INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.,

Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],

Index Copernicus Publishers Panel, Poland with IC Value of 5.09 & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than **4600 Cities** in **180 countries/territories** are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography: Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** ***anytime*** in ***M.S. Word format*** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website (***FOR ONLINE SUBMISSION, CLICK HERE***).

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____________

THE EDITOR

IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF ______________________.**

(e.g. Finance/Mkt./HRM/General Mgt./Engineering/Economics/Computer/IT/ Education/Psychology/Law/Math/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '____' for possible publication in one of your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the co-authors of this manuscript have seen the submitted version of the manuscript and have agreed to their inclusion of names as co-authors.

Also, if my/our manuscript is accepted, I agree to comply with the formalities as given on the website of the journal. The Journal has discretion to publish our contribution in any of its journals.

NAME OF CORRESPONDING AUTHOR	:
Designation	:
Institution/College/University with full address & Pin Code	:
Residential address with Pin Code	:
Mobile Number (s) with country ISD code	:
Is WhatsApp or Viber active on your above noted Mobile Number (Yes/No)	:
Landline Number (s) with country ISD code	:
E-mail Address	:
Alternate E-mail Address	:
Nationality	:

NOTES:

a) The whole manuscript has to be in ***ONE MS WORD FILE*** only, which will start from the covering letter, inside the manuscript. ***pdf. version** is liable to be rejected without any consideration*.

b) The sender is required to mention the following in the **SUBJECT COLUMN of the mail**:

New Manuscript for Review in the area of (e.g. Finance/Marketing/HRM/General Mgt./Engineering/Economics/Computer/IT/ Education/Psychology/Law/Math/other, please specify)

c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any **specific message** w.r.t. to the manuscript.

d) The total size of the file containing the manuscript is expected to be below **1000 KB**.

e) **Abstract alone will not be considered for review** and the author is required to submit the **complete manuscript** in the first instance.

f) ***The journal gives acknowledgement w.r.t. the receipt of every email within twenty four hours*** and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending a separate mail to the journal.

g) The author (s) name or details should not appear anywhere on the body of the manuscript, except the covering letter and the cover page of the manuscript, in the manner as mentioned in the guidelines.

2. **MANUSCRIPT TITLE:** The title of the paper should be **bold typed**, **centered** and **fully capitalised**.
3. **AUTHOR NAME (S) & AFFILIATIONS:** Author (s) **name**, **designation**, **affiliation** (s), **address**, **mobile/landline number** (s), and **email/alternate email address** should be given underneath the title.
4. **ACKNOWLEDGMENTS:** Acknowledgements can be given to reviewers, guides, funding institutions, etc., if any.
5. **ABSTRACT:** Abstract should be in **fully italicized text**, ranging between **150** to **300 words**. The abstract must be informative and explain the background, aims, methods, results & conclusion in a **SINGLE PARA**. ***Abbreviations must be mentioned in full***.
6. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of **five**. These should be arranged in alphabetic order separated by commas and full stop at the end. All words of the keywords, including the first one should be in small letters, except special words e.g. name of the Countries, abbreviations.
7. **JEL CODE:** Provide the appropriate Journal of Economic Literature Classification System code (s). JEL codes are available at www.aeaweb.org/econlit/jelCodes.php, however, mentioning JEL Code is not mandatory.
8. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. *It should be free from any errors i.e.* ***grammatical, spelling*** *or* ***punctuation.*** *It must be thoroughly edited at your end*.
9. **HEADINGS:** All the headings must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.
10. **SUB-HEADINGS:** All the sub-headings must be bold-faced, aligned left and fully capitalised.
11. **MAIN TEXT:**

THE MAIN TEXT SHOULD FOLLOW THE FOLLOWING SEQUENCE:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESIS (ES)

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

LIMITATIONS

SCOPE FOR FURTHER RESEARCH

REFERENCES

APPENDIX/ANNEXURE

The manuscript should preferably range from ***2000 to 5000 WORDS***.

12. **FIGURES & TABLES:** These should be simple, crystal **CLEAR**, **centered**, **separately numbered** & self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. *It should be ensured that the tables/figures are referred to from the main text*.

