ISSN:0975-9646

Effective Method of Feature Selection on Features Possessing Group Structure

Nayana Murkute[#1], Prof. Prashant Borkar[*2]

[#]*Department of computer science,*
G.H. Raisoni College of Engineering,
Nagpur University, Nagpur, India

***Abstract*— Feature selection has become an interesting research topic in recent years. It is an effective method to tackle the data with high dimension. The underlying structure has been ignored by the previous feature selection method and it determines the feature individually. Considering this fact we focus on the problem where feature possess some group structure. To solve this problem we present group feature selection method at group level to execute feature selection. Its objective is to execute the feature selection in within the group and between the group of features that select discriminative features and remove redundant features to obtain optimal subset. We demonstrate our method on benchmark data sets and perform the task to achieve classification accuracy.**

***Keywords*— feature selection, group structure, redundant, classification.**

I. INTRODUCTION

Searching hidden information and pattern from large database is the task of data mining [1]. High dimensionality is become a curse for data mining which arises problem while training the data. The curse of dimensionality can be minimizing by using feature selection. The step of searching an optimal variable subset from actual feature set is a feature selection [2]. The application in which there are large numbers of variable the feature selection is enforced to minimize the variable. The goal of feature selection is to search a relevant feature that is useful for target output. It eliminates the irrelevant and redundant feature from original feature sets. Relevant feature are those that provide useful or meaningful information and vice versa and redundant feature are those that is not useful than the selected features. So feature selection is an important process in efficient learning of large multi-feature data sets. There is some potential advantage of feature selection. It facilitate data visualization, increases data predictability and understanding .It also help to reduce the measurement and storage requirement, reduces training and processing time. Feature selection can be used in many applications such as gene selection, intrusion detection, text categorization, image retrieval, DNA microarray analysis, information retrieval etc. It enhance the literature efficiency, increases anticipating certainty and help to minimizing learned result complexity [3]. The feature selection algorithm generates an output as a subset of feature or by measuring their utility of feature with weights. The assessment of features in feature selection can be in various forms such as separability, consistency, dependency, information and training model which are generally occurred in wrapper model.

Previously feature selection method evaluates or select feature individually and avoids selecting feature from groups. It is always better to select features from group rather than selecting feature individually [4]. This help to increases accuracy and decreases computational time. The aim of a feature selection is to search the vital exploratory, whereas the vital exploratory is shown by a collection of input variables. Therefore in some situation finding a vital feature equivalent to the evaluating a group of feature. The group of variable must take an advantage of group structure while selecting an important variable.

Features can be selected from the available candidate feature set through many feature selection methods efficiently. However, they always tend to select feature at individual level with small percentage (sparsity), more preferably than the group structure. When group structure exists, it is more preferable to select features with small percentage at a group level rather than individual level. We address the problem of selecting the features from group. So we consider the problem that feature possesses some group structure, which is potent in many real world application and its common example is Multifactor Analysis of Variance (ANOVA). ANOVA is a set of learning model applied to examine the difference among group means and correlated procedures that is variation among and between the groups

Group structure can appears in different modelling goal for many reasons. Grouping can be introduced into model to take benefits of prior knowledge that is significant. Example such as, in gene expression analysis, the matches to the same categories can be known as group. In data analysis it is desirable to consider about the group structure. In some condition, the individual features in group may or me not be useful, if this features are useful then we are not interested in selecting an important features in this case group selection is our objective. But if individual features are useful then we are interested in selecting an important features and important group.

This paper developed an efficient group feature selection method; the main challenge is that they are in with group structure. In paper, we propose a new group feature selection method named as efficient group variable selection (EGVS).This comprises of two stage, within group variable selection that select discriminative features within the group. In this stage each feature is evaluated individually. After within group selection all the features are re-evaluated so far to remove redundancy this stage is known as between group variable selection.

This paper is constructed as follow, section II describe various feature selection approaches and gives review on existing literature on underlying group structure such as, group lasso. While in section III we proposed our work and section IV, we show our empirical experimental result section V proposed the conclusion.

II. FEATURE SELECTION METHODS

The feature selection method is broadly divided into three classes based on their label information is used; most commonly used method label variable. In supervised feature selection there is difficulty in acquiring the data label. In recent year unsupervised feature selection have gain more attention.

