Dimension Reduction Methodology using Group Feature Selection

Shrutika Kolhe[1], Prof. Prarthana Deshkar[2]
[1, 2] Department of Computer Science and Engineering
[1, 2] Yeshwantrao Chavhan College of Engineering, India

Abstract- *Feature selection has become a remarkable research topic in recent years. It is an efficient methodology to tackle the information with high dimension. The underlying structure has been neglected by the previous feature choice technique and it determines the feature singly. Considering this truth, we are going to focus on the matter wherever feature possess some cluster structure. To resolve this downside we are using cluster feature selection technique at cluster level to execute feature choice. Its objective is to execute the feature within the cluster and between the cluster that choose discriminative features and take away redundant options to get optimum subset. We have demonstrate our technique on benchmark knowledge sets and perform the task to attain classification accuracy.*

Keywords- Data mining, Dimension reduction, Group feature selection, KNN classifier.

I. INTRODUCTION

Searching hidden data and pattern from massive database is the task {of data|of knowledge|of data} mining [1]. High spatiality is becoming a bane for data mining that arises difficulty while training the data. The bane of spatiality will be minimizing by using feature selection. The step of searching associates best variable set from actual feature set is a feature choice [2]. The application where large number of variables are present then the feature choice is implemented to minimize the variable. The goal of feature choice is to search a relevant feature that's helpful for target output. It eliminates the orthogonal and repetitive feature from original feature sets. Relevant feature are those who give useful or significant data and repetitive feature are those who isn't helpful than the selected options. Therefore feature choice is a crucial process in economical learning of huge multi-feature information sets. There is some noumenon advantage of feature choice. It facilitate information visualization, increases information predictability and understanding. It conjointly facilitate to scale back the measurement and storage demand, reduces training and interval. Feature choice will be utilized in several applications like gene choice, intrusion detection, text categorization, image retrieval, deoxyribonucleic acid microarray analysis, information retrieval etc. It enhance the literature efficiency, increases anticipating certainty and facilitate to minimizing learned result complexity [3]. The feature choice algorithmic rule generates an output as a set of feature or by measuring their utility of feature with weights. The assessment of options in feature choice will be in numerous forms like separable, consistency, dependency, data and training model that area unit usually occurred in wrapper model.

Previously feature selection methodology evaluates or choose feature separately and avoids choosing feature from groups. It is forever higher to pick out options from cluster rather than choosing feature severally [4]. This facilitate to increase accuracy and reduce process time. The aim of a feature selection is to look the important exploratory, whereas the important exploratory is shown by a group of input variables. So in some state of affairs finding an important feature corresponding to the evaluating a bunch of feature. The group of variable should take an advantage of cluster structure while choosing a very important variable.

Features may be selected from the offered candidate feature set through several feature choice strategies efficiently. However, they tend to pick out feature at individual level with tiny percentage (sparsity), more preferably than the cluster structure. Once cluster structure exists, it's additional desirable to pick out options with tiny percentage at a group level instead of individual level. We address the matter of choosing the options from cluster. So we take into account that feature possesses some group structure, that is potent in several universe application and its common example is Multifactor Analysis of Variance (ANOVA). Analysis of variance may be a set of learning model applied to look at the distinction among group suggests that and related procedures that's variation among and between the teams.

In several modelling, it seems that there are many reasons to deal with group structure. Grouping may be introduced into model to take benefits of previous data that's important. Example like, in organic phenomenon analysis, the matches to identical classes may be referred to as cluster. In data analysis it's fascinating to think about the cluster structure. In some condition, the individual options in cluster might or may not be helpful, if this options are a unit helpful then we aren't interested in choosing a crucial options during this case group choice is our objective. However if individual options are helpful then we tend to have an interest in choosing a crucial features and vital cluster.

This paper developed an economical cluster feature selection method; the most challenge is that they're in with group structure. In paper, we tend to propose a new cluster feature selection technique named as economical cluster variable selection (EGVS).It includes 2 stages, within group variable choice that choose discriminative options within the cluster. During this stage every feature is evaluated individually. Once among cluster choice all the options are re-evaluated up till now to get rid of redundancy, this stage is known as between cluster variable choice.

II. METHODOLOGY

The main idea behind this chapter is to give brief idea about performing feature selection for the group of features. The overall design approach is basically divided into several steps. The first step is input data sets is used which is available from UCI machine learning repository datasets for feature selection. The three datasets are used i.e Ionosphere, Wdbc, Statlog (heart).The data sets which is being used have not provide any group information. Creating the group of features is the second steps. The group of features is created by dividing the feature randomly. The size of group is depending on the user choice. This step gives the group of feature.

