Original Research Article

J. Computing & Int. Systems (2017) 26–30

SACRED HEART RESEARCH PUBLICATIONS

Journal of Computing and Intelligent Systems

Journal homepage: www.shcpub.edu.in

ISSN: 2456 - 9496

A Study on the impact of Cluster analysis with Wireless Sensor Network Applications

***R Denis* [#1], *P Madhubala* [*2]**

Received on 21 Jul 2017, Accepted on 25 Jul 2017

Abstract — *Due to the advancements in electronic field more research work related with cluster analysis was developed. Each application of cluster analysis is mentioned with different goals. Cluster analysis is a data mining technique which is utilized by different streams. It plays a vital role in wireless sensor networks at different modes. For an example, in case of fuzzy logic techniques, energy scavenging sensors, cluster analysis technique is highly utilized. The different applications of cluster analysis can be categorized and resulted with good products. This paper discusses about the different comparisons of applications and cluster algorithms. Cluster analysis is a process of setting analytic procedures that decrease the complex multivariate data into subsets.*

Keywords: CA- Cluster Analysis, WSN- Wireless Sensor Networks

1 INTRODUCTION

Clustering is the job of collecting a pair of objects in a way that the objects in the identical group are more similar to each other while comparing to other groups or clusters. It is a key duty of data mining, and general technique for statistical data analysis used in many fields that includes machine learning, Image processing, management of information retrieval systems, and biomedicine. There are also terms which have similar meaning to clusters; they are automatic classification, numerical taxonomy and typological analysis.

There are numerous applications which can benefit from cluster analysis methods. Employing clusters of computers maximizes the utilization of variety of components, minimizing cost and increased scalability of the each part and the network computers.

CA is used in different applications with different methods by using WSN. The enabled research areas include cybercrime, museum, paramedical research, Market analysis, weather prediction system, climate control systems and standard settings of the ranking in the data set. For Example, the validation of any cluster solution is a critical step in drawing inferences about individual student to visualize the competent performance. Related to pharmaceutical research Belgium HTS program involved with many cluster evolution where so many research had been conducted in order to widen a strong export oriented clusters.

WSN are stepped in the research gene analogy. CA makes the cluster of dataset of people as objects with identified health problems. This necessitated the need to develop a method of comparing regions with similar socio-economic determinants of health. The sensor placed in the human body identifies and passes the signal for system and networks. The chemical industry applications also broadened to incorporate superior biological information further usually linked with bioinformatics. Clustering techniques such as co clustering or self-organizing trees in bioinformatics paves the way for the chemical industry applications. The issues such as visualization and validation of clustering outcome keep on being the challenging problems.

The outlier objects identification from the given cluster objects can be used with or without sensors. In Crime detection all objects are investigated from a small group of clusters are reported for flagging which requires further investigation. The study shows that the cluster analysis is an useful audit technology. Cluster analysis is a promising practice which can be incorporated into a schema of continuous system monitoring and assurance. Experience, judgment and monitoring procedures parameterize these clusters and categorize data.

WSN shows a significant positive reception from the stream of environmental studies where it is used to identify the weather changes and geographical movements even is used to detect the water flow of the river. CA in weather model is an important application used to save the world.

* Corresponding author: E-mail: denis@shctpt.edu, madhubalasivaji@gmail.com

[1] Assistant Professor, P.G. Dept. of Computer Science, Sacred Heart college (Autonomous), Tirupattur, Tamilnadu, India

[2] Assistant Professor, Dept. of Computer Science, Don Bosco College, Dharamapuri, Tamilnadu , India

2 RELATED WORK

It can be achieved by different algorithms that differ significantly in their conception of what constitutes a cluster and how to efficiently discover them. Popular notions of clusters include groups with small distances among the cluster objects, dense areas of the data space, intervals or particular statistical distributions. Clustering can be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings depend on the individual data set and intended use of the end results. Cluster analysis is not an automatic task, but an iterative process of knowledge discovery or an interactive multi-objective optimization that involves with trial and failure. It will often be necessary to modify the data preprocessing and model parameters until the result achieves the desired properties.

