ILLUSTRATION OF SEMI-SUPERVISED FEATURE SELECTION USING EFFECTIVE FRAMEWORKS

[1]SUJATA GAWADE, [2]DR.NAVEENKUMAR JAYKUMAR

[1]Student , Bharati Vidyapeeth Deemed University College of Engineering, Pune , Department of Computer Engineering, Pune, India

[2]Associate Professor, Bharati Vidyapeeth Deemed University College of Engineering ,Pune , Department of Computer Engineering, Pune, India

E-mail: [1]sujatagawade693@gmail.com, [2]naveenkumarjaykumar@bharatividyapeeth.edu

ABSTRACT

Semi- supervised feature selection is a supervised feature jobs as well as method that uses the unlabeled data for guiding that the small quantity of labeled dataset with a huge quantity of unlabeled dataset. Semi-supervised feature selection is drop between the unsupervised selection and supervised selection. Feature selections have been playing an essential task in the different studies as well as application area of machine learning. In this paper we have explored a three-different level framework for semi-supervised feature selection. Which are mainly feature selection methods center on discovering relevant features for optimizing high-dimensional data. In this paper, we have shown that the relevance need three essential frameworks which provide an efficient feature selection in the semi-supervised context. In the constraint selection framework they select pair wise constraints which can be extracted from the labeled part of data. The Relevance analysis framework shows original utilized realization which competently merges the direct of the restricted geometrical construction of unlabeled data with a selected constraint from the first framework. It allows us to verify the set of relevant features. For the CSFSR and efficiency framework, is to find out and supply the redundant features from the relevant ones which can be chosen from the second framework. It also shows the comparison between third framework and prims algorithm for better feature selection. Result of this proposed system is efficiency with statistical and graphical view**.**

Keywords: *Feature selection, Dimensionality reduction, Semi supervised, Constraint, Relevance, and Redundancy*

1. INTRODUCTION

Data mining is a method that helps to find out the data patterns from the huge quantity of dataset. Feature selection method is a combination of different search procedures for presenting a novel features sub datasets which come with an evaluation that determine which scores the dissimilar features sub datasets. In the redundancy relevance feature selection, a features relevancy through its redundancy in the occurrence of the further selected features. Feature selection is a significant component of machine learning. It is refer as a process that helps to reduce the input dataset to process as well as analysis plus to find out best dataset input. Recently there are many work has been done in unsupervised feature selection for classification with the different approaches.

Different Feature selections have been preoccupied in the multiplicity of a real-world system application, like natural language processing, image processing plus bioinformatics, wherever high dimensionality of data is typically experimental. Feature selection is as well utilized in distributed communication systems as well as sensor networks, wherever every mobile tools or sensor have extremely imperfect computational control. Generally, feature selection is an extremely vital technique which is often related to accumulate storage hole or decrease the computational expenditure, for troubles among also high dimensionality or incomplete capitals. Feature selection can be performed in a supervised as well as unsupervised way, in conditions of whether the label data is employed to direct the selection of relevant features. Normally, supervised feature selection technique need a huge quantity of

labeled preparation data. It still could be unsuccessful to recognize the relevant features which are discriminative to dissimilar classes, which offered the amount of labeled examples is tiny. While unsupervised feature selection technique might work through unlabeled preparation data, they avoid the label data and thus are. Frequently incapable to recognize the discriminative features. Certain the huge price in physically labeling data, and at the similar instant plentiful unlabeled data are frequently simply available, they attractive to build up feature selection techniques which are able of utilizing together labeled as well as unlabeled data. This inspire us initiate semi-supervised education into the feature selection procedure. Particularly, the methods of semi-supervised SVM through different regularization have established excellent presentation [1].

