International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 7 (2015) pp. 18299-18312

http://www.ripublication.com

Adaptive Link Based Cluster Ensemble Process for Cluster Relational Data

[1]D. Veeraiah and [2]Dr. D. Vasumathi

[1]Research Scholar, Dept of CSE, JNTUK, Kakinada, Andhra Pradesh, India
[2]Professor, Department of CSE, JNTUCEH, Hyderabad, Andhra Pradesh, India
[1]veeraiahdvc@gmail.com, [2]vasukumar_devara@yahoo.co.in

Abstract

Although efforts have been made to fix the issue of clustering particular details via group outfits, with the results being aggressive to traditional methods, it is noticed that these methods unfortunately produce any details partition based on imperfect details. The actual ensemble-information matrix provides only cluster-data point interaction, with many entries being left unidentified. The paper provides an research that indicates this issue degrades the quality of the clustering outcome, and it provides a new link-based approach, which enhances the traditional matrix by finding unidentified records through similarity between groups in an collection. In particular, an efficient link-based criterion is suggested for the actual likeness evaluation. Afterward, to obtain the ultimate clustering outcome, a chart dividing technique is used to a calculated bipartite chart that is formulated from the enhanced matrix. Trial outcomes on several real details sets recommend that the suggested link-based method almost always outperforms both traditional clustering methods for particular details and well-known group collection techniques.

Key Words: Link based Cluster Ensemble Process, Fuzzy Data sets, Principal Components Analysis of a Graph, K-Means.

1.0. INTRODUCTION

Data clustering is one of the essential resources we have for knowing the framework of a knowledge set. It performs a crucial, essential part in device studying, data exploration, information recovery, and design identification. Clustering aims to classify data into categories or categories such that the data in the same group are more just like each other than to those in different categories. Information modeling puts clustering in a traditional viewpoint based in arithmetic, research, and mathematical research. From a machine studying viewpoint groups match to invisible styles, the search for groups is without supervision studying, and the resulting system symbolizes a data concept [13][14]. From a practical viewpoint clustering plays an understanding role in data discovery programs such as scientific data discovery, information recovery and text discovery, spatial data source programs, Web research, CRM, marketing, medical diagnostics, computational chemistry, and many others. Considering these features, many well established clustering algorithms such as k-means, PAM and other clustering algorithms have been designed for numerical data whose inherit the properties can be naturally employed to measure a distance feature vectors or matched content in the uploaded data sets may appear with realistic and other commitment features in other attributes present in uploaded data sets. However, these cannot be straight used for clustering of categorical data, where sector principles are distinct and have no ordering described.

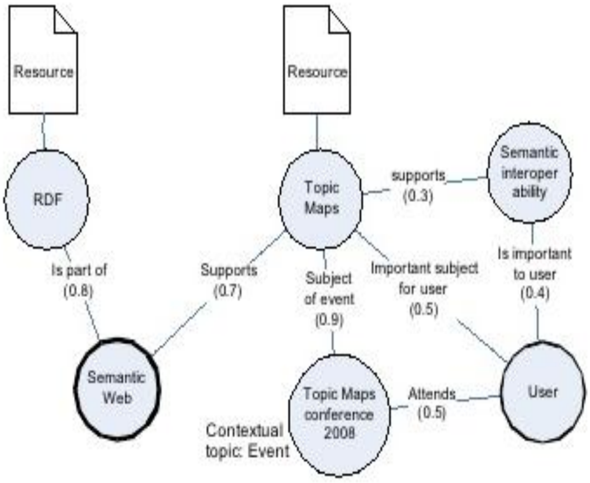

Fig 1: Record Ranking based on their data retrieval using FIS System.

