Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering pp. 359–364 DOI: 10.15439/2017R8
ACSIS, Vol. 10 ISSN 2300-5963

OFS: Online Feature Selection Based on Regression Analysis and Clustering method along with its Application

Miss. P R Vhansure[1] Prof. A A Phatak[2] Prof. A S Shimpi[3]

[1]PG Student, [2]Assistant Professor and Head of Department, [3]Assistant Professor

Computer Science & Engineering from N. B. Navale Sinhgad College of Engineering Solapur, India.

[1]priyanka.vhansure@gmail.com

[2]amol1911@gmail.com

[3] anand02cs004@gmail.com

***Abstract*—Feature Selection is one of the important techniques in the Data mining. For the purpose of reducing the computational cost and reduction of noises to improve the accuracy of classification, the feature selection is very important technique for large-scale dataset. The result of feature selection has restricted to only batch learning. Different from batch learning technique online learning has selected by a motivational scalable, well-organized machine learning algorithm which has been used for large-scale dataset. In many defined techniques are not always conveniently helpful for the large-scale dataset. The real-world applications has huge amount of data which are having very long capacity or it costly to bring the entire set of attributes. Focusing on this loophole the concept of Online Feature Selection (OFS) is established. For every occurrence the online learning technique should be retrieve complete features/ attributes from large scale dataset volume. In OFS technique it is hard to online learner to keep a classifier that consist minimum and exact number of features. The OFS technique has primary defiance that, how to make accurate prediction from a large-scale dataset of iterations by using a fixed and small number of actively working features. In this article two different ways of OFS techniques are used its main work is to acquire minimum number of features. In first task a learner has allowed with the access of all the features to elect the subset of active features, and in the second task, a learner has allowed with access of only limited number of features for every iteration. We have used Differential Evolutionary (DE) algorithm in this study. By using new techniques such as Multiclass classification, DE algorithm, Correlation and clustering method the system is implemented to solve many real-world applications, problem and give their imperial performance analysis of the large-scale dataset.**

Keywords — Online Feature Selection (OFS), Differential Evolutionry Algorithm (DE) Online Learning, Large-scale Dataset, Data Mining, Classification, Correlation, Clustering Method

I. Introduction

Feature Selection (FS) is an important step in successful data mining applications. The accomplishment of data mining applications Feature Selection (FS) has one of the intrinsic steps. In batch learning technique features selection process has increasingly used. For removing the irrelevant and redundant features of large-scale dataset the data dimensionality has reduced very adequately. Under the data mining technique the FS is a mechanism of selecting a group of original features according to certain norm; generally FS is significant which is mainly used for dimensionality subtraction or reduction technique of data mining. In many real-world applications, the dataset contains many data so their size is very large. Learning of that data is not work properly as well before eliminate the redundant features or attributes.

Feature selection can works eliminating the number of features, also eliminate irrelevant, redundant, or noise full data that carry immediate effects for large-scale dataset of many applications. Running time of a learning algorithm is automatically reduced when the number of unwanted attributes highly reduced and returns the many generic concepts [2]. For implementing the impressive prediction model the FS is very important. To choose a subset of related features that is the main goal of feature selection method. Online feature selection has increases performance analysis of the prediction model for removing unrelated and repeated features in large-scale dataset. By also reducing the curse of dimensionality, build up the generalization performance, speeding of the When problems are arises for processing of high dimensional dataset that time FS has many use for implementing the applications in different domains. The example like speeding up a mining algorithm and receive best performance of the algorithm such as approximate, predictive accuracy and also result transparency. Feature selection distributed in to 3 methods they are filter method, wrapper and embedded method. The goal of OFS is to discover the explanation of feature selection issue in an online learning form by exploring online learning method. In online learning method the attributes and features are needs at the time of