13. **EQUATIONS/FORMULAE:** These should be consecutively numbered in parenthesis, horizontally centered with equation/formulae number placed at the right. The equation editor provided with standard versions of Microsoft Word should be utilised. If any other equation editor is utilised, author must confirm that these equations may be viewed and edited in versions of Microsoft Office that does not have the editor.

14. **ACRONYMS:** These should not be used in the abstract. The use of acronyms is elsewhere is acceptable. Acronyms should be defined on its first use in each section: Reserve Bank of India (RBI). Acronyms should be redefined on first use in subsequent sections.

15. **REFERENCES:** The list of all references should be alphabetically arranged. ***The author (s) should mention only the actually utilised references in the preparation of manuscript*** and they are supposed to follow Harvard Style of Referencing. **Also check to make sure that everything that you are including in the reference section is duly cited in the paper**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parenthesis.
- ***Headers, footers, endnotes** and **footnotes** should **not be used** in the document.* However, **you can mention short notes to elucidate some specific point**, which may be placed in number orders after the references.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–23

UNPUBLISHED DISSERTATIONS

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

ENHANCEMENT OF TEMPORAL DATA CLUSTERING WITH CLIPPED GAUSSIAN DISTRIBUTION

SARALA.R
ASST. PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE
KG COLLEGE OF ARTS & SCIENCE
COIMBATORE

ABSTRACT

Temporal data clustering provides underpinning techniques for discovering the intrinsic structure and condensing information over temporal data. There have been presented a lot of temporal data clustering framework ensemble of multiple partitions produced by initial clustering analysis on different temporal data representations. Here, this paper introducing an approach of clipping the time series that reduces memory requirements and significantly speeds up clustering without decreasing clustering accuracy. This means that if the series are long enough clustering with clipped data is significantly high in accuracy than clustering with unclipped data. Clipping makes clustering more robust to outliers. Show that the clusters formed are significantly better with clipped data when there is at least a small probability of the data containing outliers. Using clipped data have rewards are that algorithms developed for discrete or categorical data can be employed and that clustering on clipped data can serve as a diagnostic method for outlier and model misspecification detection. With the help of Gaussian distribution function will be getting a smooth series of clipped data which are high in accuracy. As a result, the proposed approach provides an effective enabling technique for the joint use of different representations, which cuts the loss of information in a single representation and exploits various information sources underlying temporal data. In addition, the approach tends to capture the intrinsic structure of a data set that is the number of clusters.

KEYWORDS

temporal data, weighted clustering ensemble, clipping, gaussian distribution.

I. INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases, is a dominant fresh technology with grand potential to help companies focus on the most important information in their data warehouses. Data mining tools predict upcoming trends and behaviors. It allows businesses to make proactive and knowledge-driven decisions. Data mining tools can response business questions which usually were too time strong to resolve [1].

Temporal data clustering is to provide an effective way to discover the intrinsic structure and condense information over temporal data by exploring dynamic regularities underlying data in an unsupervised learning. In the proposed system, the clipping technique for the clustering in the time series is used [2]. This paper demonstrates that the simple process of *clipping* time series reduces memory requirements and can speed up fundamental operations on time series. This paper proposes temporal data clustering model.

Clipped Gaussian distribution approach consists of initial clustering analysis on different representations to produce multiple partitions and clustering ensemble construction to produce a final partition by combining those partitions achieved in initial clustering analysis. Initial clustering analysis can be done by any existing clustering algorithms that propose a novel weighted clustering ensemble algorithm. Then, implement the clipped Gaussian distribution concept on the clustering time series. This would reduce the cost of time. Finally the proposed system is get illustrated and compared with the conventional approaches. From the simulation results on a variety of temporal data clustering tasks and concludes that this approach is outperforming with high accuracy, lesser cost of time due to the proposed technique in this system.

II. TEMPORAL DATA CLUSTERING

Underpinning techniques provided by temporal data clustering for discovering the intrinsic structure and condensing information over temporal data. Temporal data stored in a temporal database [2] is different from the data stored in non-temporal database in that a time period close to the data expresses when it was suitable or stored in the database. Temporal data are ubiquitous in the real world and there are many application areas ranging from multimedia.

An effective way to discover the intrinsic structure and condense information over temporal data by exploring dynamic regularities underlying temporal data in an unsupervised learning way by the performance of Temporal clustering analysis. Its ultimate objective is to partition an unlabeled temporal data set into clusters so that sequences grouped in the same cluster is coherent. There are two core problems in clustering analysis, first problem is model selection and the second is grouping.

