Incremental Feature Transformation for Temporal Space

Preeti Mahadev
University of Mysore,
Mysuru, Karnataka,
India

P. Nagabhushan
University of Mysore,
Mysuru, Karnataka,
India

ABSTRACT
Temporal Feature Space generates features sequentially over consecutive time frames, thus producing a very large dimensional feature space cumulatively in contrast to the one which generates samples over time. Pattern Recognition applications for such temporal feature space therefore have to withstand the complexities involved with waiting for the arrival of new features over time and handling the knowledge hidden in large dimensions. Although, the problem of deriving the knowledge can be overcome by dimensionality reduction techniques like feature subsetting or feature transformation, the complexity due to the large dimensions still prevails. Even though the arrival of features is temporally incremental in nature, generally the pattern analysis is not carried out over time frames to enable the production of knowledge in incremental model for more effective management over time. However, temporal data in real time applications demand that the decisions be taken in the interim or at every temporal point even before all the features arrive temporally. This problem can be overcome by accumulating and building the knowledge for pattern analysis at the end of each temporal phase in an incremental mode. The temporal arrival of features would provide an environment to accumulate the knowledge in the transformed feature space at the end of every phase thereby minimizing a large dimensional space. Since the cumulative knowledge is built upon and passed on from one phase of the temporal space to the other without looking back at the previous data, the feature space required for computing at a given instant would remain fairly constant and comparatively smaller. As fewer transformed features remain in scope for processing at each phase for further reduction and knowledge extraction, computation is minimized and memory is efficiently utilized. In this proposed research, the pattern analysis and recognition of temporal space occur at every temporal point instead of at the end when all features are available. At each temporal point, the proposed model not only withstands the mandatory wait time but also works towards generating the most updated and best available transformed feature space thus far by means of continuous knowledge extraction.

General Terms
Pattern Analysis, Principal Component Analysis (PCA), Dimensionality Reduction, Feature Subsetting, Transformed Feature Space

Keywords
Big Feature Space, Incremental Dimensionality Reduction, Cumulative Variance, Optimal Feature Subset, Incremental Dimensionality Index (IDI)

1. INTRODUCTION
Growth in database conventionally implies growth in samples. Equally interesting is the evolution of database in terms of feature space over time. The inflowing information whether it is in terms of new samples or in terms of new features, tends to change the pattern structure especially in temporal data. The previously derived knowledge need to be incrementally updated with the newly arriving features in the temporal space [4, 5, 6, 7 and 8]. Specifically when feature space evolves, understanding the structure of multidimensional patterns is of fundamental importance to researchers in data mining, pattern recognition, and machine learning [9]. Many researchers have worked on dimensionality reduction but a smaller subset is incremental in nature [11, 12, 15-18]. The research that have been conducted on incremental dimensionality reduction usually deals with reducing the sample space by using supervised learning approaches for evaluation. Research also has been conducted on reducing the attribute space in a non-transformed feature space but they tend to look back at the previously stored data [11, 12 and 16]. The transformed feature space resulting from a dimensionality reduction procedure like PCA is usually employed to assimilate the knowledge contained in a multidimensional space. An interesting situation for real time applications that involve large dimensions arises when it involves reduction of feature space while the sample space remains a constant. For example in Learning Management System(LMS), when the scores of a set of students (representing sample space) for different subjects (representing feature space) arrive over time, the skillset of the students can be analyzed substantially over time as and when different subjects are evaluated for the same set of students. The aim in the proposed approach is to update and/or restructure the transformed feature space as and when the new features become available without looking back at the previous data while the sample space remains a constant. This is referred to as incremental feature transformation. When new features arrive temporally, fusing of the new knowledge derived from the recently arrived features to the existing knowledge becomes mandatory. Because of the nature of temporal arrival of features, incremental updation would involve Sequence Compulsive Incremental Learning [2]. The envisaged incremental dimensionality reduction is expected to be carried out with minimal investment of memory and time because the model proposes to process a smaller transformed feature space at a given instant of time to capture the most descriptive features incrementally without looking back at the previous data. The proposed approach embraces a ‘zero memory model’ where in it does not retain or look back for any past features in its original form in order to accumulate and build the knowledge [2].

When a large number of attributes are available for dimensionality reduction, removal of the highly correlated attributes (which do not add any value for knowledge building or decision making activity) becomes a vital task. There would be too many pairwise correlations between the variables to consider [27]. The dispersion matrix might be too large to study and interpret properly. Visualization of data will also become a challenge. With 12 variables, for example, there will be more than 200 three-dimensional scatterplots to be studied. To interpret the data in a more

meaningful form, it is therefore necessary to reduce the number of variables to a few interpretable linear combinations of the data. The two most commonly used dimensionality reduction techniques are factor analysis (FA) and Principal Component Analysis (PCA). In a very broad sense, PCA is used when visualization of summarized data using fewer dimensions is involved. FA is used when an explanatory model for the correlations among the data is involved. In this paper, PCA is the method chosen to explore the possibilities and benefits of incremental feature transformation. The central idea of PCA is to reduce the dimensionality of a data set in which there are a large number of interrelated variables, while retaining as much as possible of the variation present in the data set [34]. This reduction is achieved by transforming the original variables to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first *few* retain most of the variation present in *all* of the original variables. Computation of the principal components reduces to the solution of an eigenvalue-eigenvector problem for a positive-semidefinite symmetric matrix [23]. Each linear combination obtained will correspond to a principal component [27]. The principal components are linear combinations of the original variables weighted by their contribution to explain the variance in a particular orthogonal dimension [29]. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible [30].

2. PROBLEM FORMULATION

Traditionally, reduction of feature space is carried out once all the features are made available and subsequently the knowledge is extracted to analyze the patterns. In this paper, it is proposed to transform and reduce the feature space at every temporal point, carry forward the extracted, cumulative knowledge and merge the cumulative knowledge with the incoming batch of features. It is also assumed here that the ground truth of the temporal data is not known. At every temporal point, it has been hypothesized that the local knowledge available from the assimilated variance in the principal components provides the best possible decision making criteria for that interval. When the knowledge is incrementally enhanced this way, the proposed model provides an ideal environment for making efficient decisions at any given instant of time. The user need not wait for all the features to become available to make a decision but can make equally good decisions at every temporal point. In addition to this, the user need not look back at the previous data to build, update or accumulate the knowledge as the knowledge at hand would have enough cumulative knowledge to represent the old data sufficiently.

In the earlier research work conducted, a subset of features that encompass the maximum variance was incrementally selected from the original set of features for reduction [31]. The reduction of features was performed on a non-transformed feature space. The performance of the model was evaluated by means of an ensemble classification method, a supervised approach. But in this paper the incremental feature selection is being performed on a transformed feature space. The principal components that represent the maximum variance is selected for reduction. Here, clustering, an unsupervised grouping approach is used to evaluate the performance and to cross validate the cluster belongingness of the transformed features. In feature subsetting, the variance of the selected original features play the role of a deciding factor for building the knowledge but in incremental feature transformation, a proportion of all the original features in terms of principal components act as a deciding factor for building the knowledge. As the coverage of information is higher in incremental feature transformation for the same reason, the performance of the transformed feature space is also expected to be better than the performance of a non-transformed feature space.

