JCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-1 E-ISSN: 2347-2693

Clustering Mutual Outline for Multi Assessment Temporal Data and cancer Data

Sathishkumar K.[1*], V. Thiagarasu[2], E. Balamurugan[3], David Otto Arthur[4]

[1*] Dept. of CS, Gobi Arts and Science College (Autonomus), Gobi, India
[2] Dept. of CS, Gobi Arts and Science College (Autonomus), Gobi, India
[3]Bluecrest College, Accra, Ghana, West Africa
[4]Bluecrest College, Accra, Ghana, West Africa

[*]*Corresponding Author: sathishmsc.vlp@gmail.com*

Available online at: www.ijcseonline.org

Received: 22/Dec/2017, Revised: 31/Dec/2017, Accepted: 19/Jan/2018, Published: 31/Jan/2018

Abstract— Clustering temporal data compiled from cancer registries is a crucial problem faced by many data analyst owing to the elevated high dimensionality, weight value calculation, multi view data and multifaceted temporal correlation. This research work reveals a hypothetical effect of Temporal Clustering (TC) in various domains on cancer genome risk estimates by introducing data mining clustering algorithms. For the first time, cancer genome datasets samples were made available for the complete genome sequences consisting of point mutations and structural alternations for a huge number of cancer types which allows the variation of cancer subtypes in an exceptional excellent global analysis. In this work, TC algorithm is presented to the allocation of several time- series into a set of non-overlapping parts that fit in to k temporal clusters. The paper presents a group of clustering communal framework for multi view data, TW-K-means and an automated two-level variable clustering algorithm that can be used to calculate the weights for views and person variables. A new ATBCWCE structure is projected to improve the risk estimates in cancer genome.

Keywords — Temporal data clustering, weighted consensus function, multi view learning, k-means.

I. INTRODUCTION

In all the pattern analysis tasks, clustering is very important approach to data analysis. This is used for summarization purpose, prototype extraction, and also for the big data dimensionality reduction. K-means clustering, fuzzy c-means, k-means++ and all the other kind of variations used in clustering algorithms provides a better trade-off between the quality of the elucidation and also the difficulty of computational process [1]. On the other hand, the Dynamic Time Warping (DTW) usually using a k-means clustering of temporal data poses some difficulty due to the number of temporal kernels which requires a multiple temporal data concurrently [2-3] is difficult for the reason. In the temporal data clustering method to avoid the centroid estimation difficulties, k-means and kernel k-means [5] are commonly used [6]. And these methods are unsuitable for the capturing local temporal features [7].

Specifically, the present experimental studies expose that the temporal data clustering having major difficulties in the temporal data mining owing to the high dimensionality and multifaceted temporal correlation [2-5].

The two-step algorithms for clustering method were developed by Liao [8] for the multivariate time series of equivalent or uneven length. Initially k-means and FCM clustering algorithms are processed to find the real valued time series. This univariate discrete valued time series variable is interpreted as state variable process. The second step is used to form the transition probability matrices by grouping the univariate discrete valued time series in the k-means or FCM algorithm into some numerous clusters.

The Normalized Longest Common Subsequence (NLCS) are used to analyse the time series. This method was implemented by Dacheng [9]. This NLCS method is usually used in comparing the character sequences. In this work, Dacheng was implemented a new NLCS algorithm to estimate the comparison of time series. This algorithm is commonly used for all the unique subsequence as an alternative for a highest general sequences correctly.

The data stream clustering methods was implemented by Wu [10] and it is used to analysis the stock data. For the period of clustering process the data stream clustering method intended to keep shape and have a propensity features. There are two segments are available in this clustering process. The first one is online clustering and the second one is offline macro clustering.

The upcoming region in the machine learning application, the clustering ensemble algorithms are presently studied in the various kind of view point. For example, evidence aggregation [19], graph partitioning [18] and semi definite programming optimization [20]. The essential thought of the following clustering ensemble algorithm is grouping of multiple partitions on the similar data set to generate consent particles. This consent particle is usual to be the better to that of specified input partitions. Even though in the theoretical justification on the clustering ensemble algorithms, increasing experiential evidences maintain like thought and mention that the clustering ensemble is proficient of detecting new cluster structures [18-20]. Additionally, in the clustering ensemble the official analysis exposes beneath some particular conditions, an appropriate consent result uncovers the inherent structure of a given data set [21]. Consequently, the clustering ensemble approach gives common methods to use a various representation together for the temporal data clustering.

From the various feature spaces, the multi view data is the variable that having the multiple views such as representation or group of variables. It is the result of combination of several types of measurements on examination from various perspectives and various types of measurements can be referred as various views. In the precedent decade, the multi view data has increasing the security and it is called as multi view clustering [22]. In the clustering method, from the set of variables it takes the numerous views and also eliminates the differences between various views.