Next step is performing feature selection on group of features, We focus on the problem where feature possessing some group structure, to solve this problem we propose a framework for group feature selection it consist of two stages: intra group feature selection and inter group feature selection. The discriminative features are evaluated in intra group feature selection. The features are evaluated one at a time in this stage and the features are selected within the group. After intra group feature selection all the features are reevaluated to find the correlation between the group to find an optimal subset, namely as inter group selection. This step gives the optimal subsets of features. The validation is needed on the selected feature in order to evaluate whether the features are optimal or not classification is required. The KNN classifier is applied to evaluate the performance of selected.

A. Design Approach

The figure 3.1 gives the details scenario of proposed group feature selection system which shows the overall way to perform the feature selection and performance analysis on selected features. This approach is further divided in sub-task which is explained in the following sections.

Figure shows our proposed plan of work.

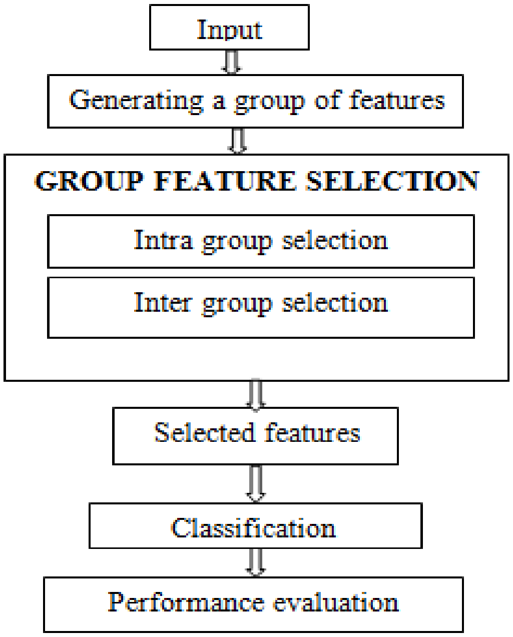


Figure 1. Proposed work models

B. Datasets Description

Our first step is input of data sets. For the feature selection three data set is used to further verify the effectiveness of our method, the datasets are ionosphere, Wdbc, Statlog(heart) are available from UCI datasets[5]

a) Ionosphere data set is a radar dataset, which consist of 34 instance, 34 attribute and 2classes this dataset shows some structure of good ionosphere and bad ionosphere. Mostly in numeric form.
b) wdbc is referred as Wisconsin diagnostic breast cancer consist 30 attributes and 569 instance. There are two classes malignant and benign.
c) The Stat-log dataset is a heart disease data set consists of 13 attribute and 270 instances there are two classes. In all three dataset there is no any grouping information is given.

Table 1. Dataset description for dataset used in EGFS

Data sets	No of classes	No of instance	No of features
Ionosphere	2	351	34
Wdbc	2	569	30
Statlog(heart)	2	270	13

C. Generating a group of features

The group of features is created by randomly dividing the feature space. The feature that are present in one group cannot be appeared in other group, each and every group have different features. The group can be user specified to minimize the time of computation. For example if we have the no of features 34 and need to create 4 groups. The formation of group is as 34mod4=2 so in each group 8 features will be assigned and remaining 2 features will be adjusted in first two group, so final groups will be G1=9, G2=9, G3=8, G4=8. This is how the group of features is created. After creating the group of features the next step is performing the features selection.

D. Group feature selection mechanism

After creating the group of features the next step is performing the features selection in this step we proposed the group feature selection method called as effective group feature selection (EGFS). The EGFS is further divided in two vital stages. The discriminative feature is determined in first stage called as intra group feature selection. In this stage each feature is evaluated one at a time in each group and performs the correlation within the features. The features are evaluated on the basis of how much information or relevance is shared by the variable or features and assigns the score. If the variable share the more information then this feature/variable is depict as the relevant one and other as irrelevant. In this stage we have used the filter approach.

After performing the intra group feature selection within the group of features there may be some probability of redundant features. So to remove or eliminate the redundant features the inter group feature selection is applied. The inter group feature selection consider about the group information which has been ignored by the intra group feature selection. It executes the feature selection between the group and aim to select the features from between the group. Determine the correlation between the groups of features. And produce the optimal features. After the complete performance of intra group feature selection and inter group feature selection it gives the final selected feature and formed the method efficient group feature selection (EGFS).

E. Classification by KNN classifier

We consider each of the characteristics in our training set as a different dimension in some space, and take the value an observation has for this characteristic to be its coordinate in that dimension, so getting a set of points in space. We can then consider the similarity of two points to be the distance between them in this space under some appropriate metric.

The way in which the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the k closest data points to the new observation, and to take the most common class among these. That is why it is called the k Nearest Neighbors algorithm.