3. INCREMENTAL TRANSFORMATION APPROACH

In conventional PCA, the entire original feature set undergoes transformation at one go and provides the percentage of variance accounted in the principal factors when all the features are available. In the proposed approach, the transformation and reduction happens at every temporal phase as and when the features are made available. It also generates the cumulative percentage of the total variance at each temporal phase incrementally. The cumulative explained variance would enable the user to make a decision in the interim and further merge the knowledge at hand with the incoming features instead of waiting for all features to arrive. In statistics, the explained variation measures the proportion to which a mathematical model accounts for the variation (dispersion) of a given data set [32]. One of the informative by products of PCA is the cumulative explained variance that quantifies the proportion of variance of the original features assimilated in the transformed feature space. This in turn could be used to measure the amount of information gain in the transformed feature space.

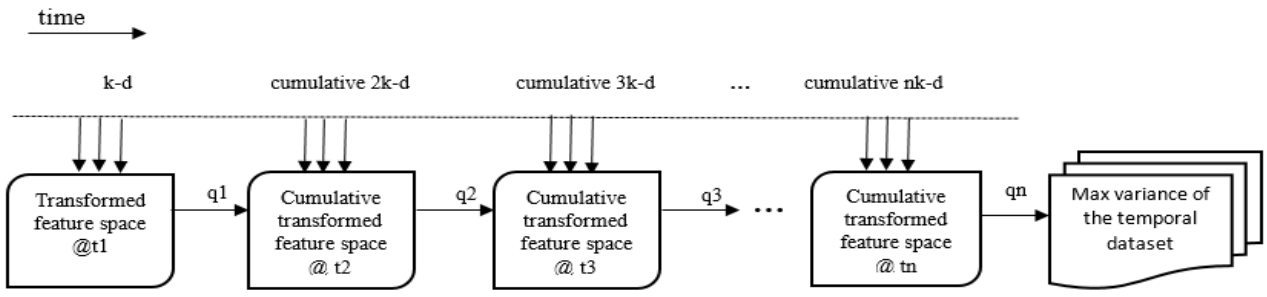

Figure 1. Incremental Feature Transformation Approach

Consider a scenario where 'm = nk' features are arriving in 'n' number of batches with 'k' features in each batch over time (see Figure 1). Here it is assumed that features f1..fk arrive in the first batch. The first temporal batch of k features are subjected to PCA based dimensionality reduction technique and a set of corresponding 'k' principal components are obtained in the transformed feature space. For the sake of illustration and for carrying out experiments, a threshold of 90% that explains the maximum variance of the original feature set has been taken into account. If a set of 'q1' principal components in the transformed feature space can explain atleast 90% of variance of the original feature space with k features (where q1 < k), those 'q1' principal components are considered to represent the best available local knowledge of that batch. This set of 'q1' transformed features is further merged with the next batch of say 'k' original features that arrive temporally. When the 'q1' principal components from the first temporal batch is merged with the next batch of 'k' original features (i.e. (q1+k) features) and is subjected to PCA , the transformation yields a new set of q2 principal components that would provide a cumulative explained variance that represents the most descriptive features available thus far. This process is carried on until there are no more new attributes for reduction and transformation. The last phase of transformation would have assimilated the cumulative variance of the entire original feature set that have arrived temporally. The last set of principal components would encapsulate the maximum cumulative explained variance of the entire temporal feature space.

3.1 An Alternative approach

An alternative approach for incremental feature transformation has been explored . It is similar to the approach discussed earlier in this paper but instead of merging the original features with the principal components incrementally, the principal components of the original features are merged with the next batch of principal components (see Figure 2) . The results from the alternate appraoch is very similar to the first one with no significant improvement in clustering performance. The alternate approach increases computation and requires more memory as it involves an additional step of transforming the features into factors at each step before merging. For a temporal data with 'n' batches, the first approach will undergo the transformations 'n' times , the alternate approach undergoes twice as much ie '2n' times. Since the results are fairly comparable and the second approach is computationally intensive ,the first approach is chosen as the model over the alternative approach.

4. ILLUSTRATIVE EXPLORATION

To illustrate the incremental transformation approach at each phase in detail, a temporal set of data known as Corn Soya dataset has been considered.

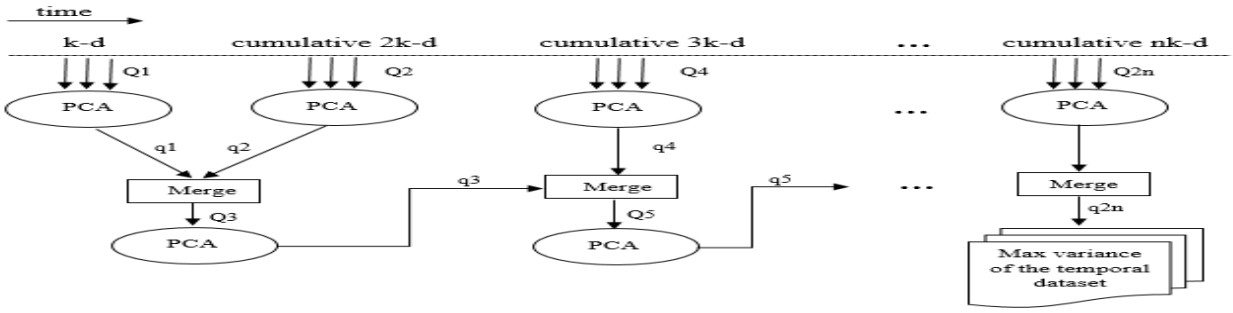

Figure 2. An alternative approach for incremental feature transformation

Corn Soya dataset from Iowa State has four batches of same feature set arriving in 4 different time slots: June 11, June 29, Jul 16, and Aug 30th in the year 1979 [26]. For illustrating some of the common traits that prove to be advantageous in the Incremental approach, 3 other datasets have also been considered. The second is a weather dataset. It has the average monthly temperatures collected over 1 year from 37 major cities across the world .The third and fourth are stock market data, Dow Jones dataset from UCI repository which has data from the first two Quarters of the year 2011. In the original dataset, Q1 and Q2 have 17 attributes for 30 stocks. The dataset has been preprocessed and rearranged to obtain 154 attributes in Q1 and 169 attributes in Q2.

Corn Soya data has 4 batches of 6 features each arriving temporally. When it is subjected to the conventional PCA with all the 24 features at once, the transformed feature space will yield 24 principal components. Among the 24 principal components, the first 4 principal components would account for > 90% of the total variance. The cumulative explained variance for Corn Soya data for Conventional PCA from the first 4 principal components will be 57.1 %,78.9 %,88.9% and 95.03% respectively (see Table 1). The first component accounts for 57.1% of the total variance, the second component accounts for 21.8 % of the total variance, the third component and the fourth component accounts for 10% and 6.13% of the total variance respectively (see Figure 3). To achieve a cumulative explained variance of 95.03%, a minimum of 4 principal components are found necessary when the Corn Soya data is subjected to conventional PCA.