Unsupervised feature selection typically select features preserving the data similarity multiple (manifold) structure [5] whereas semi supervised feature selection, make use of label information and multiple structure related to labelled data and unlabelled data .There are 3 types of approaches for feature selection, filter, wrapper, embedded method. Filter method does not involve a learning algorithm for measuring feature subset [6]. It is fast and efficient for computation. filter method can fail to select the feature that are not beneficial by themselves but can be very beneficial when unite with others. wrapper method involve learning algorithm and search for optimal attribute subset from original attribute set which discover relationship between relevance and optimal subset selection. In embedded method is a combination of wrapper method. This decreases the computational cost than wrapper approach and captures feature dependencies. It searches locally for features that allow better discrimination and also the relationship between the input feature and the targeted feature. It involves the learning algorithm which is used to select optimal subset among the original subset with different cardinality [7]. Many analysts have focuses on a feature that contain certain group structure such group lasso. The group lasso applied L2 norm of the coefficient joined in the penalty function by a collection of features. An extended form of Lasso is a group lasso. It simplifies the standard lasso. Many authors have studied the various property of group lasso by building the many approaches of lasso. Yuan and Lin [8] have demonstrated the group Lasso used to solve the problem of convex optimization that consider for size of varying group and applied Euclidean norm. This process acts as a lasso at group level, whereas if the sizes of group are same, then it reduces to the lasso. The author proposed the method for fitting the group lasso that considers the model matrices in each groups are orthonormal. Whereas in non-orthonormal case, it uses the rigid regression to handle the groups of variable. Mieere [9] proposed the method for logistic regression to extend the group lasso. Suhrid Balakrishnan and David Madigan [10] unite the idea from group lasso Yaun and Lin [8] and fused Lasso. The Bakin [11] proposed the group Lasso and computational algorithm. This method related group selection method and algorithm are further developed by Yuan and Lin [8]. Composite absolute penalty (CAP) approach developed by Zhao Rocha [12] is same as group lasso but instead of using L2 norm it uses L1 norm the group information in CAP method consider the group lasso and combine the group penalty for Lr0 norm. It does not imply any information but the grouping information. CAP method includes the group Lasso as special case. Huang Ma, Xie and Zhang [13] developed a method called Group Bridge that developed a method called Group Bridge that determines the simultaneous group and individual feature problem. Theren Bach [14] demonstrated the group Lasso for random designed model where group selection is consistent. Meinshausen and Buhlmann [15], Zhoa and Yu [16], Zou [17], Nardi and Rinaldo[18], considers the consistency of selecting a features of group lasso to under the not represent able requirement and the bounds on the prediction and to estimate the error under Eigen value condition. Wei and Huang [19] assumed the L2 bound on estimation and sparsity and error of prediction of group Lasso using sparse Riesz condition. Zhang and Huang [20] examine the feature selection and proposed the adaptive group lasso by using group lasso to estimate initially. The adaptive group lasso expressed in similar way to a standard adaptive lasso. H. Yang, Z. Xu, I. King, and M. R. Lyu [21]Have established an online group lasso algorithm to find necessary exploratory factor or variable in group manner and shows the drawback of traditional batch mode group lasso algorithm .where in advanced the data is given, and that handle the data which have several hundreds or thousands instance this limitation of batch mode of group LASSO with sparsity by selecting a feature in group level and provide a close –form solution for group lasso with L1-norm regularization. Lei Yuan, Jun Liu, and Jieping Ye [22]has propose overlapping group Lasso ,it make use of L1-norm regularization and L2-norm for group features for the overlapping group by using Accelerated gradient descent (AGD)method. This method formed a low cost processing step from the proximal operation that find and discards the null group. Volker Roth, Bernd Fischer [23] work on to handle extremely high dimensional input feature space. They presented an active set of procedure to deal with the problem of Group-LASSO estimation for Generalized Linear Model (GLM) that discover all groups which is important features for active set. The method checks the completeness and uniqueness of features. Within group sparsity doesn't field by group lasso. This selects a model that is larger than the underlying model with relatively high false positive group selection rate.

III. THE PROPOSED METHOD

An Efficient group variable selection method (EGVS) is proposed in this section for group of feature. From domain knowledge we can obtain a group structure or a user can provide the group size that help to reduce the efficiency of time. Our aim is to find an optimal subset from a group.