The above figure shows efficient data extraction from different types of records for accessing relevant data to do different activities on extracted data. Fuzzy Information System may retrieve only relevant based on fuzzy logic retrieval methods in real time application progress [18]. Fuzzy is not supported for retrieving relevant data based on links matched with searched keyword relationships for accessing relevant and other activities presented in uploaded data sets. And also it performs only SQL queries repository but they are not passed to develop real time data sets for accessing services of the progressive environment. An example of particular feature is sex = (male; female) or shape = (circle; rectangle);

For a particular information set, different methods, or even the same criteria with different factors, usually offer unique alternatives. Therefore, it is difficult for customers to choose which criteria would be the proper alternative for a given set of information [1]. Lately, cluster ensembles have appeared as an efficient remedy that is able to get over these restrictions, and enhance the robustness as well as the high top quality of clustering outcomes. The main purpose of group outfits is to merge different clustering choices in such a way as to accomplish accuracy superior to that of any personal clustering. Our proposed cluster ensemble approach follows following contributions:

- The feature-based strategy that converts the problem of group outfits to clustering categorical data (i.e., group labels) [2][3][4].
- The immediate strategy that discovers the ultimate partition through relabeling the platform clustering results.
- Graph-based methods that implement a chart partitioning methodology and the pair wise-similarity strategy that creates use of co-occurrence interaction between information points [4].

Consequently, the performance of current group collection methods may consequently be deteriorated as many matrix records are left unknown. This document presents a link-based strategy to refining these matrix, providing substantially less unidentified records. A link-based likeness evaluate is utilized to calculate unidentified principles from a link system of groups [1]. This analysis exclusively connects the gap between the process of information clustering and that of link analysis. It also increases the ability of ensemble methodology for particular information, which has not received much interest in the literary works.

The rest of this paper is organized as follows: Section II follows and explains Background approach of development and other criteria's in Fuzzy information system for processing efficient and effective data retrieval procedures with performance evaluation and experimental results. Section III explains cluster ensemble processing approach for categorical data in uploaded data sets which includes latest and other proceeding applications. Section IV explains A novel Link based approach working procedure. Section V explains experimental results when apply link based procedure on real time data sets and explains performance evaluation also. Section VI explains described conclusion with respect o Link based when compare to FIS system.

2.0. BACKGROUND APPROACH

"Fuzzy querying is just like the procedure of ordinary querying, but more buildings. ". Classical relational data source experience from a absence of flexibility in question. The given choice situation and the contents of the interaction are all sharp. A question is versatile if the following circumstances can be satisfied

1. A qualitative distinction between the chosen tuples is permitted.
2. Obscure circumstances within concerns are introduced, when the customer cannot determine his/her needs in a certain way, or when a pre specified variety of responses is preferred and therefore a edge is allowed to understand the question.

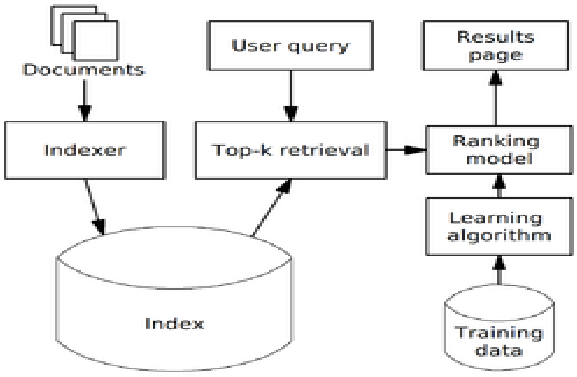

Fig 2: Procedure for Fuzzy Search queries retrieval.

A fuzzy set is almost any situation for which we have words: brief men, great females, hot day, cold climate, new developing, fresh apples, great intellect, low rate, obese, etc., where the situation can be given a value between 0 and 1. Fuzzy set ‘A' over a universe of discussion X (a limited or unlimited interval) within which the unclear set can take a value) is a set of pairs:

$$V = \{\mu V(x)/x : x \in X, \mu V(x) \in [0,1] \in R\} \quad \text{-----------} \quad (1)$$

Where $\mu V(x)$ is the membership degree of the element x to fuzzy set V. This degree ranges between the extremes 0 and 1 of the dominion of the real numbers: μV(x) = 0 indicates that x in no way belongs to the fuzzy set A, and μV(x) = 1 indicates that x completely belongs to the fuzzy set A. Note that μV(x) = 0.5 is the greatest uncertainty point. This procedure may perform individual operations on SQL data repository only, and also not perform categorical data representation in real time application development.

3.0. CLUSTER ENSEMBLE APPROACH

In this section we presents the group collection structure upon which the current analysis has been recognized [1]. The suggested link-based approach, such as the actual instinct of refining an ensemble-information matrix and details of a link-based likeness measure.