III. DESIGN & IMPLEMENTATION

This chapter includes brief explanation of our project’s implementation. We propose our efficient group feature selection method for group of feature from taking an idea from online group feature selection method. From domain knowledge we can obtain a group structure or by specifying a user specified group size to minimize the time efficiency. We have applied our method on UCI Benchmark datasets, and for classification we used KNN classifier and provide detail information in this chapter.

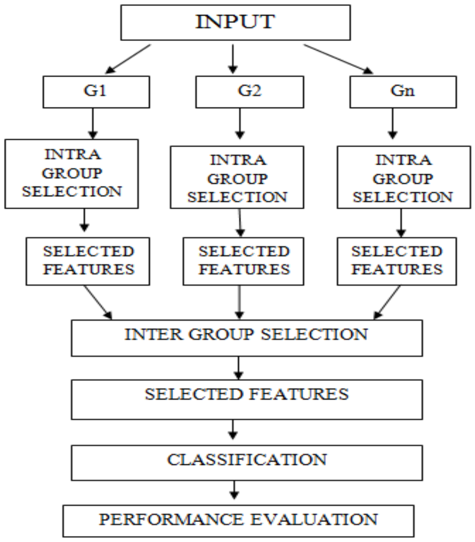


Figure 2. working flow of efficient group feature selection

A. Group feature selection method

In group feature selection method the offline method is used. Our aim is to find an optimal subset from a group. The EGFS is comprises of following stages:

a. Intra group selection

The intra group feature selection is the first stage in group feature selection method. It finds the correlation among the features and selects the discriminative features. In this stage each features is evaluated individually and assign the scores to the feature. . For intra group feature selection the weighted mutual information is applied.

Weighted Mutual Information

The weighted mutual information is derived from the feature selection method called Mutual information method. Mutual information discovers the relevance in two random variable [5]. The feature is declared to be irrelevant if their mutual information found zero and shows both the random variable are independent of each other [6]. 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, to determine the wellness of variable the score is assign to features in two substitute way :

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

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

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

The idea of a weighted form of Mutual Information is driven from mutual information method such as, For each sample sj , is a combination of input value Fi and target value T, a weight w(sj) ≥ 0 is imposed. It is given as,

$$w\, I(F_i\ ,\ T,\ W) = \log \frac{w(sj)P(fi \wedge T)}{w(sj)P(fi) \times p(T)}$$

So the feature can be estimated by using above criteria. The feature that have the higher value or weight or above the threshold value that feature will be defined as a relevant feature. In weighted mutual information if the feature have higher mutual information weight depict more information about the feature label and shows more relevance.

Algorithm for weighted mutual information:

```
Algorithm1. $S$ = wMI ($X$, $Y$)
Input: data set of observations $X$ and the corresponding labels $Y$
Output: final feature set $S$
$S \longleftarrow Null$
W $\longleftarrow$ 1 {same weight for all samples}
 While stopping criterion not true do
        $F_{max}$ = arg max [$wI(Fi,\ T,W)$] { Find feature with maximum weighted MI}
        $S \longleftarrow S\ U\ Fmax$ {Add feature to the subset
        $F \longleftarrow F \setminus F_{max}$ {Remove feature from the candidate set}
 End while
```

b. Inter Group Selection

The information of group is not considered in intra group selection and only calculates the features one at time. In intra 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.

Sparse Group Lasso

In inter group selection we have used sparse group lasso method. 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 [1]. It allows the overlaps in the groups. It generates 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 [2]. 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 inter group selection to decrease reorganization fault with sparsity constrain on coefficient of attribute is objective [3][4].The sparse group lasso is processed as

In sparse group lasso it combine the feature space with class labels and constructed the predictive variable .The sparse group lasso select the optimal by minimizing the objective function. It uses l2 norm and l1 ,the value that is obtained by the regularization parameter is required to be smaller, in general smaller the value of regularization parametr leads to the sparser model.The sparse group lasso model selects features by setting several components in to zero, then the corresponding feature fi is deemed to be irrelevant to the class label and should be discarded. Finally, the features corresponding to non-zero coefficients will be selected. After inter group selections we get the subset of optimal features. With the combination of intra group feature selection and inter group feature selections, the algorithm of efficient group feature selection is formed. The within group selection select discriminative feature and compactness control in between group feature selection can give benefit to the efficient group feature selection.

B. Classification

The K-nearest neighbors algorithm is a method for classifying objects based on closest training examples in the feature space.