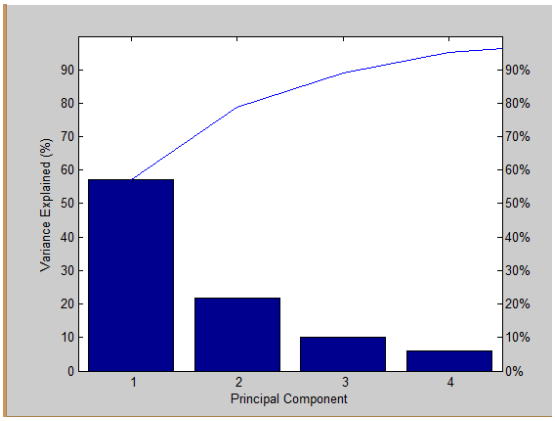

Figure 3. Principal components and % of variance explained in Conventional PCA

When the corn soya dataset is subjected to the proposed incremental feature transformation approach, seemingly efficient trends have been observed. To start with, the first batch of 6 features ie f1…f6 are transformed and reduced to achieve a minimum of 90% explained variance. The approach generates first two principal components. The first and second components account respectively for 87.5% and 95.3% of the total explained variance in the first batch of the temporal data (see Table 1). The first two principal components are merged with the next batch of features namely f7..f12 and the 8 attributes ie 6 original attributes from the second temporal batch and the 2 principal components from the first batch are subjected to incremental PCA. The reduced feature space now comprises of 2 new principal components that accounts for at least 90% of cumulative variance as shown below. The third batch when merged as per the algorithm will yield a cumulative variance of 94.9% from the first 3 principal components. At the last temporal point the reduced feature space comprises of 3 principal components that account for a total of 92.9% of the total variance of all the 4 batches of the corn soya feature space (see Figure 4). The process is carried on until there are no new features to be assimilated in the transformed feature space.

Table 1. Percentage of cumulative explained variance in the chosen principal components

Conventional PCA

		pc1	pc2	pc3	pc4	. . .	pc24
f1, f2, ..., f24	→	57.10%	78.90%	88.90%	95.03%		100%

Incremental PCA

		pc1	pc2			pc6
f1,...,f6	→	87.5	95.30%	. . .		100%
pc1,pc2 + f7,...,f12	→	pc1	pc2	. . .		pc8
		57.40%	93.20%			100%
pc1,pc2 + f13,.....,f18	→	pc1	pc2	pc3	. . .	pc8
		61.70%	84.60%	94.90%		100%
pc1,pc2,pc3+ f19,.....,f24	→	pc1	pc2	pc3	. . .	pc9
		60.40%	83.10%	92.90%		100%

When corn soya dataset is subjected to conventional PCA, it requires 4 principal components to account atleast 90% of the total variance. When the same dataset is subjected to incremental PCA by transforming, reducing and merging the features as and when they arrive temporally, it requires only 3 principal components that account for a minimum of 90% of the total variance (see Table 1). The incremental PCA uses approximately 25% less components to assimilate the same amount of information that the conventional PCA would require to use.

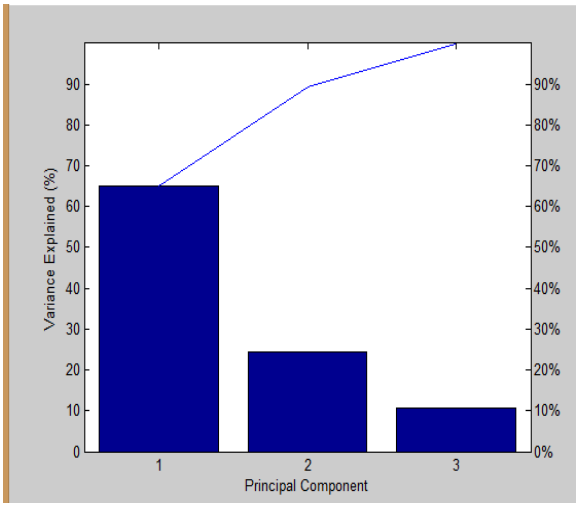

Figure 4. Principal components and % of variance explained in Incremental PCA

In the proposed approach, the percentage of cumulative variance and the intrinsic dimensionality are both found better. An Intrinsic dimensionality in the transformed feature space is the minimum number of parameters needed to account for the observed properties of data [35]. It is also defined as the minimum number of parameters required to clearly determine the belongingness of the instance to its cluster. For the purpose of illustration of the Cluster belongingness, K-means clustering method is employed owing to its simplicity and ease in analyzing the instances. K-means is one of the simplest unsupervised learning algorithms that has solved many well-known clustering problems [28].

While performing the incremental feature transformation approach, it was found that at each temporal point, the samples rearrange themselves in the transformed feature space in order to progressively converge towards the right cluster. For instance in Corn Soya dataset, the first increment (see Figure 6), it is observed that the samples are roughly distributed based on the similarities of the instances and cohesion to the centroid at that point. In the second increment (see Figure 6) the samples position themselves more closely to the right cluster pushing the outliers towards the fringes as they move closer to the centroid of the cluster they belong to. When this accumulated knowledge is merged with the third batch of features (see Figure 6) and transformed, the scatter in the cluster is reduced further indicating that the cluster becomes tighter as it progressively converges over time.At the last incremental transformation step, the clustering accuracy and sensitivity reaches its peak owing to the fact that the cumulative variance built incrementally from the first temporal phase onwards captures the most descriptive features needed to discriminate the clusters accurately. At this step, 100% of the samples distinguish themselves into the right cluster by moving away from the fringes and tightly into the cluster they belong to (see Figure 6)

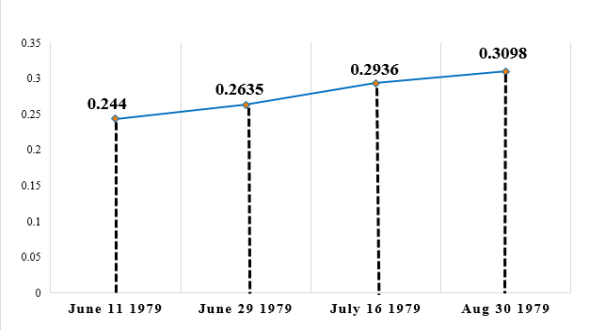

Figure 5. Average Within Cluster distance at each incremental step in Corn Soya dataset.