3.1. Problem Formation and Common Framework

Let X = (x1; . . . ; xN) be a set of N details factors and π=(π1;π2,π3,π4, . . . πn;)□ Mg be a group collection with M platform clusterings, each of which is generally known as an collection participant. Each base clustering profits a set of groups $\pi_i = \{C_1^i, C_2^i, \ldots, C_{k_i}^i\}$, such that $\bigcup_{j=1}^{k_i} C_j^i = X$, where ki is the variety of groups in the ith clustering. For each x 2 X, CðxÞ signifies the group brand to which the details factor x connected. In the ith clustering, C(x) = "j" (or "C_j^i") if $x \in C_j^i$. The issue is to discover a new partition π* of a details set X that summarizes the details from the cluster collection π [6][1].

Basically, solutions obtained from different base clustering are aggregated to form any partition. This met level technique includes two major projects of: 1) creating group collection, and 2) producing the ultimate partition, normally referred to as a agreement function.

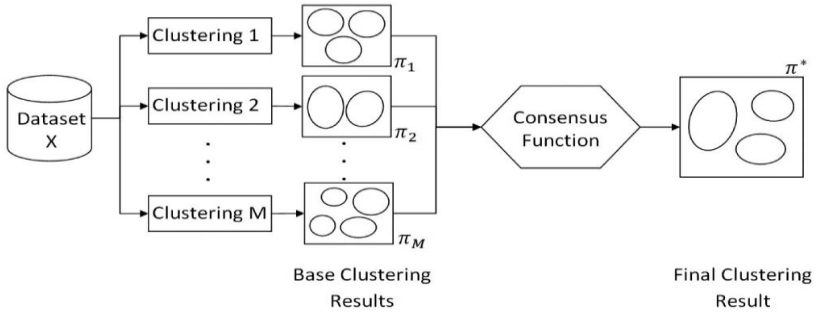

Fig 2: The primary procedure of group outfits. It first is applicable multiple base clustering's to a information set X to acquire different clustering decisions $\pi 1, \pi 2, \pi 3, ...\pi m$. Then, these alternatives are mixed to set up the final clustering outcome $\pi*$ using a agreement operate.

3.2. Ensemble Generation Methods:

It has been proven that outfits are most efficient when constructed from a set of predictors whose mistakes are dissimilar. Particularly for information clustering, the outcomes acquired with any individual criteria over many versions are usually very similar. In such a situation where all collection members agree on how a information set should be portioned, aggregating the platform clustering outcomes will display no enhancement over any of the component associates. Consequently, several heuristics have been suggested to present synthetic instabilities in clustering methods, providing variety within a cluster ensemble. Some of the consecutive functions were used for categorical data clustering specification.

3.3. Consensus Functions:

Having obtained the cluster ensemble, a variety of consensus functions have been developed and made available for deriving the ultimate data partition. Each consensus function utilizes a specific form of information matrix, which summarizes the base clustering results. In mild of this qualifications, consensus techniques can be classified as follows:

Feature-based strategy.

It converts the issue of cluster outfits to clustering particular information. Specifically, each platform clustering provides a cluster label as a new function explaining each information factor.

Direct Approach:

It is depending on relabeling π i and searching for the $\pi *$ that has the best coordinate with all $\pi 1, \pi 2, \pi 3, ...\pi m$ Conceptually, the underlying relabel procedure allows the homogeneous labels to be recognized from heterogeneous clustering decisions, where each platform clustering offers a unique set of choice labels.

Pair wise-similarity approach.
It creates a matrix, containing the pairwise similarity among data points, to which any similarity based clustering algorithm [8][9].

3.4. Cluster Ensembles of Categorical Data:

While a huge variety of group collection methods for numerical details have been put ahead in the previous decade, there are only a few research that implement such a methodology to particular details clustering. The final clustering outcome is produced using the graph-based consensus techniques.

Particular to this so-called "direct" collection creation technique, a given categorical data set can be showed using a binary cluster-association matrix. Such an details matrix is comparable to the "market-basket" numerical reflection of particular details, which has been the concentrate of conventional particular details analysis.