Clustering analysis is an extremely difficult unsupervised learning task [4]. Due to the high dimensionality and complex temporal correlation the temporal data clustering poses a real challenge in temporal data mining. Temporal data clustering algorithms as three categories: (i)temporal-proximity-based, (ii) model-based and (iii) representation-based clustering algorithms.

Temporal-proximity-based and model-based clustering algorithms directly work on temporal data [7]. Representation-based algorithm converts temporal data clustering into static data clustering. The basic idea behind clustering ensemble is combining multiple partitions on the same data set to produce a consensus partition expected to be superior to that of given input partitions. In this paper, approach to temporal data clustering with different representations to overcome the fundamental weakness of the representation-based temporal data clustering analysis.

The contributions of this paper is to develop a practical temporal data clustering model by different representations via clustering ensemble learning to overcome the fundamental weakness the representation-based temporal data clustering analysis. Proposing a novel weighted clustering ensemble algorithm, which not simply provides an enabling technique sustain this model but also can be used to merge any input partition. Formal analysis has also been done. And the third finally is the demonstrate the effectiveness and the efficiency of the model for a variety of temporal data clustering tasks as well as its easy-to-use nature as all internal parameters are fixed in the simulations. The motivation to propose the temporal data clustering model is working on different representations via clustering ensemble learning.

II.I TEMPORAL DATA REPRESENTATIONS

In this, describe the motivation to propose temporal data clustering model. Then, present a temporal data clustering model working on different representations. For illustration, perform the principal component analysis (PCA) on four typical representations of a synthetic time-series data set.

The data set is produced by the stochastic function $F(t) = A \sin(2\pi\alpha t + B) + \epsilon(t)$, where A, B, and α are free parameters and $\epsilon(t)$ is the added noise drawn from the normal distribution N(0, 1). The uses of different parameter set (A, B, α) leads to time series of four classes and 100 time series in each class [3]. Temporal data stored in a temporal database is different from the data stored in non-temporal database in that a time period attached to the data expresses when it was valid or stored in the database.

III. WEIGHTED CLUSTERING ENSEMBLE

A cluster ensemble technique is characterized by two components: the mechanism to generate diverse partitions and the consensus function to combine the input partitions into a final clustering or by applying a single algorithm with different parameter settings, possibly in combination with data or feature sampling. The *k*-means algorithm with random initializations or with random number of clusters has been widely used in the literature to generate diverse clustering.

III.I WEIGHTED CONSENSUS FUNCTION

The basic proposal of weighted consensus function is the use of the pair wise similarity between objects in a partition for evident accumulation, where a pair wise similarity matrix is derived from weighted partitions and weights are determined by measuring the clustering quality with different clustering validation criteria. Then, a dendrogram is constructed based on all similarity matrices to generate candidate consensus partitions.

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

Select three criteria of complementary nature for generating weights from different perspectives as Modified Huber's r index (MHr), Dunn's Validity Index (DVI) and Normalized Mutual Information (NMI). A high MHr value for a partition indicates that the partition has a compact and well-separated clustering structure. However, this decisive factor strongly nepotism a partition containing more clusters.

IV. CLIPPING ON TEMPORAL DATA

The simple process of clipping time series reduces memory requirements. The information discarded by clipping does not significantly decrease the accuracy of clustering algorithms. Clipped time series require much less memory to store and can be manipulated faster [5]. If the series are long enough clustering with clipped data is not significantly less accurate than clustering with unclipped data.

Advantages of using clipped data are that algorithms developed for discrete or categorical data can be employed and that clustering on clipped data can serve as a diagnostic method for outlier and model misspecification detection. Clipping could be considered a specific type of SAX (Symbolic Aggregate approXimation) transformation with two classes and no dimension reduction. The properties of clustering with clipped series are firstly, binary data can be more compactly represented and efficiently manipulated and secondly, it is possible to assess their theoretical behavior.

SAX allows a time series of arbitrary length n to be reduced to a string of arbitrary length w, ($w < n$, typically $w << n$). The alphabet size is also an arbitrary integer a, where $a > 2$. Below table summarizes the major notation used in this and subsequent sections.