It can be seen that the cluster tightness incrementally keeps increasing at every temporal phase (see Figure 5).At each incremental step of transformation, the number of principal components required to explain the amount of variance in the transformed feature space keeps decreasing. This emphasizes

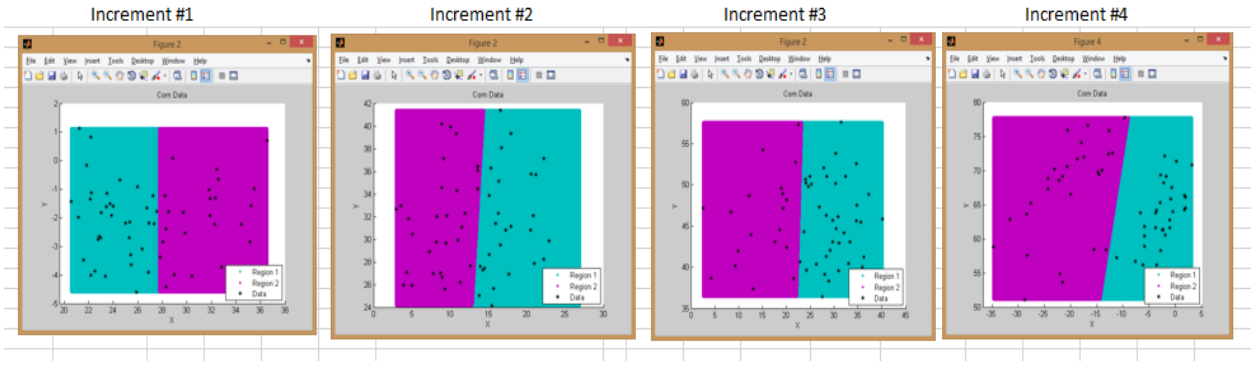

Figure 6. Incremental Convergence of instances in Corn Soya data

on the fact that the knowledge that is being built on top of the previously accumulated knowledge is condensed at the end of each phase to provide more substantial variance and contain better decision making factors at each phase of the temporal data.

Another interesting temporal dataset is the weather data from 37 cities of the world [26]. Although it is a continuous temporal dataset, for the sake of explanation, the dataset is assumed to arrive at the end every quarter when the season usually tends to change. While the dataset is subjected to incremental dimensionality reduction, it is found that the data is optimally grouped into two clusters in the first time slice (Jan., Feb., and March) (see Figure 7). The belongingness of the samples in the second time slice (April, May and June) remains unaltered. But when the time steps into the third quarter (July,Aug.,Sept.,), the cities Athens and Tehran exhibit an interesting nature of intra cluster movement in the clusters (see Figure 8). The two cities, Athens and Tehran which were clustered in the first group so far and that appeared to be in the fringes in the first 2 quarters, now slips into the second cluster. In the fourth quarter during winter (see Figure 8), Athens and Tehran move back to the first cluster where it started to exist. In retrospect, to analyze and understand the movement of instances through the incremental phases, the entire data set was analyzed as follows. A yearly mean temperature of 11.6 ºC (4.5 ºC - 17.69 ºC) are considered to be in the 'cold Cities' cluster. A yearly mean temperature of 23.7 ºC (18.05 ºC - 26.7 ºC) are considered to be in the 'hot Cities' cluster. The average mean temp of Athens and Tehran are 17.54 ºC and 17.69 ºC respectively i.e. close to the fringes of both the clusters. In the third quarter since the temperatures swing little towards a higher mercury scale, the cities Athens and Tehran move into the 'hot cities' cluster. Again in the fourth quarter when the temperature drops, both the cities are grouped in the 'cold cities' cluster.

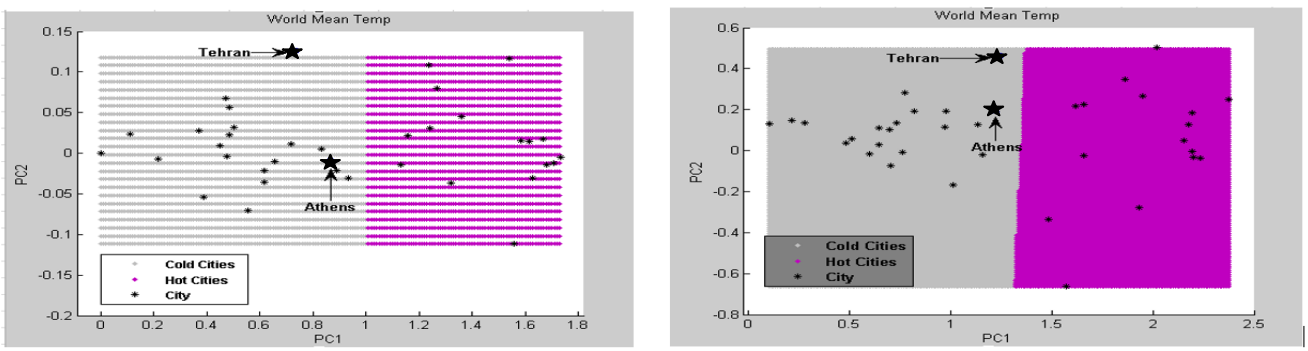

Figure 7. Intra Cluster movement of the cities Athens and Tehran in the first 2 time slices

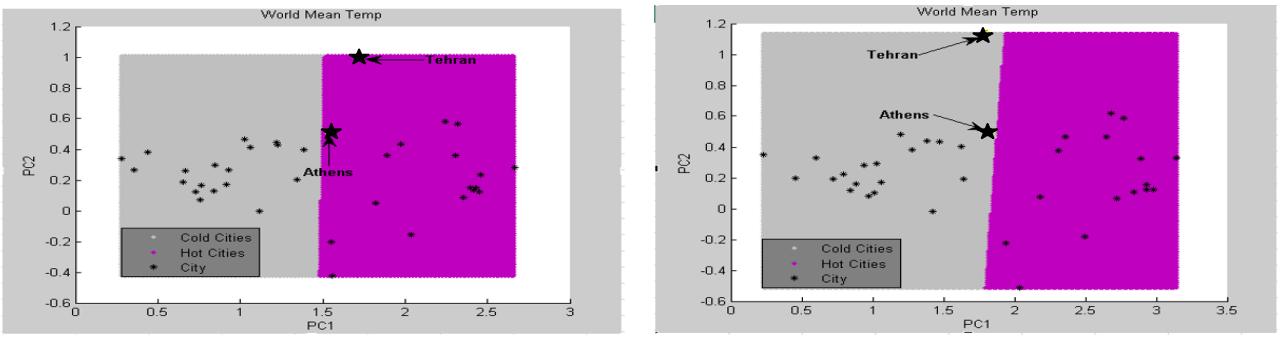

Figure 8. Inter Cluster Movement of the cities Athens and Tehran in the last 2 time slices

The intercluster movement of the two cities provides an insight on the fringe instances. The outliers are the ones that are susceptible to be inconsistent in their cluster belongingness. The majority of the instances are very consistent in the cluster belongingness throughout the phases of the incremental transformation. The significance here is that the knowledge that is accumulated at every phase is influenced by the combination of the past knowledge and the present knowledge. At every temporal point, the knowledge available to make the decision at that particular point is based on the data that is accumulated so far. Incremental transformation aid in tracing the paths of the samples over time which would provide more information on the local data that arrives temporally. This in turn would provide a stronger basis for decision making, which would be very helpful in analysing outliers and finding out the reasons for their movement in and out of the Clusters.