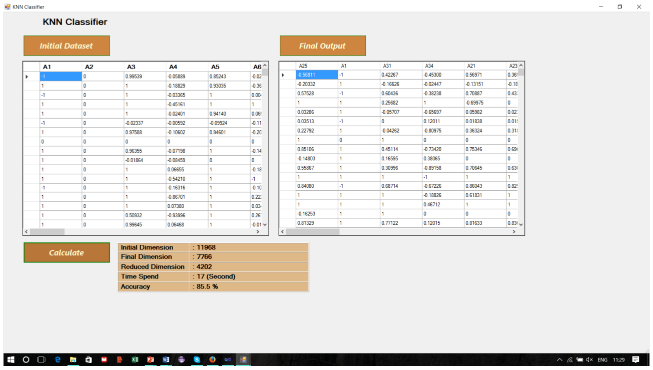
KNN algorithm:

K-Nearest Neighbors (KNN) classification divides data into a test set and a training set. For each row of the test set, the K nearest (in Euclidean distance) training set objects are found, and the classification is determined by majority vote with ties broken at random. If there are ties for the Kth nearest vector, all candidates are included in the vote.

Steps for knn classifier:

- Determine parameter K.
- Calculate the distance between the query-instance and all the training sample.
- Sort the distance and determine nearest neighbors based on the kth minimum distance.
- Gather the category of the category y of the nearest neighbors.
- Use simple majority of the category of nearest neighbors as the prediction value of instance

IV. EXPERIMENTAL RESULTS



	A1	A2	A3	A4	A5	A6
▸	-1	0	0.99539	-0.05889	0.85243	-0.02
	1	0	1	-0.18829	0.93035	-0.36
	-1	0	1	-0.03365	1	0.00
	1	0	1	-0.45161	1	1
	1	0	1	-0.02401	0.94140	0.06
	-1	0	-0.02337	-0.00592	-0.09924	-0.11
	1	0	0.97588	-0.10602	0.94601	-0.20
	0	0	0	0	0	0
	1	0	0.96355	-0.07198	1	-0.14
	1	0	-0.01864	-0.08459	0	0
	1	0	1	0.06655	1	-0.18
	1	0	1	-0.54210	1	-1
	-1	0	1	-0.16316	1	-0.10
	1	0	1	-0.86701	1	0.22
	1	0	1	0.07380	1	0.03
	1	0	0.50932	-0.93996	1	0.26
	1	0	0.99645	0.06468	1	-0.01

	A25	A1	A31	A34	A21	A23
▸	-0.56811	-1	0.42267	-0.45300	0.56971	0.36
	-0.20332	1	-0.16626	-0.02447	-0.13151	-0.18
	0.57528	-1	0.60436	-0.38238	0.70887	0.43
	1	1	0.25682	1	-0.69975	0
	0.03286	1	-0.05707	-0.65697	0.05982	0.02
	0.03513	-1	0	0.12011	0.01838	0.01
	0.22792	1	-0.04262	-0.80975	0.36324	0.31
	1	0	1	0	0	0
	0.85106	1	0.45114	-0.73420	0.75346	0.69
	-0.14803	1	0.16595	0.38065	0	0
	0.55867	1	0.30996	-0.89158	0.70645	0.63
	1	1	1	-1	1	1
	0.84080	-1	0.68714	-0.67226	0.86043	0.82
	1	1	1	-0.18826	0.61831	1
	1	1	1	0.46712	1	1
	-0.16253	1	1	0	0	0
	0.81329	1	0.77122	0.12015	0.81633	0.83

Initial Dimension	: 11968
Final Dimension	: 7766
Reduced Dimension	: 4202
Time Spend	: 17 (Second)
Accuracy	: 85.5 %

Figure 3.

We have shown the experimental result on ionosphere dataset whose dimension is very huge. Similarly, our process can be applied on any dataset whose data type is numeral and can achieve the high accuracy by using the classification with minimum process time.

V. CONCLUSION

We have presented economical cluster variable selection for group of options. Methodology focuses on the matter wherever feature comprise some cluster structure. We have a tendency to additionally offer the literature reviews on existing methodology. We have divided the efficient cluster variable choice into 2 stages, i.e., within cluster variable choice and between cluster variable selections. In inside cluster variable choice uses mutual information and introduces the thin cluster lasso to minimize the redundancy in between cluster variable selection. The inside cluster variable choice effectively select excludent feature, during this step every feature is evaluated singly. Between cluster choice controls the compactness and revaluate the options. We have additionally demonstrated the experiment on many UCI benchmark data sets. This will increase the classification accuracy and shows the effectiveness of our methodology.

VI. FUTURE WORK

In group feature selection method need to generate the group of features, in our experiment we have used data sets that provide no natural grouping information.

During our evaluation we observed that the datasets with no natural grouping for creating a group of features sometimes create some problem if the group information is not accurate. So in future we can perform our method on dataset where group structure is already generated. And the inter group feature selection compute the feature between the group optimally but it select less number of features. In future in inter group selection we can select the feature with more compactness. According to nature of feature selection the accuracy and compactness of feature are parameters plays an important role. In future we can work to increase the classification accuracy.

REFERENCES

[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. Statistic 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 “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.