Volume 9, No. 1, Jan. - 2018 (Special Issue)
International Journal of Mathematical Archive-9(1), 2018, 176-184
Available online through www.ijma.info ISSN 2229 – 5046

DISTRIBUTED INTUITIONISTIC FUZZY CLUSTERING APPROACH FOR CENSUS DATA SET

N. KARTHIKEYANI VISALAKSHI[1] & K. ARUN PRABHA[2]

[1]Assistant Professor, Department of Computer Science,
N.K.R. Government Arts College for Women, Namakkal, Tamil Nadu, India.

[21]Assistant Professor, Department of Computer Technology,
Vellalar College for Women, Erode, Tamil Nadu, India.

E-mail: karthichitru@yahoo.co.in[1] & arunjeevesh@gmail.com[2]

ABSTRACT

***T**here may be a new requirement for powerful ways to address disseminated grouping, due to explosion in the quantity of self sustaining records sources. Intuitionistic Fuzzy Set is a suitable tool to manage defectively characterized actualities and information, and additionally with uncertain learning. In this paper census data analysis can be done by using Intuitionistic Fuzzy based Distributed Fuzzy C-Means Algorithm (IF-DFCM), to group conveyed datasets, without essentially downloading every one of the information into a solitary webpage. The procedure is done in two distinct levels: neighborhood level and worldwide level. In neighborhood level, numerical datasets are transformed into intuitionistic fuzzy data and they are clustered independently from each other using modified fuzzy C-Means algorithm. In worldwide level, centroid is computed by clustering all local cluster centroids. The global centroid is again transmitted to local sites to revise and bring up to date local cluster model. The main objective is to apply and compare the results of Census dataset with Intuitionistic Fuzzy based Distributed Fuzzy C-Means Algorithm (IF-DFCM) and Intuitionistic Fuzzy based Centralized Fuzzy C-Means Algorithm (IF-CFCM). It is observed that the algorithm IF-DFCM performs better than IF-CFCM algorithm.*

Keywords: *Distributed Clustering, Fuzzy C-Means, Local Centroid, Global Centroid, Intuitionistic Fuzzy Sets.*

I. INTRODUCTION

Clustering has played an indispensable role in the field of machine learning, pattern recognition and in data mining applications. It also has a significant function in spatial database applications, web analysis, customer relationship management, marketing, bio-medical analysis and many other related areas. With the introduction of Internet into day to day lives and with the ever increasing automated business, it has become coercion for the dataset to increase in size. Moreover, many of these datasets are, in nature, geographically distributed across multiple sites. The huge size of many databases, the wide distribution of data, and the computational complexity of clustering algorithms are some of the factors found to be motivating the development of distributed clustering algorithms. Distributed clustering assumes that data to clustered are in different sites. This process is carried out in two different levels such as local level and global level. In local level, all sites carry out clustering process independently from each other. After having completed the clustering process, a local model such as cluster centroid or cluster index is determined, which should reflect an optimum trade-off between complexity and accuracy? Further, the local model is transferred to a central site, where the local models are merged in order to form a global model. The resultant global model is again transmitted to local sites to update the local models. Instead of local model, local representative samples may also be transmitted to reach global clusters [7].

The main objective of distributed data clustering algorithms is to cluster the distributed datasets without necessarily downloading all the data to the single site. The key idea of distributed data clustering is to achieve a global clustering that is as good as the best centralized clustering algorithm with limited communication required to collect the local models or local representatives into a single location, regardless of the crucial choice of any clustering techniques in local site [4]. Distributed Clustering techniques have been pertinent to a wide variety of real life research applications such as Weather Analysis, Astronomy, Biomedical and Healthcare, Retail Industry, Telecommunication Industry, E-Governance, Insurance Applications, Security and Census Analysis. The study of Census data is an important

CONFERENCE PAPER
International Conference dated 08-10 Jan. 2018, on Intuitionistic Fuzzy Sets and Systems (ICIFSS - 2018), Organized by Vellalar College for Women (Autonomous), Erode, Tamil Nadu, India.

activity for the government and business activity. They are naturally gathered from various places and it is essential to maintain integrity and reduce communication and storage cost during exploratory study. There exist a growing number of clustering applications, where the data have to be physically distributed, due to either their huge volumes or privacy concern. Distributed data clustering is a promising region for such applications. In the real world, distributed clustering applications frequently involve disparate datasets. It consists of inconsistencies or outliers, where it is difficult to obtain homogeneous and meaningful global clusters.