5. ALGORITHM

Consider a temporal dataset with totally m = n*k features arriving in n batches with k features in each batch

Input: I = {f1,f2,f3…..fk} ; The first batch of temporal dataset.

Output: O = {pc1,pc2,pc3….,pck} where q<<k ; A set of transformed features that represent the intrinsic dimensionality of the cumulatively transformed feature space.

Begin

1. Initialize the intrinsic vector to an empty vector.
2. While the number of input batches >=1 Do

a. Add the set of original features to the intrinsic vector ‘V’

b. Perform PCA on the features in the vector V

c. Calculate the cumulative explained variance of the featureset

d. Retain the principal components that account for 90% of the total variance in the intrinsic vector and discard the rest of the principal components

End Do

//This loop is executed until there are no new batches to be transformed and reduced in the temporal dataset

End

The final set of principal components present in the vector V represent the intrinsic Dimensionality of the transformed feature space of the entire data set.

The best case scenario is when the first principal component only is good enough for each of the ‘n’ batches. So the best case time complexity of the algorithm would be n * O (k)

tΩ → O (nk) where n = number of batches in the temporal dataset and t is the time taken for processing an input of n batches; k is the number of features in each batch

The worst case scenario is when all the principal components in each batch are required to account for atleast 90% of the explained variance. So the worst case time complexity of the algorithm would be n * O (2k)

t (O) →O (2nk) where where n = number of batches in the temporal dataset and t is the time taken for processing an input of n batches; k is the number of features in each batch

6. PERFORMANCE EVALUATION

The performance evaluation of the incremental feature transformation approach has been carried out in two ways. Fat first, the comparison of the algorithm is done phasewise to analyze the performance of the algorithm incrementally in each phase. Subsequently, the algorithm is compared to the other three approaches discussed in Section 4. All the 4 temporal datasets mentioned previously during illustration have been considered in both the cases: Corn Soya, Weather, Dow Jones Q1 and Dow Jones Q2 datasets.

6.1 Phase wise evaluation and analysis

The incremental transformation approach at each temporal phase is explained in detail as follows. The Corn Soya dataset has 4 temporal phases in total. At the end of each temporal phase, a set of 6 features arrive and at every temporal point, the 6 features are subjected to four different methods separately to analyze the information gain assimilated in the clusters. The four methods involved are: Original data Clustering; Incremental Feature Subsetting; Conventional PCA and Incremental PCA. Original data clustering does not involve preprocessing and the samples are naturally grouped into clusters as they arrive. Incremental Feature Subsetting is carried out by selecting a subset of features that encapsulate the maximum variance in the original set of features and clustering is done over that subset [31]. Conventional PCA is the traditional transformation technique carried out to transform the dimensions when all the features are available at hand. The proposed Incremental feature transformation technique proposed in this paper transforms and reduces the transformed feature space by incrementally merging the previously reduced feature space

with the incoming temporal features. When Corn Soya dataset is subjected to the four methods explained earlier, the feature space evolves itself differently in each method and provides a good amount of statistics to understand and

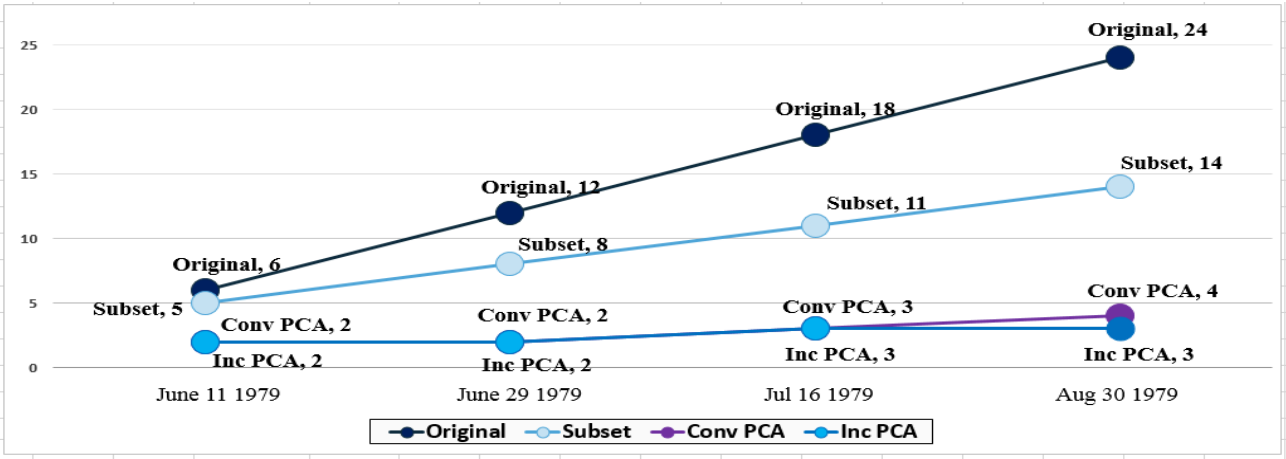

Figure 9. Comparison of the reduced feature space at each incremental step in the Corn Soya dataset

analyze the movement of samples and grouping of features through the temporal phases of the dataset. When the first set of temporal features arrive on June 11 2011, the set of 6 features are subjected to all the 4 methods separately. It is observed that the clustering achieves an accuracy of 54.09% on Original data, 55.7% on Incremental Subsetting and 57.4% each on Conventional PCA and Incremental PCA. The incremental transformation algorithm proposed in this paper is repeated when the new set of features arrive on June 29 2011, Jul 16 2011 and Aug 30th 1979 consecutively. At the end of the fourth temporal phase when the Aug 30th features are subjected to the respective methods for the last time , the final transformation achieves a Clustering Accuracy of 64.7% for original data , 95.1% for Incremental Subsetting and 100 % each for both Conventional and incremental transformation. Achieving 100% Clustering in both kind of transformation techniques gives the transformed feature space an upper hand over the non transformed feature space. One important observation here is that although both Conventional PCA and Incremental PCA approaches achieve 100% clustering accuracy at the end, the incremental approach requires 25% less factors to achieve the same accuracy when compared to the Conventional approach (see Figure 9). In addition, the cumulative explained variance for say 'n' number of principal factors in incremental approach is much more than that of the conventional approach for the same 'n' number of principal factors (see Figure 10).

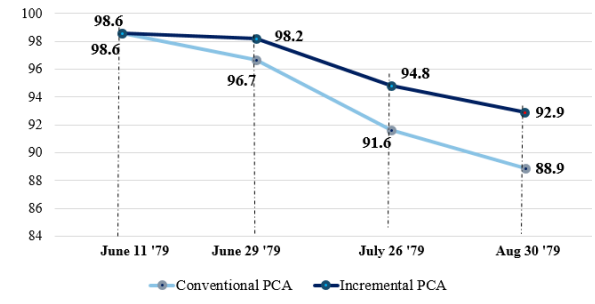

Figure 10. Comparison of explained variance (%) between conventional and incremental transformation approaches considering the first 3 principal factors in Corn Soya datsets.