Among the successful of communal networks, we contain spectator a theatrical raise of multimedia data such as video, text as well as images, that has bring rising difficulty of how to efficiently classify as well as recover those data. A simple manner is nothing but to associate the semantic ideas of multimedia data as well as labels for consequent organization responsibilities. Therefore, it is useful as well as essential to progress semantic idea analyzes methods. Generally, the above mentioned capitals are characterized via feature vectors, the measurements of which are extremely huge. Earlier research has established to simply a division of the features take the mainly discerning data as well as correctly planned feature selection is capable to get superior correctness since of its capacity of eliminating unnecessary as well as loud data in the feature representation [2]. In this supervised feature selection technique, the significance of a feature could be estimated via its association amid the class label: such as Fisher score [3], Relief and ReliefF [4], FCBF i.e. Fast Correlation-Based Filter [5], as well as SPEC i.e. SPECtrum decomposition [6]. Unsupervised feature selection is measured since a lot extra complex crisis, appropriate to the lack of class labels which could direct the investigation for significant data, Variance score [7], Laplacian score [8], SPEC [6], and HSIC i.e. Hilbert-Schmidt Independence Criterion [9]. FLD i.e. Fisher linear Discriminant [10] is a best illustration of supervised feature extraction techniques in this it could extract the essentially constructive discriminant vectors when there are

different class labels are available. The best PCA i.e. Principal Component Analysis [11] is an illustration of unsupervised feature extraction techniques which challenge to protect the worldwide covariance construction of data when class labels are not available. Further techniques could be establish in the writing selling among features removal are Locally Linear Embedding [12], Locality Preserving Projection [13] and Laplacian Eigenmap [14].

In this paper, we have presented three new frameworks for semi-supervised feature selection for high dimensionality decrease. These three frameworks are successively stands on three tasks for creating a most excellent subset of features. The first framework disquiets regarding the selection of the mostly relevant constraints that can be take it off from the labeled element of data. The second framework implies a new build up score which can be capably combine the control of the local geometrical construction of unlabeled data among the selected constraints from the first framework. It also permits us to find out the set of relevant features. Finally, the third framework is to establish as well as extract the redundant features from the relevant ones which could be selected from the second Framework.

The rest of this paper is ordered as follows: In section 3 we have analyzed the dissimilar prior successful feature selection techniques for high dimensional data. Section 4 has proposed new effective frameworks for feature selection in high dimensional data. We draw a conclusion in section 5.

2.   MOTIVATION

Semi-supervised feature selection is suffered from the problem of presence labeled and unlabelled instances together.

Labelled and unlabelled cases are predicted to represent similar objective concept. Feature subset selection is an efficient technique for decreasing dimensionality, eliminating unrelated data, rising learning accurateness, as well as civilizing outcome directness. The key thought of feature selection is to select a subset of input variables by removing features among small or no prognostic data.

In order to reveal the semi-supervised feature selection it is required to use both partitions of data.

2. LITERATURE SURVEY

M. Ramaswami and R. Bhaskaran has proposed Feature Selection methods in Educational Data Mining. Educational data mining (EDM) [15] is a novel rising study part and the core of data mining ideas are used in the educational pasture for the use of removing valuable data on the behaviors of students in the education course. In the EDM, feature selection is made for the creation of subset of applicant variables. As the feature selection controls the analytical correctness of several performance representations, it is a vital to research highly the efficiency of student presentation form in relation among feature selection methods.

Chris ding and Hanchuan Peng has presented a minimum redundancy feature selection from microarray gene expression data [16]. In this paper they have described a minimum redundancy maximum relevance (MRMR) feature selection framework. Genes are selected by means of MRMR provides an extra purpose coverage of the independence plus detain broader individuality of phenotypes.

Mohammed Hindawi et.al. have proposed Constraint Selection-Based Semi-supervised Feature Selection [17]. In this they have tackle the dare in feature selection study when managing small-labeled among huge unlabeled data example from the similar inhabitants. The management data can be used in the type of couple wise constraints; these constraints have virtually confirmed to have extremely optimistic outcomes on the education presentation. However, chosen constraints sets might contain important outcomes on education presentation. This paper has presented a new feature selection method stands on a capable selection of pair wise constraints. This aspires to grab the mainly logical constraints remove from labeled data.