4. PROPOSED APPROACH

Current group collection techniques to particular data analysis depend on the common pair wise-similarity and binary cluster-association matrices, which review the underlying collection information at a rather rough stage. Many-matrix records are remaining "unknown" and simply recorded as "0" Regardless of agreement operate; the high quality of the final clustering outcome may be deteriorated [1][5]. Consequently, a link-based method has been recognized with the capability to discover unidentified principles and, hence, enhance the accuracy of the greatest information partition.

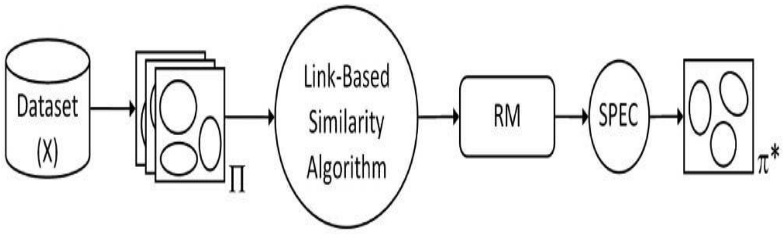

Fig 3: The link-based group collection framework:1) a cluster ensemble $\pi = \{\pi 1, \pi 2, ..., \pi m\}$ is designed from M platform clusterings, 2) a refined cluster-association matrix is then designed from the ensemble using a link-based likeness criteria, and 3) any clustering outcome $\pi *$ is designed by a agreement operate of the spectral chart dividing.

Regardless of promising findings, this preliminary structure is in accordance with the information point data point pair wise-similarity matrix, which is highly expensive to acquire. A new link-based group collection (LCE) strategy is presented herein. It is more effective than the former design, where a BM-like matrix is used to signify the collection details. The concentrate has moved from exposing the likeness among details factors to calculating those between groups. A new link-based criteria has been particularly suggested to produce such actions in an precise, affordable way.

Creating Cluster Ensemble:
Let $X = \{x1, x2, ...xn\}$ be a set of N data points, $A = \{a1, a2, ...am\}$ be a set of particular features, and $\pi = \{\pi1, \pi2, ...\pi m\}$ be a set of M partitions [8]. Each partition **πi** is generated for a specific categorical attribute $ai \in A$ **.** Clusters belonging to a partition $\pi_i = \{C_1^i, C_2^i, ...C_n^i\}$ correspond to different values of the attribute $a_i = \{a_1^i, a_2^i, ...a_n^i\}$ where $\bigcup_{j=1}^{ki} C_j^i = ai$ and ki is the variety of principles of feature ai. With this formalism, particular information X can be straight modified to a group ensemble π without actually applying any platform clustering. While single-attribute information categories may not be as precise as those acquired from the clustering of all information features, they can carry about excellent variety within an collection. Besides its performance, this collection creation technique has the prospective to cause to a high-quality clustering outcome. According to the processing of data sets presentation in real time data applications we process to generate the matrix operations.

Generating Matrix:
Each entry in this matrix BM $(x_i cl) \in \{0,1\}$ represents a crisp association degree between data point xi ∈ X and cluster $cl \in \pi$. a huge number of records in the BM are unidentified, each provided with "0." Such situation happens when interaction between different groups of a platform clustering are initially believed to be nil [6][8]. Actually, each information point can probably affiliate (to a certain level within (0; 1) to several groups of any particular clustering. These invisible or unidentified organizations can be approximated from the likeness among groups, found from a system of groups. The refined cluster-association matrix is put forward as the enhanced variation of the original BM. Its aim is to approximate the value of unknown associations ("0") from known ones ("1"), whose association degrees are preserved within the RM $BM(x_i, cl) = 1 \rightarrow RM(x_i, cl) = 1$ For each clustering $\pi_t, t = 1......M$ and their corresponding clusters $C_1^t, C_2^t, ...C_{n_t}^t$ (where kt is the number of clusters in the clustering π_t. The association degree matrix RM $(x_i, cl) \in [0,1]$ that data point $x \in X$ has with each cluster $cl \in \{C_1^t, C_1^t, ...C_{k_t}^t\}$ is approximately as follows:

$$RM(xi, cl) = \{ \begin{matrix} 1, & ifcl = C_*^t(x_i) \\ sim(cl, C_*^t(x_i)), & otherwise \end{matrix} \} \text{-------------} \quad (2)$$