TABLE 1: SUMMARIZATION OF THE NOTATION

C	A time series $C = c_1,\ldots,c_n$
$\overline{C}$	A Piecewise Aggregate Approximation of a time series $\overline{C} = \overline{c}_1,\ldots,\overline{c}_w$
$\hat{C}$	A symbol representation of a time series $\hat{C} = \hat{c}_1,\ldots,\hat{c}_w$
w	The number of PAA segments representing time series C
a	Alphabet size (e.g., for the alphabet = {**a**,**b**,**c**}, a = 3)

The discrimination procedure is unique in that it uses an intermediate representation between the raw time series and the symbolic strings. First transform the data into the Piecewise Aggregate Approximation (PAA) representation and then symbolize the PAA representation into a discrete string. There are two important advantages to doing this is dimensionality reduction and lower bounding.

IV.I CLIPPING ON CLUSTERING TIME SERIES

This paper proposed approach for the temporal clustering problem. Here, implementing a clipping technique for the data clustering. Clipping or *hard limiting*, a time series is the process of transforming a real valued time series Y into a binary series C where 1 represents above the population mean and 0 below, i.e. if μ is the population mean of series Y then

$$C(t) = \begin{cases} 1 & \text{if} \quad Y(t) > \mu \\ 0 & \text{otherwise} \end{cases}$$

Clipped time series require much less memory to store and can be manipulated sooner, yet preserve much of the fundamental structure that characterizes the real valued series. This means that if the series are long enough clustering with clipped data is not significantly less accurate than clustering with unclipped data. Clipping makes clustering more strong to outliers. This shows that the clusters formed are significantly better with clipped data when there is at least a small probability of the data containing outliers.

Additional advantages of using clipped data are that algorithms developed for discrete or categorical data can be employed and that clustering on clipped data can serve as a diagnostic method for outlier and model misspecification detection. This approach to demonstrating the benefits of clipping is to firstly specify a class of model from which the data may arise, secondly is to present the weighted clustering ensemble approach to this model and finally to experimentally demonstrate that there is significant evidence of the benefits of clipping over the class of existing approach.

IV.II CLIPPING ON CLUSTERING TIME SERIES WITH GAUSSIAN DISTRIBUTION

Here the clipping on clustering the time series is done by the help of the Gaussian distribution function which make helps to improve the representation of the cluster effective. Thus after collecting data based on the clipping technique on the different temporal data representations, this present the Gaussian distribution function to represent the data obtained. Thus by using this Gaussian distribution function will be getting a smooth series of clipped data which are high in accuracy. Next step is to implement the existing Weighted Clustering Ensemble to process the data obtained from the previous clipping technique. That is again moving on with the processing of the previous module with this clipped data represented by Gaussian distribution function. The Gaussian is a continuous distribution

μ = mean of distribution (also at the same place as mode and median)

σ^2 = variance of distribution

y is a continuous variable ($-\infty \le y \le \infty$)

Probability (P) of y being in the range [a, b] is given by an integral

$$P(a < y < b) = \int_a^b p(y)dy = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

V. A SIMPLE DIMENSIONALITY REDUCTION TECHNIQUE FOR FAST SIMILARITY SEARCH IN LARGE TIME SERIES DATABASES

Address the problem of similarity search in large time series databases. Introduce a novel-dimensionality reduction technique that supports an indexing algorithm that is more than an order of magnitude faster than the previous best known method. In addition to being much earlier our approach has many other advantages.

It is easy to know and apply, allows more stretchy distance measures as well as weighted Euclidean queries and the index can be built in linear time. Call our approach PCA-indexing (Piecewise Constant Approximation) and experimentally validate it on space telemetry, financial, astronomical, medical and synthetic data.