N.Anitha and S.Deepa has done Survey on Feature Selection in High dimensional data via Constraint, Relevance as well as Redundancy [18].feature selection is nothing the vital aspect for machine learning. Feature selection through recognizing the mainly relevant features of learning, center a learning algorithm on those features of the data mainly practical for examination for feature prediction. Feature selection as a preprocessing pace to machine learning, have efficiently in dropping dimensionality, eliminating immaterial data, rising learning correctness, as well civilizing directness. Mainly feature selection techniques center on discovering relevant features for optimizing high dimensional data. Dimensionality decrease is an important job selling by high dimensional data. Semi supervised clustering aspires to recover the clustering presentation via allowing for the pair-wise constraints. Semi supervised feature selection technique is mainly capable for discovering the relevant features, reducing the redundant features.

Mohammed Hindawi et.al has attempted to overrule several downbeat results of limitations set through the difference of their different resources [19]. They have used an ensemble method via the both pair are varieties of record as well as an accidental subspace approach. In this it has produced a inclusive levels of characteristics through the collective numerous limitation Laplacian gains on the dissimilar visions of the offered labeled plus unlabeled record set. They have also validated their approach via the observed research above the huge dimensional record sets also evaluate it among further delegate techniques.

Xiangnan Kong et. al. have performed semi-supervised feature selection for graph records in a growth mode simultaneously the associated graph characteristics mining method [20]. They have developed a feature valuation measure, which is called as a Semi that approximated the effectiveness of associated graph characteristics which is stands on the together labeled as well as unlabeled graphs. Also they have implemented another process to capably explore for finest associated graph characteristics via sensibly reducing the associated graph explore gap. They have done experimental revises on the numerous original jobs that shows semi-supervised feature selection method can successfully increase graph arrangement presentation through the semi-supervised feature selection as well as capably increasing the associated graph explore gap with all together labeled along with unlabeled graphs.

Zhihong Zhang at. et. has implemented a new hectic graph which is stand on the semi-supervised feature selection method that helps to choose related characteristics employing all together labeled as well as un- labeled record set [21]. They have done couple of major aids in this research paper. The primary is that through including multiple dimensional communication data for advanced organize matches measure; they have recognized a new hectic graph structure that is utilized for illustrating the numerous associations inside the set of

ISSN: **1992-8645**   www.jatit.org   E-ISSN: **1817-3195**

examples. Hence, the organize record dormant in the dataset that can be additional efficiently formed. Secondary, they have gain a hectic graph associate gap educational vision of feature selection that transmit the feature categorizes study into a deterioration structure which reflect on the associations amid characteristics. They have estimated joint feature permutations, relatively than individual restricted. Experimentations outcomes have expressed the efficiency of feature selection technique on a numeral of usual face record sets.

Yubo Cheng et. al. has proposed a novel feature selection method for semi-supervised characteristics selection. In this the novel method is totally stands on the earlier effort that is Logistic I-RELIEF feature selection method [22]. They have used Vital plan of Logistic I-RELIEF is to crumble an random non linear difficulty into a list of nearby linear individuals during restricted studies, also they have approximated the characteristics significance worldwide in a huge boundary structure among the normalization. In this, method tasks extraordinarily had done well in the existence of numerous of the unrelated characteristics. Also they have expanded the method for semi-supervised knowledge through integrating the limitations of unlabeled trials in the aim utility of Logistic I-RELIEF. This method otherwise calculate approximately feature loads as well as investigates for nearer of every trials till union. Their investigational outcomes have shown the viability of operating the record of unlabeled information that is used to recover the knowledge correctness in choosing applicable characteristics.

E. Elhamifar and R Vidal proposed that Sparse Manifold Clustering and Embeeding (SMCE) algorithm proposed that immediate clustering and dimensionality reduction of data which is used to find nearset neighbor and also it is used to find suitable weights [25].

H.Ward proposed that   method of finding groups that are similar with features and it creates hierarchical group and due to which data reduction is possible [26].

Z. Zhao, L. Wang, H. Liu, and J. Ye proposed Similarity Preserving Feature Selection framework which includes various feature selection methods and also remove weakness in handling feature redundancy. In this framework use of combinatorial optimization formulation for similarity preserving feature selection,   with

a sparse multiple-output regression formulation to Improve performance [27].

