International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 31 • 2017

Cluster Ensemble Approach Based Concept Summarization in Uncertain Categorical Data Streams

N. Raghavendra Sai[a] and K. Satya Rajesh[b]

[a]*Research Scholar, Bharthiar University, Coimbatore*
E-mail: nallagatlaraghavendra@gmail.com
[b]*HOD,Computer Science, SRR & CRR Govt College*
E-mail: ksatyarajesh@gmail.com

Abstract: Data summarization in unrealistic or uncertain data streams is a basic concept in relational data sources. For outstanding data summarization on uncertain data stream evaluation with jumps of data streams environments. Traditionally one-class learning concept summarization approach was introduced to define the corresponding instance and then construct Uncertain One Class Classifier (UOCC) by utilizing one class summarization effectively. This framework kernel density based method to generate possible score to obtain each attribute with feasible data maintenance; UOCC also provides support vector (SV) representation to summarization concept based on user's preferences and user's requirement in stored data source. It was generated possible score based on data instances. It is failed to support data exploration based on data attributes (characteristics) to utilize data instances with cluster relational data sets. So in this paper, we propose to develop Cluster Ensemble Approach (CEA) to define data summarization and client chunk cluster histories data exploration in uncertain data streams. CEA defines a matrix to construct unidentified records into cluster in uncertain reliable data streams with attribute partitioning and feature selection. Our experimental results show effective data summarization with uniform user's data exploration with their search histories from uncertain data streams with respect to time and other feature factors.
Keywords: *K-Means, Uncertain One Class Classifier, Cluster Ensemble Approach, Support Vector Machine, Feature Representation.*

1. INTRODUCTION

In numerous class issues, countless illustrations might be accessible in incorporation to a little drilling set. To profit by such representations, one as a rule utilizes either unquestioningly or plainly the connection between the minor strength P(x) over the cases of a classification x and the depending robustness P(y|x) speaking to the choice fringe for the brand y. For instance, high-thickness regions or gatherings in the subtle elements can be relied upon to fall totally in some classification. One system to control the minor strength P(x) between sessions is specific trying, which is a part of the dynamic examining technique. In this system, the effectiveness of classifiers is enhanced by including extra points of interest to a guiding set. By and large, there is somewhat set of checked points of interest and a colossal arrangement of unlabeled subtle elements. In incorporation, you can discover conceivable of asking a specialist (prophet) for checking more points of interest. In any case, this may not be utilized to a great degree *e.g.* for monetary reasons.

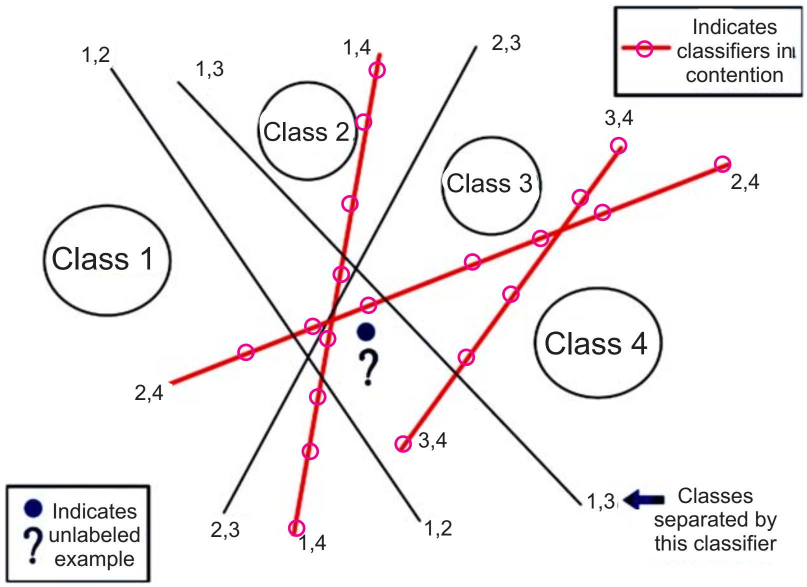

Figure 1: Illustration of separated classes in single class classification in labeled data construction

The question is: the manner by which to choose an additional piece of unlabeled points of interest with the end goal that subsequent to stamping and, for example, it in it set the effectiveness of a specific classifier builds the most. To demonstrate this system diagrammatically as appeared in figure 1. For this situation, One class studying just a single sort of illustrations is named in it organize. The checked class is commonly called the objective/positive classification, while every single other illustration not in this class are known as the nontarget classification. In some true applications, for example, variation from the norm distinguishing proof, it is anything but difficult to acquire one kind of ordinary points of interest, while gathering and checking unpredictable occurrences might be costly or unthinkable. In such cases, one-class contemplating has been considered to take in an extraordinary classifier from the marked target classification, and afterward use the found one-class classifier to choose whether a test case is one of the target class or not. To date, one-class considering has been found a huge assortment of projects from variation from the norm distinguishing proof papers classification programmed picture explanation creation affirmation, translation figure executed site recognizable proof, change ID to marker points of interest move ID.

By seeing above discussion, we address the issue of one-class learning on vague subtle elements sources and thought synopsis considering of the client from record points of interest sources. Ordinarily prescribe a structure, known as vague one-class contemplating and thought synopsis considering structure (UOLCS) on misty points of interest sources, which manages subtle elements of uncertainty and the thought rundown examining in hazy one-class subtle elements sources. UOLCS includes two sections. In the primary angle, we assemble an Uncertain One-Class Classifier (UOCC) by incorporating the hazy points of interest into the one-class SVM contemplating stage to manufacture a superior classifier. In the second viewpoint, we audit client's thought move from points of interest sources by making a bolster vectors (SVs)- based bunching procedure over the record segments. To give points of interest disclosure clients gather fixated on components and elements in dependable hazy subtle elements sources. So in this paper, we prescribe to create Cluster Ensemble Approach to characterize record joins in light of properties in indeterminate information streams with possible and ID formal parameters.

Thus, the effectiveness of current gathering accumulation methods may subsequently be disintegrated the same number of framework records are left unidentified. This paper introduces a Cluster Ensemble Way of enhancing irregular lattice to give extensively less unidentified records. A connection based similarity assess is used to figure unidentified standards from a connection system of gatherings. This exploration only associates the hole between the procedure of data bunching and that of web connection investigate. It additionally expands the capacity to accumulation system for specific data, which has not acquired much consideration in the artistic works. Strategy of the bunch gathering approach appeared in figure 2 with relative components in group social information bases.

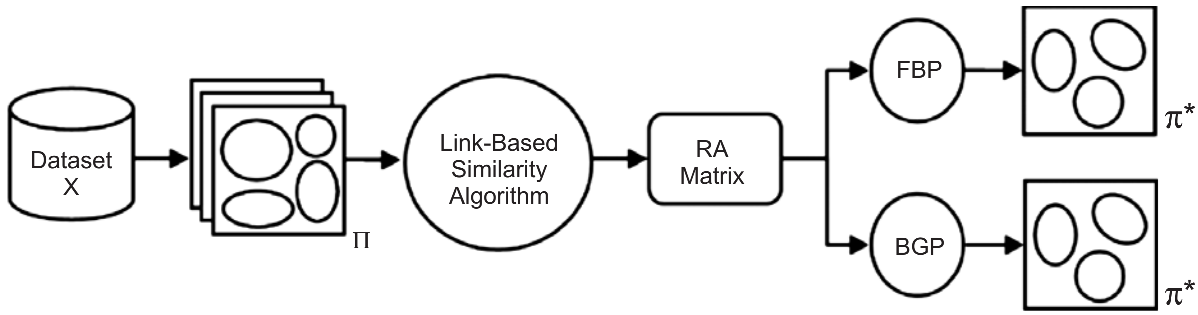

Figure 2: Procedure of the cluster ensemble for feature links in uncertain cluster data

Notwithstanding the issue of grouping specific data that is analyzed thus, the proposed structure is by and large with the end goal that it can likewise be effectively used to other data sorts.

The fundamental commitments of our proposed approach as takes after:

1. The component based procedure that changes over the issue of gathering outfits to bunching absolute information (*i.e*., aggregate marks)
2. The quick procedure that finds a definitive segment through relabeling the base clustering comes about
3. Graph-based techniques that utilization a diagram apportioning strategy
4. The sets insightful similitude methodology that uses co-event communication between data focuses.

Remaining of this organized as follows: Section 2 related work to define one class classification procedure to define attributes based on instances. Section 3 formalizes problem definition in one class classification in uncertain data streams. Section 4 defines cluster ensemble approach to define relations between selected features in cluster relational databases. Section 5explains experimental evaluation with respect to UOCC and CEA based on selected features. Sections 6 concludes overall conclusion about CEA approach to construct data summarization based on user group based on instances.

2. RELATED WORK

In this area, we rapidly overview past business identified with our survey. Starting late, many pushed strategies have been made to accumulate and store broad measures of data; while a couple records in the data might be spoiled on account of tumult, the precision of equipment or diverse components, which prompts to absent or mostly whole data. In like manner, the data articles may be quite recently disastrously shown and considered unverifiable in their portrayal. To date, various figurings have been made to address questionable data. We rapidly review the past work from bundling, gathering, and other order. For the gathering methods with vague data, they for the most part extend the principal bundling methodologies to adjust to data of shakiness. We rapidly overview some of them as takes after FDBSCA, made on DBSCAN probabilistically decides the faulty partitions between articles. OPTICS familiarizes a cushy partition work with measure the closeness between questionable data on top of the different leveled thickness based batching computation. UKmeans allots a challenge the gathering whose delegate has the most diminutive expected partition from the question.

Aggarwal uses approach based approach to manage handle screw up slanted and missing data. The system focuses the issue of packing vague articles whose ranges are delineated by probability thickness limits and uses Voronoi diagrams and R-Tree rundown to amass vague data. For the game plan techniques on unverifiable data, the standard combined SVM is contacted manage questionable data, which gives a geometric estimation by improving the probabilistic division between the two classes on both sides of the point of confinement. In like manner, Gao and Wang mine discriminative cases from questionable data and build up an unmistakable classifier for unverifiable data. Tsang et al. propose a movement of pruning techniques for decision tree to collect classifier for questionable data. Moreover, visit configuration mining on questionable data is investigated [13], in which the probability of a thing having a place with a particular trade is frequently illustrated. The work in [9] concentrates the disclosure of ceaseless cases and alliance rules from the probabilistic data under the possible world semantics. Besides, exemption acknowledgment with vague data has been analyzed in [6], which draws various cases from the data and procedures the division of the examples. The figurings for mining unverifiable chart data are moreover made [5]. Murthy et al. propose combination work in probabilistic databases while Yuen et al. use the nearest neighbor look for on flawed spatial databases.

The above systems on questionable data are made for static vague data. Since data is always gotten in stream condition, unverifiable data streams have been inspected. The technique for bundling questionable data streams has been discussed in [5], which wires bumble bits of knowledge and the little scale gathering thought into learning stage [3]. The similarity join taking care of is created on questionable data streams [5]. Besides, sub diagram configuration looks for over certain and questionable graph streams are furthermore made [12]. Despite much progress on unverifiable data mining, by far most of the past work has not unequivocally overseen one-class learning on flawed data. This paper proposes an uncertain one-class learning and thought summation learning structure to adjust to the data of powerlessness and the thought plot learning. In this framework, flawed one-class classifier builds up the standard one-class SVM for unverifiable data. Despite the way that UOCC is a support vector procedure, it is not exactly the same as uncertain twofold SVM [7], [8], [3]. At first, we propose a close-by piece thickness based technique to deliver a set out score toward every event. Second, we enhance our progression issue (8) into a standard QP (quadratic programming) streamlining issue (9) by considering the typical for one-class acknowledging, which has the refinement from questionable combined SVM [8]. Third, we put forth thought summation learning in the one-class data streams to pack the possibility of the customer.

3. BACKGROUND APPROACH

In one-class-based data streams, subject to testing oversights or contraption surrenders, the case might be spoiled and starting there is seen as questionable in its portrayal. Another recognition is that we may need to gather the thought buoy of a customer over the data streams. To deal with the one-class slanting and thought abstract learning on flawed data streams, we propose the uncertain one class learning and thought layout framework, as spoke to in Fig. 3..

UOLCS structure comprise of two sections, the initial segment is to develop dubiously one-class classifier from unverifiable information streams, the second part is idea outline learning over the history information streams. Two modules used in this scenario, they are 1) One Class Learning 2) Concept Summarization Learning.

3.1. One Class Learning

One class learning approach defines three main modules in developing application for uncertain data streams with feasible data streams. For generate threshold score for instance based with local behavior using local kernel density based for threshold generation in uncertain data streams. In second step, incorporate generated threshold score into learning phase to identify features instantly using uncertain one class classifier construction in uncertain data streams. After that classify classified features with relative data dimensionality based on uncertain one class classifier representation to extract data effectively from relative uncertain data sets.

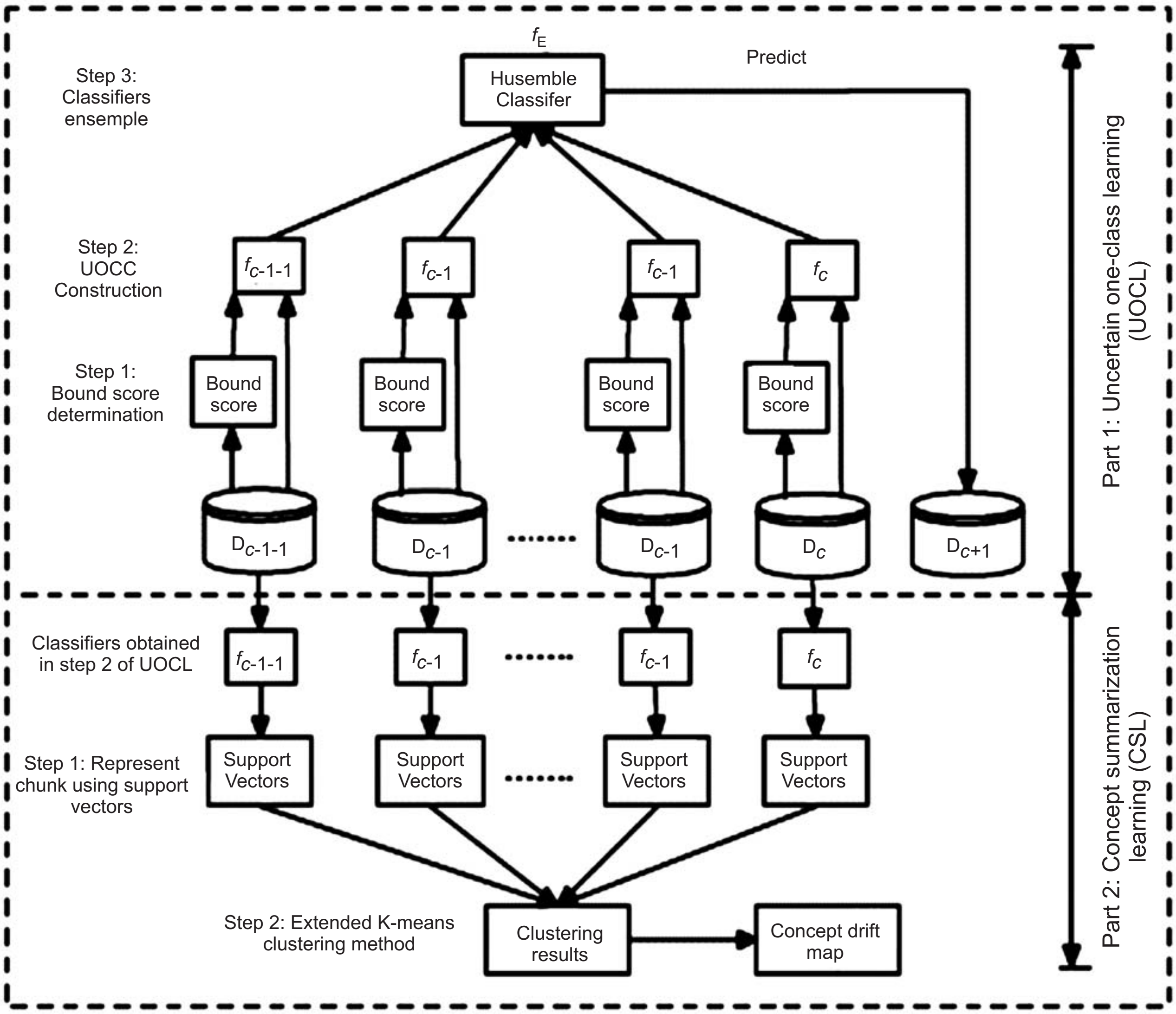

Figure 3: Concept summarization and one class learning in cluster data sets

3.2. Concept Based Summarization Learning

In information streams learning, it is important to know the ideas and their relations of the client from history pieces. In this area, we will advance our bolster vector based grouping strategy for idea synopsis gaining from information streams. Naturally, we can view the information streams in general and lead grouping calculations on the stream, and each bunch means one idea of the utilization. From that point forward, we can abridge the idea of the client by exploring which lumps have a similar idea of the client. Be that as it may, this may set aside an excess of time for learning in general information streams, and information stream learning is continually requiring ones scanning of the information streams without alluding to verifiable information. Another approach utilizes highlight based grouping method to condense idea of the client. It first extricates highlights from an information lump and considers this piece as a virtual specimen spoke to by the separated components, hence, the entire information streams is introduced by a virtual specimen set, in which each virtual example speaks to one information piece.

These two steps are used to define one class classification procedures for threshold score calculation and define summarization based on classification with processing instances. This procedure achieves one class classification based on instances only. So a better system is required for classify with preferable summarization attributes with characteristics with reliable uncertain data streams. So next section defines those relations with realistic summarization from real data sets.

4. 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.

4.1. Problem Formation and Common Framework

Let C = (c1; c2; ... ; cN) be a set of M details factors and α = (α1, α1, ..., αn) Mg be a team selection with N system bunching, each of which is referred to as an selection individual. Each platform clustering earnings a set of categories $\pi_i = \{C^i_1, C^i_2, C^i_3, \ldots C^i_n\}$, such that $\bigcup_{j=1}^{k_i} C^i_j = C$, where *ki* is the variety of groups in the i^{th} clustering. For each x 2 X, CðxÞ signifies the group brand to which the details factor x connected. In the i^{th} clustering C(c) = " j " (or "C^i_j")$if\ x \in C^i_j$. 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].

Generally, alternatives acquired from different platform clustering are aggregated to form any partition. This met level technique contains two major tasks of: 1) creating group selection, and 2) generating the final partition, normally known as contract function.

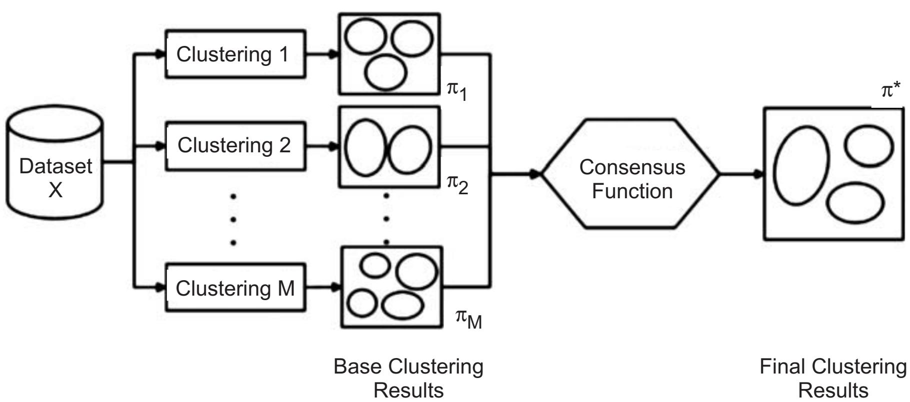

Figure 4: The main process of team clothing. It first is appropriate several platform clustering's to a information set X to obtain different clustering choices π1, π2, π3, ... πm. Then, these solutions are combined to set up the ultimate clustering result π* using a contract operate

4.2. Ensemble Creation Methods

It has been confirmed that clothing are most effective when made with a set of predictors whose errors are different. Particularly for details clustering, the effects acquired with any individual requirements over many editions are usually very similar. In such a situation where all selection members believe the fact on how a details set should be portioned, aggregating the system clustering results show no improvement over any of the element affiliates. Consequently, several heuristics have been recommended to present artificial instabilities in clustering techniques, offering wide range within a group selection. Some of the successive features were used for particular information clustering requirements.

4.3. Consensus Functions

Having acquired the group selection, a number of agreement features have been designed and made available for drawing the greatest information partition. Each agreement operate uses a specific form of details matrix, which summarizes the platform clustering results. In light of this credentials, agreement techniques can be categorized as follows:

Feature-based technique. It transforms the issue of group clothing to clustering particular details. Particularly, each system clustering provides a group brand as a new operate describing each details factor.

Direct Approach: It is depending on relabeling π_i and searching for the π^* that has the best coordinate with all $\pi 1$, $\pi 2$, $\pi 3$, ... πm Conceptually, the actual efficient process allows the homogeneous brands to be identified from heterogeneous clustering choices, where each system clustering provides a exclusive set of choice brands.

Pair wise-similarity strategy. It makes a matrix, containing the couple sensible likeness among information points, to which any likeness centered clustering criteria [8][9].

4.2. Cluster Outfits of Particular Data

While a vast number of team selection methods for mathematical information have been put forward in the past several years, there are only a few research that apply such a strategy to particular information clustering. The last clustering result is created using the graph-based agreement techniques.

Particular to this so-called "direct" selection development strategy, a given express information set can be revealed using a binary cluster-association matrix. Such an information matrix is related to the "market-basket" mathematical representation of particular information, which has been the focus of traditional particular information analysis

5. PERFORMANCE EVALUATION

This area provides the evaluation of the recommended CEA, using a number of reliability robots and real details places. The top quality of details groups created by this method analyzed against those designed by different particular details clustering methods and group collection techniques.

Table 1
Different types of information places relevant to draw out the finish process based on process of connection of each information point

Dataset	*N*	*D*	*A*	*K*
Zoo	102	19	48	18
Lymphography	153	35	73	20
Soybean	308	45	160	28
20 News Group	1001	6,085	12,157	3
KDDCup99	100,11	56	150	24

5.1. Datasets Retrieval

The test evaluation is conducted over nine information locations. The "20Newsgroup" details set is a section of the well known written text details collection—20-Newsgroups,2 while the others are obtained from the UCI Program Studying Database. Their details are described in Desk 1.

Missing concepts (denoted as "?") in these details locations are merely managed as a new particular value. The "20Newsgroup" information set contains 1,000 information from two newsgroups, each of which is described by the circumstances of 6,084 different circumstances. In particular, the regularity (*f* 2 *f* 0; 1; . . . ;1*g*) that a key phrase seems to be in each papers is customized into a cost-effective value: "Yes" if *f* > 0, "No" otherwise [3][5]. Moreover, the "KDDCup99" information set used in this evaluation is a randomly selected section of the exclusive details. Each details aspect (or record) suits to a process connection and contains 42 attributes: some are cost-effective and the rest are continuous.

6. EXPERIMENTAL RESULTS

In compliance with the course perfection, Table 2 examines the efficiency of different clustering methods over examined details locations [7]. Notice that the offered activities of group collection methods that apply the selection Type-II and Type-III are the income across 50 functions. Moreover, even is recognizable "N/A" when the clustering result is not accessible. For each details set, the greatest five CA-based principles are defined in boldface.

Table 2
Accuracy Results of traditional and proposed techniques

Data Set	*Uncertain One Class Classification*	*Cluster Ensemble Approach*
Accident	0.55	0.53
Diabetes	0.75	0.43
Economy Rating	0.33	0.27
Marks	0.02	0.003

The outcomes confirmed in this desk indicate that the CEA methods usually perform better than the examined assortment of team choice methods and clustering methods for particular details [12]. CEA is also appropriate to a large details set such as KDDCup99, for which several team choice methods (CO + SL, CO + AL, and CSPA) are negligible.

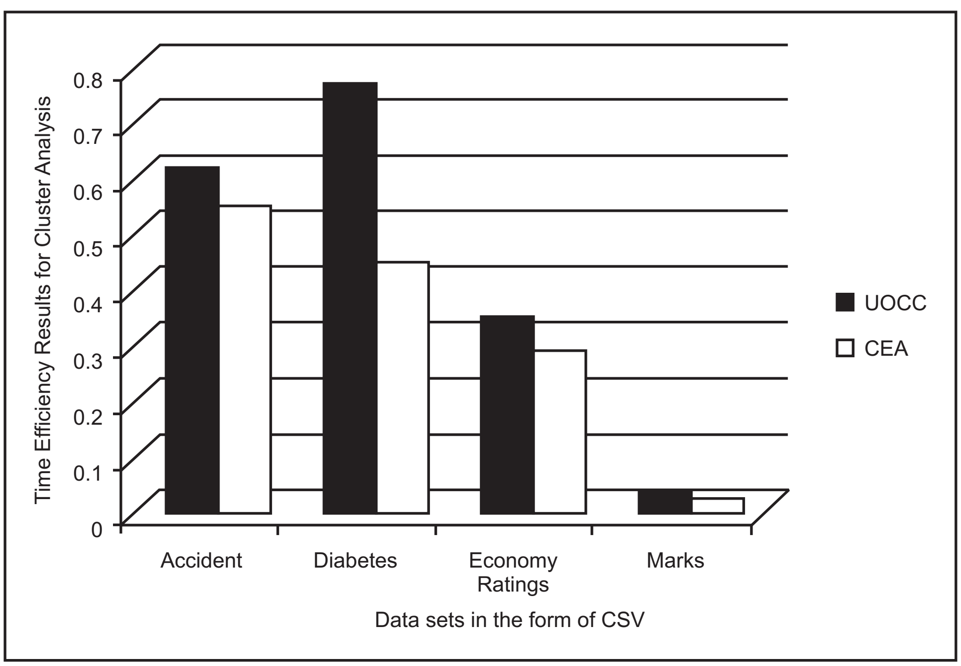

Figure 5: Experimental results comparison of both UOCC and CEA in real time data sets

With those things of CEA styles being mostly higher than those of the corresponding guide alternatives (Base), the high company's RM seems to be considerably better than that of the initial, binary distinction. The level of assurance of the express information as proven in below with element of the handling information points results in real-time information places as follows with contains durations $[X_{L_{C(i,\beta)}}, U_{L_{C(i,\beta)}}]$ for the mean as $L_{C(i,\beta)}$ with validity criteria C as follows:

$$X_{L_{C(i,\beta)}} = L_{X(i,\beta)} - 1.89\frac{S_{X(i,\beta)}}{\sqrt{n}} \quad (1)$$

$$U_{L_{C(i,\beta)}} = L_{X(i,\beta)} + 1.89\frac{S_{X(i,\beta)}}{\sqrt{n}} \quad (2)$$

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.

$$B_{C(i)} = \sum_{\forall\beta\in DT}\ \sum_{\forall i^*\in CN,\, i^*\neq i} \text{better}_X^{\beta}(i, i^*), \quad (3)$$

$$\text{Better}^{\beta}_{C}(i, i^*) = \begin{Bmatrix} 1 & \text{if } L_{XC(i,\beta)} > U_{XC(i^*,\beta)} \\ 0 & \text{otherwise} \end{Bmatrix} \quad (4)$$

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

$$W_{C(i)} = \sum_{\forall\beta\in DT}\ \sum_{\forall i^*\in CM,\, i^*\neq i} \text{worse}_C^{\beta}(i, i^*), \quad (5)$$

$$\text{Worse}^{\beta}_{C}(i, i^*) = \begin{Bmatrix} 1 & \text{if } U_{XC(i^*,\beta)} < L_{XC(i,\beta)} \\ 0 & \text{otherwise} \end{Bmatrix} \quad (6)$$

Table 3
For each assessment catalog, "B" and "W" signify how frequently that a particular technique works considerably "better" and "worse" than the others

Ensemble Type	*Method*	*CA*		*NMI*		*AR*	
		B	*W*	*B*	*W*	*B*	*W*
1.	CEA	171	35	141	70	151	61
	UOCC	137	78	134	70	142	73
2. Fixed	CEA	218	8	212	45	204	18
	UOCC	138	64	141	31	116	82
3. Random	CEA	219	35	209	15	208	34
	UOCC	122	52	209	12	208	15

Using these assessment formalism, Desk 3 shows for each method the wavelengths of important better (B) and important more intense (W) efficiency, which are sorted based on the assessment spiders [19][20]. The efficiency of the both FIS and CEA shown in Desk 3, CEA give best efficiency than previously strategy provided in information process immediately database integration. Despite the point that many clustering methods and CEA are designed with the capability of evaluating function concepts in mind, they achieve the recommended measurement in a different way, using particular information styles [16]. CEA exclusively and clearly styles the natural issue as the assessment of link-based similarity among graph vertices, which take a position for particular function concepts or created groups.

Furthermore, CEA works constantly better than its competitors with all different selection measurements, while CO + SL appear to be the least effective. Realize that a bigger selection outcomes in an enhanced perfection, but with the trade-off of runtime.

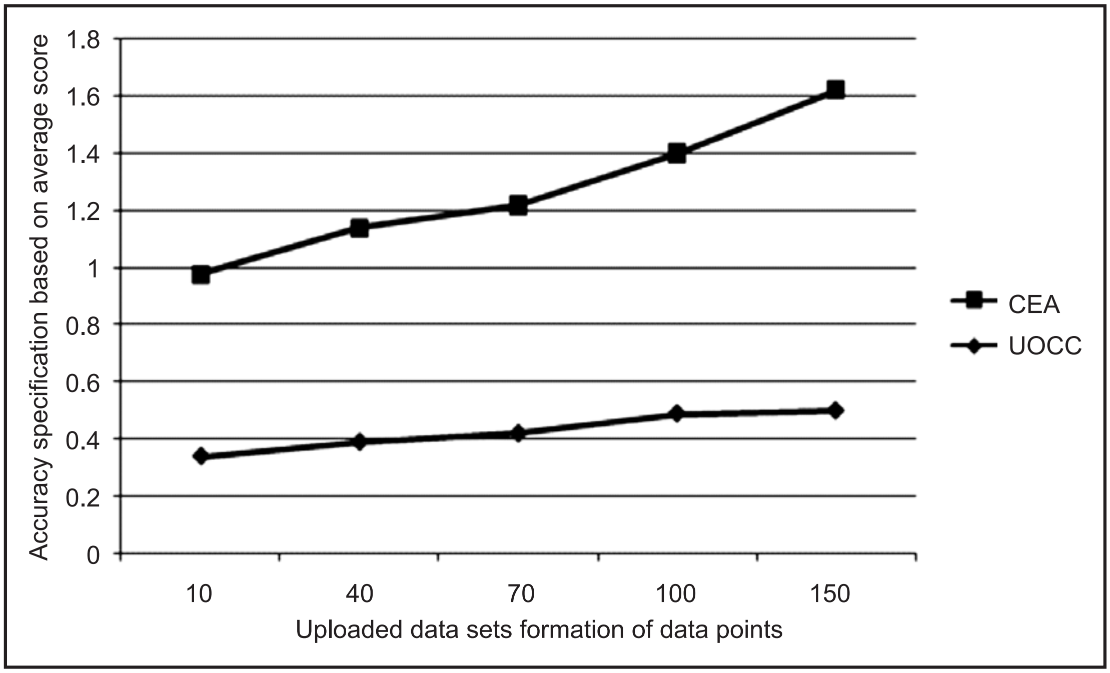

Fig 6: Accuracy measure based on generated data point processing real time datasets using two relational approaches

The results shown in this table indicate the excellent efficiency of the suggested link-based methods, as compared to other clustering techniques included in this research. To start with, we make the limited score to capture the area uncertainty in light of every illustration's close by information perform, and after that produce a doubtful one-class classifier by combining the uncertainty data into a CEA with SVM-based studying framework. Second, we make enhance vectors-based collection technique to summarize the understanding of the consumer over the history pieces. Wide assessments have revealed that our unverifiable one category studying can get an excellent performance and is less sensitive to fuss in connection with the standard one-class SVM. The assessments additionally illustrate that the support vectors-based collection technique can well reduce the understanding of the consumer in connection with emphasize centered collection way of concept summary learning.

7. CONCLUSION

This paper shows a novel, significantly practical relationship centered collection group way to deal with all-out details bunching. It changes the first unmitigated details lattice to a data defending mathematical variety, to which a highly effective plan apportioning system can be straight, linked. The issue of creating the RM is efficiently resolved by the similitude among unmitigated represents (or clusters), utilizing the Calculated Triple-Quality similitude computation. The observational evaluation, with various clothing types, authenticity actions, and informative selections, suggests that the suggested interface centered strategy, for the most part, achieves unmatched collection comes about compared with those of the traditional unmitigated details computations and standard collection clothing systems

REFERENCES

[1] Bo Liu, Yanshan Xiao, Philip S. Yu, "Uncertain One-Class Learning and Concept Summarization Learning on Uncertain Data Streams", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 2, FEBRUARY 2014.

[2] C.C. Aggarwal, Y. Xie, and P.S. Yu, "On Dynamic Data-Driven Selection of Sensor Streams," Proc. 17th ACM SIGKDD Int'l Conf.Knowledge Discovery and Data Mining (KDD), pp. 1226-1234, 2011.

[3] C.C. Aggarwal and P.S. Yu, "A Survey of Uncertain Data Algorithms and Applications," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 5, pp. 609-623, May. 2009.

[4] F. Bonchi, M.V. Leeuwen, and A. Ukkonen, "Characterizing Uncertain Data Using Compression," Proc. SIAM Conf. DataMining, pp. 534-545, 2011.

[5] F. Bovoloa, G. Camps-Vallsb, and L. Bruzzonea, "A Support Vector Domain Method for Change Detection in Multitemporal Images," Pattern Recognition Letters, vol. 31, no. 10, pp. 1148-1154, 2010.

[6] L. Chen and C. Wang, "Continuous Subgraph Pattern Search over Certain and Uncertain Graph Streams," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 8, pp. 1093-1109, Aug. 2010.

[7] B. Liu, Y. Xiao, L. Cao, and P.S. Yu, "Vote-Based LELC for Positive and Unlabeled Textual Data Streams," Proc. IEEE Int'l Conf. Data Mining Workshops (ICDM), pp. 951-958, 2010.

[8] R. Murthy, R. Ikeda, and J. Widom, "Making Aggregation Work in Uncertain and Probabilistic Databases," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 8, pp. 1261-1273, Aug. 2011.

[9] L. Sun, R. Cheng, D.W. Cheung, and J. Cheng, "Mining Uncertain Data with Probabilistic Guarantees," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 273-282, 2010.

[10] M. Takruri, S. Rajasegarar, S. Challa, C. Leckie, and M. Palaniswami, "Spatio-Temporal Modelling-Based Drift-Aware Wireless Sensor Networks," Wireless Sensor Systems, vol. 1, no. 2, pp. 110-122, 2011.

[11] S. Tsang, B. Kao, K.Y. Yip, W.S. Ho, and S.D. Lee, "Decision Trees for Uncertain Data," IEEE Trans. Knowledge and Data Eng., vol. 23, no. 1, pp. 64-78, Jan. 2011.

[12] L. Trung, T. Dat, N. Phuoc, M. Wanli, and D. Sharma, "Multiple Distribution Data Description Learning Method for Novelty Detection," Proc. Int'l Joint Conf. Neural Networks (IJCNN), pp. 2321-2326, 2011.

[13] S.M. Yuen, Y. Tao, X. Xiao, J. Pei, and D. Zhang, "Superseding Nearest Neighbor Search on Uncertain Spatial Databases," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 7, pp. 1041-1055, July 2010.

[14] X. Zhu, W. Ding, P.S. Yu, and C. Zhang, "One-Class Learning and Concept Summarization for Data Streams," Knowledge and Information Systems, vol. 28, no. 3, pp. 523-553, 2011.

[15] Z. Zou, H. Gao, and J. Li, "Discovering Frequent Subgraphs over Uncertain Graph Databases under Probabilistic Semantics," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 633-642, 2010.

[16] Natthakan Iam-On, Tossapon Boongoen, Simon Garrett, and Chris Price, "Link-Based Cluster Ensemble Approach for Categorical Data Clustering", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 3, MARCH 2012.

[17] T. Boongoen, Q. Shen, and C. Price, "Disclosing False Identity through Hybrid Link Analysis," Artificial Intelligence and Law, vol. 18, no. 1, pp. 77-102, 2010.