VI. SOLVING CLUSTER ENSEMBLE PROBLEMS BY BIPARTITE GRAPH PARTITIONING

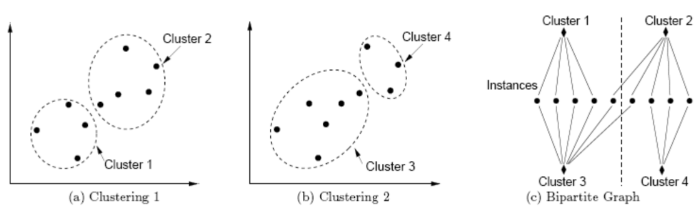
A critical problem in cluster ensemble research is how to combine multiple clustering's to yield a final superior clustering result. Leveraging advanced graph partitioning technique, solve this crisis by sinking it to a graph partitioning problem. Introduce a new reduction method that constructs a bipartite graph from a given cluster ensemble. The resulting graph models similarly instances and clusters of the ensemble concurrently as vertices in the graph. This approach retains all of the information provided by a given ensemble, allowing the resemblance with instances and the similarity among clusters to be considered jointly in forming the final clustering. Further, the resulting graph partitioning crisis can be solved skillfully. Empirically estimate the planned approach against two commonly used graph formulations and show that it is more robust and achieves comparable or better performance in comparison to its competitors.

Clustering for unsupervised data discovery and examination has been investigated for decades in the statistics, data mining, and machine learning communities. A critical problem in designing a cluster ensemble system is how to combine a given ensemble of clustering in order to produce a final solution, referred to as the cluster ensemble problem here. The approach of this problem is by reducing it to a graph partitioning problem. In graph partitioning, the input is a graph that consists of vertices and weighted edges. The goal is to partition the graph into K roughly equal-sized parts with the objective of minimizing the cut.

This paper proposes a new graph formulation that simultaneously models both instances and clusters as vertices in a bipartite graph. Such a graph retains all the information of an ensemble, allowing both the similarity among instances, and the similarity among clusters to be considered collectively to construct the final clusters. Moreover, the resulting graph partitioning problem can be solved efficiently. The experiment compares the proposed graph formulation to the instance-based and cluster-based approaches on five data sets.

The graph partitioning based approaches appear to be very competitive compared with other techniques.

FIGURES: 1 TO 3



An example of the graph formulation

The above figure shows the cluster ensemble with the help of bipartite graph. Figure (a) and (b) depict two different clustering of nine instances and Figure (c) shows the graph constructed by Hybrid Bipartite Graph Formation (HBGF), in which the diamond vertices represent the clusters and the round vertices represent the instances. An edge between an instance vertex and a cluster vertex indicates that the cluster contains the instance. All the edges in the graph have weight one.

Shown in Figure (c) as a dashed line, a separation of the bipartite graph partitions the cluster vertices and the instance vertices concurrently. The separation of the instances can then be output as the ending clustering.

VII. GENERATING CLUSTER ENSEMBLES

Cluster ensembles can be generated in different ways. The resulting ensembles may differ and the same approach for solving the ensemble problems may perform differently accordingly. It is thus important for this experiment to consider different ways to generate cluster ensembles. Experiments use two approaches, random sub sampling and random projection, to generate the ensembles. Note that for both approaches, K-means is used as the base clustering algorithm and the number K is pre-specified for each data set and remains the same for all clustering runs. Also examined a third approach, randomly restarting K-means, and it produced similar results to those of random sub sampling.

VII.I GRAPH PARTITIONING ALGORITHMS

The goal is to evaluate different graph formulation approaches. To reduce the innocence of any chosen graph partitioning algorithm on this evaluation, use two well-known graph partitioning algorithms that differ with respect to their search for the best partition.

1. Spectral graph partitioning is a well studied area with many successful applications. Then, choose a popular multi-way spectral graph partitioning algorithm proposed
2. Metis (Karypis & Kumar, 1998), a multilevel graph partitioning system, approaches the graph partitioning problem from a different angle. It separates a graph with three basic steps: Coarsen the graph by collapsing vertices and edges, partition the coarsened graph and refine the partitions.

Comparison with other graph partitioning algorithms, Metes is highly efficient and achieves competitive performance.

VIII. COMBINING MULTIPLE CLUSTERING USING EVIDENCE ACCUMULATION

Explore the idea of evidence accumulation (EAC) for combining the results of multiple clustering. First, a clustering ensemble - a set of object partitions, is produced. Given a data set (*n* objects or patterns in *d* dimensions), special way of produce data partitions are (1) Applying different clustering algorithms and (2) Applying the same clustering algorithm with different values of parameters or initializations.

The final data partition of the *n* patterns is obtained by applying a hierarchical agglomerative clustering algorithm on this matrix [7]. Stability of the results is evaluated using bootstrapping techniques. Investigational results of the planned method in several synthetic and real data sets are compared with other combination strategies, and in the company of individual clustering results shaped by well known clustering algorithms.