Intuitionistic Fuzzy Sets (IFS) [17] are generalized fuzzy sets that are useful in coping with the uncertainty originating from faulty and vague information. Recently, limited attention has been paid in proposing intuitionistic fuzzy based clustering for centralized environment [27][28]. Fuzzy sets are designed to represent or manipulate data and information possessing non-statistical uncertainties. Since L. A. Zadeh [20] introduced the concept of fuzzy sets, various notions of high-order fuzzy sets have been proposed. Among them, IFSs introduced by Krassimir T. Atanassov [17], have captured the attention of many researchers in the last decades. This is mainly due to the fact that IFSs are consistent with human behavior, by reflecting and modeling the hesitancy present in real-life situations. Since IFSs can present the degrees of membership and non-membership with a degree of uncertainty, the knowledge and semantic representation become more meaningful and applicable. Most of the distributed clustering algorithms found in the literature [18][19] seek to cluster numerical data. Karthikeyani *et al.* [16] proposed a novel method of distributed fuzzy clustering using intuitionistic fuzzy set theory. The same authors[15] proposed modified distributed combining algorithm to cluster disparate data sources having diverse, possibly overlapping set of features and also need not share objects. Both K-Means and Fuzzy C-Means algorithm is used for local clustering.

The Census data is the mixture of numeric and categorical data attributes. But clustering categorical data is an important and challenging data analysis task. Hence, it is required to have appropriate numerical representation scheme for categorical data. Thangavel *et al.* proposed an algorithm to convert the attribute set in the categorical domain to numerical domain and it is referred as Weighted ASCII Representation (WAR). This algorithm estimates the numerical value of the categorical data using ASCII (American Standard Code for Information Interchange) value of each character and its positional weight. It converts categorical data into numerical form, before clustering census data. The attributes with the categorical data are converted into number format using WAR, and independent evaluation is performed on the existing fuzzy distributed clustering algorithm. Hence, it is required to integrate Intuitionistic Fuzzy approach with Distributed Fuzzy Clustering to deal with uncertainty among the Census data objects and obtain effective and efficient fuzzy clusters in distributed environment. In this approach, Intuitionistic Fuzzy based Distributed Fuzzy Clustering (IFDFCM) algorithm and confirms its superior performance by combining distributed fuzzy clustering using intuitionistic fuzzy set and WAR algorithms.

The rest of this paper is organized as follows: Section 2 discusses the related works. Section 3 and 4 presents Distributed data Clustering and Intuitionistics fuzzy clustering respectively. Section 5 describes distributed fuzzy clustering of IF data. Section 6 summarizes the experimental analysis performed with census datasets. Finally, Section 7 concludes the paper.

II. RELATED WORKS

There are various distributed clustering solutions proposed in the literature and their comprehensive survey can be obtained from [26] [13]. This section reviews the recent research works on distributed clustering and intuitionistic fuzzy based centralized clustering. The P2P K-Means algorithm is proposed [26] for distributed clustering of data streams in a peer-to-peer sensor network environment. Jin R. *et al.* [12] presented distributed version of Fast and Exact K-Means (FEKM) algorithm, which collected sample data from each data source, and communicated it to the central node. The main data structure of FEKM, the cluster abstract table is computed and sent to all data sources to get global clusters. Lamine M. Aouad *et al.* [18] proposed a lightweight distributed clustering technique based on a merging of independent local sub clusters according to an increasing variance constraint. The key idea of this algorithm is to choose a relatively high number of clusters locally, or an optimal local number using an approximation technique, and to merge them at the global level according to an increasing variance criterion which requires a very limited communication overhead. Cormode *et al.* [1] have introduced the problem of continuous, distributed clustering, and given a selection of algorithms, based on the paradigms of local vs. global computations, and furthest point or parallel guessing clustering. In their experimental evaluation, the combination of local and parallel guessing addressed the least communication cost. Le-Khac N. *et al.* [19] presented an approach for distributed density-based clustering. The local models are created by DBSCAN at each node of the system and these local models are aggregated by using tree based topologies to construct global models. In [3] P-SPARROW algorithm is proposed for distributed clustering of data in peer-to-peer environments. The algorithm combined a smart exploratory strategy based on a flock of birds with a density-based strategy to discover clusters of arbitrary shape and size in spatial data.

