VOL. 14, NO. 12, JUNE 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A NOVEL INDEX BASED PROCEDURE TO EXPLORE SIMILAR ATTRIBUTE SIMILARITY IN UNCERTAIN CATEGORICAL DATA

Srinivas Kolli and M. Sreedevi
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India
E-Mail: kollisreenivas@gmail.com

ABSTRACT

In knowledge discovery and data mining, clustering is an aggressive concept to explore different attributes with different relations, because each data type has its own and unique challenge to achieve relative data based on partitioning of homogeneous data. In knowledge discovery categorical data clustering is an essential and challenging task because of special characteristics. So to arrange attributes in systematic manner for uncertain categorical data indexing approach is required. In this paper we propose and introduce A Novel Fuzzy based Partitioned Genetic Algorithm (NFPGA) for uncertain categorical data. This novel approach consists two phases to explore and process categorical data. In first stage partition data set with maximum number of clusters then combine all the clusters generated in first phase. This procedure repeated until number of clusters equal to pre-defined clusters present in data set. This proposed approach i.e. NFPGA is implemented on synthetic data sets which are available UCI repository, novel fitness function; cross-over and mutation operations are evaluated on categorical data based on parallel partitioning procedure. Performance of proposed approach has been crossover with different existing clustering related approaches with objective functionalities and similarity index measures, from this proposed approach gives better and excellence performance.

Index terms: fuzzy clustering, genetic algorithm, cross-over, mutation and fitness function, categorical data, partitioning data, similarity index and feature relation.

1. INTRODUCTION

Group examination is one of the "super problems" in information mining. As a rule, grouping is dividing information focuses into instinctively comparable gatherings (Saxena *et al.* 2017). This definition is straightforward and does not consider the difficulties that happen while applying bunch investigation to genuine world datasets. By and by, this sort of examination is basic in various fields, e.g. content mining, promoting research, client conduct investigation, monetary market investigation. In these days, different grouping calculations have been produced in the writing. Every one of them has its points of interest and drawbacks. In addition, as the information come in various structures, for example content, numeric, all out, picture, the calculations perform diversely in various situations. At the end of the day, the execution of a specific grouping calculation relies upon the structure of the information under thought.

Bunch investigation of numeric information is moderately all around concentrated in the writing. Different methodologies are executed for example, agent based, progressive, thickness based, chart based, display based, network based (Sajana *et al.* 2016). Of late, expanding consideration has been paid to apportioning non-numeric sorts of information. A vital point is the grouping of straight out information. The issue is that the most widely recognized grouping calculations for absolute information are alterations of the ones presented for numeric information. For example, K-modes (Huang 1997) is a model of the K-implies (MacQueen 1967) calculation. Be that as it may, a few specialists have created calculations explicitly for all out information (for example Nguyen and Kuo 2019, Chen *et al.* 2016, Yanto *et al.* 2016); however there is still much space for new methodologies. The primary issue in apportioning clear cut information is that the usage of the standard tasks utilized in grouping calculations has a few constraints. For example, the meaning of separation between two items with absolute highlights isn't as direct similarly as with numeric highlights, in light of the fact that all out information takes as it were discrete qualities which don't have any request, in contrast to numeric information. The most straightforward arrangement is to change the unmitigated information into twofold information and afterward apply one of the normal grouping calculations. Among the diverse strategies connected to fuzzy grouping that can be found in the writing, we center around those based on the fuzzy C-implies (FCM) calculation [13], portion techniques [14, 15], factual strategies [16], clonal determination hypothesis [17], rule-based grouping [18-20], and a wide range of heuristic and meta heuristic methodologies [11- 15]. Meta-heuristic calculations have been completely connected to fuzzy grouping in the most recent years because of their better properties of vigor and combination than close ideal arrangements at a moderate.