Where $C_*^t(x_i)$ is corresponding to a particular cluster of the clustering π_i to which data point xi belongs. In addition, $sim(Cx, Cy) \in [0,1]$ denotes the similarity between any two clusters $Cx, Cy,$ which can be discovered using the following link-based algorithm [1]. Notify the procedure as follows for cluster formation as follows:

$\pi_t \in \pi, 1\angle \sum \forall C \in \pi tRM(x_i, C \subseteq k_t)$ Unlike the measure of fuzzy membership, the typical constraint of $\sum \forall C \in \pi_t RM(xi, C) = 1$ is not appropriate for rescaling associations within the RM. Actually, this regional normalization will considerably change the real semantics of known associations (“1”), such that their magnitudes become dissimilar, different from one clustering to another. According to the scientific research, this fuzzy-like enforcement decreases the high company's RM, and hence, the performance of the causing group collection technique.

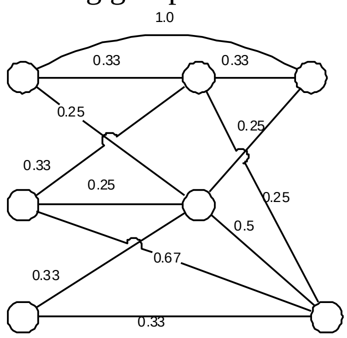

Figure 4: An example of a group system, where each advantage is noticeable with its bodyweight.

The above figure shows efficient data presentation based on their proceeding events real time application process. Based on above graph we present to classify and perform measure of clusters $Cx, Cy \in y$ with respect to each triple weight based on calculation of Weighted Triple Quality (WTQ) algorithm:

$$WTQ(G, Cx, Cy)$$
$$G = (v, w); aweightedgraph, whereCx, Cy \in V;$$
$$Nk \subset V, asetofadjacent, neighbor, ofCk \in V;$$
$$Wk = \sum \forall C_t \in N_k w_{tk};$$
$$WTQ_{xy}, the, WTQ, measureof, Cx\{and\}Cy;$$
$$1.WTQ_{xy} \leftarrow 0$$
$$2.Foreachc \in Nx$$
$$3.if, c \in Ny$$
$$4.WTQ_{xy} \leftarrow WTQ_{xy} + \frac{1}{Wc}$$
$$5.return, WTQ_{xy}$$

Fig 5: Procedure for processing clusters based on their allocated weights.

By using the procedure of the WTQ the similarity cluster in between Cx,Cy as follows:

$$sim(Cx, Cy) = \frac{WTQ_{xy}}{WTQ_{\max}} XDC \text{ ....} \quad (3)$$

a graph-based partitioning method is exploited to obtain the final clustering. The equ (3) shows maximum process utilization of the cluster formation based on above cluster procedure in recent application process.

5. PERFORMANCE EVALUATION

This area provides the assessment of the suggested linkbased method (LCE), using a wide range of credibility spiders and real information places. The high quality of information categories produced by this strategy is evaluated against those designed by different categorical information clustering methods and group ensemble techniques.

Table 1: Different types of data sets related to extract the complete process based on procedure of relationship of each data point.

Dataset	N	D	A	K
Zoo	101	18	37	8
Lymphography	152	20	63	19
Soybean	307	34	150	24
20 News Group	1000	6,084	12,168	2
KDDCup99	100,00	46	148	26

5.2. Datasets Retrieval:

The trial assessment is performed over nine data sets. The “20Newsgroup” information set is a part of the wellknown text information collection—20-Newsgroups,2 while the others are acquired from the UCI Device Learning Repository. Their information are described in Table 1.

Missing principles (denoted as “?”) in these information places are simply handled as a new particular value. The “20Newsgroup” data set contains 1,000 records from two newsgroups, each of which is described by the situations of 6,084 different conditions. In particular, the frequency (f 2 f0; 1; . . . ;1g) that a key term seems to be in each document is modified into a affordable value: “Yes” if $f > 0$, “No” otherwise [3][5]. Moreover, the “KDDCup99” data set used in this assessment is a arbitrarily chosen part of the unique information. Each information factor (or record) matches to a system relationship and contains 42 attributes: some are nominal and the relax are ongoing.