The Improved Distributed Combining Algorithm (IDCA) [15] is a refined version of Distributed Combining Algorithm [5], designed for distributed hard clustering. The process of centroid mapping is performed effectively, with the support of Hungarian method of unbalanced assignment problem, when each dataset produces different number of clusters. R. Kashef and M. S. Kamel [16] proposed a Distributed Cooperative Hard-Fuzzy Clustering (DCHFC) model for document clustering. This model is based on the intermediate cooperation between the hard distributed K-Means and fuzzy distributed C-Means to enhance the performance of the K-Means reduce the computational time taken by the fuzzy algorithm and produce a better global solution. In [27], Vicenc Torra *et al.* introduced a method to define intuitionistic fuzzy partitions from the result of different fuzzy clustering algorithms such as FCM, entropy based FCM and FCM with tolerance. In this approach, the intuitionistic fuzzy partition permits to cope with the uncertainty present in the execution of different fuzzy clustering algorithms with the same data and with the same parameterization.

In [23], N. Pelekis *et al.* investigated the issue of clustering intuitionistic fuzzy representation of images. For this, they proposed a clustering approach based on the FCM algorithm utilizing a novel similarity metric defined over IFS. The performance of the modified FCM algorithm is evaluated for object clustering in the presence of noise and image segmentation. It is proved that clustering intuitionistic fuzzy image representations is more effective, noise tolerant and efficient as compared with the conventional FCM clustering of both crisp and fuzzy image representations.N. Karthikeyani Visalakshi, K. Thangavel, and R. Parvathi [16] introduced a novel intuitionistic fuzzy based distributed clustering algorithm, to cluster distributed datasets without necessarily downloading all the data into a single site.

III. DISTRIBUTED DATA CLUSTERING

The key idea of distributed data clustering is to achieve a global clustering with limited communication required to collect the local models or local representatives into a single location, regardless of the crucial choice of any clustering techniques in local site [13]. Most of the applications that deal with time-critical distributed data are likely to benefit by paying careful attention to the distributed resources for computation, storage, and communication cost. Moreover, there exist a growing number of clustering applications, where the data have to be physically distributed, due to either their huge volumes or privacy concern. Distributed data clustering is a promising candidate for such applications.

3.1. Taxonomy of distributed clustering

Distributed clustering algorithms can be classified into different independent dimensions [25][26]. For instance, data distribution, data communication, quality of global model, and architecture usually lead to different taxonomies of distributed clustering algorithms. Different properties of distributed clustering algorithms can be described as follows:

Homogeneous vs. Heterogeneous Distributed Data

A common classification based on data distribution is that which applies to homogeneously distributed (horizontally partitioned) or heterogeneously distributed (vertically partitioned) data. Homogeneous datasets contain the same set of attributes across distributed data sites. Examples include local weather databases at different geographical locations and market-basket data collected at different locations of a grocery chain. Heterogeneous data model supports different data sites with different schemata. For example, a disease emergency detection problem may require collective information from a disease database, a demographic database, and biological surveillance databases [24].

Multiple Communication Round vs. Centralized Ensemble-based

The multiple communication round algorithm consists of methods requiring multiple rounds of message passing between different sites during the process of clustering. These methods require a significant amount of synchronization, whereas the centralized ensemble-based algorithm works asynchronously, by first generating the local clusters and then combining those at the central site. Both the ensemble approach and the multiple communication round-based clustering algorithms usually work efficaciously than their centralized counterparts in a distributed environment [25].