A considerable lot of these methodologies depend on transformative variations of the C-implies calculation [16, 17] or just on direct fuzzy grouping calculations dependent on hereditary and developmental methodologies [18-19], multi-target calculations [12], differential advancement [13], molecule swarm meta-heuristics [14], or transformative programming approaches [15]. In any case, in spite of the examination movement contributed on various meta-heuristic methodologies connected to fuzzy grouping, a few cutting edge calculations have not been investigated yet completely for fuzzy grouping issue. In particular, this paper proposes and introduces A Novel Fuzzy based Partitioned Genetic Algorithm (NFPGA) for

uncertain categorical data. The Partitioned Genetic Algorithm (PGA) [16, 17] is a class of transformative calculations whose encoding method is particularly intended to manage gathering issues. It has been effectively connected to an assortment of issues including gathering of things at the same time, shockingly, its execution has not been evaluated yet while handling fuzzy grouping issues. For this reason, this paper expands upon starter work in [18] by showing a novel gathering encoding, an altered target capacity, and hybrid and transformation administrators explicitly adjusted to fuzzy grouping issues handled by means of PGA heuristics. So as to additionally improve the execution of the gathering hereditary methodology the proposed plan too joins a nearby hunt organize and a parallelization of the PGA utilizing the outstanding island display, which can be both considered as extra novel fixings as for [8]. Reproduction results are displayed to evaluate the execution of the proposed plan in various application situations, in light of which it is closed that the PGA-based strategy here exhibited outflanks customary fuzzy C-implies strategies.

2. REVIEW OF RELATED WORK

An extensive segment of the standard gathering computations are expected to base either on numeric data or on supreme data. The assembled data in evident as often as possible contain both numeric and hard and fast properties. It is troublesome for applying customary gathering estimation direct into these sorts of data. Regularly, when people need to apply standard partition based grouping counts to store up these sorts of data, a numeric regard will be consigned to each class in this characteristics. A couple of out and out characteristics, for example "low", "medium" and "high", can without a lot of a stretch be transformed into numeric characteristics. Nevertheless, if obvious properties contain the qualities like "red", "white" also, "blue" ... etc., it can't be asked for ordinarily. In light of the refinements in their features, with an explicit ultimate objective to total these fluctuating data, it respects misuse the gathering system which uses split and association approach to manage clarify this issue. For clustering mixed compose properties in [1] Ming-Yi Shih displayed another two-advance grouping method is acquainted with find clusters on Blended Categorical and Numeric Data. Things in full scale credits are taken care of to construct the likeness or associations among them in perspective of the considerations of co-occasion; by then every out and out quality can be changed over into numeric properties in light of these manufactured associations. Finally, since each straight out datum are changed over into numeric, the present gathering estimations can be associated with the dataset without torment.

Jongwoo Lim *et. al.* [2] proposed a gathering structure that support clustering of datasets with mixed attribute make (numerical, straight out), while in the meantime restricting information mishap in the midst of clustering. They at first utilize an entropy based proportion of outright characteristics when in doubt work for closeness. Second, in light of the eventual outcomes of entropy based resemblance, they separate candidate gather numbers and affirm their weighting plan with pre-gathering occurs. Finally, they pack the mixed trademark form datasets with the expelled contender aggregate numbers and the loads. Zhexue huang [6] showed a k models figuring which relies upon the k-infers perspective yet clears the numeric data limitation while protecting its viability. In the figuring, objects are grouped against k models. A method is created to intensely invigorate the k models remembering the true objective to support the intra assemble closeness of things. Exactly when associated with numeric data the computation is undefined to the k-implies. To help interpretation of gatherings we use decision tree enrollment figurings to make rules for packs.

These standards, together with various estimations about cluster, can help data excavators to appreciate and recognize interesting gatherings. Jamil Al-Shaqsi and Wenjia Wang [7] show a grouping gathering system in perspective of a novel three-masterminded gathering count. Their social occasion is produced with a proposed grouping estimation as a middle showing technique that is used to deliver a movement of gathering comes to fruition with different conditions for a given dataset. By then, a decision combination instrument, for instance, casting a ballot is used to find a joined package of the particular gatherings. The casting a ballot framework pondered in a manner of speaking test comes about that convey intra-resemblance regard higher than the ordinary intra- closeness regard for a particular between times. The purpose of this strategy is to find a gathering result that constrains the amount of contrasts between different groupings comes to fruition.