Another interesteing trend observed is the number of principal components required over time while subjecting the Dow Jones dataset to the proposed approach. The stockmarket data from the second quarter of 2001 was assumed to arrive in 12 seperate time slices incrementally and was subjected to incremental feature transformation likewise. To start with, the minimum number of components required to account for a minimum of 90% of the explained variance of the first batch of temporal features is the first four principal components. As time passes by, the number of principal components required to assimilate atleast 90 % cumulative variance keeps on decreasing incrementally (see Figure 11) and finally ends up with 2 principal components.

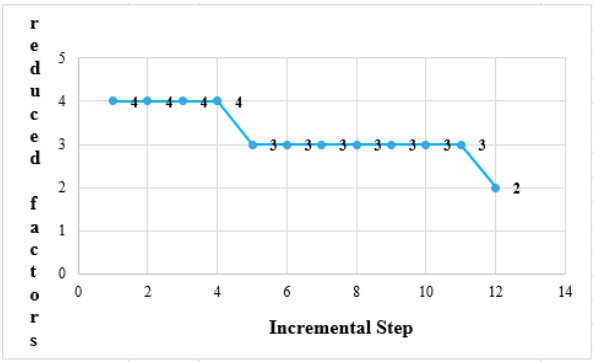

Figure 11. Incremental steps in Dow Jones data and the corresponding reduced principal factors required at each step

This shows that knowledge with the most descriptive features is being distilled upon the previous knowledge to provide a concentrated essence of the features. The least descriptive features are gradually reduced/removed in the process.

6.2 Evaluation of Incremental feature transformation Vs other approaches

There are a myriad of metrics that are available to evaluate the quality of a cluster. All the metrics aid in measuring either the intra cluster distance or inter cluster distance. Intra cluster distance signifies the tightness of samples in the cluster and inter cluster distance signifies the separation of clusters from each other. The various metrics used for the

performance evaluation of the approaches are discussed in the following section with examples.

6.2.1 Silhouette Index

The silhouette value for each point is a measure of how similar that point is to points in its own cluster, when compared to points in other clusters. The silhouette value for the ith point, S_i, is defined as

$$S_i = (b_i - a_i) / \max(a_i, bi)$$

where a_i is the average distance from the ith point to the other points in the same cluster as i, and b_i is the minimum average distance from the ith point to points in a different cluster, minimized over clusters [36].

The average S_i over all data of a cluster is a measure of how tightly grouped all the data in the cluster are. Thus the average S_i over all data of the entire dataset is a measure of how appropriately the data has been clustered [36]. The silhouette criteria lies between 1 and -1; If the value is near to +1, it means that the cluster is tightly arranged and if the value is towards -1 , it means that the samples are loosely arranged and is not well clustered [36]. The distance metric used in this case is Euclidean distance. The Silhouette index can be used to measure the quality of the clusters.

The Silhouette index for the 4 datasets considered is as below. The Silhouette criteria is highest for Incremental Transformation approach among all the datasets considered. This supports the fact that the clusters obtained using the Incremental approach yields the best quality of clusters.

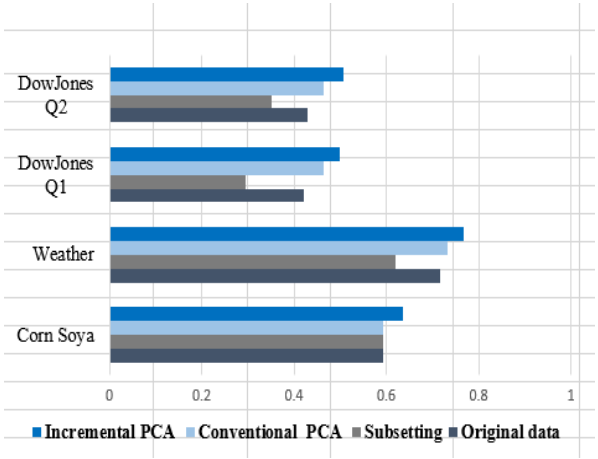

Figure 12. Comparison of Sihoutte Criteria

6.2.2 Scatter in the Cluster

The standard deviation among the samples in a cluster is the least in the proposed approach. It is also referred to the scatter in the clusters and quantitatively measures how adjacent samples are spatially related (see Figure 13). Scatter is the least when incremental transformation approach is employed indicating the effectiveness of clustering and the tightness of samples in the clusters.

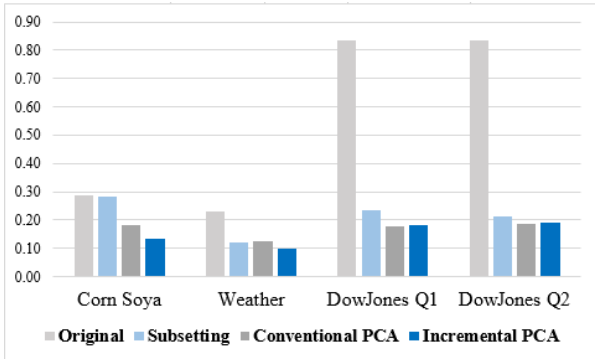

Figure 13.Comparison of scatter among the various approaches

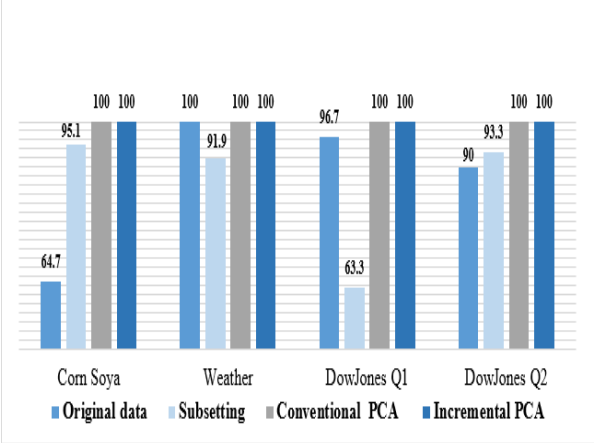

Figure 14. Comparison of Clustering Accuracy

6.2.3 Clustering Accuracy

When different experiments are conducted on the 4 temporal datasets, it was observed that although maximum clustering accuracy is obtained in both conventional and incremental transformation approaches, the number of factors required to achieve the accuracy in the incremental approach is much better than the conventional approach. Also the amount of cumulative variance captured in the transformed feature space is higher in incremental transformation when compared to the conventional feature space. While performing PCA, instead of trying to interpret each factor, McCabe has suggested finding the principal variables [34].