Exact vs. Approximate

This classification is based on the quality of global result obtained from distributed clustering. Exact distributed clustering algorithms produce a global model identical to a hypothetical model generated by a centralized approach having access to the full dataset. The exact algorithm works as if the local datasets at each node are bought together into one dataset first, then a centralized clustering algorithm is performed on the whole dataset. The clustering solutions are then distributed again by intersecting the local datasets with the global clustering solutions. Approximate distributed clustering algorithms on the other hand, produce a model that closely approximates a centrally-generated model. Most distributed data clustering research focuses on approximate algorithms as they tend to produce comparable results to exact algorithms with far less complexity.

IV. INTUITIONISTIC FUZZY SETS

Fuzzy sets are designed to represent or manipulate data and information possessing non-statistical uncertainties. Since L. A. Zadeh [23] introduced the concept of fuzzy sets, various notions of high-order fuzzy sets have been proposed. Since IFSs can present the degrees of membership and non-membership with a degree of hesitancy, the knowledge and semantic representation become more meaningful and applicable.

4.1. Intuitionistic Fuzzy Sets

An intuitionistic fuzzy set is defined as a generalization of a fuzzy set.

Definition: An IFS *A* is an object of the form

$$A=\left\{\left\langle x,\mu_A(x),\nu_A(x)\right\rangle \middle| x\in E\right\} \tag{3.1}$$

where $\mu_A : E \rightarrow [0,1]$ and $\nu_A : E \rightarrow [0,1]$ define the degree of membership and non-membership, respectively, of the element $x \in E$ to the set $A \subset E$. For every element $x \in E$, it holds that $0 \le \mu_A(x)+\nu_A(x) \le 1$.
For every $x \in E$, if $\nu_A(x) = 1-\mu_A(x)$, then *A* represents a fuzzy set. The function

$$\pi_A(x)=1-\mu_A(x)-\nu_A(x) \tag{3.2}$$

represents the degree of hesitancy of the element $x \in E$ to the set $A \subset E$.

4.2. Intuitionistic Fuzzy Representation of Numerical Data Objects

The analysis of Census data played important role in administration and industry applications. The census data are naturally collected from different geographical places and it is essential to maintain integrity and reduce communication and storage cost during exploratory analysis. Clustering categorical data is an important and challenging data analysis task. But, the census data naturally contain mixed of numerical and categorical attributes. Most of the distributed clustering algorithms found in the literature seek to cluster numerical data. The proposed algorithms are directly not endurable for datasets containing categorical attributes, because they use the local clustering algorithm as K-Means or its variants and Euclidean distance for the computation of local and global centroid. Hence, it is required to have appropriate numerical representation scheme for categorical data, so that the same algorithms can also be applied for census data segmentation in distributed environment.

The proposed IFDFC algorithm requires that each element is to be converted into a pair of membership and non membership values. A new procedure for intuitionistic fuzzy representation of numeric data is derived, by modifying the definition for intuitionistic fuzzy representation of digital image [24]. In this process, the crisp dataset is first transferred to fuzzy domain and sequentially into the intuitionistic fuzzy domain, where the clustering is performed. The following Section describes numerical representation scheme for categorical data.

4.3. Weighted ASCII Representation

Thangavel *et al.* [30] proposed an algorithm to convert the attribute set in the categorical domain to numerical domain as in Figure 1 and it is referred as weighted ASCII representation. This algorithm estimates the numerical value of the categorical data using ASCII (American Standard Code for Information Interchange) value of each character and its positional weight. For instance, in order to convert the categorical data “RED” into corresponding numerical value, first the length of the word, say *l* is to be found. Here *l* is equal to 3. Then, the sum of ASCII value is to be computed after multiplying its positional weight such as $l * ASC('R')+(l-1)* ASC('E')+(l-2)* ASC('D')$, where the function *ASC (y)* gives the ASCII value of “*y*”. The positional weights are attached to avoid the duplicate values.

```
Algorithm: WAR
Input    : List of categorical values y and number of values n
Output   : List of numerical values Y
Procedure
Step-1: Repeat step 2 to step 5, for each categorical value $y_i$, $i=1,2,...,n$
Step-2: Find the length of categorical value $y_i$, $l = length(y_i)$ and $s=0$
Step-3: Repeat step 3 for $j=1,2,...,l$
Step-4: Find the ASCII value of each character and multiply with corresponding positional weight,
        $s=s+(l-j+1) * ASC(x_{ij})$
Step-5: Replace categorical value of $y_i$ with corresponding numerical value, $Y_i = s$
```

Figure-1: Weighted ASCII Representation

V. INTUITIONISTIC FUZZY BASED FCM CLUSTERING

The proposed methodology for the centralized robust fuzzy clustering of numerical datasets, based on the concept of IFSs involves two stages. In the first stage, intuitionistic fuzzification is done to convert the real scalar values into intuitionistic fuzzy values. Secondly, the modified FCM algorithm based on IF similarity measure is used to cluster intuitionistic fuzzy data.