From the multiple views, multi view clustering reveals the information and get the dissimilar between the various views into deliberations in order to generate an exact outcome and vigorous data partitioning. In cluster analysis process the variable weighting clustering has been significant examine subject [23-24]. In this case, it is robotically calculate a weight for every variable and detect the imperative variables and unimportant variables all the way through variable weight. There are two levels of variables available in the multi view data. In the clustering method in the multi view data, the comparison of the views and the major role of individual variables in every view added to the dataset. In the variable weighting clustering methodology, only calculate the weights for individual variables and eliminate the comparison in views in the multi view data.

This future work having the initial clustering analysis on various representations to generate a multiple partitioning process and clustering ensemble algorithm is used to generate a final partition process by grouping those partitions to gather an initial clustering process. At the same time in the initial clustering analysis can also be used to the previous clustering algorithms and intend a new weighted clustering ensemble algorithm of a two-step process. In the ATBCWCE structure, a weighting consensus function adjusting the input partition to person concurrence partitions based on different clustering validation. In the multi view temporal data for estimating the weight the Big Bang big Crunch is implemented using Automated TW-k means Weighting Clustering Ensemble Structure. It is known as ATBCWCE. The common analysis has also been done by using the ATBCWCE algorithm.

MOTIVATION AND PROBLEM SPECIFICATION

In the classification method the ensemble algorithm is used [25], ensemble methodology is also used in a pattern categorization and data mining process too. Clustering ensemble algorithm is used to grouping the multiple partitioning of the data to get a single clustering outcome. There are two phases available in the clustering ensemble algorithm. The first phase is producing a number of clustering result and the second phase is known as ensemble creation, at the same time the second phase is also known as clustering ensemble trouble. In common, the individual clustering result can be produced from various perspectives, like various data sample subsets [26].

The above-mentioned partitioning methodology is not suitable for the multi view temporal data. It is the outcome of integration of multiple types of dimensions on annotations from various point of view and dissimilar types of estimation can be measured as various views. For example, the blood cell data can be partitioned into views of density, geometry, colour and texture. Each and every one is representing a view of the significant evaluation on the blood cell. In the previous method, the multi view data are used in the multi view clustering [28-30]. Compared to the other clustering method this method is used to get the multiple views as a set of variables and eliminate a dissimilar content between the various views. But the other algorithms in [28-30] have major difficulties that these algorithms are not suitable for the temporal datasets. To overcome these problems the TW-k-means methodology are proposed and also the two-level variable weighting k-means clustering algorithms for the multi view data. The TW-k-means algorithms [31] is used to compare the impacts of various views and various variables in the clustering process, the weights of views and individual variables are implemented to the distance function.

The temporal clustering analysis process gives the essential path to determine the inherent structure and compact data in excess of temporal data by reveal changeable characteristics beneath temporal data in an unauthorized path. In common, in the clustering analysis process there are two troubles is available. One is model selection and another one is grouping. Some of the inherent clusters are not covered underlying temporal dataset, at the same time the coherent sequences are combined together to create a cluster matching. Specifically, present experiential topic [26] exposes that temporal data clustering focus a major trouble in temporal data mining owing to the high dimensionality and complex temporal correlation. For that only the TW-k-means algorithm is introduced to overcome this trouble.

II. PROPOSED METHODOLOGY

Cluster ensemble algorithm produce numerous of various clustering and grouped together automatically and the exact consensus clustering. Specifically, [26] exposes that temporal data clustering focus a trouble on the temporal data mining owing to the high dimensionality and complex temporal correlation. However, in the TW-k-means methodology the estimation of weight values and the clustering ensemble methodology is not suitable for all the work. For that reason, the Big Bang Big Crunch is implemented to calculate the weight value in the Automated TW-k means Weighting Clustering Ensemble structure for multi view temporal data namely ATBCWCE. At first, a simple clustering ensemble structure to demonstrate the compassion of the ATBCWCE algorithm.

Temporal Data Representations

The temporal data can be written as,

$$\{x(t)\}_{t=1}^{T}$$

In this equation explains with a length of T temporal data points, using two representations namely Piecewise Local Statistics and Piecewise Discrete Wavelet Transform are implemented in the previous methods [32] and also the two-classical global representation are available namely Polynomial Curve Fitting and Discrete Fourier Transforms.

Weighted Consensus Function

The common thought of the weighted consensus function is the used to get the similarity between the objects in a partition for evident accumulation. This similarity matrix is determined from the weighted partitions and weights are calculated by measuring the clustering quality with various clustering process. after that, a clustering is implemented based on the entire similarity matrices to produce a person consensus partitions.

