International Journal of Engineering & Technology, 7 (3.29) (2018) 688-693

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Enhanced Feature Selection Clustering Algorithm for Attribute Similarity in High Dimensional Data

Deena Babu Mandru[1*], Y.K. Sundara Krishna [2]

[1] *Research Scholar (PhD), Depatment of Computer Science, Krishna University, Machilipatnam, A.P.*
[2]*Principal,Krishna University, Machilipatnam, A.P.*
**Corresponding Author E-mail: dbmandru@gmail.com*

Abstract

Data collection is aggressive concept in data mining which is based on various attributes from dissimilar data sets. For some real world data, real time dataset portioning with abnormal behavioral class label instances is expensive and impossible to data presentation. Through user preferences, now a day's data summarization based on clustering with different attributes is another aggressive concept. Traditionally clustering with multi-attribute framework was introduced to group multiple attributes to explore uncertain data for reliable data sets. In multi attribute similarity measure for uncertain data, feature selection is the factor to provide most matched and most useful features which produces compatible results from original set of features present in data sets. So feature selection algorithm is required to evaluate efficiency to form subset of features with respect to quality assurance for subset of features. In this paper, we proposed and implemented Enhanced Feature Selection based Clustering (EFSC) algorithm to evaluate above considerations. Our proposed method consists of two stages in implementation. In first stage, classify features into clusters using graph based theoretic approach. In second stage, identify most representative attribute which is most relate to selected attribute from each cluster to sub set of features. In this paper, we use Minimum spanning tree (MST) for effective clusters formation with respect to subset of features. EFSC is compare with some existing algorithms like FCBF, ReliefF, CFS, Consist, and FOCUS-SF with respect to chosen classifiers prior to and later than feature selection from subset of features. Our experimental results performed on company statistical data with text, image orientated data, and EFSC produces small subset of features with high accuracy and less time efficiency for real time data sets.

Keywords: *Multi attribute clustering, minimum spanning tree, feature selection, theoretic graph clustering and sub set of features.*

1. Introduction

Data exploration is a strongly idea in information recovery depending on different features from different data resources. For efficient data gathering from sources, with regard to relevant data single class learning is required to perform marked centered category with individual training series on features. For some real world data sourcing, for real-time data set portioning with irregular behavior class complete instances with dearly-won not possible knowledge demonstration. To find out these forms of combined series in period of time data set techniques to reason target knowledge into distinctive classifier knowledge techniques. For form of totally different programs abnormality recognition, papers class image annotation and content necessities for various data formations. In real time applications, an enhanced system is needed to process on totally dissimilar attributes to extend multi-attribute label presentation on high dimensional data. Clustering with Multi-Attribute Framework (CMAF) is that the procedure to boost irregular lattice to relinquish extensively low level data set illustration. It is a connection-based way to access irrelevant data in classification with completely different attributes and supports similarity options. This examination solely associates the outlet between the procedures of dataset which of web connection investigate. In addition, it expands the capability to accumulation system for explicit data and not non-heritable much thought for real time data sets.

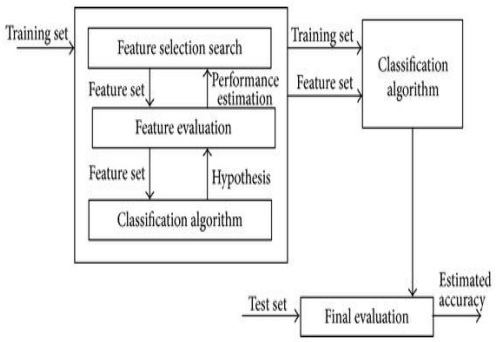

Fig-i: Basic selection procedure for subset features based on classification

Choosing subset of most effective features with respect to relative target class, feature selection rooted on subset demonstration is an effective way to reduce feature dimensionality, irrelevant data removal, improve learning accuracy and increase comprehensive results based on subset features shown in fig-1. Traditionally some of the feature selection based approaches and algorithms were proposed and studied in machine learning applications. Cluster analysis for various application implement has been defined to be