VIII.I IMPLEMENTATION AND EVALUATION

Data clustering or unsupervised learning is an important but an extremely difficult problem. The objective of clustering is to partition a set of unlabeled objects into homogeneous groups or clusters. Many application areas use clustering techniques for organizing or discovering structure in data, such as data mining, information retrieval, image segmentation, and machine learning. In real world problems, clusters can show with different shapes, sizes, data sparseness, and degree of separation. Additionally, noise in the data can mask the true underlying structure present in the data.

Both agglomerative and divisive approaches have been proposed; different algorithms are obtained depending on the definitions of similarity measures between patterns and between clusters. The single-link (SL) and the complete-link (CL) hierarchical methods are the best known techniques in this class, emphasizing, respectively, connectedness and compactness of patterns in a cluster. Prototype-based hierarchical methods, which define similarity between clusters based on cluster representatives, such as the centroid, emphasize compactness.

Among the various clustering methods, the K-means algorithm, which minimizes the squared-error criteria, is one of the simplest clustering algorithms. Extensions of the basic K-means algorithm include: use of Mahalanob is distance to identify hyper-ellipsoidal clusters.

While hundreds of clustering algorithms exist, it is difficult to find a single clustering algorithm that can handle all types of cluster shapes and sizes. And even decide which algorithm would be the best one for a particular data set.

VIII.I.I PRODUCING CLUSTERING ENSEMBLES
Clustering ensembles can be generated by following two approaches: (1) choice of data representation, and (2) choice of clustering algorithms or algorithmic parameters. In the first approach, different partitions of the objects under analysis may be produced by
(a) Employing different pre-processing and/or feature extraction mechanisms, which ultimately lead to different pattern representations (vectors, strings, graphs, etc.) or different feature spaces?
(b) Exploring sub-spaces of the same data representation, such as using sub-sets of features
(c) Perturbing the data, such as in bootstrapping techniques (like bagging), or sampling approaches, as for instance, a set of prototype samples using to represent huge data sets.
The second approach can generate clustering ensembles by: (i) applying different clustering algorithms, (ii) using the same clustering algorithm with different parameters or initializations and (iii) exploring different dissimilarity measures for evaluating inter-pattern relationships, within a given clustering algorithm. A combination of these two main mechanisms would produce clustering ensembles leads to exploration of distinct views of inter-pattern relationships. From a computational perspective, clustering results produced in an independent way facilitate efficient data analysis by utilizing circulated computing, and reuse the results obtained previously.

VIII.I.II COMBINING EVIDENCE
The Co-Association Matrix In order to cope with partitions with different numbers of *clusters*, propose a voting mechanism to combine the clustering results leading to a new measure of similarity between patterns.

FIG. 4

DATA CLUSTERING USING EVIDENCE ACCUMULATION (USING SL).

Input: n - number of patterns $n \times p$ nearest neighbor matrix p - nearest neighbor index

N - number of clusterings $\mathbb{P} = \{P^1, \ldots P^N\}$ - clustering ensemble

Output: P^* - Combined data partition.

Initialization: Set the $n \times p$ co-association matrix, $\mathcal{C}(.,.)$, to a null matrix.

1. For each data partition $P^l \in \mathbb{P}$ do:

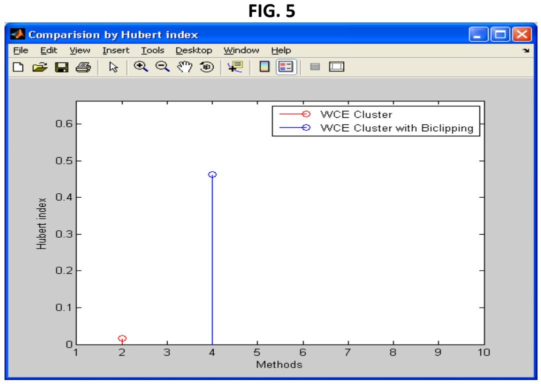
1.1. Update the co-association matrix: for each pattern pair (i, j) in the pth neighbor list, that belongs to the same cluster in P^l, set $\mathcal{C}(i, j) = \mathcal{C}(i, j) + \frac{1}{N}$.