Partition Weighting Scheme

Assume that $X = \{x_n\}_{n=1}^{N}$ is a data set of n=1 to N objects and there are M= 1 to m partitions$P = \{P_m\}_{m=1}^{M}$, where the cluster number in M partitions could be vary. This information is gathered from the initial clustering analysis. Partition weighting method fixes a weight w_m^{π} to every P_m in terms of a clustering validation criterion, and weights for the entire partitions based on the criterion selectively form a weight vector $w = \{w_m^{\pi}\}_{m=1}^{M}$ for the partition collection P. In the partition weighting method, describe a weight

$$w_m^{\pi} = \frac{\pi(P_m)}{\sum_{m=1}^{M} \pi(P_m)} \qquad (1)$$

where $w_m^{\pi} > 0$ and $\sum_{m=1}^{M} w_m^{\pi} = 1$. $\pi(P_m)$ Is the clustering validity index value in term of the criterion π . instinctively, the weight of a partition would express its contribution to the grouping in terms of its clustering quality measured by the clustering validation criterion π.

In the above equation (1), assume the weight values depending on single view only not for multi view so it is comprehensive to multi view data clustering.

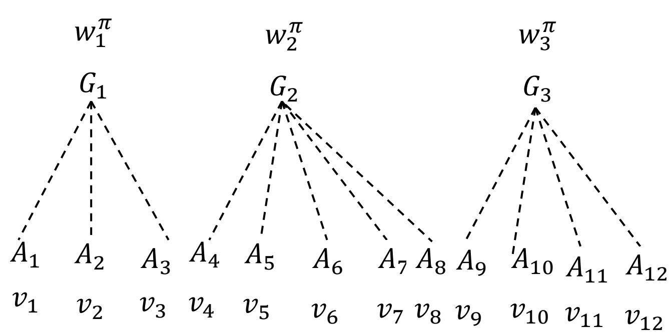


Figure 1. Example of ATBCWCE framework

Figure 1 demonstrates the ATBCWCE structure. The set of partition matrix M with N objects represented by the set A of variables $P = (P_{1n}, P_{2n}, \dots . P_{mn})$. Consider A is partitioned to T views $\{G_t\}_{t=1}^{T}$. In this case, $G_t \cap G_s = \emptyset$ for $s \neq t$ and $\bigcup_{t=1}^{T} G_t = A$. Allow $W = \{w_1, w_2, \dots w_T\}$ be a set of T view weights. In this case, w_t is referred to as weight and this value is assigned to the t[th] view and $\sum_{t=1}^{T} w_t = 1$. In the set of m variable weights allow$V = \{V_j\}$. In this case, v_j is referred to as the weight and it is assigned to the j[th] variable and $\sum_{j \in G_t} v_j = 1, (1 \leq t \leq T), \sum_{j=1}^{o} v_j = T$. Consider k cluster available in X. determine the set of k clusters from G. determine the set of k cluster from G. and also validate the views from the view weight matrix $W = \{w_1, .. w_T\}$ and validate the variable from the variable weight matrix $V = [v_j]_o$. suppose the set of individual variables in data X considered as G and this trouble is corresponding to the individual variable weighting. Consequently, the two level variable weighting methodology as a simplification of the present variable weighting methods.

In the two levels variable weighting methodologies, the variable weights V are used to validate the specific variables in every view and the view weights W are used to validate the compact cluster structure.

Problem formulation

The clustering process to partition P_{mn} into k clusters with weights for both views and individual variables is modelled as minimization of the following objective function

$$F(U,Z,V,W) = \sum_{i=1}^{m}\sum_{j=1}^{o}\sum_{t=1}^{T}\sum_{l=1}^{k} u_{i,l}\, w_t v_j d(P_{i,j}, z_{l,j}) + \eta \sum_{j=1}^{o} v_j \log(v_j) + \lambda \sum_{t=1}^{T} w_t \log(w_t) \quad (2)$$

Focus to

$$\begin{cases} \sum_{l=1}^{p} u_{i,l} = 1\,, u_{i,l} \in \{0,1\}, 1 \le i \le m \\ \sum_{t=1}^{T} w_t = 1\ , 0 \le w_t \le 1 \\ \sum_{j=1}^{o} v_j = 1\,, 0 \le w_t \le 1, 1 \le t \le T \end{cases} \quad (3)$$

Where

U is referred to as $m \times k$ partition matrix whose essentials $u_{i,l}$ are binary where $u_{i,l} = 1$ indicates that object i is owed to cluster l

$Z = \{Z_1, Z_2, \dots Z_k\}$ s a set of k vectors demonstrating the centres of the k clusters;

$W = \{w_1, w_2, \dots w_T\}$ are T weights for T views;

$V = \{v_1, v_2, \dots, v_o\}$ are o weights for o variables;

$\lambda > 0, \eta > 0$ are two given parameters;

$d(P_{i,j}, z_{l,j})$ is a distance or dissimilarity measure on the j^th^ variable among the i^th^ object and the centre of the l^th^ cluster. If the variable is arithmetical, then

$$d(P_{i,j}, z_{l,j}) = (P_{i,j} - z_{l,j})^2 \quad (4)$$

If the variable is categorical, then

$$d(P_{i,j}, z_{l,j}) = \begin{cases} 0, (P_{i,j} = z_{l,j}) \\ 1, (P_{i,j} \ne z_{l,j}) \end{cases} \quad (5)$$

In the first equation is estimating the addition of the cluster dispersions. The second and third equation is the two negative weight entropies. For the more number of views and variables to manage the incentive for clustering process two positive parameters λ and η is used. There are four steps are used to solve the minimization trouble by iteratively based on (2).