H. Peng, F. Long, and C. Ding proposed minimal-redundancy-maximal-relevance method for feature selection [28].They used two stage feature selection algorithm minimal redundancy maximal relevance and other refined feature selectors. By using maximal stastical dependency on similar type of information selection of features is possible. MRMR proves better performance on feature selection and classification accuracy compare to other algorithms.

B. Auffarth, M. Lopez, and J. Cerquides proposed Comparison of redundancy and relevance measures for feature selection in tissue classification of CT   images [29].They have used

Value difference metric for relevance score and find out relation between features and create relevant set of features which avoids data redundancy.

J. B. MacQueen proposed classification methods And analysis if multivariate observations [30].They used process of partitioning into N dimensional sets. They have divided into k sets which partitioned by k-means algorithm and it also varied into classes.

Kais Allab Khalid Benabdeslem proposed a SOM Self Organizing Map method for clustering [31].They have created clusters with the help of constraints and coherence characteristics for semi- supervised learning. They   Proposed algorithm for constraint selection.

D. Zhang, S. W. M. Rand proposed objective criteria for clustering data [32]. This method had been used for retrieval of internal structure and to resample into new data. This method has been used for measure same data between two different clusters.

Chen, and Z. Zhou proposed mehod for feature selection with pairwise constraints.[33].In this they had implemented preprocessing method of high dimensional data. They have implemented concept of must-link constraints and cannot-link constraints. Must-link constraint means same class constraints and cannot-link constraint means different class constraints .They have been implemented algorithm which is known as Constraint Score. Constraint Score algorithm also compared with Laplacian Score algorithm.

I. Davidson, K. Wagstaff, and S. Basu, proposed method for measure constraint-set utility for partitional   clustering algorithm [34] .They had been implemented two measures, informa-

tiveness and coherence that can be used to select required constraint sets.in this algorithm they add constraints in clusters to improve efficiency.

Haytham Elghazel , Khalid Benabdeslem and Alain Dussauchoy proposed clustering with constraints based on graph[35].They have been proposed method which is suitable for unsupervised features. Authors proposed modified version of COBWEB for background information to create constraints and then formation of cluster.

4.   PROPOSED SYSTEM

The proposed system has presented a new method for the Semi-Supervised feature selection with constraint selection framework, relevance selection framework as well as redundancy selection framework [23], [24]. This paper has proposed a three level framework for feature selection. Firstly Constraint selection framework , in this it aids to   choose the majority relevant constraints from the labeled feature set, the secondly relevance selection framework indicates a new extended increase which knowledgeably combine the ability of the local geometrical construction of unlabeled data through the selected constraints from the first constraint selection framework. It also permits us to find out the set of relevant features. Lastly, redundancy selection framework is to find out as well as eliminate the redundant features from the relevant ones which could be selected from the second relevance selection framework.
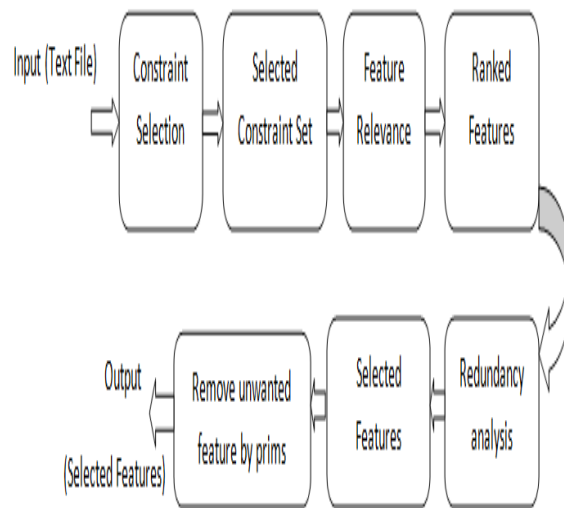

Fig1. Proposed System Workflow

There are three different frameworks for feature selections are as follows:

1. Constraint Selection Framework

Constraint selection framework offers a GUI. User has to present through input as dataset and view results. User can select dataset as well as obtain relevance dataset. Then we will get chosen subset of constraint from input dataset.