An Efficient group variable selection framework consists of two stages: within group variable selection and between group variable selection. At first we need to create a group of features, the group of feature generated by randomly dividing the feature space. Within group variable selection aim to find discriminative feature, the feature are determined one at a time. After within group variable selection all the features are re-evaluated to find the correlation between the groups to find an optimal subset, namely as between group variable selection.

In figure1 we have shown the overview of efficient group variable selection (EGVS). So in further section we provide the information of procedure.

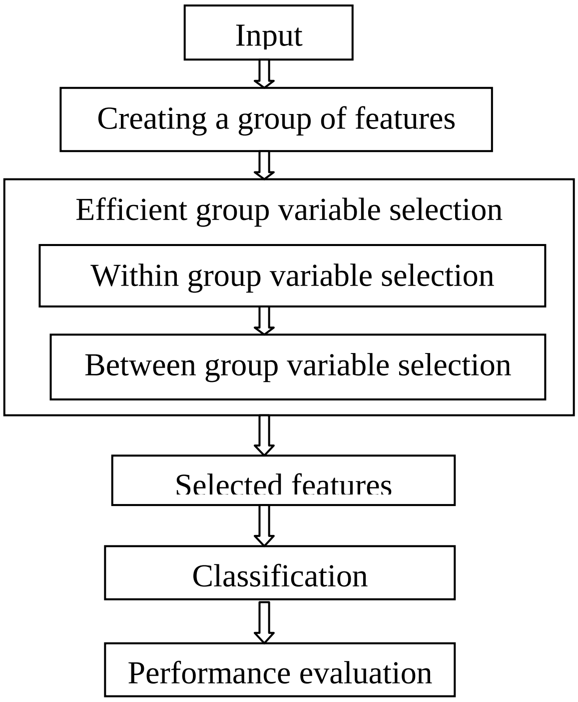

Fig 1. Flow of work for efficient group variable selection approach.

A. *Within Group Variable Selection*

In within group selection, to select a discriminative feature from group we have used mutual information method; mutual information discovers the relevance in the two random variable [24]. The feature is declared to be irrelevant if their mutual information found zero and shows both the random variable are independent of each other [25]. To gain the correlation among the features mutual information applied correlation coefficient of feature and then it gives the scores to the features. The feature that have the higher value or score or above the threshold value that feature will be defined as a relevant feature. In mutual information if the feature has higher mutual information scores depict more information about the feature label and shows more relevance.

Let assume term Fi and the T is target and the X is the number of count Fi and T co-occurs, Y is the number of count the Fi occurs without T, C is the number of count T occurs without Fi .N is total features. The Mutual information for between one feature Fi and T is considered as.

$$I\,(Fi, T) = \log \frac{P(Fi \wedge T)}{P(Fi) \times P(T)} \qquad (1)$$

evaluated as,

$$I\,(Fi, T) = \log \frac{X \times N}{(X+C) \times (X+Y)} \qquad (2)$$

If the value of I(Fi, T) is zero then it shows that they shares no information and considered as irrelevant, in feature selection to evaluate the goodness of variable, the score is assign to features in two substitute way:

$$I_{avg}\,(Fi) = \sum_{i=1}^{m} P(Ti)\, I(Fi, Ti) \qquad (3)$$

$$I_{max}\,(Fi) = \underbrace{\max^{m}}_{i=1} \{\, I(Fi, Ti) \,\} \qquad (4)$$

From equation (3) and equation (4) the feature that have the maximum scores is depict as the relevant in mutual information.

B. *Between Group Variable Selections*

The information of group is not considered in within group variable selection and only calculates the features one at time. In within group selection it select relevant feature from every group but there may be probability of consisting redundant feature so to remove redundant feature the between group feature selection re-evaluated the entire feature and find the optimal subset. The sparse group lasso is used to minimize the error and penalty. The unique case of group lasso is sparse group lasso that place an additional penalty 1-norm of the coefficient vector. It allows the overlaps in the groups.it generate the sparse set of groups. The models in the group if included then all the variables in the group become non-zero. Sometimes we like to have both the sparsity of group and within each group, i.e. between and within the group. Such as example, in genes expression we like to identify particular “useful” genes among the number of genes, this is the focus of sparse group lasso. The sparsity is exists in two types, first is group wise sparsity and the other is within group sparsity. The number of group with least one nonzero coefficient is referred in group wise sparsity and the number of nonzero coefficient within every nonzero group is referred in within group sparsity. The sparse group lasso uses L1 L2 norm regularization penalty function. It considers the global group information and performs the feature selection. In sparse group lasso feature space is important for underlying group structure and considers the correlation structure in feature space. In between group variable selection to decrease reorganization fault with sparsity constrain on coefficient of attribute is objective.