1. Problem Prb_1: Fix $Z = \widehat{Z}, V = \widehat{V}$ and $W = \widehat{W}$, and solve the reduced problem $P(U, \widehat{Z}, \widehat{V}, \widehat{W})$;

2. Problem Prb_2: Fix $U = \widehat{U}, V = \widehat{V}$, and $= \widehat{W}$, and solve the reduced problem $P(\widehat{U}, Z, \widehat{V}, \widehat{W})$;

3. Problem Prb_3: Fix $U = \widehat{U}, Z = \widehat{Z}$ and $= \widehat{W}$, and solve the reduced problem $P(\widehat{U}, \widehat{Z}, V, \widehat{W})$;

4. Problem Prb_4: Fix $U = \widehat{U}, Z = \widehat{Z}$, and $V = \widehat{V}$, and solve the reduced problem $P(\widehat{U}, \widehat{Z}, \widehat{V}, W)$;

Let $U = \widehat{U}, Z = \widehat{Z}$, and $W = \widehat{W}$ be fixed. $P(\widehat{U}, \widehat{Z}, V, \widehat{W})$is minimized if and only if

$$v_j = \frac{\exp\left\{\frac{-E_j}{\eta}\right\}}{\sum_{h \in G_t} \exp\left\{\frac{-E_h}{\eta}\right\}} \quad (6)$$

$$E_j = \sum_{l=1}^{k}\sum_{i=1}^{m} \hat{u}_{i,l}\, \hat{w}_t\, d(P_{i,j}, \hat{z}_{l,j}) \quad (7)$$

In the above equations t is referred to as the index of the view that the j^th^ variable is assigned to the process.

Let $U = \widehat{U}, Z = \widehat{Z}$, and $V = \widehat{V}$ be fixed. $P(\widehat{U}, \widehat{Z}, \widehat{V}, W)$is minimized if and only if

$$w_t = \frac{\exp\left\{\frac{-D_t}{\eta}\right\}}{\sum_{h=1}^{T} \exp\left\{\frac{-D_h}{\eta}\right\}} \quad (8)$$

$$D_t = \sum_{l=1}^{k}\sum_{i=1}^{m}\sum_{j \in G_t} \hat{u}_{i,l}\, \hat{v}_j\, d(P_{i,j}, \hat{z}_{l,j}) \quad (9)$$

To manage the allocation of the two types of weights V and W the input parameters λ and η are used. If $\lambda = 0$ and $\eta = 0$ means it is very easy to validate the objective functions with respect to the V and W.

Big Bang–Big Crunch (BB–BC)

Randomness can be seen as corresponding to the energy debauchery at the same time as convergence to a local or global optimum point. This is known as gravitational attraction. The GA and BB-BC methodologies are similar and both are used to generate an initial population arbitrarily. In this proposed work the weights of views and individual variables are considered as inputs to the k cluster. The generation of initial population arbitrarily from the weights values of the view and individual variables is known as Big Bang phase. In these particular steps, the weight values of the persons are broadening the entire search space unvaryingly. The Big Bang phase method is working on the basis of Big Crunch phase. The Big Crunch is one of the convergence operators that have more number of weights for views and individual variables as input but only one clustering outcome can be generated beneath k clusters and it is known as centre of mass. By using this centre of mass value, the cluster can be estimated. The following formula is used to estimate the optimal weight calculation and it is represented by x_c

$$x_c = \frac{\sum_{i=1}^{n} \frac{1}{f^i} x_i}{\sum_{i=1}^{n} \frac{1}{f^i}} \tag{10}$$

In the above equation xi is referred to as the selected weights for both views and individual variables surrounded by an n-dimensional search space produced, the fitness value of the data is represented by f^i , N is represented as the population size in Big Bang phase. To evaluate the centre of gravity the convergence operator is used by choosing the two dissimilar weight values. In the Big-Crunch phase both the views and individual variables weights are selected as members and it is used as a contraction operator. After this Big Crunch phase process, the novel weights for both views and individual variables are generated. Subsequently this value can be used the next iteration process.

The second detonation subsequent to, the fitness value of every centre and cluster data point is re-estimated. These consecutive detonation and reduction steps are conceded continuously in anticipation of a stopping criterion has been met. For both the views and individual variables the elapsed iteration generates good weights and reduces a value for the partition. In the iteration process, the partition will reach zero value means this partition process goes to infinity stage.

In terms of various kinds of partitions, a weighted similarity matrix S is used to replicate the collective relationship between the entire data. This matrix is usually tended to the collect confirmation for the clustering quality and for this reason the entire partitions are treated in different manner.