training instance. The online learning method OFS goal is to choose a smallest exact number of features for multiclass classification. The framework of algorithm has two aspects. We first work on every iteration by unconstrained gradient descent step. After that we solve an optimization problem that is for minimization of a regularization term at the same time the result of the first phase to keep closeness to each other. In large-scale dataset the feature selection takes only labeled data. For batch penalized risk minimization and online learning the simple valid algorithms are used [11]. Such examples medical diagnosis, forensic science, fraud detection they are cannot taken for granted for real world applications because their labeled data are very hard to recognize. So, this issue finds the "semi-supervised feature selection" to the optimal subset of features or attributes gives both types of data that is labeled and unlabeled problems and their result gives exact classifier for many learning algorithm [12]. OFS has to gives two different category in different environment: the first task of OFS by learning with full input data, this task contains learning with all data in to the high dimensional dataset and find the subset of effective features, the second task contains OFS by learning with partial input data, this task contains the access of limited features of large dataset.

Differential Evolutionary (DE) is one of the algorithms for optimization which is constructing to use distinct number of logic, distinct model and also their own merits and demerits. To increase the solution with regard to a given measure of aspect the differential evolution is a technique for optimizing a problem by iterative form. In very large space of dataset many techniques are generally known as not assumption round the problem being optimized. FS has been found wide applications in various fields, mainly for the problems having large-scale approach for high dimensional data. To collect the all information of training data such suppositions may not always correct for real-world applications. Bioinformatics is one of the examples of FS, which consist set of features or attributes and for each iteration it is expensive because high cost of conducting experiments. Finding the relationship between two or more feature here the correlation is best. Correlation method is one of statistical class of statistical relationship which involves dependence between common features. For clustering purpose Nearest Neighboring Algorithm is used it is easy to implement and executes quickly. In present research data mining is most popular method for the purpose of analyzing read accident data.

The aim of this study is to find out effective way to select the best suitable distance metric to cluster the series of counts is expected that to provide a better clustering result. It is observed that in many practical applications online and batch algorithms has heavy data with missing features. Online setup has comparison hypothesis which is fixed throughout any given iteration and extending the standard setting. In case of batch setup imputation function which are used to fill in value of missing features with classification hypothesis are find by convex relaxation of non-convex problem.

II. RELATED WORK

This research work is mainly related to studies of online learning of two tasks and their feature selection. The reviews on related important work are as follows. First OFS: Online Feature Selection based on Regression Analysis and Clustering Method with its application. Second Online Feature Selection with its applications [1]. Recently, many numbers of online learning algorithms has proposed. Here Correlation has used to find the relation between two or more features. The relation between pair of features to each other can be studied by correlation statistical method. Consider one example i.e. height and weight both are related to each other, taller people look heavier than shorter people, this relation is perfect by correlation. For finding the predictive relationship between features the correlation is mainly used, i.e. for example production of electricity may get vary depends on whether conditions. The more or heavy use of electricity is for heating or cooling is depend on extreme weather conditions. In this example it shows that the correlation is not sufficient to demonstrate presence of casual relationship. In mathematical conditions/methods of probabilistic independence do not satisfy random variables of any conditions referred by the dependence. Correlation works on many random variables from independence relationship, and also works on several types of relationship arise in between mean values of variables. For combination of most related features there is clustering technique used which is contain group of related attributes. Nearest neighbor clustering algorithm used because it takes nearest related attribute for clustering .Nearest neighboring is a part of supervised learning that has used in so many applications in the field of data mining, pattern recognition, image processing and many other applications. OFS technique has two different types of tasks which are as follows:

1) Using OFS, learner can read with full input of dataset and
2) Using OFS, learner can read with partial input of dataset.

In the first task of OFS, learner can access all the features which are involved in to the dataset of training data, and find the exact number of features that are used for accurate prediction. The second task is also same but in this task learner is allowed to access only minimum and fixed number of features for each training data instance of large dataset to find the subset of related features [2]. Clustering method based on a fitness function that relies on a distance measure and usually tries to develop "tight" clusters. Nearest neighboring algorithm has simple and powerful rule. It runs fast and gives proper output for the clustering. It has requires lot of training data and reduces the noisy data. Also reduce the redundant data from large dataset. After the clustering of related features has done then evolutionary optimization has done. The algorithms that allow optimization of fitness function of different variables. Online learning algorithms have become especially popular in natural language processing for tasks including classification, tagging, and parsing [5].