In sparse group lasso, feature space is divided in G groups and consists of regularization parameter. The sparsity of selected feature is modulated by regularization parameter. In natural way the regularization factor will lead sparser model by decreasing its value. After between group variable selections subset of optimal features is obtained. By the merging of within group variable selection and between group variable selections, the procedure of efficient group variable selection is formulated. The within group variable selection select discriminative feature and compactness control in between group variable selection can give benefit to the efficient group variable selection.

IV. EXPERIMENT

We have shown the effectiveness of our method in this section. We have performed the experiment on datasets to determine the effectiveness of our procedure the datasets are taken from UCI machine learning repository. The employed evaluation matrices are compactness and classification accuracy. Percentage of selected features defines the compactness. The factor of accuracy is shown on the basis of classification for the dataset depends on selected feature space. The dataset, in which feature with no grouping then we randomly partitioned the feature space like F= [G1, G2…Gi] with the dimension. This experiment helps us to measure the effectiveness of efficient group variable selection when the no information of natural group is there.

TABLE I

DESCRIPTION OF UCI DATASETS

Data sets	No. of classes	No. of instance	Features
Ionosphere	2	351	34
Wdbc	2	569	31
Statlog(heart)	2	270	13

TABLE II

EXPERIMENTAL RESULT BY PERFORMING EFFICIENT GROUP VARIABLE SELECTION METHOD ON UCI DATA SETS

Data sets	No. of instance	Features	Selected features by EGVS
Ionosphere	351	34	18
Wdbc	569	31	12
Statlog(heart)	270	13	08

TABLE III

EXPERIMENTAL RESULT BY PERFORMING CLASSIFICATION

Data sets	Classification accuracy on all features	Classification accuracy on selected features
Ionosphere	52%	62.85%
Wdbc	91%	92.3%
Statlog(heart)	77.02%	77.77%

A. *Experimental Result on data sets*

Table 1 provides the description of the datasets which we have taken from UCI machine learning i.e. ionosphere and Wdbc and Statlog(heart). In this dataset information of group is not given we have divided the feature space randomly to create a groups. In table 2 we have shown the result of efficient group variable selection method. After the feature selection on the final selected feature the Neuro-fuzzy classifier is used to measure the performance of feature space. In neuro-fuzzy classifier the half part of both the classes are given for training. And rest of the part is given for testing [29]. By applying the neuro-fuzzy classifier on all 34 features it achieve 52% accuracy where as it give 62.85% accuracy on selected features. In Wdbc data set the final selected features is 12 from original 30 features, it give the 91.19% accuracy on original features and on final selected features it gives 92% accuracy. And in Statlog (heart) dataset from 13 features 8 features are selected and achieves 77.02% accuracy on all features 77.77% accuracy is acquire on selected features. The performance analysis is based on compactness and classification accuracy.. Table 3 shows the classification accuracy by using neuro-fuzzy classifier. The Neuro-fuzzy classifier gives the accuracy as the final result. On this data set effective group variable selection gives the better compactness and also able to find the discriminative features. On this three data set our method obtain gain accuracy, because the previously selected features are revaluated that provide us the optimal subset of features with discriminative features. It also provides the better correlation of features and more effective and obtains relatively better classification accuracy.

V. CONCLUSION

We have presented efficient group variable selection for group of features. Method focuses on the problem where feature comprise some group structure. We also provide the literature reviews on existing method. We divided the efficient group variable selection into two stages, i.e., within group variable selection and between group variable selections. In within group variable selection uses mutual information and introduces the sparse group lasso to minimize the redundancy in between group variable selection. The within group variable selection effectively select discriminative feature, in this step each feature is evaluated individually. Between group selection controls the compactness and revaluate the features. We have also demonstrated the experiment on several UCI benchmark data sets. This increases the classification accuracy and shows the effectiveness of our method.

REFERENCE

[1] X. Wu, X. Zhu, G.Q. Wu, and W. Ding, "Data mining with big data," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107, 2014.