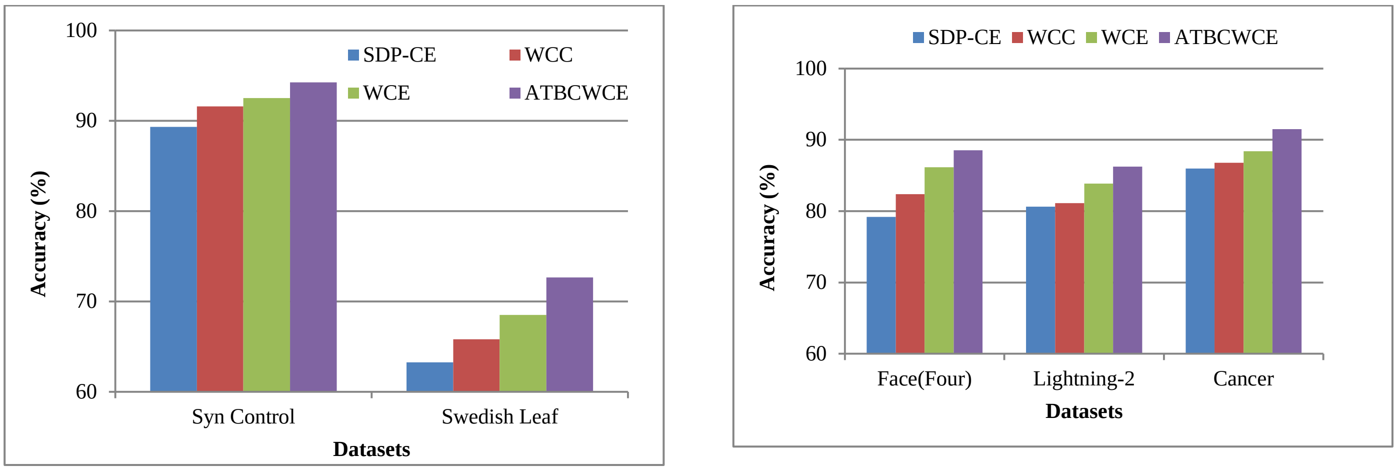
III. EXPERIMENTATION RESULTS

For assessment purpose, the ATBCWCE methodology is applied to the gathering of time-series benchmarks for temporal data mining [35]. To estimate a time series categorization and clustering algorithms in the context temporal data mining there are sixteen synthetic or time series datasets are gathered [35]. In this anthology, by using the already available information like the number of classes K and the time series class label in a data set auxiliary divided namely training and testing subsets. This training and testing subsets are used to estimate the classification algorithm. In this process, the entire sixteen datasets are used. Both the training and test subsets are available in this dataset. The classic temporal data clustering algorithms are directly applied on the time series to extract their performance on the benchmark collection. This is yielded by temporal proximity and model based algorithm.

It's make use of the three state algorithms implemented from various points of view for comparison. These are Semi Definite Programming based Clustering Ensemble (SDP-CE) algorithm [20], Weighted Clustering Ensemble (WCE) algorithm [37], Weighted Consensus Clustering (WCC) algorithm [36] and ATBCWCE algorithm. In this work, for the first clustering analysing process the k-mean step is used. By using the given information, the three algorithms are implemented without addressing the model selection. And in the k-mean step the accurate cluster number of every data set $K^*$is used. In the process there are ten partitions are produced for a single representation. So, totally 40 partitions are produced and these partitions are grouped together for every dataset. These similar methods are used in the WCE also for k-mean to generate a partition. And the K values are chosen arbitrarily from $K^* - 2 \leq K \leq K + 2 (K > 0)$. After the twenty trails, the average and standard deviation of classification accuracy rates is stored in the dataset. In this work, the four dataset samples are used for the clustering process namely Syn Control, Face (four), Swedish Leaf, Two patterns and Lightning-2.

Table 1: Clustering Accuracy (in Percent) of Different Clustering Algorithms on Time-Series Benchmarks [35]

Dataset	Accuracy (%)				Error (%)			
	Different representations				Different representations			
	SDP-CE	WCC	WCE	ATBCWCE	SDP-CE	WCC	WCE	ATBCWCE
Syn Control	89.32	91.58	92.51	94.26	10.68	8.42	7.49	5.74
Swedish Leaf	63.28	65.81	68.52	72.65	36.72	34.19	31.48	27.35
Face(Four)	79.21	82.38	86.14	88.53	20.79	17.62	13.86	11.47
Lightning-2	80.63	81.14	83.85	86.26	19.37	18.86	16.15	13.74
Cancer	85.96	86.78	88.41	91.51	14.04	13.22	11.59	8.49



(a). Accuracy of different clustering methods to Syn Control and Swedish Leaf

(b). Accuracy of different clustering methods to Face (Four), Lightning-2 and Cancer

Figure 2: Clustering Accuracy (in Percent) of Different Clustering Algorithms on Time-Series Benchmarks

Table 1 illustrate the collective list of each and every outcome of the clustering methods. First of all for the clustering process the accurate cluster number for four datasets have been calculated. This cluster number was mentioned in Table 1. It indicates the clustering temporal data of high dimension.