In feature selection technique, how the algorithm and training set interact with each other, is depends on the best performance analysis with a particular learning algorithm on a particular training set of the large-scale dataset. The optimal feature subset selection method and related wrapper method both are related to each other and having in good relationship. Both methods are used to search an optimal feature subset to a particular algorithm and particular domain [7]. The large applications methods that have already been implemented in the machine learning and so many data mining fields, but some particular applications such as bioinformatics have to be implemented in a wealth of newly proposed method [8]. Budgeted learning has three variants, such settings in which the learner allowed to access a small and exact number of features from training data in large-scale dataset. The first setting is "local budget", in which the design of an efficient algorithm for linear predictors that actively work for selecting the features of each training instances. In the second setting contains the "global budget", in that overall numbers of features are taken for accessing the training data [10].

A. *Differntial Evolutionary (DE) Algorithm*

In high dimensional dataset at the time of analysis the objective of feature selection is to select the features from dataset which has no target variables or features. All selected features are found at the possible targets which are derived from the complete large dataset and also from subset of it.
The accuracy and complexity of prediction model is based on the dataset which are reduced by using feature selection or any other technique, and comparison is done with in both technique and also those complete large-scale dataset. The final analysis of dataset shows the similarity in the correlated features which was chosen by using the supervised (filter) technique and change the consistency those which are selected by the clustering techniques i.e. nothing but unsupervised method. In large part of dataset the number of features distribution is based on correlated features from many features groups and also learning algorithms selected by as per the user need [9].

The Differential Evolutionary (DE) algorithm is mainly used for large population based dataset i.e. also same like genetic algorithm for optimization, both are use the same operators such as crossover, mutation and selection process. Finding the better solution for the problem is common in both algorithms but genetic algorithms based on crossover function only and DE is based on mutation operator. The main work of both algorithms is based on the differences in between randomly selected pairs of solutions in large-scale dataset. DE algorithm uses the purpose of a search method i.e. mutation operation and for purpose of direct search in to the specified area i.e. for use the selection operation. The components of the present dataset members that are construct the many trial vectors; in that crossover is the operator for recombination. It shuffles data from search space for finding the best solution. For representation of D-dimensional search space the optimization process consist of D parameters. In DE algorithm the solution of data is successfully improved by using mutation operator, crossover and selection operators. For using crossover operator that can take child node parameter from one parent node i.e. is done from many others node.

DE algorithm has some important steps they are as follows:

1. Initialization of data from Evaluation
2. Repeat above step
3. Mutation Recombination process
4. Evaluation Selection Until target termination criteria are met

The real-world applications such as problems have many objective functions i.e. non-linear, noise full data, flat and multi- dimensional problems they are difficult to solve analytically but by using global optimization this can be done very efficiently. For finding the approximate solutions in high dimensional dataset DE is mainly used:

i. DE is an Evolutionary Algorithm for optimization
ii. This class also includes Genetic Algorithms which is used for optimization, Evolutionary Strategies also Evolutionary Programming

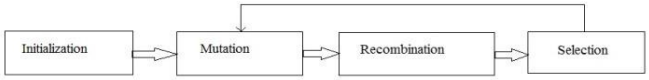

Fig 1: General Evolutionary Algorithm Procedure

B. *How To Calculate Correlation Matrix*

Correlation matrix provides the relations between features it is a type of matrix, which is also provides the correlation between whole pairs of data sets in a matrix. It should use at the time of optimization in DE algorithm.

Sum of squared matrix

$$1/(a-1)\begin{vmatrix} PP_{XX} & PP_{XY} & PP_{XZ} \\ PP_{YX} & PP_{YY} & PP_{YZ} \\ PP_{ZX} & PP_{ZY} & PP_{ZZ} \end{vmatrix}$$