[2] Guyon and A. Elisseeff. "An introduction to variable and feature selection," Journal of Machine Learning Research, 3:1157–1182, 2003.

[3] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," The Journal of Machine Learning Research, vol. 5, pp. 1205–1224, 2004.

[4] Haiguang Li, Xindong Wu, Zhao Li, Wei ding"Group feature selection with streaming features," IEEE 13th international conference on data mining. 2013.

[5] Jennifer G. Dy, Carla E. Brodley "Feature Selection for Unsupervised Learning," Journal of Machine Learning Research, 845–889.2004.

[6] H. Liu and H. Motoda, "Computational methods of feature selection," CRC Press, 2007.

[7] Daphne Koller, Mehran Sahami, "Toward Optimal Feature Selection," Computer Science Department, Stanford University, Stanford, CA 94305-9010.1996.

[8] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," Journal of the Royal Statistical Society, vol. 68, no. 1, pp. 49–67, 2006.

[9] Meier L., Van De Geer, S., & Buhlmann P. "The Group Lasso for Logistic Regression," J. Roy. Stat. Soc.B, 70, 53–71.2008.

[10] Suhrid Balakrishnan and David Madigan, "Finding predictive runs with LAPS" 7TH IEEE conference on Data mining, 2007.

[11] S.Bakin. “Adaptive regression and model selection in data mining problems,” Ph.D. thesis, Australian National Univ., Canberra. 1999.
[12] Zhao, P., Rocha, G. and Yu, B. “The composite ab-solute penalties family for grouped and hierarchical variable selection,” Annals of Statistics , Vol. 37, No. 6A, 3468-3497.2009.
[13] Huang, J., Ma, S., Xie, H. and Zhang, C.-H “A group bridge approach for variable selection,” Biometrika, 96 339–355. 2009.
[14] Bach F. R. “Consistency of the group lasso and multiple kernel learning,” Journal of Machine Learning Res. 9 1179–1225.2009.
[15] N. Meinshausen and P. Buhlmann “High-dimensional graphs and variable selection with the lasso,” Annal of Statistic.,34 1436–1462.2006.
[16] Zhao.P. and Yu.B. “On model selection consistency of Lasso,” Journal of Machine Learning,” Res. 7 2541–2563. 2006
[17] H. Zou “The adaptive lasso and its oracle properties”.J. Amer. Statist. Assoc.2006.
[18] Wei. F. and Huang. J. “Consistent group selection in high-dimensional linear regression,” Bernoulli 16 1369–1384. 2010.
[19] Zhang, C.-H. and Huang, J. “sparsity and bias of the LASSO selection in high-dimensional linear regression,” The Annals of Statistic. 36 1567–1594.2008.
[20] Lei Yuan, Jun Liu, and Jieping Ye, “Efficient method for overlapping group lasso,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 9, September 2013.
[21] S. Xiang, X. T. Shen, and J. P. Ye, “Efficient sparse group features election via nonconvex optimization,” in ICML, 2012.
[22] Seyoung Kim, Eric P. Xing, “Tree-Guided Group Lasso for Multi-Task Regression with Structured Sparsity,” in ICML, 2010.
[23] Roth.V. and Fischer. B. “The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms,” In ICML, pp. 848–855, 2008.
[24] Hanchuan Peng, Fuhui Long, and Chris Ding, "Feature selection based on mutual information criteria of Max dependency, max relevance and min redundancy,” IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 8, august 2015.
[25] Erik Schaffernicht and Horst-Michael Gross “Weighted Mutual Information for Feature Selection”21 international conference on artificial neural network(ICANN 2011),Espoo, Finland, LNCS 6792,pp. 181-188, Springer 2011.
[26] Noah Simon, Jerome Friedman, Trevor Hastie, and Rob Tibshirani. “A sparse-group lasso,” Journal of Computational ssand Graphical Statistics. May 2011.
[27] S. Xiang, X. T. Shen, and J. P. Ye, “Efficient sparse group features election via non convex optimization,” in ICML, 2012
[28] UCI Machine Learning Repository [Online]. Available: http://archive.ics.uci.edu/ml/datasets.html.
[29] Prashant Borkar, M.V.Sarode,Latesh Malik, “Modality of adaptive neuro-fuzzy classifier for acoustic signal-based traffic density state estimation employing linguistic hedges for feature selection” , Int. J. Fuzzy Syst, Springer 2015