Specifically, the SDP-CE algorithm took a more time to complete a comparison process with other algorithms including the clustering ensemble algorithm. The ATBCWCE algorithms produce a better outcome when compared to the other algorithms. From observing the Table 1, the ATBCWCE is usually better than the SDP-CE algorithm on the suitable representation space and the preeminent parameter setup.

For example, considering the Swedish Leaf dataset the ATBCWCE produce 72.65% outcome correctness. When this outcome is compared to the SDP-CE, WCC and WCE methods it is 9.37%, 6.84% and 4.13% higher respectively. The final outcome of five different dataset and their clustering methods are illustrated in Figure 2 (a) and (b). For example, considering the cancer dataset samples proposed ATBCWCE produces 91.51% accuracy which is higher when compared to other existing algorithms.

The measuring the clustering accuracy used Precision, Recall, F-measure, accuracy and average cluster entropy to evaluate the results.

Table 2: Summary of Different Clustering Algorithms on five datasets by four parameters

Dataset	Evaluation	SDP-CE	WCC	WCE	ATBCWCE
Syn Control	**Precision (%)**	86.58	88.63	90.75	92.18
	Recall (%)	89.52	90.25	91.56	92.84
	F measure (%)	88.02	89.43	91.15	92.50
	ACE	2.85	1.92	1.58	1.23
Swedish Leaf	**Precision (%)**	67.53	68.81	69.71	72.95
	Recall (%)	69.82	71.25	72.83	74.18
	F measure (%)	68.65	70.01	71.24	73.56

	ACE	1.57	1.28	1.15	0.76
Face(Four)	**Precision (%)**	80.23	81.56	84.19	86.52
	Recall (%)	81.29	82.53	84.23	87.46
	F measure (%)	80.76	82.04	84.21	86.98
	ACE	1.63	1.42	1.38	1.26
Lightning-2	**Precision (%)**	80.52	81.47	82.58	83.79
	Recall (%)	81.26	82.18	83.45	84.15
	F measure (%)	80.88	81.82	83.01	83.96
	ACE	1.93	1.67	1.28	1.09
Cancer	**Precision (%)**	86.93	87.53	88.19	91.58
	Recall (%)	89.32	91.51	92.37	93.81
	F measure (%)	88.125	89.52	90.28	92.695
	ACE	1.23	1.15	1.08	0.95

The total clustering outcome is shown in Table 2. According this outcome, the ATBCWCE considerably to another three algorithms in nearly too every outcome, specifically on the sample from the five data set. Even though ATBCWCE is the elaboration to TW –k means and WCE methodology, the weights on views increased the outcome of this process. In every five data sets the SDP-CE generate a reasonable outcome of this process. One of the very imperative scrutinise is the estimation of the clustering fault by beneath the result into different clusters. At last, in the ATBCWCEW process generate a smaller amount of clustering fault when compare to the previous methodologies because this ATBCWCE process is without any difficulties it is relevant to the temporal and multi view data.

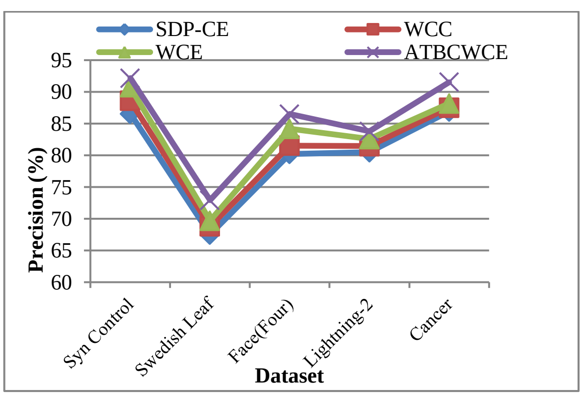


Figure 3: Precision of Different Clustering Algorithms on Time-Series Benchmarks datasets

When compared to the SDP-CE, WCC and WCE methodologies the ATBCWCE with multi view and temporal data achieve a better result. The result is 92.18% clustering precision which is 25.6%, 3.55% and 1.43% increased with respect to the data set samples.

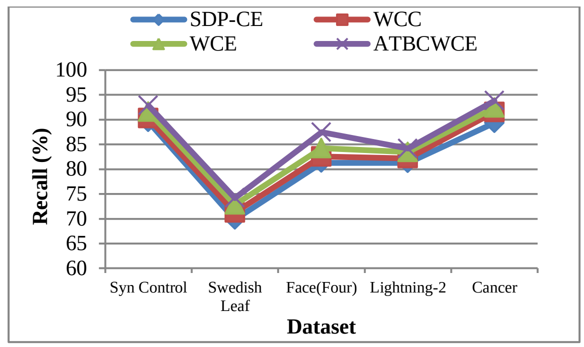


Figure 4: Recall of Different Clustering Algorithms on Time-Series Benchmarks datasets

When compared to the SDP-CE, WCC and WCE methodologies the ATBCWCE with multi view and temporal data achieve a better result. The result is 92.84% clustering precision evoke value which is 3.32%, 2.59% and 1.28% increased with respect to the data set samples.