5.1. Intuitionistic Fuzzification

A new procedure for intuitionistic fuzzification of numerical dataset is derived. In this process, the crisp dataset is first transferred to fuzzy domain and sequentially into the intuitionistic fuzzy domain, where the clustering is performed. Let *X* be the dataset of *n* objects and each object contains *d* features. The proposed IF data clustering requires that each data element x_{ij} belongs to an IFS X' by a degree $\mu_i(x_j)$ and does not belong to X' by a degree $\nu_i(x_j)$, where *i* and *j* represent objects and features of the dataset respectively.

A membership function $\overline{\mu_i}(x_j)$ for intermediate fuzzy representation is defined by

$$\overline{\mu_i}(x_j) = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}, \text{ where } i = 1, 2, ..., n \text{ and } j = 1, 2, ..., d \quad (3.3)$$

The intuitionistic fuzzification based on the family of parametric membership and non-membership functions, used for clustering, are defined respectively by

$$\mu_i(x_j;\lambda) = 1 - (1 - \overline{\mu_i}(x_j))^{\lambda} \quad (3.4)$$

and

$$\nu_i(x_j;\lambda) = (1 - \overline{\mu_i}(x_j))^{\lambda(\lambda+1)}, \text{ where } \lambda \in [0,1] \quad (3.5)$$

The intuitionistic fuzzification converts crisp dataset $X(x_{ij})$ into intuitionistic fuzzy dataset $X'(x_{ij}, \mu_i(x_j), \nu_i(x_j))$.

5.2.Fuzzy Clustering of Intuitionistic Fuzzy Data

In this stage, Pelekis's modified FCM is applied to cluster IF data. The procedure used in modified FCM is same as conventional FCM, except in similarity measure used to compute the membership degree of the object to cluster. Instead of Euclidean distance in conventional FCM, the modified FCM described in Figure 2 applies IF similarity measure. Initially, *K* numbers of centroids are randomly selected from the IF data objects, which contain both membership and non-membership values. Next, the membership degree of each object to each cluster U_{ij} is computed using IF similarity measure. The centroids are then updated using cluster membership matrix U_{ij} and corresponding membership and non-membership degrees of centroids C_i are also computed. The above two steps are repeated, until it reaches convergence.

```
Algorithm:  MFCM
Input       : IF dataset of n objects with d features, number of clusters K and fuzzification value m > 1
Output      : Membership matrix $U_{ij}$ for n objects and K clusters
Procedure
Step-1:  Declare a membership matrix U of size $n \times K$
Step-2:  Generate K cluster centroids randomly within the range of the data or select K objects randomly as initial
         cluster centroids. Let the centroids be $c_1, c_2, \ldots, c_K$.
Step-3:  Compute similarity between each object $x_i$, $i = 1, 2, \ldots, n$ and cluster centroids $c_j$, $j = 1, 2, \ldots, K$, using IF
         similarity measure $S_{ij}$
```

Step-4: Compute membership matrix U_{ij},

$$\underset{\substack{1\le i\le n\\ 1\le j\le K}}{\forall} U_{ij}=\begin{cases}\dfrac{\left(s_{ij}\right)^{\frac{1}{1-m}}}{\sum\limits_{l=1}^{K}\left(s_{il}\right)^{\frac{1}{1-m}}}, & I_i=\phi\\ 0, & j\notin I_i\\ \sum\limits_{j\in I_i}U_{ij}=1, & j\in I_i,\ I_i\ne\phi\end{cases}\qquad \text{where } \underset{\forall 1\le i\le n}{I_i}=\left\{j\,\middle|\,1\le j\le K\,;s_{ij}=0\right\}$$

Step-5: Compute new cluster centroid c_j

$$\underset{1\le j\le K}{\forall} c_j=\frac{\sum\limits_{i=1}^{n}(U_{ij})^m x_j}{\sum\limits_{i=1}^{n}(U_{ij})^m}$$

Step-6: Compute membership and non-membership degrees of c_j

Step-7: Repeat step 2 to step 4 until converges.