5.3. Experimental Results:

In accordance with the category precision, Desk 2 analyzes the performance of different clustering techniques over analyzed information places [7]. Observe that the provided actions of cluster ensemble techniques that apply the selection Type-II and Type-III are the earnings across 50 operates. Moreover, a measure is noticeable “N/A” when the clustering result is not obtainable. For each information set, the biggest five CA-based values are outlined in boldface.

Table 2: Accuracy Results of traditional and proposed techniques.

Data Set	Fuzzy Information System	Link based Clustering analysis
Accident	0.55	0.53
Diabetes	0.75	0.43
Economy Ratings	0.33	0.27
Marks	0.02	0.003

The results proven in this table indicate that the LCE techniques usually execute better than the investigated collection of group selection techniques and clustering algorithms for particular information [12] . LCE is also appropriate to a huge information set such as KDDCup99, for which several group collection techniques (CO+SL, CO+AL, and CSPA) are negligible.

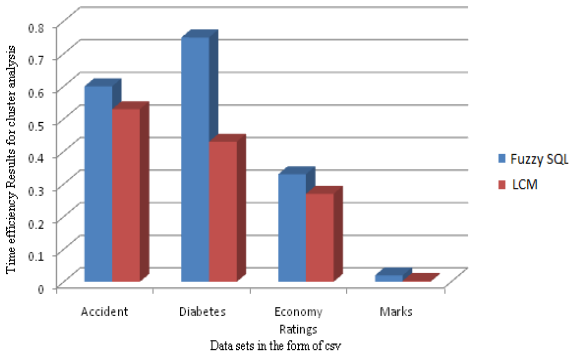

Fig 5: Experimental results comparison of both fuzzy information search and other proceedings in real time data sets.

With the actions of LCE designs being mostly higher than those of the corresponding guideline counterparts (Base), the high company's RM seems to be significantly better than that of the unique, binary difference. The confidence level of the categorical data as shown in below with requirement of the processing data points results in real time data sets as follows with includes intervals $\left[L_{X_{C(i,\beta)}},\ U_{X_{C(i,\beta)}}\right]$ for the mean as $X_{C(i,\beta)}$ with validity criteria C as follows:

$$L_{X_{C(i,\beta)}} = X_{C(i,\beta)} - 1.89\frac{S_{C(i,\beta)}}{\sqrt{n}} \text{-----------} \quad (3)$$

$$U_{X_{C(i,\beta)}} = X_{C(i,\beta)} + 1.89 \frac{S_{C(i,\beta)}}{\sqrt{n}} \text{----------} \qquad (4)$$

As shown in the above figure $S_{C(i,\beta)}$ is standard deviation of the validity index C cross n runs for a clustering method I and data set β. Compare to the processing of earlier techniques and proposed application development calculated by using better performance when forms clusters.

$$B_{C(i)} = \sum_{\forall \beta \in DT} \sum_{\forall i^* \in CM, i^* \neq i} better_C^{\beta}(i, i^*), \text{------} \qquad (5)$$

$$better_C^{\beta}(i, i^*) = \{ \begin{matrix} 1 & if L_{X\,C(i,\beta)} > U_{XC(i^*,\beta)} \\ 0 & otherwise \end{matrix} \} \text{-------} \qquad (6)$$

Similarly, the number of times that one method i < CM is significantly worse than its competitors, WC(i), in accordance with the validity criterion C, can be computed as:

$$W_{C(i)} = \sum_{\forall \beta \in DT} \sum_{\forall i^* \in CM, i^* \neq i} worse_C^{\beta}(i, i^*), \text{------} \qquad (7)$$

$$worse_C^{\beta}(i, i^*) = \{ \begin{matrix} 1 & if U_{XC(i^*,\beta)} < L_{X\,C(i,\beta)} \\ 0 & otherwise \end{matrix} \} \text{--------} \qquad (8)$$

Using these evaluation formalism, Table 3 demonstrates for each method the wavelengths of significant better (B) and important worse (W) performance, which are classified depending on the evaluation indices [19][20]. The performance of the both FIS and LCE illustrated in Table 3, LCE give best performance than earlier technique presented in data process in real time application development. Despite the point that many clustering methods and LCE are designed with the ability of comparing attribute principles in mind, they accomplish the preferred metric differently, using particular information designs [16]. LCE uniquely and clearly designs the inherent issue as the evaluation of link-based likeness among chart vertices, which take a position for particular feature principles or produced clusters .