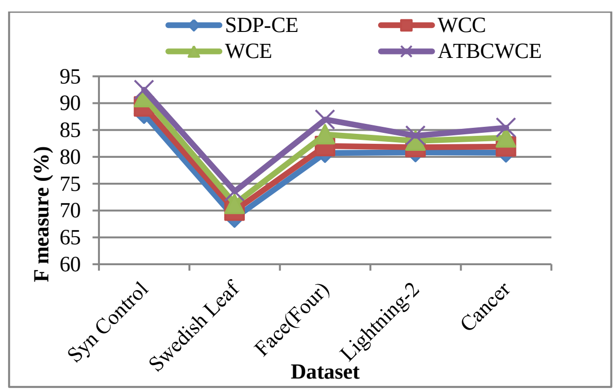


Figure 5: F-measure of Different Clustering Algorithms on Time-Series Benchmarks datasets

When compared to the SDP-CE, WCC and WCE methodologies the ATBCWCE with multi view and temporal data achieve a better result. The result is 92.5% F-measure which is 4.48%, 3.07% and 1.35% increased with respect to the data set samples.

IV. CONCLUSION

The ATBCWCE algorithm has a huge potential to cluster cancer genomics data repositories, along with cluster subtypes to cluster and analyse these data. This method estimates the weights for both views and individual variable from the multiple view temporal data at the same time in the clustering process. By using the compact views, weight values and significant variables can be detected whiles low quality views eliminate the noise variables. In this proposed work the ATBCWCE calculates the multiple view weights. These weights produce a novel weighting method beneath a data set in common and expose the convergence possessions of the view weights. When compared to the three clustering algorithms and ATBCWCE on five temporal data sets, the outcome revealed that the ATBCWCE produced a better result as compared to the other three clustering algorithms in the four assessment indices like Recall, F-measure, Accuracy and Precision. In upcoming work, the two-level variable weighting method and another method like fuzzy methodology, semi-supervised methodology and subspace clustering methodology are grouped together automatically.

REFERENCES

[1]. Arthur D., S. Vassilvitskii , K-means++: the advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’07, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2007, pp. 1027–1035 .

[2]. Cuturi M., Fast global alignment kernels, in: Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, pp. 929–936 .

[3]. Cuturi M., J.-P. Vert , Ø. Birkenes , T. Matsui , A kernel for time series based on global alignments, in: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 11, 2007, pp. 413–416.

[4]. Petitjean F., A. Ketterlin , P. Gançarski , A global averaging method for dynamic time warping, with applications to clustering, Pattern Recognit. 44 (3) (2011) 678–693 .

[5]. Dhillon I.S., Y. Guan , B. Kulis , Kernel k-means: spectral clustering and normalized cuts, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2004, pp. 551–556 .

[6]. Liao W., Clustering of time series data – a survey., Pattern Recognit. 38 (2005) 1857–1874 .

[7]. Frambourg C., A. Douzal-Chouakria , E. Gaussier , Learning multiple temporal matching for time series classification, in: A. Tucker, F. Höppner, A. Siebes, S. Swift (Eds.), Intelligent Data Analysis, Springer Berlin Heidelberg, London, 2013, pp. 198–209 .

[8]. Liao T.W., Mining of vector time series by clustering, Working paper, 2005.

[9]. Dacheng N.; F. Yan; Z. Junlin; F. Yuke; X. Hu; , "Time series analysis based on enhanced NLCS," Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on, vol., no., pp.292-295, 23-25 June 2010.

[10]. Wu X.; D. Huang;, "Data stream clustering for stock data analysis," Industrial and Information Systems (IIS), 2010 2nd International Conference on , vol.2, no., pp.168-171, 10-11 July 2010.

[11]. Huang X., H.-l. LI,” Research on Predicting Agricultural Drought Based on Fuzzy Set and RlS Analysis Model”, 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE),pp 186-189.

[12]. Lin Y.; Y. Yang;, "Stock markets forecasting based on fuzzy time series model," Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on, vol.1, no., pp.782-786, 20-22 Nov. 2009.

[13]. Gao, S., He, Y., & Chen, H. (2009). Wind speed forecast for wind farms based on ARMA-ARCH model. International Conference on Sustainable Power Generation and Supply, 2009. SUPERGEN'09, pp. 1-4.

[14]. Jixue D.; , "Data Mining of Time Series Based on Wave Cluster," Information Technology and Applications, 2009. IFITA '09. International Forum on , vol.1, no., pp.697-699, 15-17 May 2009

[15]. Powell N.; S.Y. Foo; M. Weatherspoon; "Supervised and Unsupervised Methods for Stock Trend Forecasting," System Theory, 2008. SSST 2008. 40th Southeastern Symposium on , vol., no., pp.203-205, 16-18 March 2008.