The below table summarizes the proximity formula calculations:

Table 1 – proximity measure with distance and domains

Proximity Measure	Distance Formula	Used Domains
Euclidean	$D_{ij} = \left(\sum_{l=1}^{d} \lvert x_{il} - x_{jl} \rvert^{1/2}\right)^2$ D- Distance, X- objects representation in the dataset	Fuzzy Logic Techniques
Manhattan	$d1(p,q) = \lvert\lvert p - q \rvert\rvert_1 = \sum_{i=1}^{n} \lvert pi - qi \rvert$ D- Distance, X- objects representation in the dataset	K- Means algorithm, Agglomerative clustering algorithm, Divisive clustering algorithm
Mahanolobis	$D_{ij} = (x_i\text{-}x_j)^T S^{-1}(x_i\text{-}x_j)$ where S is within group covariance matrix. A-Distance, X- objects representation in the dataset	Ellipsoidal art & hypersoidal art
Minkowski	$D_{ij} = \left(\sum_{l=1}^{d} \lvert x_{il} - x_{jl} \rvert^{1/n}\right)^n$ D- Distance, X- objects representation in the dataset	Fuzzy logic techniques
City block distance	$D_{ij} = \sum_{l=1}^{d} \lvert x_{il} - x_{jl} \rvert$ D- Distance, X- objects representation in the dataset	Fuzzy logic techniques
Sup Distance	$D_{ij} = \sum_{l=1}^{d} \lvert x_{il} - x_{jl} \rvert$ D- Distance, X- objects representation in the dataset	Fuzzy logic techniques
Correlation	$D_{ij} = (1 - r_{ij})/2$, where $r_{ij} = \frac{\sum_{l=1}^{d} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j)}{\sqrt{\sum_{l=1}^{d} (x_{il} - \bar{x}_i)^2 \sum_{l=1}^{d} (x_{jl} - \bar{x}_j)^2}}$ D- Distance, X- objects representation in the dataset	Genetical clusters
Cosine similarity	$D_{ij} = (1 - r_{ij})/2$, where $r_{ij} = \frac{\sum_{l=1}^{d} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j)}{\sqrt{\sum_{l=1}^{d} (x_{il} - \bar{x}_i)^2 \sum_{l=1}^{d} (x_{jl} - \bar{x}_j)^2}}$ D- Distance, X- objects representation in the dataset r- correlation factor	Document clusters
Categorical distance	D_{ij} = (Number of x_i-y_i)/N D- Distance, X, Y- objects representation in the dataset N- Number of dataset	Fuzzy logic techniques
Chebychev distance	$D_{ij} = \max \lvert x_j - x_i \rvert$ D- Distance, X- objects representation in the dataset	Fuzzy logic techniques

Table 1 explained the different formulas and their related applications. In general most of the applications are developed using Euclidean distance and manhattan distance. The manhattan distance is easy to do the calculations but still the better results can be achieved from other formulas like categorical distance and chebychev distance. The below table summarizes the different algorithms and their procedures.

Table 2 – Cluster Analysis with Applications

Area (Domain)	Algorithm used	Efficiency	Future Enhancements
Medical	K – Means	98%	Genetically analysis Research
Standard Settings	Agglomerative method	76.25%	Standardization for Academic Board
Museum & Library research	Divisive method	92%	Research on Interested group Alignment
Environmental	DEEP	76%	Geographical analysis for Tsunami Detection in coastal regions
Market Analysis	LEACH	80%	Research on Industrial Business Market analysis
Cybercrime Detection	LEACH	81%	Research on Forensics data

Table 2 explores the different cluster analysis algorithms and their different application domains. It is also proved that with the results of their efficiency against the domain areas. The future enhancements of the applications on WSN were also discussed.

Table 3 explained the different cluster analysis algorithms and their different procedures, The percentage of utilization among these algorithms with different application systems are listed. The below table summarizes the applications of the cluster analysis with WSN and their efficiency levels.