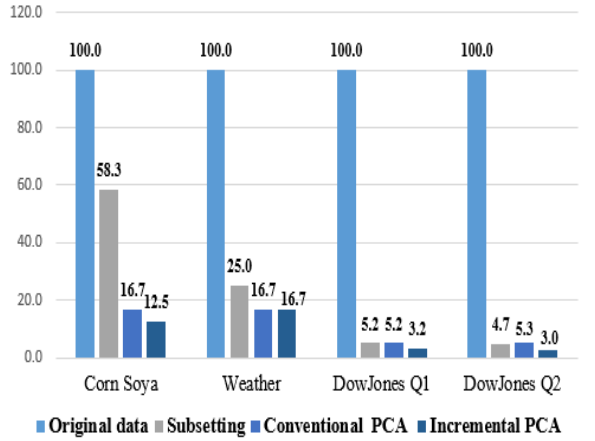

Figure 15. Comparison of the percentage of factors required to achieve the Clustering Accuracy

It has been observed before in Corn Soya dataset that if 3 principal components of the transformed feature space are to be considered at each temporal phase, the incremental transformation assimilates higher cumulative variance at each temporal point (see Figure 10) than the conventional transformation. In Dow Jones data for Q1, it requires 37.5% lesser principal factors and for Q2 it requires 44% lesser principal factors than what a conventional transformation would require to achieve the same clustering accuracy (see Table 2). This shows that the most descriptive features are retained effectively in the incremental transformation approach.

6.2.4 Compactness

The other metrics that quantitatively measure the intra class similarity are Compactness, Associativity and Disassociativity. A cluster can be classified as a good cluster if it has maximum intra cluster tightness and maximum inter cluster separation. Compactness represents the density of the cluster. Compactness of a cluster is calculated as below [1]:

Compactness = 1/D

$D = (1/N) \sum d_k$ = Average distance in a cluster

Variance = $\sigma^2 = (1/N) \sum (d_k - D)^2$

Where N = number of samples in the class considered

d_k = distance of kth sample in the class from its class mean.

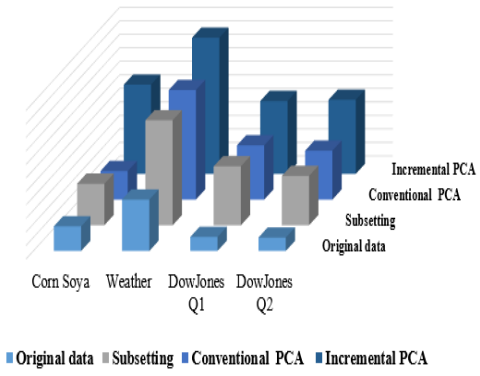

Figure 16. Comparison of Compactness

Compactness of a cluster can be analyzed in terms of Associativity. Associativity measures the tightness of samples in the cluster and Disassociativity measures the distance between one cluster to another. They can be calculated as shown below [26].

Associativity = Compactness / Scatter

Disassociativity = $D_{ij} / ((D_i + D_j)/2)$

Where D_{ij} = Distance between the means of the ith and the jth clusters

Di = Average distance of all samples belonging to ith class from its class mean and

Dj= Avearge distance of all the samples belonging to jth class from its class mean

Compactness value is the highest in the proposed approach yet again supporting the method of convergence of samples into the right cluster it belongs to (see Figure 16).

6.2.5. Incremental Dimensionality Index

Both associativity and disassociativity among similar samples should be high for well defined clusters .While traversing through different phases of time, few samples might change their course temporarily thereby identifying themselves as outliers. This would cause the disassociativity to decrease but the associativity metric needs to be weighed in parallel to dwell on a better conclusion about the clustering accuracy during such times. Hence the ratio of associativity to disassociativity is considered and is termed as *Intrinsic Dimensionality index (IDI)*. This index is the combination of intra cluster distance and inter cluster distance measures and the index value should be high for a cluster to be classified as a good cluster. It is observed that among all the 4 datasets considered, IDI for the incremental feature transformation approach is highest; indicating that the decision making design is indeed better in the proposed approach when compared to others (see Figure 17).

IDI = Associativity / Disassociativity

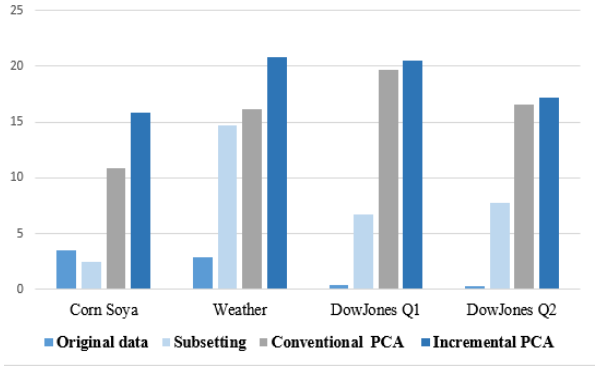

Figure 17. Comparison of Intrinsic dimensionality reduction index

In Summary, it is observed from the various experiments conducted (see Table 2) that the metrics to measure the intra cluster distance like scatter, compactness, associativity and metrics to evaluate inter cluster distance like Silhouette criteria and incremental dimensionality index exhibit the best scores for the proposed incremental transformation approach. Not only is the proposed approach the most effective one for making decisions when all the features arrive at the end but is also a feasible design that performs best while one has to make decisions in the interim. The maximum variance of the original dataset and the most descriptive features that represent the entire feature set are transformed and the accumulated knowledge is kept ready dynamically for making decisions at every temporal phase. The proposed algorithm is based on the context that the arrival of features in the temporal feature space happens to occur with a visibly wider gap or a period such that the pattern analysis and decision making gets warranted. On the contrary, if the features were to arrive in quick succession, then the intermediate pattern analysis at every temporal stage would neither be feasible nor expected and would predominantly make the process offline. However this model still would be useful to accomplish effective feature reduction although it appears to have happened after the arrival of all features.

7. CONCLUSION AND FUTURE WORK

After a thorough evaluation (see Table 2) of the four different approaches has been conducted using temporal data, the incremental formulation of the best clustering accuracy and generation of well-defined clusters were obtained while performing the proposed approach.