[16]. Wu J. X., J. L. Wei, “Combining ICA with SVR for prediction of finance time series”, Proceedings of the IEEE International Conference on Automation and Logistics August 18 - 21, 2007, Jinan, China, pp 95-100.

[17]. Verdoolaege G. and Y. Rosseel,” Activation Detection In Event-Related FMRI Through Clustering Of wavelet Distributions”, Proceedings of 2010 IEEE 17th International Conference on Image Processing, September 26-29, 2010, Hong Kong, pp 4393-4395.

[18]. Fern X. and C. Brodley, “Solving Cluster Ensemble Problem by Bipartite Graph Partitioning,” Proc. Int’l Conf. Machine Learning, pp. 36-43, 2004

[19]. Gionis A., H. Mannila, and P. Tsaparas, “Clustering Aggregation,” ACM Trans. Knowledge Discovery from Data, vol. 1, no. 1, article no. 4, Mar. 2007.

[20]. Singh V., L. Mukerjee, J. Peng, and J. Xu, “Ensemble Clustering Using Semidefinite Programming,” Advances in Neural Information Processing Systems, pp. 1353-1360, 2007.

[21]. Topchy A., M. Law, A. Jain, and A. Fred, “Analysis of Consensus Partition in Cluster Ensemble,” Proc. IEEE Int’l Conf. Data Mining, pp. 225-232, 2004.

[22]. Chaudhuri K., S. Kakade, K. Livescu, and K. Sridharan, “Multiview Clustering via Canonical Correlation Analysis,” Proc. 26th Ann. Int’l Conf. Machine Learning, pp. 129-136, 2009

[23]. Tsai C.-Y. and C.-C. Chiu, “Developing a Feature Weight Self-Adjustment Mechanism for a k-Means Clustering Algorithm,” Computational Statistics and Data Analysis, vol. 52, no. 10, pp. 4658- 4672, 2008.

[24]. Deng Z., K. Choi, F. Chung, and S. Wang, “Enhanced Soft Subspace Clustering Integrating Within-Cluster and Between-Cluster Information,” Pattern Recognition, vol. 43, no. 3, pp. 767- 781, 2010.

[25]. Parvin H, Minaei-Bidgoli B, Parvin S, Alinejad H (2012b) A New Classifier ensemble methodology based on subspace learning. J Exp Theor Artif Intell. doi:10.1080/0952813X.2012.715683

[26]. Parvin H, Minaei-Bidgoli B, Alinejad H (2013) Data weighing 1201 mechanisms for clustering ensembles. Comput Electr Eng. http:// 1202 dx.doi.org/10.1016/j.compeleceng.2013.02.004

[27]. Fred A, Jain AK (2002a) Data clustering using evidence accumulation. In: Proceedings of the 16th international conference on pattern recognition, pp. 276–280

[28]. Tzortzis G. and C. Likas, "Multiple View Clustering Using a Weighted Combination of Exemplar-Based Mixture Models," IEEE Trans. Neural Networks, vol. 21, no. 12, pp. 1925-1938, Dec. 2010.

[29]. B. Long, P. Yu, and Z. Zhang, "A General Model for Multiple View Unsupervised Learning," Proc. Eighth SIAM Int'l Conf. Data Mining (SDM '08), 2008.

[30]. Greene D. and P. Cunningham, "A Matrix Factorization Approach for Integrating Multiple Data Views," Proc. European Conf. Machine Learning and Knowledge Discovery in Databases, pp. 423-438, 2009.

[31]. Chen, X., Xu, X., Huang, J. Z., & Ye, Y. (2013). TW-k-means: automated two-level variable weighting clustering algorithm for multiview data. IEEE Transactions on Knowledge and Data Engineering, 25(4), 932-944.

[32]. S. Wang and K. Chen, "Ensemble Learning with Active Data Selection for Semi-Supervised Pattern Classification," Proc. Int'l Joint Conf. Neural Networks, 2007.

[33]. W. Chen and S. Chang, "Motion Trajectory Matching of Video Objects," Proc. SPIE/IS&T Conf. Storage and Retrieval for Media Database, 2000.

[34]. C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast Subsequence Matching in Time-Series Databases," Proc. ACM SIGMOD, pp. 419-429, 1994.

[35]. E. Keogh, Temporal Data Mining Benchmarks, http://www.cs.ucr.edu/~eamonn/time_series_data, 2010.

[36]. H. Kien, A. Hua, and K. Vu, "Constrained Locally Weighted Clustering," Proc. ACM Int'l Conf. Very Large Data Bases (VLDB), pp. 90-101, 2008

[37]. Yang, Y., & Chen, K. (2011). Temporal data clustering via weighted clustering ensemble with different representations. IEEE Transactions on Knowledge and Data Engineering, 23(2), 307-320.