2. Relevance Selection Framework
This relevance selection framework includes relevance algorithm that gives relevant subset of feature.

3. CSFSR and Efficiency Framework
This CSFSR and efficiency framework illustrates output is feature set that is relevant and redundancy free.

- **Prims Algorithm:**
 Once discovering dissimilar features with admiration to dissimilar frameworks, we have also used new algorithm called prims algorithm. In this algorithm we have reduced the entire unwanted edges from the tree T, once reducing the unwanted edges a forest is find. Each spitted tree T_j that is forest which signifies a cluster which shown as $V(T_j)$. The cluster is a group of vertex set of tree T_j. The features are available in every cluster are unnecessary so that every cluster will prefer a representative features whose relevance is maximum. The processes for the prims algorithm are as follow:

Choose the h best relevant features order as per the relevance algorithm.

- create the graph Gh (V_h,E_h)
- Get the minimum spanning tree $Gh(V_h,E_h)$ from Gh via prims algorithm.
- Reduce unwanted edges from tree T.
- A forest is discovering from tree T_j.
- Presenting a cluster $V(T_j)$.
- Choose most representative feature from cluster.

Following are the different results of proposed system which show different output for different three frameworks by giving input as a text file. The frameworks are constraint selection

framework Relevance analysis Framework as well as Redundancy analysis framework. Here new prims algorithm is compared with the third framework to get the better result of feature selection.

In this following result, we have given a dataset as an input to calculate the constraint selection. From the given dataset it calculates total constraints as well as selected constraints.
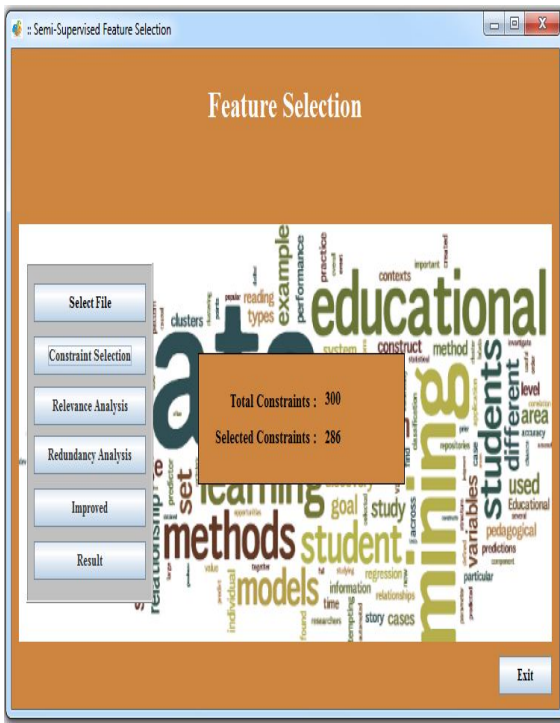

Fig.2 Result of Constraint selection Framework

In this following result we have given dataset as an input to find out nearest neighbor. For this we use relevance analysis framework. This result shows the nearest neighbor from given dataset to find out the relevance.
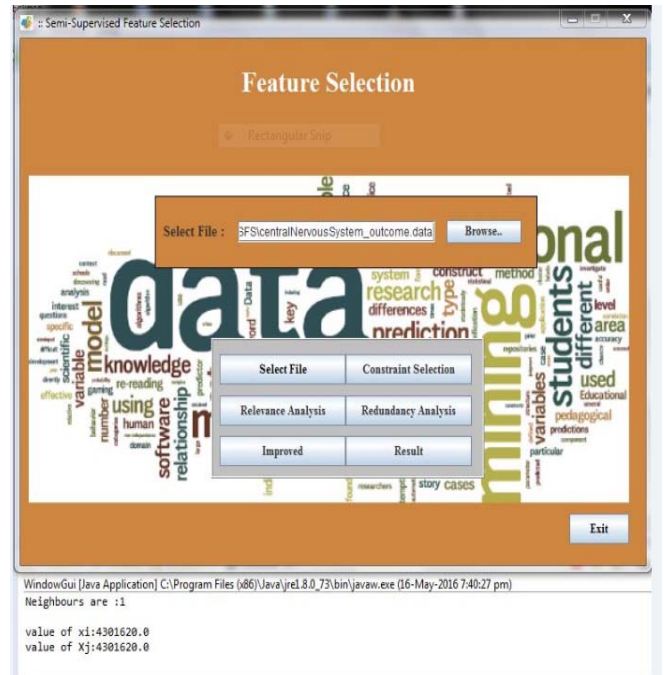