more effective than existing feature selection algorithms. For cluster analysis graph based theoretic approach used in many type of applications and it is simple and find nearest neighbor selection of instances, then delete any edge from graph that much longer/shorter than neighbors data. In implementation, use graph based theoretic approach on features to form cluster based on subsets. We adopt minimum spanning tree (MST) clustering algorithm, because they don't assume data points are clustered based on regular geometric curves have been implemented in practical scenario. Based on this method, we proposed and implemented Enhanced Feature Selection based Clustering (EFSC) algorithm.

The EFSC algorithm has two steps. First, highlights are partitioned into clusters by utilizing graph based theoretic approach. Secondly, the most helpful feature that is forcefully recognized with target classes is selected from each group to shape the last subset of highlights. Highlights in various clusters are moderately autonomous; the group based method of EFSC has a high probability of creating a subset of helpful and free highlights. The proposed highlight subset determination algorithm EFSC was tried upon 35 openly accessible picture, small scale cluster, and content informational collections. The trial comes about demonstrate that, contrasted and other five unique kinds of highlight subset selection algorithms; the proposed algorithm lessens the quantity of highlights, as well as enhances the exhibitions of the four surely understood distinctive sorts of classifiers.

2. Review of Related Literature

Highlight subset choice can be seen as the procedure of recognizing and expelling the all unimportant and excess includes as could be allowed. This is on the grounds that 1) unessential highlights don't add to the prescient precision [3], what's more, 2) repetitive highlights don't redound to getting a better indicator for that they give for the most part data which is as of now exhibit in different feature(s). Of the numerous element subset determination calculations, some can successfully dispense with unessential highlights yet neglect to handle repetitive highlights [3], [1], [7], [4], [5], [9], however some of others can kill the unimportant while taking care of the excess highlights [5], [9], [4], [6]. FAST algorithm comees under second category. Customarily, highlight subset choice research has concentrated on scanning for important highlights. A notable case is Relief [4], which measures each component agreeing to its capacity to separate occurrences under various targets in light of separation based criteria work. Be that as it may, Alleviation is ineffectual at evacuating repetitive highlights as two prescient however exceedingly associated highlights are likely both to be exceedingly weighted [6]. Help F [7] expands Relief, empowering this technique to work with uproarious and fragmented informational collections and to manage multiclass issues, yet at the same time can't distinguish repetitive highlights. In any case, alongside unessential highlights, excess includes likewise influence the speed and exactness of learning calculations, and consequently ought to be disposed of also [1], [2], [1]. CFS [3], FCBF [4], and CMIM [5] are illustrations that mull over the repetitive highlights. CFS [2] is accomplished by the speculation that a decent component subset is one that contains includes profoundly associated with the objective, yet uncorrelated with each other. FCBF ([8], [1]) is a quick channel technique which can recognize significant highlights and also excess among pertinent highlights without pairwise relationship examination. CMIM [2] iteratively picks highlights which expand their shared data with the class to foresee, restrictively to the reaction of any component as of now picked. Unique in relation to these calculations, our proposed the Quick calculation utilizes the grouping based strategy to pick highlights. As of late, various leveled grouping has been received in word choice with regards to content grouping (e.g., [5], [4], and [18]). Distributional grouping has been utilized to cluster words into clusters construct either with respect to their investment specifically syntactic relations with different words by Pereira et al. [5] or on the appropriation of class marks related with each word by Baker and McCallum [4]. As distributional grouping of words are agglomerative in nature, and result in problematic word clusters and high computational cost, Dhillon et al. [8] proposed another data theoretic troublesome calculation for word clustering furthermore, connected it to content characterization. Butterworth et al. [8] proposed to cluster highlights utilizing an uncommon metric of Barthelemy-Montjardet separation, and after that makes utilization of the subsequent cluster chain of importance to pick the most significant traits. Tragically, the group assessment measure in view of Barthelemy-Montjardet remove does not distinguish a component subset that permits the classifiers to enhance their unique execution exactness. Further all the more, even contrasted and other component determination strategies, the acquired exactness are lower. Progressive clustering additionally has been utilized to choose includes on ghastly information. Van Dijck and Van Hulle [4] proposed a half and half channel/wrapper include subset determination calculation for relapse. Krier et al. [8] introduced a system joining progressive compelled clustering of phantom factors and determination of clusters by shared data. Their element clustering technique is like that of Van Dijck and Van Hulle [6] with the exception of that the previous powers each group to contain back to back highlights as it were. The two techniques utilized agglomerative various leveled grouping to evacuate excess highlights.