Table 3 - Cluster analysis algorithms and utilization with applications

Name of the algorithm	Classification	Procedure	Percentage of Utilization	Distance Measure Formula, Application
K- Means	Partitioned Clustering Approach	1. Select some data objects 2. Make them as centroids 3. Find the distance between them 4. Allocate them to the nearest clusters 5. Perform the same until reach centroids	98%	Manhattan, Market Based Analysis
Agglomerative Method	Hierarchical Approach	1. Starting with n clusters 2. Distance matrix for all pairs of cluster objects 3. Sort distance in ascending order 4. Find smallest one, remove it and merge them 5. Continue until finding as a single cluster	76.25%	Manhattan, Weather Prediction Systems
Divisive Method	Hierarchical Approach	1. Starting with n clusters 2. Distance matrix for all pairs of cluster objects 3. Find largest one, remove it and merge them 4. Continue until finding as a single cluster	92%	Manhattan, Climate Control System
Bisecting K-Means	Partitioned Clustering Approach	1. List out the number of clusters 2. Select a cluster 3. Divide the cluster using K-Means 4. Add the cluster based on the Bisection with the lowest distance 5. Until the last cluster	76%	Euclidean, Pharmaceutical Research Labs
DBSCAN	Density Based method	1. Select the neighborhood objects 2. Finalize the core object 3. Find the proximity between the objects 4. Find out the boarder objects 5. Continue until all objects are processed	80%	Euclidean, Museum Research
CLIQUE	Grid Method	1. Partitioning the data objects in non overlapping intervals 2. Embed the data objects as cells 3. Identify the least points with density threshold 4. Joins the two K dimensional dense cells	81%	Manhattan, Climate Control System
CHAMELON	Hierarchical Approach	1. Find the data sets 2. Construct a sparse matrix graph 3. Partition the graph by K-nearest neighbor method and cut the edges 4. Merge all the sub partition data by using agglomerative method	83%	Minkowski, Cybercrime Detection
BIRCH	Hierarchical Approach	1. Select some data objects 2. Make them as centroids 3. Find the distance among them 4. Allocate them to nearby clusters 5. Perform the same until the centroids changes	85%	Minkowski, Water flow Research
OPTICS	Density Based method	1. Starting with n clusters 2. Generate the distance matrix for all pairs of cluster objects 3. Find the largest one, remove it and merge them 4. Continue them until finding as a single cluster	83%	Manhattan, Climate Control System
CLARA	Partitioned Clustering Approach	1. Select some data objects 2. Make them as centroids 3. Find the distance among them 4. Allocate them to nearby clusters 5. Perform the same until the centriod changes	83%	Categorical distance, Market based Analysis

CLARANS	Partitioned Clustering Approach	1. A graph abstraction, $G_{n,k}$ 2. Each node is collection of k medoids 3. |S1 S2|=k-1 4. Each node has k(n-k) neighbors 5. Cost of each node is total dissimilarity of objects to their mediods 6. PAM searches whole graph 7. CLARA searches sub graph	83%	Minkowski, Market Based Analysis
DENCLUE	Density Based Method	1. Starting with n clusters 2. Generate the distance matrix for all pairs of cluster objects 3. Find the largest one, remove it and merge them 4. Continue them until finding as a single cluster	83%	Manhattan, Climate Control System

5 CONCLUSION

As a whole this paper summarizes the information about the different applications related with cluster analysis. It explains the views with the different CA which is useful for building the real time and conceptual applications oriented with clusters. Clusters are useful to identify the similar interest and to discover the outliers. Compared with other data reduction methods CA is the best mechanism for the health sciences, water research, chemical industry or even in the development of various domains. It has become a more widely used analytic tool because of the ease of use and efficiency.

REFERENCES

[1] Andrew McCallum, Kamal Nigam and Lyle Ungar, *Nova Southeastern University College of Pharmacy, Fort Lauderdale*, FL 33328, USA.

[2] Regan, E. "Rethinking museum visitors: Using K-means cluster analysis to explore a museum's audiences." 2011.

[3] Sutapat Thiprungsri, Miklos A. Vasarhelyi "Cluster Analysis for in Accounting Data: An Audit Approach Rutgers University" USA 2010.

[4] Krantz, A., Korn, R., & Menninger, M. "Rethinking museum visitors: Using K-means cluster analysis to explore a museum's audiences". Curator: The Museum Journal, 52(4), 363–374. (2009).

[5] Sanja Scitovski University of Osijek Trg Sv. Trojstva. "Cluster analysis of periodic data with application to investigating and forecasting a high water level of river flow".