Where,
a=N * N Matrix value

$$PP_{XX} = \Sigma (X_i - \bar{X})^2 \quad (1)$$

$$PP_{XY} = \Sigma (X_i - \bar{X}) * \Sigma(Y_i - \bar{Y}) \quad (2)$$

Similarly

$$PP_{XZ} = \Sigma (Y_i - \bar{Y}) * \Sigma(Z_i - \bar{Z}) \quad (3)$$

Correlation Matrix

$$\begin{vmatrix} 1 & S_{XY} & S_{XZ} \\ S_{YX} & 1 & S_{YZ} \\ S_{ZX} & P_{ZY} & 1 \end{vmatrix}$$

Where,
a = N * N

$$S_{XY} = PP_{XY} / \sqrt{(PP_{XX} * PP_{YY})} \quad (4)$$

Based on eq. (1), (2), (3) and eq. (4) the correlation matrix is calculated that can be used for at the time of DE optimization. That contains the relations between variables or features. In many research areas there are many real-world applications for which has large datasets that contains hundreds or thousands of features, so the research focuses mainly on

feature selection. The research areas of online feature selection include the text processing of internet documents, microarray gene expression analysis, and also combinational chemistry. The main goal of feature selection is done for three reasons:

1. For improving the prediction performance of the predictors.
2. Provides faster and more cost-effective data for predictors.
3. Also provides a better understanding of the process that generates the data [6].

C. Online Feature Selection (OFS)

The real-world applications has deal with series of training data with high dimensionality at that time the OFS technique is very important and useful. The online spam classification tasks, for this example the traditional batch feature selection technique cannot be use directly. This research introduces the problem of online feature selection for multiclass classification datasets, by using correlation, clustering method and also DE algorithm. Feature selection can be done by using two tasks of OFS they are as follows:

1) OFS: Learning with Full Input data

This task of OFS technique contains the DE algorithm. In this the learner can access all the input from start to end and select the features from large-scale approach with in high dimensional dataset. The primary goal is to find an exact number of relevant features for accurate prediction or analysis and removes the noisy, unwanted data from the dataset.

2) OFS: Learning with Partial Input data

This second task of OFS technique, the learner is works same like the first task i.e. learning with full input, but in this task learner is allowed to access partial input only. This task contains partial data of the each training instance. In the form of percentage the data is given that shown in the following example.

III. RESULTS

This result section contains an extensive set of real-world examples to perform the proposed OFS technique and DE algorithms. OFS tasks works on several applications from machine learning repository which has multiclass classification dataset. In this the multiclass stock market data set has considered for demonstrating the result of OFS technique. The dataset has related to stock market in that daily up, down values stored. The data should be in numerical form. That data will distributed in four fields. The class values are hold, buy and sell. There are two tasks where the OFS runs:

A. OFS Learning with partial input data from the dataset

This task contains all input data, so the user or learner can enter any percentage of data. After entering any percentage of data we get approximate same accuracy.