3. Background Work

Below procedure specify multi-attribute clustering specification which is listening with dissimilar attributes.

3.1. Basic Procedure for Data Summarization

Consider $C = (c1; c2; \ldots; cN)$ is a combination of datasets with N information factors and $\gamma = (\gamma 1, \gamma 2, ..., \gamma n)$ Ng is a group selection with M group's study, each of which is used as selection individual. Every stage clustering earnings a combined with categories, $\pi_i = \{X_1^i, X_2^i, X_3^i, ...X_n^i\}$, such that $\bigcup_{j=1}^{k_i} C_j^i = C$, where k_i is dissimilar choice of cluster with unlike parameters. For all x 2 C, X(c) distinguishes the collective type connection with factor c with cluster series. In I^{th} clustering, $X(x) = "\ j"(or"\ X_j^i\ ")ifc \in X_j^i$. This separation gives main assets π^* of a entire set C, which consists of combined attributes with similar attributes π [6][1].

Thus the fundamental cluster pattern from dissimilar attribute groups with appropriate data with compromise learning functions based on outcomes with related attributes process shown in figure 2.

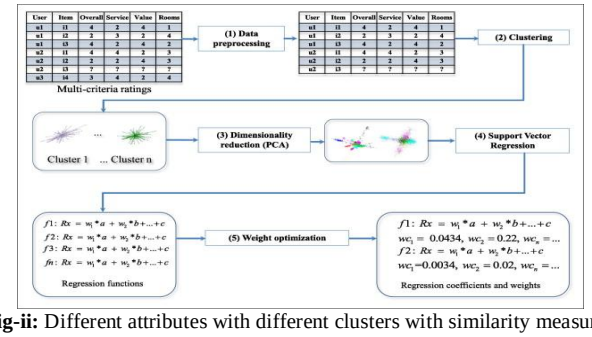

Fig-ii: Different attributes with different clusters with similarity measure.

3.2. Group Creation

It is the primary idea to create totally dissimilar attributes as one with similar associations. In cluster, individual attributes over extra data streams. Chosen attributes believe several conditions with related features based on client needs. In this case, selected clients carry out the final system improvement based on cluster result. Automatically, a number of attributes suggested as gift attributes in grouping methods with range of exact consecutive relational attributes. Lastly, consecutive features were used to explain grouping necessities with dissimilar multi-objectives.

3.3. Consensus Methods

Out of all attributes, indiscriminately choose classified options are designed for accessible data with attribute partition. Using Markov-chain matrix model same attributes ordered in cognitive functions. Basically, number of feature- based techniques for cluster study transforms in operational attributes in real time data streams to elaborate categorization. Accordingly, a matrix is formatted with direct and indirect labeled formations.

3.4. Direct Technique

In direct approach, depending attributes are unit individuals for selecting relative label i.e. number of attributes in i with multi objectives in relations for different formations using consensus function. To provide related attributes with groups for random choice from dissimilar data sets. In Markova-chain model, Euclidian distance is used to create matrices between all attributes in data streams [8][9].