Figure-2: Modified Fuzzy C-Means Algorithm

5.3. INTUITIONISTIC FUZZY BASED DFCM ALGORITHM

The step by step procedure of proposed IF-DFCM algorithm for homogeneously distributed datasets is described in Figure3. First, minimum and maximum values of each feature vectors are extracted from all local datasets and transmitted to central place, where global minimum and maximum values are identified. These two values are used to convert real scalar values of local datasets into pair of global IF data objects using the Equations 3.4 and 3.5. Next, the IF objects of local datasets are clustered using modified FCM to obtain membership matrix and local centroids in terms of membership and non-membership. All local centroids in terms of membership and non-membership are separately merged into a pair of datasets and clustered using the same modified FCM algorithm at the central place to obtain global centroids. The membership matrices of local datasets are then updated using global centroids to obtain global fuzzy clusters of distributed datasets.

Algorithm: IF-DFCM
Input : Homogeneous *r* datasets, each with *d* dimensions
Output : Global fuzzy clusters of *r* datasets
Procedure
Step-1: Find maximum and minimum values of each feature from each local dataset and transmit them into central place
Step-2: Compute global maximum and minimum value at central place
Step-3: Convert real scalar values of local datasets into IF values using Equation 3.5 and Equation 3.6
Step-4: Cluster each local IF dataset by modified FCM algorithm and obtain membership matrix and IF form of cluster centroids
Step-5: Transmit IF form of centroid matrix into a central place
Step-6: Merge membership and non membership values of cluster centroids of local datasets into a pair of centroids datasets
Step-7: Cluster centroids datasets using modified FCM to obtain global centroids
Step-8: Update membership matrix using global centroids to obtain global fuzzy clusters

Figure-3: Intuitionistic Fuzzy based Distributed Fuzzy C-Means Algorithm

VI. EXPERIMENTAL ANALYSIS

Census dataset [21] was extracted from the US Census Bureau database after identifying a set of clean records by Barry Becker, and is again available in the UCI Machine Learning Repository. This dataset contains 48,842 instances with 14 attributes. The instances with missing values are removed and the remaining 45,222 instances are considered for analysis. The attributes with the categorical data are converted into number format WAR.

6.1. Fuzzy Cluster Evaluation

In this section, empirical evidence is provided for distributed fuzzy clustering, that the high quality global cluster models is obtained with limited communication overhead and high level of privacy. All experiments are conducted with the assumption of having non-overlapping objects with same set of features in distributed datasets, for uniform type of data distribution.

Table-1: Fuzzy Cluster Evaluation based on WAR

Index	Algorithm	*K*=2	*K*=3	*K*=4	*K*=5	*K*=6	*K*=7
FDB Index	**DFCM**	0.6028	0.5366	0.5431	0.5655	0.5181	0.5488
	IF-DFCM	**0.5928**	**0.5226**	**0.5313**	**0.5498**	**0.5012**	**0.5217**
	CFCM	0.6028	0.5366	0.5433	0.5625	0.5179	0.5315
	IF-CFCM	0.5920	0.5221	0.5311	0.5422	0.5014	0.5211
XB Index	**DFCM**	0.1173	0.1425	0.1169	0.1546	0.0893	0.0976
	IF-DFCM	**0.1032**	**0.1387**	**0.1027**	**0.1418**	**0.1099**	**0.1134**
	CFCM	0.1173	0.1427	0.1154	0.1517	0.1224	0.1254
	IF-CFCM	0.1020	0.1325	0.1092	0.1418	0.0984	0.1100

The Intuitionistic Fuzzy based Distributed Fuzzy C-Means Algorithm (IF-DFCM) is applied on census data evaluated using two fuzzy validity indices, Fuzzy DB (FDB) index and Xie-Beni (XB) index. Table 1 illustrate the average fuzzy results of census data based on WAR in terms of FDB index and XB index, for different values of *K*. From the Table 1 the algorithms DFCM and IDC-FCM yield almost equal results, in terms of FDB index and XB index. It is also observed that the algorithm IF-DFCM performs better than the other two algorithms, for all the values of *K*.