	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO	DP
1	Shree Cer	Godrej	Consumer Pr	Dabur Ind	Eicher Mo	Vedanta	Tech Mah	YES Bank (	Bajaj Fina	Siemens L	Cipla	Ambuja C	JSW Steel	capital market	financial market	securities market	economy
2	16.53	8.45	24.2	16.43	16.41	16.19	24.57	16.81	32.59	24.09	8.26	32.76	32.82	hold	sell	sell	buy
3	16.32	32.76	8.24	8.41	24.76	24.21	8.94	24.4	24.11	32.11	16.57	24.75	8.34	hold	sell	buy	buy
4	18.69	9.76	18.95	36.67	36.5	18.03	36.33	27.74	27.53	9.61	36.04	36.09	9.26	buy	hold	buy	hold
5	32.22	32.11	24.8	8.21	8.19	24.08	16.48	24.5	24.44	16.61	32.06	32.11	8.36	buy	hold	hold	sell
6	32.57	24.37	8.2	8.14	32.9	24.21	8.96	8.06	8.27	24.57	16.61	16.61	16.53	hold	sell	buy	sell
7	36.44	27.46	36.04	36.5	28	36.48	36.33	36.9	18.39	36.23	36.49	27.79	36.55	sell	sell	hold	hold
8	18.52	27.62	27.18	9.3	27.76	27.16	18.44	27.39	9.17	36.41	9.74	36.87	36.68	buy	hold	sell	hold
9	18.01	9.43	9.18	18.34	36.58	18.54	36.68	27.57	18.35	36.13	36.93	18.76	9.6	buy	hold	buy	hold
10	8.08	16.37	8.91	16.29	32.04	32.79	24.39	24.33	16.43	16.74	24.07	24.41	8.21	sell	hold	hold	sell
11	24.94	24.41	24.75	16.48	8.46	16.74	24.23	8.52	32.28	16.42	32.94	32.33	24.67	buy	buy	hold	sell
12	8.82	24.14	24.26	24	32.44	32.74	8.69	32.57	16.83	24.53	8.94	16.61	24.4	hold	sell	sell	buy
13	8.61	24.91	24.57	32.54	8.82	16.28	32.49	24.98	16.03	24.48	8.41	24.23	32.32	sell	sell	sell	buy
14	32.39	24.52	24.41	16.45	24.58	8.9	16.3	32.28	32.74	8.45	8.47	32.84	8.92	buy	hold	buy	sell
15	27.54	18.98	27.62	9.58	27.27	18.71	9.19	16.94	18.01	36.49	36.25	9.83	9.74	sell	hold	sell	hold
16	24.89	8.74	16.47	8.39	24.35	32.22	8.42	16.69	16.3	8.59	16.82	32.86	8.6	sell	hold	buy	sell
17	24.31	16.01	8.38	8.76	16.94	32.85	16.06	32.85	16.02	16.16	8.31	16.68	24.19	buy	buy	sell	buy
18	8.44	24.59	8.44	8.13	8.64	8.68	8.82	32.99	16	8.35	16.82	24.44	8.52	hold	sell	sell	buy
19	18.79	9.73	36.36	9.33	36.7	36.91	18.81	36.67	9.38	9.85	36.56	18.24	18.65	buy	sell	sell	hold
20	24.02	16.55	32.28	24.68	16.22	24.57	32.45	24.46	24.34	24.58	16.73	16.24	32.32	hold	sell	hold	sell
21	9.58	36.45	36.87	18.88	27.75	18.9	9.34	9.07	9.29	27.79	36.58	36.93	9.82	sell	buy	buy	hold
22	8.54	16.12	24.4	32.08	32.9	32.02	24.3	8.79	24.87	32.03	24.32	8.99	8.79	buy	hold	buy	sell
23	32.76	8.54	16.06	16.81	8.37	32.27	8.26	24.94	8.45	32.62	24.16	32.35	32.28	sell	sell	hold	sell
24	24.21	16.51	24.26	8.03	32.97	8.78	8.5	24.76	16.42	16.94	8.18	32.8	32.68	buy	hold	buy	sell
25	24.79	16.25	24.2	24.06	32.01	24.8	8.58	8.16	16.94	32.2	16.33	8.8	8.95	sell	hold	sell	buy

Fig.2: Stock market dataset

.

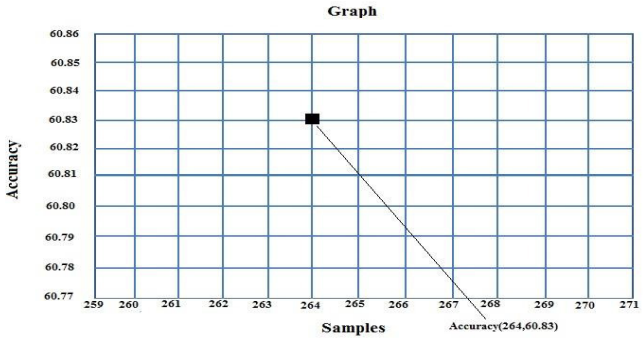

Fig.3: Accuracy graph for 90% of data

In Fig.3 shows the accuracy of selected features within stock market dataset. The data that is under the 90% is considered 264 rows which has taken for the processing the OFS technique. Correlation finds the relation between the dependent variable after that nearest neighboring algorithm used which finds near value for creating the cluster depends on clusters DE algorithm optimize the values with in large dataset. The following Table 1 shows the different percentage of data gives the same accuracy so selected features accuracy has fixed for any number percentage of data.