3.5. Outlier Clustering Groups

From the direct technique a matrix is formatted and attributes arranged with similar attributes in relations. Outliers are created depends on attributes with several properties in dissimilar consensus for combining selected features in recent attributes to detect outlier from relations.

4. Methodology Implementation

4.1. Basic Definitions

Irrelevant feature removal along with repeated attributes, definably accuracy of the different machine learning approaches, Thus, feature subset selection could be recognize and remove as much of the unrelated and repetitive information as possible. Moreover, "good function subsets have features extremely associated with the course, yet contrast with each other."

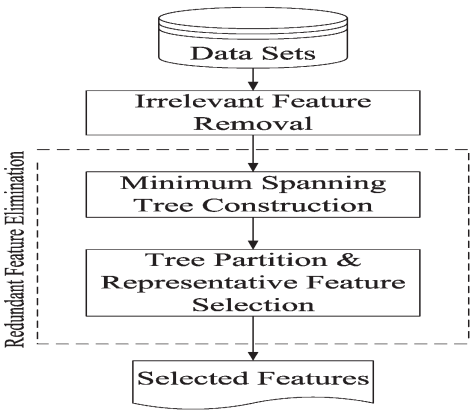

Fig iii: Implementation procedure of proposed approach.

We build up a novel calculation which can proficiently and viably manage both unessential and excess highlights, and get a decent element subset. We achieveed it from another component choice system (appeared in Fig. 3) which made out of the two associated parts of insignificant element cleaning and repeated component disposal. The previous acquires emphasizes appropriate to the actual idea by taking out insignificant ones, and the last mentioned expels excess highlights from appropriate ones by means of choosing agents from various component groups, and in this way produces the last subset. The unessential element expulsion is direct once the correct importance measure is characterized or chosen, while the repetitive element end is a touch of complex. In our proposed EFSC algorithm 1) the development of the base spreading over tree from a weighted finish graph; 2) the dividing of the MST into forests with each tree speaking to a group; and 3) the choice of delegate highlights from the groups. With a specific end goal to all the more definitely present the calculation, and on the grounds that our proposed include subset choice system includes insignificant element evacuation and excess element end.
Assume F to be the full arrangement of highlights, Fi ∈ F be a component, Si = F – {Fi} and S0 I < Si. Let Si' be an esteem task of all highlights in Si', fi an esteem task of highlight Fi, and c an esteem task of the objective idea C. Definitions formalized as takes after:
4.1.1. Relevant Feature: Fi is applicable to the objective idea C if and only if there prevails some s0 i, fi, and c, such that, for possibility

$$p(S_i^1 = s_i^1, F_i = f_i) > 0$$

$$p(C = c \mid S_i^1 = s_i^1, F_i = f_i) \neq p(C = c \mid S_i^1 = s_i^1)$$

Otherwise Fi is an irrelevant feature. Most of the details found in repetitive features already exist in other functions. As a outcome, redundant features do not support getting better interpreting ability to the focus on idea.

4.1.2. Markov Blanket:

Given a feature $F_i \in F$, let $M_i \subset F(F_i \notin M_i)$, Mi be the markov blanket for fi if and only if

$$p(F - M_i - \{F_i\}, C \mid F_i, M_i) = p(F - M_i - \{F_i\}, C \mid M_i)$$

4.1.3. Redundant Feature:

Let S be a set of functions, a feature in S is repetitive if and only if it has a Markov Blanket within S.
Appropriate functions have powerful connection with target concept so is always needed for a finest part, while redundant functions are not because their principles are completely correlated with each other. Thus, thoughts of feature redundancy and have importance are normally in conditions of feature connection and feature-target idea connection.
The Symmetric Uncertainty (SU) relies on the common simple elements by reducing it to the entropies of emphasize requirements or potential requirements and emphasize on classes, and has been utilized to study the advantages of capabilities for depiction by a combined bag of scientists. Accepting their undertaking the symmetrical uncertainty is recognized as takes after:

$$SU(X \mid Y) = \frac{2 \times Gain(X \mid Y)}{H(X) + H(y)}$$

Where,. H(X) is the entropy of a distinct unique varying X. Suppose p(x) is the before possibilities for all values of X, H(X) is determined by

$$H(x) = -\sum_{x \in X} p(x) \log_2 p(x)$$

Gain(X|Y) is the quantity by which the entropy of Y decreases. It shows the extra information about Y offered by X and is known as the information gain which is given by

$$Gain(X \mid Y) = H(X) - H(X \mid Y)$$

Where H(X|Y) is the based upon entropy which quantifies the remaining entropy (i.e., uncertainty) of a exclusive different X given that the value of another exclusive different Y is known. The computations of SU concepts for T-Relevance and F-Correlation, which has directly range complexness with regards to the range of conditions in a given details set.

4.2. Algorithm Implementation

The proposed EFSC algorithm consists 3 steps 1) irrelevant feature removing 2) MST construct based on relative attributes 3) select appropriate or representative attributes.

Inputs: D (F1, F2, .Fm, C) is a Dataset and θ is the T-Relevance threshold

Output: S-Selected feature Subset

//===Part 1: Irrelavant Feature Removal======

1. for i=1 to m do
2. T-Relevance=S U (Fi, C)
3. if T-Relevance > 0 then
4. S=S U { Fi };

//Part 2: Minimum Spanning Tree Construction

5. G=NULL; // G is a Complete Graph
6. foreach pair of features { F'i,, F'j } ⊂ S do
7. F-Correlation=SU {F'i, F'j}
8. Add F'2, and F'3 to G with F-Correlation as the weight of the corresponding edge
9. minSpanTree=Prim (G); **// Using Prim's to generate the minimum spanning tree**
10. Forest=minSpanTree
11. for each edge Eij ∈ Forest do
12. if SU(F'i,F'j) < SU(F'i, , C) ∩ SU (F'i, F'j) < SU(F'j, C) then
13. Forest=Forest- Eij
14. S= ∅
15. for each tree Ti ∈ Forest do
16. F0 R= argmax F'k ∈Ti SU(F'k; C)
17. S=S U {F j R};
18. return S

Algorithm 1.EFSC Algorithm Procedure

For a knowledge set D with m functions F = {F1; F2; . . . ; Fm} and category C, we estimate the T-Relevance SU (Fi; C) value for each function $F_i(1 \le i \le m)$ in the first thing. The features whose S U (Fi; C) principles are higher than a predefined threshold consist of the target-relevant function subset. In second step, we calculate f-correlation for different pair of features F and $F_i^{'}$. In third step, the finish chart G shows the connections among all the target-relevant functions. Unfortunately, graph G has k vertices and k(k – 1)/2 sides. For high-dimensional data, it is intensely heavy and the sides with different loads are strongly inter weaved. Moreover, the breaking down of complete chart is NP-hard. Thus for graph G, we build an MST, which joins all vertices such that the sum of the loads of the sides is the lowest, using the well known Prim algorithm. This is the proposed implementation for sub feature selection from different data sources.

5. Experimental Results

5.1. Data Source

For the motivations behind assessing the execution and adequacy of our proposed EFSC calculation, confirming regardless of whether the strategy is possibly helpful practically speaking, and enabling different specialists to affirm our outcomes,15 openly accessible data sets were utilized. The quantities of highlights of the 35 informational collections change from 37 to 49, 52 with a mean of 7,874. The dimensionality of the 54.3 percent informational indexes surpasses 5,000, of which 28.6 percent informational collections have in excess of 10,000 highlights. The 15 informational indexes cover a scope of utilization spaces, for example, content, picture and bio smaller scale cluster information grouping. Table 1 demonstrates the relating measurable data. Note that for the informational collections with consistent esteemed highlights, the outstanding off-the-rack MST strategy was utilized to discretion the persistent qualities.