[6] McCallum, A.; Nigam, K.; and Ungar L.H.Matching", Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pg:169-178, 2000.

[7] Stefano Serafin, Alessio Bertò, Dino Zardi , "Application of cluster analysis techniques to the verification of quantitative precipitation forecasts", pg:77, I-38050 Trento, Italy.

[8] Maria halkid,Yannis,Michallis Athens University of Economics & Business, "On Clustering Validation Techniques", Patision 76, 10434, Journal of Intelligent Information Systems, 17:2/3, 107–145, 2001.

[9] Rui Xu,Donald Wunch II,"Survey of Clustering Algorithms", IEEE Transactions on Neural Networks,vol 16,No 3, pg:648-649, May, 2005.

[10] Marte Abrahamsen, Ozan Acar,Dany Bahar,Ben Brinded,Vered Rainisch,"The Belgian Pharmaceutical Cluster", Institute for strategy and competitiveness, Harvard business school,2005.

[11] Muhammad Naveed khalid,"Cluster analysis- A standard setting technique in measurement and testing", University of Cambridge ESOL Examinations, UK, JAQM VOL 6, NO 2,summer 2011.

[12] Jian Wei , Li ; Pinghua Chen, "The application of Cluster analysis in Library system", Knowledge Acquisition and Modeling Workshop, 2008. KAM Workshop 2008.

[13] LingxiangZhu ; LiangZou ,"Application of Cluster Analysis Neural Network in the Safety Evaluation for the Urban Underground Gas Pipeline ",Intelligent Control and Automation, 2006. WCICA 2006. 12884 Publication Year: 2006 , Page(s): 2845 – 2849

[14] TangZhi-hang, "Investigation and Application of Cluster Analysis in Service Industries", Information Engineering and Electronic Commerce, 2009. IEEC '09.

[15] Skormin,V.A. ; Popyack,L.J. ; Gorodetski,V.I. ; Araiza,M.L. ;Michel,J.D. ,"Applications of cluster analysis in diagnostics-related problems ,"Aerospace Conference, 1999.

[16] Kuan-Ching Li ; Hsiao-Hsi Wang ; Chiou-Nan Chen ;Chun-Chieh Liu ; Chia-Fu Chang ; Chia-Wen Hsu ;Sheng- ShiangHung ,".Design issues of a novel toolkit for parallel application performance monitoring and analysis in cluster and grid environments", Parallel Architectures, Algorithms and Networks, 2005.

[17] Jianwei Niu ; Yiling He ; Muyuan Li ; Xin Zhang ; Linghua Ran; Chuzhi Chao ; Baoqin Zhang ," A comparative study on application of data mining technique in human shape clustering: Principal component analysis vs. Factor analysis" Industrial Electronics and Applications (ICIEA).

[18] Hai-Dong Meng ; Yu-Chen Song ; Fei-Yan Song ;Hai-Tao Shen"Application research of cluster analysis and association analysis " ,Software Engineering and Data Mining (SEDM), 2010 2nd International Conference on Publication Year: 2010 , Page(s): 597 – 602.

[19] Liu Na ; Gao Wensheng ; Tan Kexiong ; Wang Xiaoning",Application of a combinatorial neural network model based on cluster analysis in transformer fault diagnosis ", TENCON '02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering olume: 3, DigitalObjectIdentifier,10.1109/TENCON.2002.1182702 , Publication Year: 2002 , Page(s): 1873 - 1876 vol.3.

[20] Tumeo,A. ; Villa.O. ".AcceleratingDNA analysis applications onGPUclusters",Application Specific Processors(SASP), 2010 IEEE 8th Symposium DigitalObjectIdentifier: 10.1109/SASP.20 10.5521145 ,PublicationYear:2010 ,Page(s):71-76 Cited by: Papers (1).

[21] Guanji Hou ; Vittal, V. ," Cluster Computing-Based Trajectory Sensitivity Analysis Application to the WECC System", Power Systems, IEEE Transactions on Volume:27, DigitalObjectIdentifier: 10.1109/TPWRS.2011.2160373 PublicationYear:2012 ,Page(s):502-509 ,Cited by: Papers (1).