Table 1: Accuracy and time

Percentage of data given for OFS.	No. of rows within dataset taken for processing the OFS.	Accuracy of selected features.	Time(ms) for selected features.
90%	264	60.83	70ms
45%	140	60.81	25ms
48%	142	60.81	26ms
50%	139	60.27	29ms
80%	225	60.57	55ms

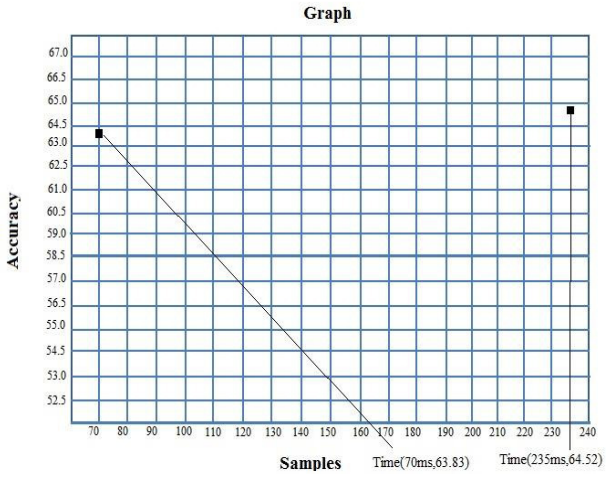

Fig.4: Displaying selected or remaining all features accuracy and their processing time.

The Fig.4 shows the accuracy of selected features and their processing time. In graph shows two plots first contains the time and accuracy for selected features and second plot shows the time and accuracy for all features with in the dataset. This above graph shows the comparison of selected features accuracy and all features accuracy which is approximately near to each other. Here using OFS technique if their accuracy is nearly same then, user can easily select the minimum features and that used for any purpose. By using this technique, improving performance of prediction model and also increasing the speed of processing model is easy. User also reduces complexity of prediction model.

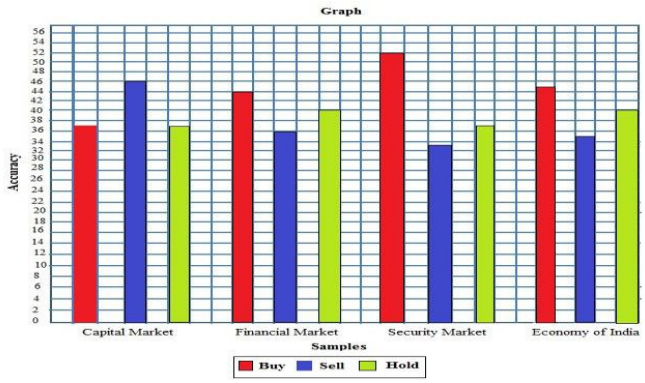

Fig.5: Selected features for analysis.

The Fig.5 shows the use of selected features. For any purpose that is analysis, prediction or any application development selected features used. In stock market, daily updates are arrived for every company. The values changed tries in day. When we have to predict or do some analysis for future work and we need to calculate the value for buy, sell and hold in minimum time for the daily stock that time we can easily use this OFS method for multiclass classification shown in Fig.5.
In this by using two task of OFS method we can select the minimum features that is nothing but minimum companies and calculate their accuracy if their accuracy is near to 100%. Then it gives approximate prediction on Buy, Sell and Holding of the data which has shown in Fig 4. The Table 1 contains the percentage of data and their no. of rows within the dataset which is use for processing the OFS techniques. The minimum average accuracy for any percentage of data that is near to 60% and their processing time is also near to each other which calculated in macro second.

B. OFS Learning with Full input data from the dataset

In task the working of OFS is similar to the partial input learning it takes the total percentage of data that is 100%. Afterward it will show the total number of features and their counts. It also shows the selected features accuracy and their time.

Table 2.: Comparison of Selected Features Accuracy and All Features Accuracy.

Percentage of data given for OFS.	Number of selected features	Accuracy of selected features	Time (ms) for selected features	Accuracy of all features	Time (ms) for selected features
90%	55	60.83	70ms	64.55	235ms
94%	49	64.28	68ms	64.52	429ms
78%	62	64.150	67ms	64.525	258ms
68%	50	64.28	44ms	64.52	177ms

In Table 2. Shows the comparison between selected features accuracy and all features accuracy or their processing time.
The accuracy of selected features or all features is approximate same so by using OFS technique user can easily do their future work in minimum time for large-scale dataset. In case of large dataset user have to more time for processing the dataset but by use of OFS technique user can take minimum time for finish their work or processing the large-scale dataset.

IV. CONCLUSION

This research work implements, online feature selection (OFS) which selects a small and fixed number of features for multiclass classification. The OFS technique has two ways in two different directions which are used for selecting the features in online fashion of the dataset. 1) OFS technique by learning with full inputs of all the features/attributes, and 2) OFS technique by learning with partial inputs of the features in large scale approach for high dimensional dataset. The OFS algorithms are used to solve each of the OFS tasks in multiclass classification data, and gives theoretical analysis on the numerical data from the high dimensional dataset. It also extensively shows proposed OFS techniques are mainly works for solving the real-world applications problem such as neural network, microarray gene expression, CPU Performance: introducing Numeric Prediction in computer vision, and also works on microarray gene expression analysis in bioinformatics.

In this its result analysis shows that algorithm is effectively works for feature selection tasks of many real-world online applications, this OFS algorithm is scalable and capable for handling large dataset. Future work can extend this framework to other settings, such as online feature selection for numerical, textual classification and it can also be extending for image classification

References

[1] "Online Feature Selection and Its Applications" Jialei Wang, Peilin Zhao, Steven C. H. Hoi, Member, IEEE, and Rong Jin

[2] M. Dash and H. Liu, "Feature Selection for Classification," Intelligent Data Analysis, vol. 1, nos. 1-4, pp. 131-156,1997.

[3] S. C. H. Hoi, J. Wang, P. Zhao, and R. Jin, "Online Feature Selection for Mining Big Data," Proc. First Int'l Workshop Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine '12), pp. 93-100, 2012.

[4] S. C. H. Hoi, J. Wang, and P. Zhao, "LIBOL: A Library for Online Learning Algorithms," Nanyang Technological Univ., 2012.

[5] M. Dredze, K. Crammer, and F. Pereira, "Confidence-Weighted Linear Classification," Proc. 25th Int'l Conf. Machine Learning (ICML '08), pp. 264-271, 2008.

[6] Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," J. Machine Learning Research, vol. 3, pp. 1157-1182, 2003.

[7] R. Kohavi and G.H. John, "Wrappers for Feature Subset Selection," Artificial Intelligence, vol. 97, nos. 1-2,pp.273-324,1997.

[8] Y. Saeys, I. Inza, and P. Larran˜ aga, "A Review of Feature Selection Techniques in Bioinformatics," Bioinformatics, vol. 23, no. 19,pp. 2507-2517, 2007.

[9] "The Use of Various Data Mining and Feature Selection Methods in the Analysis of a Population Survey Dataset" Ellen Pitt and RichiNayak, Faculty of Information Technology, Queensland University of Technology George Street, Brisbane, 4001, Queensland.

[10] N. Cesa-Bianchi, S. Shalev-Shwartz, and O. Shamir, "Efficient Learning with Partially Observed Attributes," J. Machine Learning Research, vol. 12, pp. 2857-2878, 2011.

[11] J. Duchi and Y. Singer, "Efficient Online and Batch Learning Using Forward Backward Splitting," J. Machine Learning Research, vol. 10, pp. 2899-2934, 2009.

[12] J. Ren, Z. Qiu, W. Fan, H. Cheng, and P.S. Yu, "Forward Semi-Supervised Feature Selection," Proc. 12th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD '08), pp. 970-976, 2008.

[13] Kumar, S.; Toshniwal, D. Analysis of road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPCC). Journal of Big Data,3, 13:1-11.

[14] A. Rostamizadeh, A. Agarwal, and P.L. Bartlett, "Learning withMissing Features," Proc. Conf. Uncertainty in Artificial Intelligence(UAI '11), pp. 635-642, 2011.