Table 3: For each evaluation index, “B” and “W” denote the number of times that a particular method performs significantly “better” and “worse” than the others.

Ensemble Type	Method	CA		NMI		AR	
		B	W	B	W	B	W
I	LCE	171	35	141	70	151	61
	FIS	137	78	134	70	142	73
II FIXED	LCE	218	8	212	45	204	18
	FIS	138	64	141	31	116	82
III-RANDOM	LCE	219	35	209	15	208	34
	FIS	122	52	209	12	208	15

Furthermore, LCE performs continually better than its opponents with all different collection dimensions, while CO+SL appear to be the least efficient. Observe that a larger collection results in an improved precision, but with the trade-off of runtime.

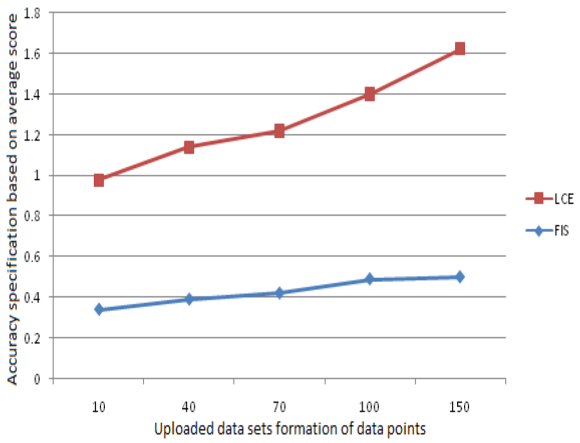

Fig 6: Accuracy measure based on generated data point processing real time datasets using two relational approaches.

The results shown in this table indicate the superior effectiveness of the proposed link-based methods, as compared to other clustering techniques included in this experiment.

6. CONCLUSION

A novel, impressive link-based cluster collection strategy to particular information clustering. It transforms the unique particular information matrix to an information-preserving mathematical difference (RM), to which an efficient chart dividing strategy can be directly applied. The issue of building the RM is efficiently resolved by the likeness among particular brands (or clusters), using the Calculated Triple-Quality similarity algorithm. The scientific research, with different ensemble types, credibility actions, and information places, indicates that the proposed link-based strategy usually accomplishes superior clustering outcomes in comparison to those of the traditional categorical information methods and standard group ensemble techniques. The popular upcoming work contains an extensive research regarding the actions of other link-based similarity actions within this issue perspective. Also, the new strategy will be used to particular websites, including tourism and medical information places.