Table –i: Different datasets used in our implementation.

Data set	EFSC	FCBF	CFS	Relief	Consist	FOCUS-SF
Chess	106	60	352	12660	2000	653
M feat Fourier	1475	716	940	13918	3229	660
Coil 2000	866	885	1453	30412	53850	1281
Elephant	790	320	910	20991	2500	1100
Fqs-nowe	977	97	736	1075	1400	1040
B-cell	160	250	103895	930564	4556	5230
B-cell2	626	1620	1097122	7001	4700	2500

5.2. Experimental Setup

The performance of proposed approach is evaluated by comparing with some of the existing algorithms like FCBF, Relief, CFS , Consist and FOCUS-SF respectively. To prove our approach to support efficient attribute formation with feature selection, we take B-Cell Company statistical data set, which consists of 12 types of attributes with high dimensions. Based on above data set exploration from different data sources after collection of selected sub set attributes and then they can be represented as shown in fig-4.

DataSet Prepared For DataMining Process

ProductID,ProductName,CategoryID,Beverages,Condiments,Confections,Dairy Products,Grains/Cereals,Meat/Poultry,Produce,Seafood,ProductName,UnitPrice
1,Chai,1,YES,NO,NO,NO,NO,NO,NO,NO,Cote de Blaye,263.50
2,Chang,1,YES,NO,NO,NO,NO,NO,NO,NO,Thuringer Rostbratwurst,123.79
3,Aniseed Syrup,2,NO,YES,NO,NO,NO,NO,NO,NO,Mishi Kobe Niku,97.00
4,Chef Anton's Cajun Seasoning,2,NO,YES,NO,NO,NO,NO,NO,NO,Sir Rodney's Marmalade,81.00
5,Chef Anton's Gumbo Mix,2,NO,YES,NO,NO,NO,NO,NO,NO,Carnarvon Tigers,62.50
6,Grandma's Boysenberry Spread,2,NO,YES,NO,NO,NO,NO,NO,NO,Raclette Courdavault,55.00
7,Uncle Bob's Organic Dried Pears,7,NO,NO,NO,NO,NO,NO,YES,NO,Manjimup Dried Apples,53.00
8,Northwoods Cranberry Sauce,2,NO,YES,NO,NO,NO,NO,NO,NO,Tarte au sucre,49.30
9,Mishi Kobe Niku,6,NO,NO,NO,NO,NO,YES,NO,NO,Ipoh Coffee,46.00
10,Ikura,8,NO,NO,NO,NO,NO,NO,NO,YES,Rossle Sauerkraut,45.60
11,Queso Cabrales,4,NO,NO,NO,YES,NO,NO,NO,NO,,
12,Queso Manchego La Pastora,4,NO,NO,NO,YES,NO,NO,NO,NO,,
13,Konbu,8,NO,NO,NO,NO,NO,NO,NO,YES,,
14,Tofu,7,NO,NO,NO,NO,NO,NO,YES,NO,,
15,Genen Shouyu,2,NO,YES,NO,NO,NO,NO,NO,NO,,
16,Pavlova,3,NO,NO,YES,NO,NO,NO,NO,NO,,
17,Alice Mutton,6,NO,NO,NO,NO,NO,YES,NO,NO,,
18,Carnarvon Tigers,8,NO,NO,NO,NO,NO,NO,NO,YES,,
19,Teatime Chocolate Biscuits,3,NO,NO,YES,NO,NO,NO,NO,NO,,
20,Sir Rodney's Marmalade,3,NO,NO,YES,NO,NO,NO,NO,NO,,
21,Sir Rodney's Scones,3,NO,NO,YES,NO,NO,NO,NO,NO,,
22,Gustaf's Knackebrod,5,NO,NO,NO,NO,YES,NO,NO,NO,,
23,Tunnbrod,5,NO,NO,NO,NO,YES,NO,NO,NO,,

Fig-iv: Data set processing with selected attributes.

