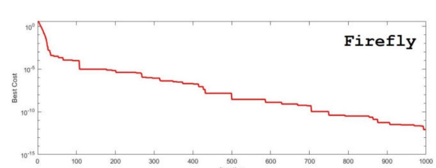

Fig. 8 Firefly algorithm training stage over 1000 iterations based on Table 3 parameters

Table 6. CC and MSE comparison for fuzzy regression and fuzzy firefly regression in three ranks of features on four datasets

Features = DMUs	**Clickstream**	**Daily Demand**	**Online News**	**Statlog**
		Fuzzy Regression		
Rank 1 = 75 % of Features	CC = 0.836 MSE = 0.285	CC = 0.882 MSE = 0.154	CC = 0.944 MSE = 0.081	CC = 0.970 MSE = 0.022
Rank 2 = 50 % of Features	CC = 0.779 MSE = 0.300	CC = 0.875 MSE = 0.181	CC = 0.937 MSE = 0.099	CC = 0.969 MSE = 0.082
Rank 3 = 25 % of Features	CC = 0.783 MSE = 0.319	CC = 0.850 MSE = 0.187	CC = 0.912 MSE =0.058	CC = 0.920 MSE = 0.061
		Fuzzy Firefly Regression		
Rank 1 = 75 % of Features	CC = 0.936 MSE = 0.211	CC = 0.928 MSE = 0.119	CC = 0.991 MSE = 0.004	CC = 0.998 MSE = 0.001
Rank 2 = 50 % of Features	CC = 0.889 MSE = 0.260	CC = 0.905 MSE = 0.117	CC = 0.980 MSE = 0.059	CC = 0.983 MSE = 0.036
Rank 3 = 25 % of Features	CC = 0.849 MSE = 0.297	CC = 0.900 MSE = 0.162	CC = 0.987 MSE = 0.032	CC = 0.996 MSE = 0.007

4. Conclusion

By empowering DEA with nature-inspired algorithms, it is possible to achieve more efficient results compare to using traditional DEA. Evolutionary feature selection (here BBO) could remove more inefficient DMUs rather than mathematical PCA or Lasso algorithms which are not intelligent. Furthermore, to have better understanding of relation between variables after DUM selection, a nature-inspired regression (here Firefly) in combination with fuzzy logic which could provide higher correlation coefficient versus old techniques, is used. Overall, by employing mentioned techniques, more efficient and precise result returned in comparison with other methods; however, due to using optimization algorithm, runtime is sacrificed. Using more related business and management datasets plus comparing the propose system with deep learned feature selection and regression methods is of future works. For this research the system had 7 cores of CPU, so it is suggested to run the proposed system on over 5000 iterations and more than 50 populations, as optimization algorithms demand more hardware resource for higher parameter values and obviously, they will return better results in that condition. The proposed system could be used as a decent tool for resource allocation in business and management applications, when number of DMUs are so high.

References

1. Charles, Vincent, and Mukesh Kumar, eds. Data envelopment analysis and its applications to management. Cambridge Scholars Publishing, 2013.
2. Mousavi, Seyed Muhammad Hossein, Vincent Charles, and Tatiana Gherman. "An evolutionary Pentagon Support Vector finder method." Expert Systems with Applications 150 (2020): 113284.
3. Ghosh, Manosij, et al. "Binary genetic swarm optimization: A combination of GA and PSO for feature selection." Journal of Intelligent Systems 29.1 (2020): 1598-1610.

4. Simon, Dan. "Biogeography-based optimization." IEEE transactions on evolutionary computation 12.6 (2008): 702-713.
5. Mousavi, Seyed Muhammad Hossein, S. Younes MiriNezhad, and Atiye Mirmoini. "A new support vector finder method, based on triangular calculations and K-means clustering." 2017 9th International Conference on Information and Knowledge Technology (IKT). IEEE, 2017.
6. Charles, Vincent, and Tatiana Gherman. "Big data analytics and ethnography: Together for the greater good." Big data for the greater good. Springer, Cham, 2019. 19-33.
7. Abdi, Hervé, and Lynne J. Williams. "Principal component analysis." Wiley interdisciplinary reviews: computational statistics 2.4 (2010): 433-459.
8. Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society: Series B (Methodological) 58.1 (1996): 267-288.
9. Zhang, Yishi, et al. "Feature selection using data envelopment analysis." Knowledge-Based Systems 64 (2014): 70-80.
10. "Feature selection using PSO-SVM." International Journal of Computer Science (2007).
11. Babatunde, Oluleye H., et al. "A genetic algorithm-based feature selection." (2014).
12. Yang, Xin-She. Nature-inspired metaheuristic algorithms. Luniver press, 2010.
13. Charnes, Abraham, William W. Cooper, and Edwardo Rhodes. "Measuring the efficiency of decision making units." European journal of operational research 2.6 (1978): 429-444.
14. Farrell, Michael James. "The measurement of productive efficiency." Journal of the Royal Statistical Society: Series A (General) 120.3 (1957): 253-281.
15. Zadeh, Lotfi A. "Fuzzy logic." Computer 21.4 (1988): 83-93.
16. Bezdek, James C., Robert Ehrlich, and William Full. "FCM: The fuzzy c-means clustering algorithm." Computers & geosciences 10.2-3 (1984): 191-203.
17. Likas, Aristidis, Nikos Vlassis, and Jakob J. Verbeek. "The global k-means clustering algorithm." Pattern recognition 36.2 (2003): 451-461.
18. Łapczyński, Mariusz, and Sylwester Białowąs. "Discovering Patterns of Users' Behaviour in an E-shop-Comparison of Consumer Buying Behaviours in Poland and Other European Countries." Studia Ekonomiczne 151 (2013): 144-153.
19. Ferreira, R. Pinto, et al. "Study on daily demand forecasting orders using artificial neural network." IEEE Latin America Transactions 14.3 (2016): 1519-1525.
20. Fernandes, Kelwin, Pedro Vinagre, and Paulo Cortez. "A proactive intelligent decision support system for predicting the popularity of online news." Portuguese conference on artificial intelligence. Springer, Cham, 2015.
21. Quinlan, J. Ross. "Simplifying decision trees." International journal of man-machine studies 27.3 (1987): 221-234.
22. Zuo, Hua, et al. "Fuzzy regression transfer learning in Takagi–Sugeno fuzzy models." IEEE Transactions on Fuzzy Systems 25.6 (2016): 1795-1807.
23. Murphy, Allan H. "Skill scores based on the mean square error and their relationships to the correlation coefficient." Monthly weather review 116.12 (1988): 2417-2424.