Fig.3 Result of Relevance analysis Framework

In the following result it helps to calculate the total selected features using redundancy analysis from the given dataset.
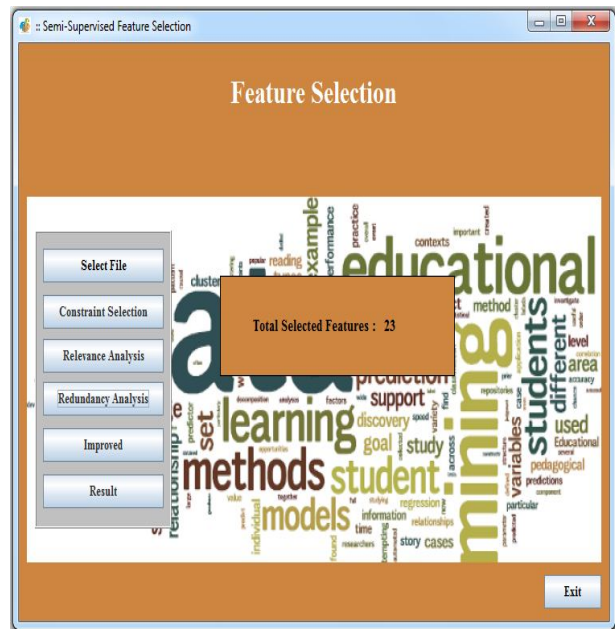

Fig. 4 Result of Redundancy analysis Framework

In this result, it calculates the total selected features from given dataset using improved algorithm called as prims algorithm for better performance.
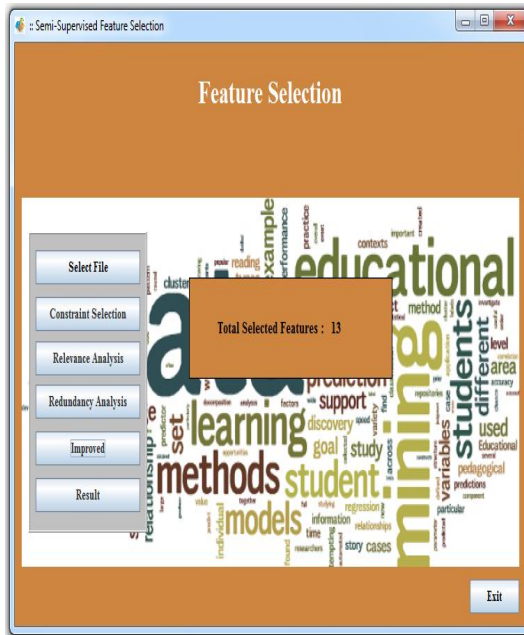

Fig.5 Result of Prims Algorithm

This result shows the comparison between semi-supervised feature selection framework and improved semi-supervised feature selection using prims algorithm. Prim's algorithm provide more refined data.
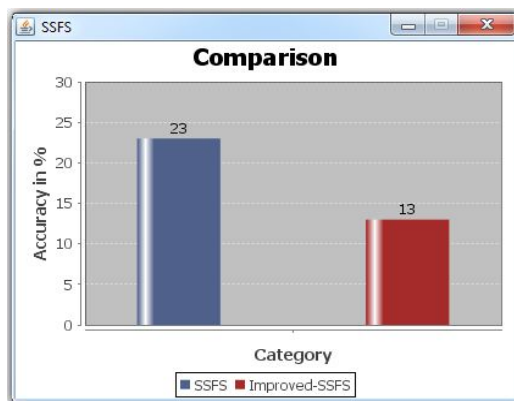

Fig.6 Comparison between Third framework and Prims Algorithm

- **Advantages:**

It decreases the dimension of massive dataset. It builds up pair constraint selections, which results in a rational constraint subset removed from the labeled data. It holds feature redundancy with a new graph stands on three frameworks, which iteratively take away redundant features amid relevant ones.