After collecting selected sub attributes from overall data set using theoretic graph based clustering to group different features, then apply MST on these attributes, result can be shown in figure 5 with high, medium and low combination of different selected features.

Fig-v: Top products feature selection representation with high matched attribute.

For example, from b-cell data sets, we select top products as a sub feature from overall data set then top matched or useful target attribute shows high resolution shown in figure 5. Accuracy to represent different attributes with different data relational representation shown in fig-6.

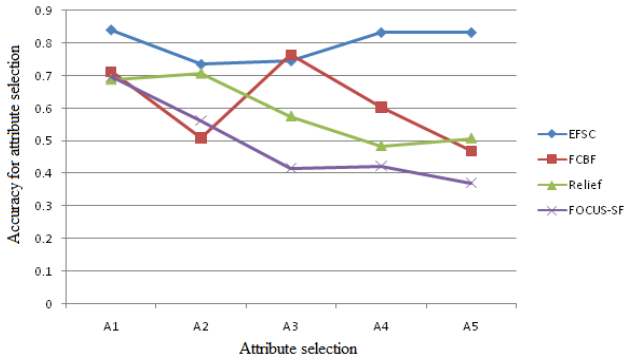

Fig-vi: Accuracy representation for selected attributes from overall data sets.

Accuracy for different feature selection after performing MST on each subset attributes from different data sources. Time efficiency for feature dimensionality for different subset attributes shown in fig-7.

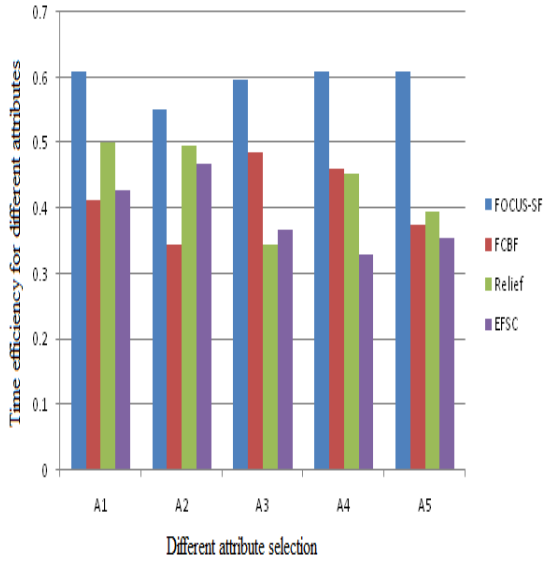

Fig- vii: Time efficiency results for different selected attributes from overall data sources.

Based on above experimental setup or feature selection from different data relations, our proposed approach gives better performance when compare to different existing algorithms.

6. Conclusion

In this paper, we proposed a novel clustering approach, Enhance Feature Selection Clustering (EFSC) to evaluate high dimentional data for feature selection . This approach involves irrelevant feature removal, constructs MST and selected with relative attributes and partitioning attributes with MST on selected features. In our proposed approach each cluster consist set of attributes, and each cluster define each feature is defined as each dimensionality with reduction of selection of attributes. Our proposed approach gives high performance when compare to existing clustering approaches by means of precision, time on some real time synthetic data set present in real time environment. Our experimental results perform on company statistical data with text, image orientated data, and EFSC produces small sub set of features with high precision and low time efficiency for real time data sets. Further improvement of proposed approach is to extend and support multi-view representation of multiple attributes using some machine learning algorithms.

References

[1] Qinbao Song, Jingjie Ni, and Guangtao Wang, "A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 1, JANUARY 2013.

[2] Natthakan Iam-On, Tossapon Boongoen, Simon Garrett, and Chris Price, "A Link-Based Cluster Ensemble Approach for Categorical Data Clustering", In IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 3, March 2012.