Table 2. Evaluation metrics for the four datasets

Data Set	*Approach*	*Factors considered*	*SD*	*Compactness*	*Associativity*	*Disassociativity*	*Intrinsic Dim Index*	*Sensitivity*	*Specificity*	*Accuracy*
CORN	Original data	24	0.29	0.91	3.18	0.89	3.56	68.75	57.89	64.71
	Incremental Subset	14	0.28	1.54	5.43	2.21	2.45	100.00	89.66	95.08
	Conventional PCA	4	0.18	1.05	5.77	0.53	10.86	100.00	100.00	100.00
	Incremental PCA	3	0.14	3.31	25.01	1.58	15.85	100.00	100.00	100.00
WEATHER	Original data	12	0.23	1.90	8.40	2.88	2.92	100.00	100.00	100.00
	Incremental Subset	3	0.12	3.89	34.55	2.35	14.70	95.24	87.50	91.89
	Conventional PCA	2	0.13	4.06	33.78	2.10	16.11	100.00	100.00	100.00
	Incremental PCA	2	0.10	5.04	50.70	2.44	20.79	100.00	100.00	100.00
DOWJONES Q1	Original data	154	0.83	0.52	0.62	1.64	0.38	91.67	100.00	96.67
	Incremental Subset	8	0.24	2.18	9.25	1.37	6.76	100.00	38.89	63.33
	Conventional PCA	8	0.18	2.01	13.16	0.67	19.69	100.00	100.00	100.00
	Incremental PCA	5	0.18	2.70	17.62	0.86	20.46	100.00	100.00	100.00
DOWJONES Q2	Original data	169	0.84	0.50	0.59	1.80	0.33	75.00	100.00	90.00
	Incremental Subset	8	0.21	1.83	8.68	1.12	7.74	100.00	88.89	93.33
	Conventional PCA	9	0.19	1.81	10.49	0.63	16.55	100.00	100.00	100.00
	Incremental PCA	5	0.19	2.74	15.51	0.90	17.19	100.00	100.00	100.00

Not only does it exhibit the best possible performance at the end but it also exhibits best performance at every interval of time in the lifecycle of temporal data. This model is proven to provide the best environment for making decisions at any instant because at every incremental temporal phase, the design assimilates the most descriptive features , captures the best proportion of all the original features available for reduction .It also eliminates the need for looking back at the previous data and reduces the searchspace for accumulating the knowledge. The proposed model successfully preserves as much discriminatory information as possible at every incremental step and speeds up the convergence of clusters as time goes by. At any instant of time, the model provides an efficient design for making the best decisions based on the knowledge accumulated over time and the information available at hand. While dealing with temporal data, since the arrival of features is sequence compulsive, the exploration of the proposed approach was also sequence compulsive. If the features arrive from a distributed or a multisensory environment, the exploration of the incremental feature transformation approach need not be sequence compulsive. Further research can be performed to explore if and how the order and flow of features can affect the incremental accumulation of knowledge. Additional work can be carried out to extract the contribution of the original features in the transformed feature space by means of backtracking the explained variance in the principal components.

8. REFERENCES

[1] P. Nagabhushan, An efficient method for classifying remotely sensed data (incorporating dimensionality reduction), Ph.D thesis, Universityof Mysore, 1988

[2] Syed Zakir Ali., P Nagabhushan., Pradeep Kumar R, Incremental datamining using Clustering Intelligent Methods of Fusing the Knowledge During Incremental Learning via Clustering in A Distributed Environment , PhD Thesis, 2010

[3] Syed Zakir Ali., P Nagabhushan., Pradeep Kumar R, Regression based Incremental Learning through Cluster Analysis of Temporal data, International Conference on Data Mining (DMIN) 2009

[4] Martin H.C. Law, Anil K. Jain, Incremental Nonlinear Dimensionality Reduction by Manifold Learning, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 3,2006

[5] Chowdhury.F et,al, Single-pass incremental and interactive mining for weighted frequent patterns, 2012.

[6] S. Kotsiantis., K. Patriarcheas., and M. Xenos. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education, 2010.

[7] T.Gharib et,al, An efficient algorithm for incremental mining of temporal association rules, 2010.

[8] D.Dudek, RMAIN:Association rules maintenance without reruns through data, 2009

[9] Law, Martin HC, Nan Zhang, and Anil K. Jain. "Nonlinear Manifold Learning for Data Stream." SDM. 2004.

[10] Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28.12 (2006): 2037-2041.

[11] Jieping Ye., IDR/QR: an incremental dimension reduction algorithm via QR decomposition, Knowledge and Data Engineering, IEEE Transactions , Vol17, Issue09, 2005

[12] Balsubramani, S Dasgupta, The fast convergence of incremental pca Advances in Neural Information Processing Systems, 2013

[13] J. Han, M. Kamber, "Data Mining: Concepts and Techniques," Third Edition, Elsevier Inc., Raj Kamal Electric Press, 2011.

[14] Anil K. Jain's talk: Clustering Big Data, University of Notre Dame, Nov. 29, 2012

[15] Feng Wang , Jiye Liang , Yuhua Qian, Attribute reduction: A dimension incremental strategy, Knowledge-Based Systems, 39, p.95-108, February, 2013

[16] GF Lu, J Zou, Y Wang - Pattern Recognition, , Incremental complete LDA for face recognition, Elsevier, 2012

[17] Balsubramani, S Dasgupta , Freund The fast convergence of incremental PCA, Advances in Neural Information Processing Systems 26, 2013.

[18] J Feng, H Xu, S Mannor, S Yan, Online-PCA for contaminated data , Advances in Neural Information Processing Systems 26, 2013

[19] I Mitliagkas, C Caramanis, P Jain - Memory Limited, Streaming PCA, Advances in Neural Information 26, 2013

[20] http://in.mathworks.com/help/stats/clustering.evaluation.silhouetteevaluation-class.html

[21] http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html

[22] http://www-users.cs.umn.edu/~han /dmclass/ cluster_ survey_ 10_02_00.pdf

[23] Principal Component. Analysis, Second Edition. I.T. Jolliffe. Springer, NewYork, 2002

[24] Raducanu, Bogdan, and Fadi Dornaika. "A supervised non-linear dimensionality reduction approach for manifold learning." *Pattern Recognition*45.6 (2012): 2432-2444

[25] Geraldine, J. Mercy, E. Kirubakaran, and S. Sathiya Devi. "WEIGHTED TEMPORAL PATTERN MINING WITH DIMENSIONALITY REDUCTION USING MODIFIED AFCM." *Int J AdvEngg Tech/Vol. VII/Issue I/Jan.-March*565 (2016): 571

[26] Rangarajan, Lalitha, and P. Nagabhushan. "Dimensionality reduction of multidimensional temporal data through regression." Pattern recognition letters 25.8 (2004): 899-910

[27] https://onlinecourses.science.psu.edu/stat505/node/49

[28] http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html

[29] https://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf

[30] http://www.doc.ic.ac.uk/~dfg/ProbabilisticInference/old_IDAPILecture14.pdf

[31] P.Nagabhushan, Preethi Mahadev.,"Incremental Feature Subsetting useful for Big Feature Space Problem", International Journal of Computer Applications,Volume 97, Issue12,2014

[32] https://en.wikipedia.org/wiki/Explained_variation

[33] Huan Liu et al.,Incremental Feature selection, Journal Applied Intelligence, Vol 9 Issue 3 ,pp 217-230, 1998

[34] http://www.ncss.com/wp-content/themes/ncss /pdf/ Procedures/NCSS/Principal_Components_Analysis.pdf

[35] Introduction to statistical pattern recognition, second edition, Keinosuke Fukunaga, Academic Press Professional, Inc. San Diego, CA, 1990

[36] https://en.wikipedia.org/wiki/Silhouette_(clustering)