- Disadvantages:

The neighborhood option is not properly specified; therefore the range of (k) options has major causes on the outcomes. Constraint score totally is relying on the selected constraint subset, the variety of constraints is yet a difficult problem that automatically decreases the presentation of the feature selection method.

5. CONCLUSION

This paper has explored three frameworks for feature selection such as constraint selection, relevance selection as well as redundancy elimination for semi-supervised dimensionality reduction and the relevance of features based on both, the locally geometrical architecture of unlabeled data and the constraint preserving ability of labeled data. And the redundancy elimination is providing to verify and eliminate the redundant features from the relevant one. It also shows efficiency result with statistical and graphical view. The proposed three frameworks have several advantages. It decreases the dimension of huge dataset. It develops a couple constraint selections, which results in a logical constraint subset extracted from the labeled data. It handles feature redundancy with a novel graph based on three frameworks, which iteratively removes redundant features among relevant ones. Prims algorithm is used to calculate features from dataset for better performance. We have also shown the comparative analysis between Prims algorithm and redundancy analysis framework. In this proposed system, we have used numerical as well as text file as dataset only. At a time only one dataset is given for feature selection process.

6. FUTURE WORK

Future research in the feature selection can be the growth of the unconventional of label set from which the constraints are created. Further perspective could be the grouping of diversity amid subsets of features as a alternate by them in one set, which is difficult for high-dimensional data. Also discover dissimilar varieties of correlation measures, plus study various formal properties of feature space.

REFERENCES

[1] Zenglin Xu, Rong Jin, Michael R. Lyu, Irwin King, "Discriminative Semi-Supervised Feature Selection via Manifold Regularization"Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09).

[2] Xiaojun Chang, Feiping Nie, Yi Yang and Heng Huang, "A Convex Formulation for Semi-Supervised Multi-Label Feature Selection", IIS-1344152, Australian Research Council Project DE130101311. Copyright 2014.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. New York, NY, USA: Wiley Interscience, 2000.

[4] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of relief and relieff," Mach. Learn., vol. 53, no. 1–2, pp. 23–69, 2003.

[5] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," J. Mach. Learn. Res., vol. 5, pp. 1205–1224, Oct. 2004.

[6] Z. Zhao and H. Liu, "Spectral feature selection for supervised and unsupervised learning," in Proc. 24th Int. Conf. Mach. Learn., Corvallis, OR, USA, 2007.

[7] C. M. Bishop, Neural Networks for Pattern Recognition. New York, NY, USA: Oxford University Press, 1995.

[8] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in Proc. NIPS, Vancouver, BC, Canada, 2005.

[9] L. Song, A. Smola, A. Gretton, J. Bedo, and K. Borgwardt, "Feature selection via dependence maximization," J. Mach. Learn. Res., vol. 13, no. 1, pp. 1393–1434, Jan. 2012.

[10] R. Fisher, "The use of multiple measurements in taxonomic problems," Ann. Eugen, vol. 7, no. 2, pp. 179–188, Sept. 1936.

[11] I. Jolliffe, Principal Component Analysis. New York, NY, USA: Springer, 2002.

[12] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by local linear embedding," Science, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[13] X. He and P. Niyogi, "Locality preserving projections," in Proc. NIPS, 2004.

[14] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in Proc. NIPS, 2002.

[15] M. Ramaswami and R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining", journal of computing, volume 1, issue 1, December 2009, issn: 2151-9617.