[3] Gionis, H. Mannila, and P. Tsaparas, "Clustering Aggregation," Proc. Int'l Conf. Data Eng. (ICDE), pp. 341-352, 2005.

[4] N. Nguyen and R. Caruana, "Consensus Clusterings," Proc. IEEE Int'l Conf. Data Mining (ICDM), pp. 607-612, 2007.

[5] A.P. Topchy, A.K. Jain, and W.F. Punch, "Clustering Ensembles: Models of Consensus and Weak Partitions," proceedings in IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 12, pp. 1866-1881, Dec. 2005.

[6] T. Boongoen, Q. Shen, and C. Price, "Disclosing False Identity through Hybrid Link Analysis," Artificial Intelligence and Law, vol. 18, no. 1, pp. 77-102, 2010.

[7] F. Fouss, A. Pirotte, J.M. Renders, and M. Saerens, "Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation," in IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3, pp. 355-369, Mar. 2007.

[8] P. Chanda, Y. Cho, A. Zhang, and M. Ramanathan, "Mining of Attribute Interactions Using Information Theoretic Metrics," Proc. IEEE Int'l Conf. Data Mining Workshops, pp. 350-355, 2009.

[9] S. Chikhi and S. Benhammada, ,"ReliefMSS: A Variation on a Feature Ranking Relieff Algorithm," proceedings in Int'l J. Business Intelligence and Data Mining, vol. 4, nos. 3/4, pp. 375-390, 2009.

[10] J. Demsar, in J. "Statistical Comparison of Classifiers over Multiple Data Sets," Machine Learning Res., vol. 7, pp. 1-30, 2006.

[11] G. Van Dijck and M.M. Van Hulle, ,"Speeding Up the Wrapper Feature Subset Selection in Regression by Mutual Information Relevance and Redundancy Analysis," in Proc. Int'l Conf. Artificial Neural Networks, 2006.

[12] C. Sha, X. Qiu, and A. Zhou, "Feature Selection Based on a New Dependency Measure," Proc. Fifth Int'l Conf. Fuzzy Systems and Knowledge Discovery, vol. 1, pp. 266-270, 2008.

[13] Z. Zhao and H. Liu, "Searching for Interacting Features," in Proc. 20th Int'l Joint Conf. Artificial Intelligence, 2007.

[14] Z. Zhao and H. Liu, J. "Searching for Interacting Features in Subset Selection," Intelligent Data Analysis, vol. 13, no. 2, pp. 207-228, 2009.

[15] S. Das, "Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection," Proc. 18th Int'l Conf. Machine Learning, pp. 74-81, 2001.

[16] M. Dash and H. Liu, "Consistency-Based Search in Feature Selection," Artificial Intelligence, vol. 151, nos. 1/2, pp. 155-176, 2003.
[17] J. Demsar, "Statistical Comparison of Classifiers over Multiple Data Sets," J. Machine Learning Res., vol. 7, pp. 1-30, 2006.
[18] I.S. Dhillon, S. Mallela, and R. Kumar, "A Divisive Information Theoretic Feature Clustering Algorithm for Text Classification,"J. Machine Learning Research, vol. 3, pp. 1265-1287, 2003.
[19] E.R. Dougherty, "Small Sample Issues for Microarray-Based Classification," Comparative and Functional Genomics, vol. 2, no. 1,pp. 28-34, 2001.
[20] U. Fayyad and K. Irani, "Multi-Interval Discretization of Continuous- Valued Attributes for Classification Learning," Proc. 13th Int'l Joint Conf. Artificial Intelligence, pp. 1022-1027, 1993.
[21] D.H. Fisher, L. Xu, and N. Zard, "Ordering Effects in Clustering," Proc. Ninth Int'l Workshop Machine Learning, pp. 162-168, 1992.
[22] F. Fleuret, "Fast Binary Feature Selection with Conditional Mutual Information," J. Machine Learning Research, vol. 5, pp. 1531-1555,2004.