3. NOVEL INDEX BASED CLUSTERING APPROACH DESIGN AND IMPLEMENTATION

As discussed in section 1 and 2, partition based genetic algorithm is evaluated on novel encoding strategy especially to define grouping of designed clustering problems. It was first proposed by Falkenauer *et al* [16], who explored that conventional hereditary calculations experienced issues when connected to gathering issues. In PGAthe encoding technique and hybrid and change administrators of conventional GAs are altered to yield a minimal algorithm, with enhanced execution in grouping based issues. In light of their beating conduct concerning its conventional partners, gathering hereditary calculations have so far been effectively connected to assorted issues, including fresh bunching. This paper joins the upsurge of research floating on PGAs by adjusting this heuristic to fuzzy bunching issues. This segment examines a few alterations we have conceived towards further upgrading the execution of GGAs in fuzzy grouping, counting our adjustments in the encoding procedure, the target work, and the cross-over and mutation operators.

A. Encoding Procedure: The proposed PGA for fuzzy bunching is a variable-length hereditary calculation, with a novel encoding to manage this particular problem. The encoding is done by part every chromosome in the calculation (or then again comparably, its relating

individual or applicant arrangement) into two sections: c = [U | g]. The initial segment is the component area formed by the segment network U, while the second part is indicated as the gathering area of the person. Following this documentation, someone in particular for a fuzzy bunching issue with N components (objects or perceptions) and k bunches can be communicated as:

$$\begin{matrix} u_{11}, u_{12}, \ldots\ldots u_{1n} & \\ \cdot & \mid g_1, g_1, \ldots\ldots, g_k \\ \cdot & \\ u_{k1}, u_{k2}, \ldots\ldots u_{kn} & \end{matrix}$$

Where it is worth noting that each factor $u\ i$, represents the level of account of jth statement to i^{th} group, whereas the team area keeps a record of labels associated with each of the groups of the perfect remedy is. Also realize that in this encoding; both the team and the factor area have a variable duration, since the amount of groups is also a variable of the issue. For the benefit of quality, let us believe the following individual:

$$\left[\begin{matrix} 0.6 & 0.0 & 0.0 & 0.8 & 0.0 & 1.0 & 0.6 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.4 & 1.0 \\ 0.0 & 0.0 & 1.0 & 0.2 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.2 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.0 & 0.9 & 0.0 & 0.0 & 0.8 & 0.0 & 0.4 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.3 & 0.0 & 0.1 & 0.0 & 0.0 & 0.2 & 1.0 & 0.0 & 0.0 \end{matrix} \mid 1,2,3,4 \right]$$

This chromosome encodes an individual (competitor arrangement) for a basic bunching issue with N = 15 objects: X ={x1, . . . , x15}. Note that the gathering area encodes an answer with 4 bunches, marked "1," "2," "3," and "4," separately. Any of the segments in the component segment demonstrates to what degree any item xj has a place with a bunch Ci, that is, the partition network component $u\ ij$. For example, the primary segment in the component area encodes an applicant fuzzy arrangement in which the article x1 has a place with bunch C 1 with a level of participation u 1, 1 = 0.6 and has a place with C 4 with u 4,1 = 0.4. Remembering this, the previously mentioned chromosome encodes a person that speaks to an answer with 4 bunches, where perceptions x 2, x3, x5, x6, x8, x10, x11, x13, and x15 have a place with a solitary group, perceptions x1, x4, x9, and x12 have a place with two unique groups with various degrees of enrollment, lastly perceptions x7 and x14 have a place with three unique groups.

B. Objective Index Procedure: The proposed GGA will be kept running with various target (wellness) capacities to lead the inquiry. In particular, and for near purposes, we will utilize a portion of the established target capacities for fuzzy grouping. In this paper we propose an adjustment of the outstanding Davis-Bouldin file (utilized in fresh grouping issues) to the fuzzy case which, to the best of our insight, is novel in fuzzy bunching. We will demonstrate that the utilization of this adjusted file renders better outcomes for the PGA than the other existing assessment records. The possibility of the Davis-Bouldin file for fresh bunching issues is to limit the intra-cluster separations while at the same time expanding the separations among the distinctive groups, yielding:

$$\mathrm{DB}(\mathrm{U}) = \frac{1}{k}\sum_{i=1}^{k}\max_{i\neq j}\left\{\frac{\sum_{x\in C_i} d^2(\mathrm{x},\mu_i) + \sum_{x\in C_j} d^2(\mathrm{x},\mu_j)}{d^2(\mu_i,\mu_j)}\right\}$$

In the above articulation take note of that little estimations of the traditional DB record relate to reduced and very much isolated clusters. The adjustment of the DB list for fuzzy bunching proposed in this work is communicated as:

$$\mathrm{MDB}(\mathrm{U},d) = \frac{1}{k}\sum_{i=1}^{k}\max_{i\neq j}\left\{\frac{\sum_{t=1}^{N} u_{i,t}^{\alpha} d^2(\mathrm{x}_t,\mu_i) + \sum_{t=1}^{N} u_{j,t}^{\alpha} d^2(\mathrm{x}_t,\mu_j)}{d^2(\mu_i,\mu_j)}\right\},$$

where $\mu\ i$ represents the centroid related with bunch $C\ i$, determined by considering the normal of every perception weighted by the level of enrolment to group $C\ i$. Note that the proposed MDB record expressly relies upon the specific definition considered for the separation d. For instance, on the off chance that we consider the GK remove and dependent on the covariance frameworks of the bunches, the DB record for fuzzy grouping issues will be given by:

$$\mathrm{MDB}(\mathrm{U},d_{\mathrm{GK}}) = \frac{1}{k}\sum_{i=1}^{k}\max_{i\neq j}\left\{\frac{\sum_{t=1}^{N} u_{i,t}^{\alpha} d^2_{|\Sigma_i|^{1/d}\Sigma_i^{-1}}(\mathrm{x}_t,\mu_i) + \sum_{t=1}^{N} u_{j,t}^{\alpha} d^2_{|\Sigma_j|^{1/d}\Sigma_j^{-1}}(\mathrm{x}_t,\mu_j)}{\min\left\{d^2_{|\Sigma_i|^{1/d}\Sigma_i^{-1}}(\mu_i,\mu_j), d^2_{|\Sigma_j|^{1/d}\Sigma_j^{-1}}(\mu_i,\mu_j)\right\}}\right\}.$$

C. Cross-over & Mutation Operators: The cross-over and mutation implemented in the collection inherited criteria used in this paper is an improved edition of the one originally suggested by Falkenauer *et al*, but with the added reward of being adapted to the unclear clustering problem. These are the main steps followed in the cross-over function. (1) Select two people randomly and choose two traversing points in their group part. (2) Place the sun and rain that belong to the chosen categories of the first personal into the children. (3) Allocate the quality of account of the placed components similar to the first personal. (4) Place the sun and rain that belong to the chosen categories of the second personal into the children. (5) Allocate the quality of account of the placed components in the following way. First, the staying level account after a job of the aspects of the first personal is measured. This staying level account is then proportionally shared among the aspects of the second personal. (6) Eliminate vacant groups, if any. (7) Change the brands of the current categories in the children in order to numerate them from 1 to k. A simple yet illustrative enough example follows. Let us consider two different

people $\xi 1$ that have been randomly chosen among everyone in a given PGA population so as to perform cross-over on them. The categories chosen to carry out the procedure are noticeable in boldface:

$$\xi_1 = \begin{bmatrix} 0.6 & 0.0 & 0.0 & 0.8 & 0.0 & 1.0 & 0.6 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.4 & 1.0 \\ 0.0 & 0.0 & 1.0 & 0.2 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.2 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.0 & 0.9 & 0.0 & 0.0 & 0.8 & 0.0 & 0.4 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.3 & 0.0 & 0.1 & 0.0 & 0.0 & 0.2 & 1.0 & 0.0 & 0.0 \end{bmatrix} \Big| 1,2,3,4$$

Cluster formation after arrange different data elements with respect to different attribute relations:

$$O = \begin{bmatrix} 0.0 & 0.0 & 1.0 & 0.2 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.2 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.0 & 0.9 & 0.0 & 0.0 & 0.8 & 0.0 & 0.4 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.9 & 0.0 & 0.0 & 0.0 & 0.0 & 0.12 & 1.0 & 0.35 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.8 & 1.0 & 0.5 & 0.0 & 0.0 & 0.1 & 0.0 & 1.0 & 0.08 & 0.0 & 0.05 & 1.0 \end{bmatrix} \Big| 1,2,3,4$$

Based on cluster index representation of different notation shown in above equations cluster representation as follows:

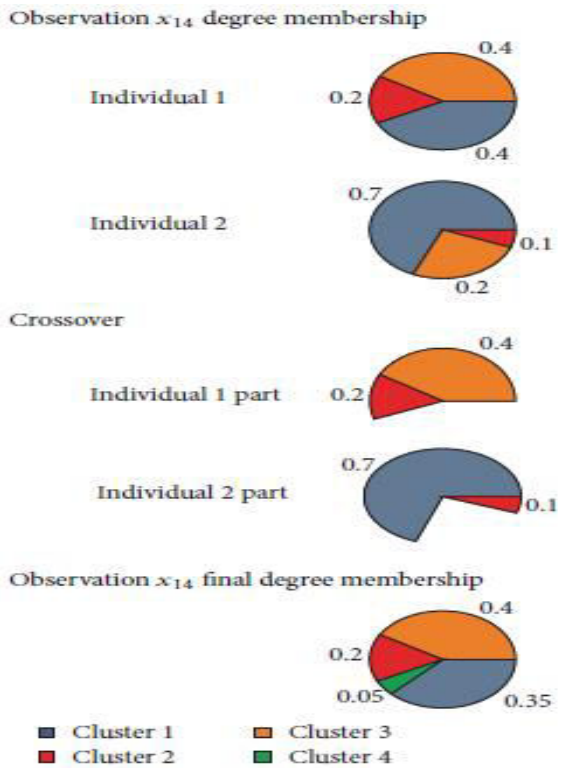

Figure-1. Cluster representation in proposed approach based on cross-over and mutation operations.

D. Search Relevant Items: We utilize a neighborhood look strategy to attempt to discover nearby optima in a nearby neighborhood of guaranteed person. The proposed neighborhood seeks depends on minor changes of the present individual, to the extent they deliver an expansion of the related target work: the neighborhood seek changes the degree of membership of the perceptions, beginning by one arbitrarily chosen. The changes in the degree of enrolment are haphazardly created. We at long last keep the task with the biggest target work. Since this nearby inquiry methodology is a tedious activity, it is connected to a given individual with a little likelihood, p , that is changed between an underlying and last an incentive in the calculation similarly that the hybrid likelihood is changed.

4. EXPERIMENTAL SETUP

This section condenses and discusses about the trial work we have done so as to survey the execution of our proposed PGA approach. We have investigated a number of varieties of the proposed PGA (by joining unique separations and additionally target capacities) in an assortment of fuzzy grouping situations (which, as will be appeared, show an expanding level of unpredictability).

Table-1. Parameter description users in experimental setup.

Parameter	Meaning	Value
Ps	Population size	20
S	Number of subpopulations	4
TG	Maximum number of generations	400
P_{ci}	Initial crossover probability	0.8
P_{cf}	Final crossover probability	0.6
P_{mi}	Initial mutation probability	0.05
P_{mf}	Final crossover probability	0.1
P_{bi}	Initial local search probability	0.1
P_{bf}	Final local search probability	0.05
P_e	Probability of migrating (islands model)	0.03
α	Fuzziness degree	2

Table-1 lists the estimations of the PGA parameters utilized in every one of the re-enactments completed in this paper. These qualities have been observed to be the most suitable after various side investigations, not appeared the purpose of brevity. The algorithm presented here is looked at with the fuzzy C-implies (FCM) [3] algorithm because it has been effectively connected to numerous genuine bunching issues what's more, applications described by various dimensions of multifaceted nature.

Table-2. Highest index representation for different approaches.

Algorithm	Number of clusters	Rand index
Proposed GGA (MDB index)	8	**0.9937**
Proposed GGA (XB index)	8	0.9805
Proposed GGA (FS index)	9	0.9874
GGA from [8] (MDB index)	8	0.9918
GGA from [8] (XB index)	8	0.9785
GGA from [8] (FS index)	9	0.9847
FCM	8	0.9712

We have connected to this problem a number of designs of the proposed PGA-with MDB, XB, and FS objective (wellness) capacities-and the FCM algorithm with the genuine number of groups as from the earlier data. Now it is essential to underscore that the proposed GGA is capable to gather the quantity of bunches inside the issue, while the FCM requires this parameter to be set before execution (to be specific, C in the above depiction of FCM). To evade this issue, side reproductions have been kept running for FCM and the considered situation by shifting C over a wide scope of whole number values, from which the esteem rendering the best measurement esteem has been chosen for correlation. Additionally included is the PGA come closer from [8] so as to survey the effect of the novel parts of the island-based GGA proposed here.

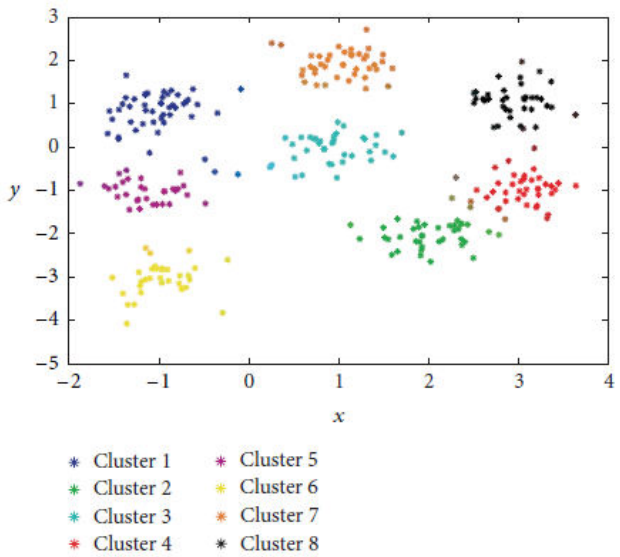

Figure-2. Best result of proposed approach with different cluster formations.

Having said this, Table-2 shows the directed assessment of the outcomes acquired by the previously mentioned calculations. Note that the proposed PGA with the three diverse target capacities gets preferred outcomes over the FCM calculation. In specific, our GGA with the MDB list displays the best conduct (R = 0.9937), higher than that of the customary FCM calculation (R = 0.9712) and the GGA with MDB list from [8] (R = 0.9918). What's more, take note of that the PGA with MDB and XB lists accomplishes the arrangement with the ideal number of groups (i.e., 8). There was different techniques available in traditionally to support efficient clustering based on

Indexing of different attributes with respect to precision, recall and accuracy with existing approaches (shown in following Tables).

Techniques Comparison Table for Different Approaches

Table-3. Accuracy values comparison formation with different datasets.

Accuracy				
Datasets	**NFPGA**	**K-Means**	**ACO**	**Fuzzy C-means**
Dataset 1	0.84	0.712	0.689	0.696
Dataset 2	0.736	0.51	0.708	0.562
Dataset 3	0.746	0.764	0.576	0.415
Dataset 4	0.832	0.604	0.484	0.423
Dataset 5	0.832	0.470	0.508	0.371

Table-4. Precision presentation values for different datasets.

Precision				
Datasets	**NFPGA**	**K-Means**	**ACO**	**Fuzzy C-means**
Dataset 1	0.56	0.342	0.4245	0.352
Dataset 2	0.42	0.301	0.415	0.401
Dataset 3	0.472	0.399	0.28	0.338
Dataset 4	0.445	0.444	0.392	0.345
Dataset 5	0.472	0.333	0.335	0.345

Table-5. Recall values with different data sets.

Recall				
Datasets	**NFPGA**	**K-Means**	**ACO**	**Fuzzy C-means**
Dataset 1	0.61	0.412	0.501	0.427
Dataset 2	0.552	0.346	0.496	0.468
Dataset 3	0.598	0.486	0.346	0.367
Dataset 4	0.61	0.462	0.454	0.329
Dataset 5	0.61	0.376	0.396	0.356

So as to more readily depict the conduct of the best calculation (the PGA with our MBD list), it would be exceptionally fascinating to have a more intensive take a gander at Figures-2.

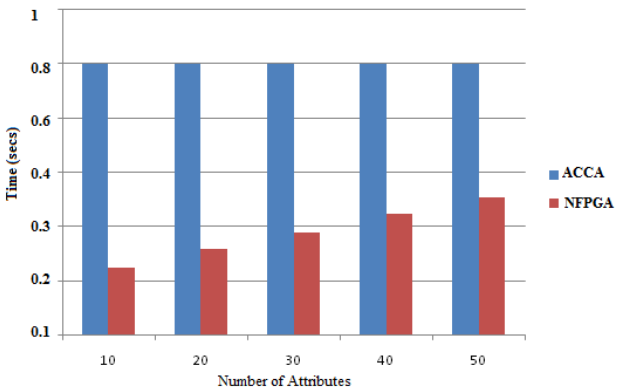

Figure-3. Time for indexing attributes with different data sets.

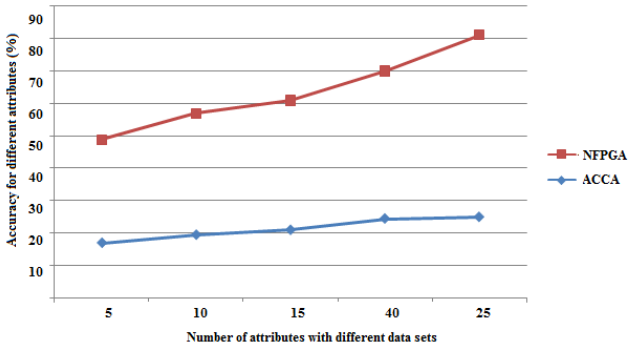

Figure-4. Accuracy of proposed approach with traditional approach.

Based on results present in Figures 2-4, proposed approach gives better results with respect to cluster formation based on indexing with different attribute relations for different synthetic data sets.

5. CONCLUSIONS

In this paper, we present a Novel Fuzzy based Partitioned Genetic Algorithm (NFPGA) for uncertain categorical data and fuzzy based index cluster problems. In this approach, mainly presents three steps 1. Novel encoding approach for individual attributes evaluation process to partition values with elements which can form by clustering. 2. Novel fitness functions for evaluating index (Davis-Bouldin index) for calculation of efficiency of clustering problems which enables arrangement of attribute values in semantic manner and search local elements which are used to improve performance of proposed approach. Experimental results of proposed approach describe efficient clustering results with respect to fitness, index and cluster formation with comparison of existing approaches. Further improvement of this approach is to enable services relevant multi attribute similarity with respect to multi attribute representation.

REFERENCES

[1] Srinivas Kolli, M. Sreedevi. 2018. Adaptive Clustering Approach to Handle Multi Similarity Index for Uncertain Categorical Data Streams. Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 04-Special Issue, 2018.

[2] Natthakan Iam-On, Tossapon Boongoen, Simon Garrett and Chris Price. 2012. A Link-Based Cluster Ensemble Approach for Categorical Data Clustering. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING. 24(3).

[3] D. Cristofor and D. Simovici. 2002. Finding Median Partitions Using Information-Theoretical-Based Genetic Algorithms. J. Universal Computer Science. 8(2): 153-172.

[4] S. Salcedo-Sanz, 1 J. Del Ser, 2 and Z.W. Geem. 2014. An Island Grouping Genetic Algorithm for Fuzzy Partitioning Problems. Hindawi Publishing Corporation the Scientific World Journal Volume 2014, Article ID 916371, 15 pages http://dx.doi.org/10.1155/2014/916371.

[5] J. Ji, W. Pang, C. Zhou, X. Han and Z. Wang. 2012. A fuzzy kprototypeclustering algorithm for mixed numeric and categorical data. Knowledge-Based Systems. 30: 129-135.

[6] S.Mitra and H. Banka. 2006. Multi-objective evolutionary biclustering of gene expression data. Pattern Recognition. 39(12): 2464-2477.

[7] P. Scheunders. 1997. A genetic C-means clustering algorithm applied to color image quantization. Pattern Recognition. 30(6): 859-866.

[8] V. M. Gomez-Mu˜noz and M. A. Porta-G´andara. 2002. Local wind patterns for modeling renewable energy systems by means of cluster analysis techniques. Renewable Energy. 25(2): 171-182.

[9] X. Guo, Z. Zhu and J. Shi. 2012. A corporate credit rating model using support vector domain combined with fuzzy clustering algorithm. Mathematical Problems in Engineering. 2012: 20.

[10] D. Graves and W. Pedrycz. 2010. Kernel-based fuzzy clustering and fuzzy clustering: a comparative experimental study. Fuzzy Sets and Systems. 161(4): 522-543.

[11] Y. Zhong and L. Zhang. 2011. A new fuzzy clustering algorithm based on clonal selection for land cover classification. Mathematical Problems in Engineering. 2011: 21.

[12] S. Salcedo-Sanz, L. Carro-Calvo, J. A. Portilla-Figueras, L. Cuadra and D. Camacho. 2013. Fuzzy clustering with grouping genetic algorithms. in Intelligent Data Engineering and Automated Learning, vol. 8206 of Lecture Notes in Computer Science, pp. 334-341, Springer, Berlin, Germany.

[13] L. E. Agustn-Blas, S. Salcedo-Sanz, S. Jim´enez-Fern´andez, L. Carro-Calvo, J. del Ser and J. A. Portilla-Figueras. 2012. A new grouping genetic algorithm for clustering problems. Expert Systems with Applications. 39(10): 9695-9703.

[14] J. Yu and M.-S. Yang. 2007. A generalized fuzzy clustering regularization model with optimality tests and model complexity analysis. IEEE Transactions on Fuzzy Systems. 15(5): 904-915.

[15] W. Yaonan, L. Chunsheng and Z. Yi. 2009. A selection model for optimal fuzzy clustering algorithm and number of clusters based on competitive comprehensive fuzzy evaluation. IEEE Transactions on Fuzzy Systems. 17(3): 568-577.

[16] C. Hwang and F. C. Rhee. 2007. Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means. IEEE Transactions on Fuzzy Systems. 15(1): 107-120.

[17] M. Lee and W. Pedrycz. 2009. The fuzzy C-means algorithm with fuzzy P-mode prototypes for clustering objects having mixed features. Fuzzy Sets and Systems. 160(24): 3590-3600.

[18] V. S. Alves, R. J. Campello and E. R. Hruschka. 2007. A fuzzy variant of an evolutionary algorithm for clustering. in Proceedings of the IEEE International Conference on Fuzzy Systems, pp. 375-380, Imperial College, London, UK.

[19] C. Chun-Hao, V. S. Tseng and H. Tzung-Pei. 2008. Cluster-based evaluation in fuzzy-genetic data mining. IEEE Transactions on Fuzzy Systems. 16(1): 249-262.

[20] M. K. Pakhira, S. Bandyopadhyay and U. Maulik. 2005. A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification Fuzzy Sets and Systems. 155(2): 191-214.

[21] G. Gan, J. Wu and Z. Yang. 2009. A genetic fuzzy k-Modes algorithm for clustering categorical data. Expert Systems with Applications. 36(2): 1615-1620.

[22] A. Mukhopadhyay, U. Maulik and S. Bandyopadhyay. 2009. Multi-objective genetic algorithm-based fuzzy clustering of categorical attributes. IEEE Transactions on Evolutionary Computation. 13(5): 991-1005.

[23] Vijay Kumar, G. Valli Kumari, V. 2013. Incremental mining for regular frequent patterns in vertical format. International. 5(2): 1506-1511.

[24] Vijay Kumar, G. Valli Kumari, V. 2012. Sliding window technique to mine regular frequent patterns in data streams using vertical format. IEEE International Conference on Computational Intelligence and Computing Research, ICCIC.