[16] Chris ding and hanchuan peng," minimum redundancy feature selection from microarray gene expression data Chris ding", journal of bioinformatics and computational biology vol. 3, no. 2 (2005) 185–205

[17] Mohammed Hindawi, Kaïs Allaband Khalid Benabdeslem, "Constraint Selection based Semi supervised Feature Selection", https://hal.archives-ouvertes.fr/hal 00874960 Submitted on 21 Oct 2013.

[18] N. Anitha1 and S.Deepa2." Survey on Feature Selection in High dimensional data via Constraint, Relevance and Redundancy", nternational Journal of Application or Innovation in Engineering & Management , Volume 03, Issue 09, September 2014.

[19] Mohammed Hindawi, Haytham Elghazel ,Khalid Benabdeslem, "Efficient semi-supervised feature selection by an ensemble approach" , 2013/04/copem2013,www.ecmlpkdd2013. org.

[20] Xiangnan Kong, Philip S. Yu, "Semi-Supervised Feature Selection for Graph Classification", Copyright 2010 ACM 978-1-4503-0055-110/07

[21] Zhihong Zhang , Edwin R. Hancock and Xiao Bai ," Hypergraph Spectra for Semi-supervised Feature Selection " , Lecture Notes in Computer Science: Authors' Instructions, www.cs.bris.ac.uk/~flach/ECMLPKDD201 2papers.

[22] Yubo Cheng, Yunpeng Cai, Yijun Sun, Jian Li, "Semi-Supervised Feature Selection Under Logistic I-RELIEF Framework", 978-1-4244-2175-6/08/$25.00   ©2008 IEEE

[23] Khalid Benabdeslem and Mohammed Hindawi," Efficient Semi-Supervised Feature Selection: Constraint, Relevance, and Redundancy," IEEE transactions on knowledge and data engineering, VOL. 26, NO. 5, MAY 2014.

[24] Gawade, M.S.S. and Kumar, N., "Three Effective Frameworks for semi-supervised feature selection" , International journal of research in management & Technology, Vol. 6, No. 2, Mar-Apr 2016, ISSn:2249-9563.

[25] E. Elhamifar and R. Vidal, "Sparse manifold clustering and embedding," in Proc. NIPS, 2011, pp. 55–63.

[26] J. H. Ward, "Hierarchical grouping to optimize an objective function,"J. Amer. Statist. Assoc., vol. 58, no. 301, pp. 236–244, 1963.

[27] Z. Zhao, L. Wang, H. Liu, and J. Ye, "On similarity preserving feature selection,"IEEE Trans. Knowl. Data Eng., vol.25, no.3,pp. 619–632, Mar. 2013.

[28] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance,and min-redundancy,"IEEE Trans. Pattern Anal. Mach. Intell.,vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[29] B. Auffarth, M. Lopez, and J. Cerquides, "Comparison of redun-dancy and relevance measures for feature selection in tissueclassification of CT images," inProc. 10th ICDM, Berlin, Germany

[30] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," inProc. 5th Symp. Math. Statist.. probab, Berkley, CA, USA, 1967, pp. 281–297

[31] Kais Allab Khalid Benabdeslem"Constraint selection for semi-supervised topological clustering"

[32] W. M. Rand, "Objective criteria for the evaluation of clustering method,"J. Amer. Statist. Assoc., vol. 66, no. 336, pp. 846–850, Dec.1971

[33] D. Zhang, S. Chen, and Z. Zhou, "Constraint score: A new filter method for feature selection with pairwise constraints," PatternRecognition , vol. 41, no. 5, pp. 1440–1451, 2008.

[34] I. Davidson, K. Wagstaff, and S. Basu, "Measuring constraint-set utility for partitional clustering algorithms," in prac ECML/PKDD , 2006.

[35] Haytham Elghazel,, Khalid Benabdeslem, and Alain Dussauchoy "Clustering With Constraints Using Graph Based Approach "LIESP (ex. PRISMa) Laboratory, Claude Bernard University of Lyon I, France {elghazel,kbenabde,dussauchoy}@bat710. univ-lyon1.fr