7. REFERENCES

[1] "A Link-Based Cluster Ensemble Approach for Categorical Data Clustering",

by Natthakan Iam-On, Tossapon Boongoen, Simon Garrett, and Chris Price, In IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 3, March 2012.
[2] "Clustering Aggregation," by A. Gionis, H. Mannila, and P. Tsaparas, Proc. Int'l Conf. Data Eng. (ICDE), pp. 341-352, 2005.
[3] "Consensus Clusterings," by N. Nguyen and R. Caruana, Proc. IEEE Int'l Conf. Data Mining (ICDM), pp. 607-612, 2007.
[4] "Clustering Ensembles: Models of Consensus and Weak Partitions," by A.P. Topchy, A.K. Jain, and W.F. Punch, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 12, pp. 1866-1881, Dec. 2005.
[5] "Disclosing False Identity through Hybrid Link Analysis," by T. Boongoen, Q. Shen, and C. Price Artificial Intelligence and Law, vol. 18, no. 1, pp. 77-102, 2010.
[6] F. Fouss, A. Pirotte, J.M. Renders, and M. Saerens, "Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3, pp. 355-369, Mar. 2007.
[7] E. Minkov, W.W. Cohen, and A.Y. Ng, "Contextual Search and Name Disambiguation in Email Using Graphs," Proc. Int'l Conf.Research and Development in IR, pp. 27-34, 2006.
[8] P. Reuther and B. Walter, "Survey on Test Collections and Techniques for Personal Name Matching," Int'l J. Metadata, Semantics and Ontologies, vol. 1, no. 2, pp. 89-99, 2006.
[9] M. Dutta, A.K. Mahanta, and A.K. Pujari, "QROCK: A Quick Version of the ROCK Algorithm for Clustering of Categorical Data," Pattern Recognition Letters, vol. 26, pp. 2364-2373, 2005.
[10] E. Abdu and D. Salane, "A Spectral-Based Clustering Algorithm for Categorical Data Using Data Summaries," Proc. Workshop Data Mining using Matrices and Tensors, pp. 1-8, 2009.
[11] C. Domeniconi and M. Al-Razgan, "Weighted Cluster Ensembles: Methods and Analysis," ACM Trans. Knowledge Discovery fromData, vol. 2, no. 4, pp. 1-40, 2009.
[12] H. Xue, S. Chen, and Q. Yang, "Discriminatively Regularized Least-Squares Classification," Pattern Recognition, vol. 42, no. 1, pp. 93-104, 2009.
[13] M.J. Zaki and M. Peters, "Clicks: Mining Subspace Clusters in Categorical Data via Kpartite Maximal Cliques," Proc. Int'l Conf.Data Eng. (ICDE), pp. 355-356, 2005.
[14] "A Literature Overview of Fuzzy Database Models*",by Z. M. M and LI YAN, Journal Of Information Science And Engineering 24, 189-202 (2008).
[15] "A new direction in AI : Toward a computationa ltheory ofperceptions",by 12. Zadeh L. A. Technologies for Constructing Intelligent Systems I,Physica-Verlag Heidelberg New York, pp 3-20 (2002).
[16]. Zimmerman J. [2001], "Fuzzy Set Theory – And It's Applications", KluwerAcademic Publishers, Norwell, Massachusetts, U.S.A.
[17] M. Saerens and F. Fouss, "HITS Is Principal Component Analysis," Proc.

2005 IEEE/WIC/ACM Int'l Joint Conf. Web Intelligence, pp. 782-785, 2005.

[18] M. Saerens, F. Fouss, L. Yen, and P. Dupont, "The Principal Components Analysis of a Graph, and Its Relationships to Spectral Clustering," Proc. 15th European Conf. Machine Learning (ECML '04), pp. 371-383, 2004.

[18] L. Yen, F. Fouss, C. Decaestecker, P. Francq, and M. Saerens, "Graph Nodes Clustering Based on the Commute-Time Kernel," Proc. 11th Pacific-Asia Conf. Knowledge Discovery and Data Mining (PAKDD '07), 2007.

[19]. "A Link Analysis Extension of Correspondence Analysis for Mining Relational Databases", by Luh Yen, Marco Saerens, in IEEE Transactions On Knowledge And Data Engineering, VOL. 23, NO. 4, APRIL 2011.

[20] "The Application Of Fuzzy Logic To The Construction Of The Ranking Function Of Information Retrieval Systems", By N.O. Rubens**,** *Computer Modelling And New Technologies, 2006, Vol.10, No.1, 20-27 Transport And Telecommunication Institute, Lomonosov 1, Lv-1019, Riga, Latvia*

About Authors

D. Veeraiah received his B.Tech and M.Tech from JNT University Hyderabad, India in 2005 and 2009 respectively and currently working toward PhD degree from JNT University Kakinada. Currently working as an Associate Professor in Department of CSE, LakiReddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh, India. His areas of research interests are Data Mining and Network Security. He is a life member for CSI and ISTE.

Dr.D. Vasumathi received her B.Tech, M.Tech and Ph.D from JNT University Hyderabad, India in 1999, 2005 and 2011 respectively. Currently working as Professor in Department of CSE, JNTU College of Engineering, Hyderabad, India. Her areas of research interests are Data Mining and Network Security. She is a life member for CSI, ISTE and IEEE. She has authored a good number of research papers published in International Journals and Conference Proceedings.