6.2. Scalability Measurement

In order to evaluate the scalability of the distributed clustering algorithms, the Census dataset is divided into different number of data sources and clustered. Tables 2 show the results of the IDC-FCM and IF-DFCM distributed clustering algorithms based on WAR, for different values of *r* (number of data sources). From these results, it is proved that the proposed algorithms are consistent, independent of the number of subsets.

Table-2: Scalability Measurement based on WAR

K=3				
Algorithm	***r* = 5**	***r* = 7**	***r* = 10**	***r* = 12**
IDC-FCM	0.5366	0.5366	0.5366	0.5366
IF-DFCM	0.5226	0.5227	0.5238	0.5219

From the Tables 1 and 2, distributed clustering algorithms provide equal Performance as its corresponding centralized clustering algorithms for census data, they outperform centralized clustering in terms of communication overhead, space complexity, and privacy maintenance.

VII. CONCLUSION

The distributed fuzzy clustering algorithms IDC-FCM, and IF-DFCM are applied on census data. After evaluation, it is observed that the performance of distributed fuzzy clustering algorithms based on WAR is improved. It is also observed that the algorithm IF-DFCM performs better than the other three algorithms, for all the values of *K*. While comparing the results of IF-DFCM with IF-CFCM, it is satisfied that the intuitionistic fuzzy based distributed fuzzy clustering algorithms achieve almost equal performance as intuitionistic fuzzy based centralized fuzzy clustering algorithm.