2. Compute the SL dendrogram of $\mathcal{C}$; the final partition, P^*, is chosen as the one with the highest lifetime.

The proposed method for combining various clustering partitions of a given data set in order to obtain a partition that is better than individual partitions. These entity partitions could have been obtained either by applying the similar clustering algorithm with different initialization of parameters or by different clustering algorithms applied to the given data. A K-means based evidence accumulation technique was analyzed with light of the proposed optimality criteria. Outcome obtained on both synthetic and real data sets illustrate the ability of the evidence accumulation technique to identify clusters with arbitrary shapes and arbitrary sizes.
Experimental results were compared with individual runs of well known clustering algorithm, and also with other cluster ensemble combination methods.
The main principle of the temporal data clustering is to provide an effective way to discover the intrinsic structure and condense information over temporal data by exploring dynamic regularities underlying data in an unsupervised learning [6]. Its ultimate objective is to partition an unlabeled temporal data set into clusters so that sequences grouped in the same cluster are coherent. In order to have an effective performance in our proposed system are proposing a clipped Gaussian distribution technique to the clustering time series which could increase the accuracy and reduces the cost of time.

IX. PERFORMANCE MEASURES
Evaluating and comparing proposed approach of temporal data clustering with clipped Gaussian distribution with the previous approach. Here are evaluating with the classification accuracy rate with the missing data and is found that proposed approach performs better than the existing approach.

FIG. 5



The above figure was compared with weighted clustering ensemble and weighted clustering ensemble with clipped gaussian distribution are compared by Dunns Index.

X. CONCLUSION
The conclusion of this imply that a temporal data clustering approach with clipped Gaussian distribution on different representations are further propose a useful measure to understand clustering ensemble algorithms based on a formal clustering ensemble analysis. Simulations show that this approach yields favorite results for a variety of temporal data clustering tasks in terms of clustering quality and model selection.
As a generic framework, this weighted clustering ensemble approach allows other validation criteria to be incorporated directly to generate a new weighting scheme as long as they better reflect the intrinsic structure underlying a data set. Thus, this approach provides a promising yet easy-to-use technique for real world applications.

XI. FUTURE ENHANCEMENT

The clipping with Gaussian on temporal data clustering can be presented for the future work by focusing to more methodological settings in which the benefits of clipping data when clustering time series. With most real world time series clustering problems space and time will be a genuine constraint on the mining process. Clipping time series will allow for more series to be stored in main memory, which in itself will increase the speed of mining. Specific algorithms for binary series can lead to improved time complexity.

When clustering, the information discarded by clipping does not decrease the accuracy if the series are long enough if the data series are long enough then clipping does not significantly decrease clustering accuracy; and if the data contains outliers, the clustering accuracy on clipped data is significantly better. Thus our information when clustering time series is to start with clipped data, then examine any results from more sophisticated transformations in relation to the results obtained after clipping, particularly if the series are long and time and space are important considerations.

REFERENCES

1. Alon, J., Sclaroff, S., Kollios, G., & Pavlovic, V. (2003). Discovering clusters in motion time-series data. In *IEEEComputer Vision and Pattern Recognition Conference (CVPR)*.
2. Bagnall, A. J., Janacek, G., de la Iglesia, B., & Zhang, M. (2003). Clustering time series from mixture polynomial models with discredited data. In *Proceedings of the Second Australasian Data Mining Workshop* (pp. 105–120).
3. Chu, S., Keogh, E., Hart, D., & Pazzani. M. (2002). Iterative deepening dynamic time warping for time series. In *Proceedings of the 2nd SIAM International Conference on Data Mining*.
4. J. Kleinberg, “An Impossible Theorem for Clustering,” Advances in Neural Information Processing Systems, vol. 15, 2002.
5. E. Keogh and S. Kasetty, “On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Study,” Knowledge and Data Discovery, vol. 6, pp. 102-111, 2002.
6. K. Murphy, “Dynamic Bayesian Networks: Representation, Inference and Learning,” PhD thesis, Dept. of Computer Science, Univ. of California, Berkeley, 2002.
7. R. Xu and D. Wunsch, II, “Survey of Clustering Algorithms,” IEEETrans. Neural Networks, vol. 16, no. 3, pp. 645-678, May 2005.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application & Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail **infoijrcm@gmail.com** for further improvements in the interest of research.

If you have any queries, please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-
Co-ordinator

DISCLAIMER

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals