OFS: Online Feature Selection based on Correlation and Clustering method along with its application

Prof. A S Shimpi[1]; Prof. S A Awatade[2]

[1]Department of Computer Science & Engg, PAHS University, Solapur, India,

[2]Department of Computer Science & Engg, PAHS University, Solapur

Corresponding Author Email: aanandshimpi@gmail.com

Abstract—Feature Selection is one of the important techniques in Data mining. At the time of dimension reduction for reducing the computational cost and also reduction of noises to improve the classification accuracy the feature selection is an essential technique. In this, its result shows, that almost all learning of feature selection has been restricted to batch learning. Dissimilar to existing batch learning methods, online learning has been chosen by an encouraging familiar of well-organized and scalable machine learning algorithms for large-scale approaches such existing technique is not always appropriate and useful for real-world applications when data instances are of high dimensionality or very expensive to acquire the full set of attributes/features. To highlight this limitation, here found the problem of Online Feature Selection (OFS). The large-scale dataset quantity of online learning needs to retrieve all the features/attributes of occurrence. In OFS it is hard for the online learner to keep a classifier that involves a minimum and fixed or exact number of features. The major challenge of OFS is how to make exact predictions for an iteration using a small number of active features. This article shows two distinct tasks of OFS. The first one is learning with full input in this a learner is allowed to access all the features to decide the subset of active features, and the second is learning with partial input in this only a limited number of features is allowed to be accessed for each iteration by the learner. We have used the Differential Evolutionary (DE) algorithm in this study. The proposed system represents novel techniques such as Multiclass classification, Correlation, and clustering methods to clear up each of the problems and give their performance analysis.

Keywords: Feature Selection, 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. In the Feature Selection process batch learning is continuously used. It has effectively reduced data dimensionality by discarding the irrelevant and the redundant features. FS is a process of choosing a subset of original features according to certain criteria; it is an important and often-used dimensionality reduction method for data mining. It reduces the number of features, removes unrelated or, redundant, noisy data, and brings instant effects for applications: In many applications, the size of a dataset is so large that learning might not work as well before removing the unwanted features. Reducing the number of irrelevant/redundant features drastically reduces the running time of a learning algorithm and yields a more general concept [2]. The objective of feature selection is to select a subset of relevant features for building effective prediction models. Removing irrelevant and redundant features, OFS has improved the performance of the prediction model by alleviating the effect of the curse of dimensionality, enhancing the generalization performance, speeding up the algorithm, and improving the model interpretability. Feature selection has found applications in many domains, especially for problems involving high-dimensional data [3]. Speeding up a data mining algorithm, and getting the best mining performance such as predictive accuracy and result comprehensibility. Feature selection has been separated into 3 categories: filter, wrapper, and embedded. The purpose of online feature selection is to find solutions to feature selection problems in an online fashion by effectively exploring online learning techniques.

Online learning needs all the attributes or features of training instances. The Online Feature Selection aims to select a minimum and fixed number of features for multiclass classification in an online learning fashion [4].
The algorithmic framework alternates between two phases. On each iteration, we first perform an unconstrained gradient descent step. We then cast and solve an instantaneous optimization problem that uses the minimization of a regularization term while keeping proximity to the result of the first phase. This view yields a simple yet effective algorithm used for batch-penalized risk minimization and online learning [11]. The feature selection methods work directly on labeled examples. The accessible examples cannot be taken for granted for many real-world applications, such as medical diagnosis, forensic science, fraud detection, etc, where labeled examples are hard to find. This problem calls for the need for "semi-supervised feature selection" to find the optimal set of features given both labeled and unlabeled problems that return the most accurate classifier for a learning algorithm [12].
OFS has to give two different types of tasks in distinct settings: The first task is OFS by learning with full inputs, in this task/method learner is allowed to access full features to determine the subset of active features, and in the second task is OFS by learning with partial input, in this method learner has allowed to access only limited features for each instance. Differential Evolutionary (DE) algorithm is designed to use different reasons, and different models and also its advantages and disadvantages.

Differential evolution is a method that optimizes a problem iteratively to try to improve a solution about a given measure of quality. Such methods are commonly known as no supposition about the problem being optimized and can find a very large area of dataset. Feature selection has found many applications or uses in many domains or fields, especially for problems involving high-dimensional data. Such suppositions may not always be appropriate for real-world examples in which training examples arrive in order it is expensive to gather the all information of training data.

Bioinformatics is another example of feature selection, where acquiring the complete set of features/attributes for every iteration is expensive due to the high cost of conducting experiments. Finding the relationship between two or more features here the correlation is best. It is one of the statistical classes of statistical relationship involving dependence though in common usage. For clustering purposes Nearest Neighboring Algorithm is used it is easy to implement and executes quickly. Data mining is also a frequently used method for analyzing road accident data in present research. The motive of this study is to provide an efficient way to choose the best suitable distance metric to cluster the series of counts data that provide a better clustering result [13]. The online and batch algorithms are robust to data with missing features, a situation that arises in many practical applications. In the online setup, the comparison hypothesis changes as a function of the subset of features that views on any given round, extending the standard setting where the comparison hypothesis is fixed throughout. In the batch setup, a convex relaxation of a non-convex problem to estimate an imputation function is used to fill in the values of missing features, along with the classification hypothesis [14].

II. RELATED WORK

This work is closely related to the studies of online learning of two task and feature selection in literature. Here below is a review of important related works in both fields. One OFS: Online Learning 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 is used to find the relation between two or more features. Correlation is a statistical method that can show whether and how strongly a pair of features is related to each other. For example, height and weight both were related; taller people look after being heavier than shorter people. The relationship is perfect. Correlation is useful because it can indicate a predictive relationship that is derived in practice. For example, an electrical utility may produce less power on a clement day based on the correlation between electricity demand and weather. These examples are related to each other, because extreme weather causes people to use more electricity for heating or cooling; however, correlation is not enough to demonstrate the presence of such a causal relationship (i.e., correlation does not imply causation). The dependence refers to any condition in which random variables do not satisfy mathematical conditions or methods of probabilistic independence.

In loose consumption, correlation can refer to any departure of many more random variables from independence, but technically it refers to any of several particular types of relationship between mean values. There are several of correlation coefficients, often denoted r, measuring the degree of correlation.

The common of the Pearson Correlation coefficient is quick response only to a linear relationship between one or two features(which may also exist even if one is a nonlinear function of the other). Alternative correlation coefficients have been developing more robust than Pearson correlation – that is, more sensitive to nonlinear relationships.

For a combination of most related features there is clustering technique used which contains 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. There are two different types for online feature selection tasks

1. OFS by learning with full input reading, and
2. OFS by learning with partial input reading.

For the first task, here consider that learner can access all the features of training instances, and our goal is to effectively identify a fixed number of related features for accurate prediction. In the second task, we assume more challenging scenario where the learner has allowed accessing a fixed minimum number of features for each training data instance to find the subset of relevant 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]. To achieve the best performance with a particular learning algorithm on a particular training set, a feature selection method should consider how the algorithm and the training set interact. The relation between optimal feature subset selection and relevance wrapper method searches for an optimal feature subset tailored to a particular algorithm and a domain [7]. In addition to the large application of techniques that have already been implemented in the machine learning and data mining fields, particular applications in bioinformatics have led to a wealth of newly proposed techniques [8]. The three variants of budgeted learning, are a setting in which the learner is allowed to access a limited number of features from training or test examples. In the “local budget” setting, where a constraint is imposed on the number of available features per training example, the design and analysis of an efficient algorithm for learning linear predictors that actively sample the attributes of each training instance. The second is the “global budget” setting, where the overall number of accessible training features is constrained. The third one is the “prediction on a budget” setting [10].

III. DIFFERENTIAL EVOLUTIONARY ALGORITHM

The objective of feature reduction in the analysis of a population-based dataset for which there were no specific target variables or features. All attributes or features were found as possible targets in models derived from the full dataset and from subsets of it. The feature selection methods used were of the filter type and wrapper type as well as clustering techniques.

The predictive accuracy and complexity of models based on the reduced datasets for every technique are compared both amongst the technique and with those of the complete dataset. Analysis shows a marked similarity in the correlated features chosen by the supervised (filter) technique and moderate consistency in those chosen by the clustering techniques (unsupervised). The breadth of distribution of the correlated features amongst the feature groups is related in large part to the number of features selected by the given algorithm or elected by the user [9]. The Differential Evolutionary (DE) algorithm is a population-based algorithm similar to genetic algorithms and they use the same operators; crossover, mutation, and selection. The main difference in finding the better solution is that genetic algorithms rely on crossover function and DE relies on mutation operation. This main work is based on the differences between randomly selected pairs of solutions in the large data. The algorithm uses mutation operation as a search method and selection operation to direct the search toward the prospective area in the search space. By using the components of the current data members to create trial vectors, the crossover is nothing but recombination. The Crossover operator shuffles data or information about successful combinations, enabling the search for the best solution space. An optimization process consisting of D parameters represent by a D-dimensional search space. In DE, a data of NP solution data is randomly created at the start. This solution is successfully improved by applying mutation, crossover, and selection operators. DE algorithm also uses a crossover that can take child node parameters from one parent node more often than it is done from others. The important steps of the DE algorithm are given below:

1. Initialization Evaluation
2. Repeat
3. Mutation Recombination
4. Evaluation Selection Until (termination criteria are met)

Global optimization is needed in fields such as engineering, statistics, and finance but many practical problems have objective functions that are no differentiable, non-continuous, non-linear, noisy, flat and multi-dimensional or have many local minima, constraints or stochasticity some problems are difficult if not impossible to solve analytically. DE is used to find approximate solutions to such large-scale problems:

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

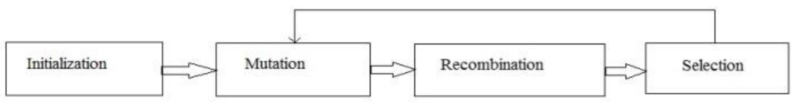

Fig 1: General Evolutionary Algorithm Procedure

3.1 HOW TO CALCULATE CORRELATION MATRIX

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

The sum of the squared matrix

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

Where,

n=N * N Matrix value

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

$SS_{XY} = \Sigma(X_i - \overline{X}) * \Sigma(Y_i - \overline{Y})$ (2)

Similarly

$SS_{XZ} = \Sigma(Y_i - \overline{Y}) * \Sigma(Z_i - \overline{Z})$ (3)

Correlation Matrix

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

Where,

$n = N * N$

$P_{XY} = SS_{XY} / \sqrt{(SS_{XX} * SS_{YY})}$ (4)

Based on eq. (1), (2), (3) and eq. (4) the correlation matrix is calculated that can be used at the time of DE optimization. That contains the relations between variables or features. Variable and feature selection have become the focus of much research in areas of application for which datasets with hundreds or thousands of variables are available. These areas include text processing of internet documents, gene expression array analysis, and combinatorial chemistry. The objective of variable selection or feature selection comes for three reasons: to improve the prediction performance of the predictors, to provide faster and more cost-effective predictors, and to provide a better understanding of the underlying process that generated the data [6].

3.2 ONLINE FEATURE SELECTION

Online feature selection is particularly important and needed when a real-world application has to deal with a series of training data with high dimensionality, such as online spam classification tasks, where traditional batch feature selection techniques cannot be applied directly. In this paper, we understand the problem of online feature selection for Multiclass Classification. Using the DE algorithm, the feature can selected into two tasks.

3.2.1 OFS: LEARNING WITH FULL INPUT

To motivate this algorithm, first present a simple but ineffective algorithm that simply select the features. The failures of this simple algorithm induce us to develop effective algorithms for OFS. The goal is to efficiently find a fixed number of relevant features for accurate prediction.

3.2.2 OFS: LEARNING WITH PARTIAL INPUT

In the second task of the OFS algorithm, a more challenging structure has been understood where the learner is allowed to access a fixed number of features for each instance to find the subset of relevant features. This task contains partial data for each training instance. In the form of percentages, the data is given as shown in the following example.

IV. RESULTS

In this section, an extensive set of experiments to evaluate the performance of the proposed online feature selection algorithms. First, evaluate the online predictive performance of the two OFS tasks on several data sets from the machine learning repository. In this, the multiclass stock market data set has been considered for demonstrating the result of the OFS technique. The dataset is related to the stock market in that daily up, and down values are stored. The data should be in numerical form. That data will distribute in four fields. The class values are hold, buy, and sell. There are two tasks where the OFS runs:

	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 Mahi	YES Bank I	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	16.97	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	9.16	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	36.94	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

4.1. OFS LEARNING WITH PARTIAL INPUT

In this task, the user or learner can enter any percentage of data. After entering any percentage of data we get approximate same accuracy.

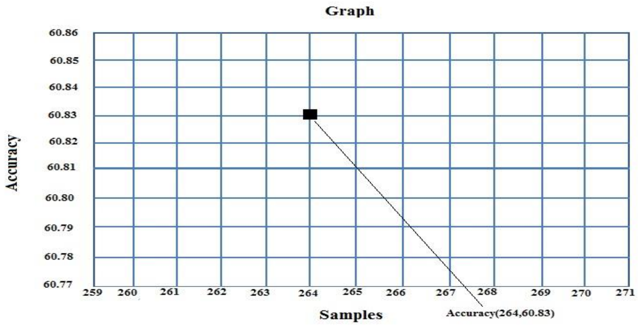

Fig.3: Accuracy graph for 90% of data

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

In Fig.3 shows the accuracy of selected features within the stock market dataset. The data that is under 90% is considered 264 rows which has been taken for 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 depending on clusters DE algorithm optimizes the values within in large dataset. The following Table 1 shows the different percentage of data that gives the same accuracy so selected features accuracy is fixed for any number percentage of data. The Fig.5 shows the use of selected features. For any purpose that is analysis, prediction or any application development selected features are used.

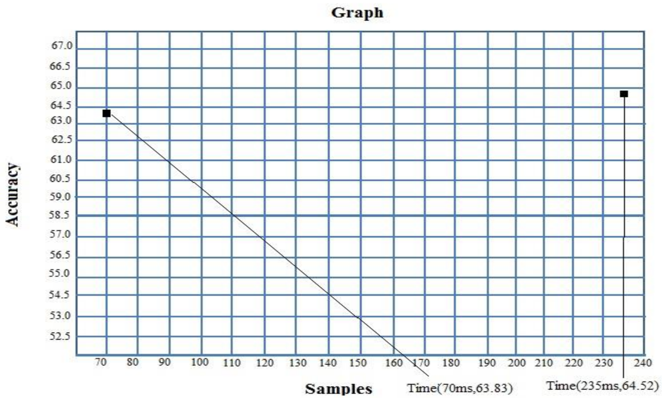

Fig.4: Displaying selected or remaining all features accuracy and their processing time.

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 the second plot shows the time and accuracy for all features within the dataset. The above graph shows the comparison of selected features' accuracy and all features accuracy which is approximately near to each other. Here using the OFS technique if their accuracy is nearly the same then, the user can easily select the minimum features that are used for any purpose.

By using this technique, improving the performance of the prediction model and also increasing the speed of the processing model is easy. User also reduces the complexity of the prediction model.

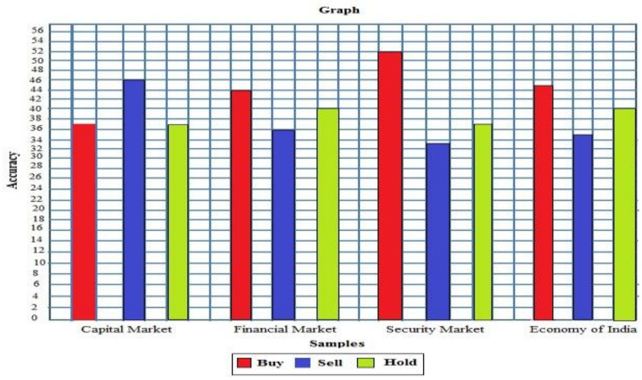

Fig.5: Selected features used for analysis.

In the stock market, daily updates are arrived for every company. The values changed tries in the 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 tasks of the OFS method we can select the minimum features that are nothing but minimum companies and calculate their accuracy if their accuracy is near to 100%. Then it gives approximate prediction on buying, Sell, and Holding of the data which is shown in Fig 4. Table 1 contains the percentage of data and the no. of rows within the dataset that is use for processing the OFS techniques. The minimum average accuracy for any percentage of data that is near 60% and their processing time is also near to each other which is calculated in macro seconds.

4.2. OFS Learning with Full Input

This task, like partial input learning 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 feature's 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.

V. CONCLUSION

This paper introduces a research problem, online feature selection which selects a small and fixed number of features for multiclass classification. In particular, it accepts two types of OFS tasks in two different settings: 1) OFS by learning with full inputs of all the features/attributes, and 2) OFS by learning with partial inputs of the attributes in large-scale approach. The novel OFS algorithms to solve each of the OFS tasks, and offered theoretical analysis on the numerical data bounds of the proposed OFS algorithms. It also extensively examined their performance and applied the proposed techniques to solve real-world applications: neural network, microarray gene expression, CPU Performance: Introducing Numeric Prediction in computer vision and microarray gene expression analysis in bioinformatics. The result analysis shows that algorithm is fairly effective for feature selection tasks of many online applications, and more efficient and scalable as compare to batch feature selection technique. Future work could extend this framework to other settings, such as online feature selection for numerical, textual classification and it also extends for image classification.

If desired, a section of acknowledgment can be included following the conclusion.

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 with Missing Features," Proc. Conf. Uncertainty in Artificial Intelligence(UAI '11), pp. 635-642, 2011.

Prof. A. S. Shimpi working as an Assistant Professor, at the Department of Computer Science and Engineering, BMIT Solapur Email Id - aanandshimpi@gmail.com

Prof. S. A. Awatade working as an Assistant Professor, at the Department of Computer Science and Engineering, BMIT Solapur. Email Id -sharad.awatade@gmail.com