REFERENCES

1. Cormode G., Muthukrishnan S., Zhuang W., Conquering the divide: continuous clustering of distributed data streams, In IEEE 23rd International Conference on Data Engineering, Turkey, (2007), 1036-1045.
2. Dimitrios K. Iakovidis, Nikos Pelekis, Evangelos E. Kotsifakos, Ioannis Kopanakis, Intuitionistic fuzzy clustering with applications in computer vision. In Advanced Concepts for Intelligent Vision Systems. LNCS, Blanc-Talon *et al.* (eds.), Springer Berlin/ Heidelberg, (2008), 764-774.
3. Folino G., Forestiero A., Spezzano G., Swarm-based distributed clustering in peer-to-peer systems, In Artificial Evolution, Lecture Notes in Computer Science, Talbi E. et al. (Eds.), Springer-Verlag, (2006), 37-48.
4. Ghosh J., Merugu S., Distributed clustering with limited knowledge sharing, In Proceedings of the 5th International Conference on Advances in Pattern Recognition, Calcutta, India, (2003), 48-53.
5. Hore P., Lawrence O. Hall, Scalable clustering: A distributed approach, In IEEE International Conference on Fuzzy Systems, Hungary, (2004), 25-29.
6. Hore P. Lawrence O. Hall, Dimitry B. Goldgofz, A Cluster ensemble framework for large datasets, In Proceedings of IEEE Conference on Systems, Man Cybernetics B, Taiwan, 4 (2006), 3342-3347.
7. Hore P., Lawrence O. Hall, Dimitry B. Goldgof, A scalable framework for cluster ensembles, Pattern Recognition, 42(5) (2008), 676-688.
8. Halkidi M., Batistakis Y., Vazirgiannis M., Cluster validity methods: part II, ACM SIGMOD Record, 31(3) (2002), 19-27.
9. Hui Xiong., Junjie Wu., Jian Chen., K-means clustering versus validation measures: a data distribution perspective. ACM SIGKDD international Conference on Knowledge Discovery and Data Mining. Philadelphia, PA, USA, (2006), 779-784.
10. Ioannis K., Vlachos, George D. Sergiadis., The role of entropy in intuitionistic fuzzy contrast enhancement. In Foundations of fuzzy logic and soft computing, LNCS, Melin P. et al. (eds.), Springer Berlin/Heidelberg, (2007), 104-113.
11. Jain A. K., Murthy M. N., Flynn P. J., Data clustering: A review, ACM Computing Surveys, 31(3) (1999), 265-323.
12. Jeong J., Ryu B., Shin D., Shin D., Integration of distributed biological data using modified k-means algorithm, In Emerging Technologies in Knowledge Discovery and Data Mining, LNCS, Washio T. et al. (eds.), Springer-Berlin, (2007), 469-475.
13. Jin R., Goswami A., Agarwal G., Fast and exact out-of-core and distributed K-Means clustering, Knowledge and Information Systems, 10 (1) (2006), 17-40.
14. Ji Genlin, Ling Xiaohan, Ensemble learning based distributed clustering, In Emerging Technology and Knowledge Discovery and Data Mining, LNCS, Washio T. et al. (eds.), Springer-Verlag, (2007), 312-321.
15. Karthikeyani Visalakshi N., Thangavel K., Alagambigai P., Ensemble approach to distributed clustering, In Mathematical and Computational Model, Natarajan et al. (eds.), Narosa Publishing House, New Delhi, (2007), 252-261.
16. Karthikeyani Visalakshi, Thangavel K., Parvathi R., An Intuitionistic Fuzzy Approach to Distributed Fuzzy Clustering , International Journal of Computer Theory and Engineering, 2(2) (2010), 1793-8201 .
17. Krassimir T., Atanassov., Intuitionistic fuzzy sets: past, present and future. In Proceedings of the 3rd Conference of the European Society for Fuzzy Logic and Technology. Germany, (2003), 12-19.
18. Lamine M. A., Le-Khac N., Tahar M. K., Lightweight clustering technique for distributed data mining applications. In Advances in data mining, Theoretical aspects and applications, LNCS, Perner P. (Ed.), Springer, (2007), 120-134.
19. Le-Khac N., Lamine M A., Tahar M K., A new approach for distributed density based clustering on grid platform, In Data Management, Data every where, LNCS, Kooper R. Kennedy J. (eds.), Springer-Berlin, (2007), 247-258.
20. Lotfi A. Zadeh., Fuzzy sets, Information and Control, 8(3) (1965), 338-353.
21. Merz C. J., Murphy P. M., UCI repository of machine learning databases, Irvine, University of California, < http://www.ics.uci.eedu/~mlearn/>, (1998).
22. Nguyen ThoThongLe HoangSon., HIFCF: An effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis, Expert Systems with Applications, 42(7) (2015), 3682-3701
23. Nikos Pelekis, Dimitrios K., Iakovidis, Evangelos E., Kotsifakos, Ioannis Kopanakis., Fuzzy clustering of intuitionistic fuzzy data, International Journal of Business Intelligence and Data Mining, 3(1), 45-65.
24. Pang-Ning Tan, Steinbach M., Kumar V., Cluster analysis: basic concepts and algorithms in Introduction to Data Mining, Pearson Addison Wesley, Boston, (2006).

25. Park B., Kargupta H., Distributed data mining, The Hand Book of Data Mining, Nong Ye,(ed.), Lawrence Erlabum Associates, Publishers, Mahwah, Newjersey, (2003), 341-358.
26. Sanghamitra B., Giannella C., Maulik U., Kargupta H., Liu. K., Datta S., Clustering distributed data streams in peer-to-peer environments, Information Science, 176 (4) (2006), 1952-1985.
27. Torra., Miyamoto S., Endo Y., Domingo-Ferrer J., On intuitionistic fuzzy clustering for its application to privacy. In Proceedings of IEEE International Conference on Fuzzy Systems,, Hong Kong, China, (2008), 1042-1048.
28. Zeshui Xu, Jian Chen, Junjie Wu., Clustering algorithm for intuitionistic fuzzy sets, Information Sciences, 178(19), (2008), 3775-3790.
29. Zhou A., Cao F., Yan Y., Sha C C., He X., Distributed data stream clustering: a fast EM-based approach, In IEEE 23rd International Conference on Data Engineering, Turkey, (2007), 736-745.

Source of support: Proceedings of UGC Funded International Conference on Intuitionistic Fuzzy Sets and Systems (ICIFSS-2018), Organized by: Vellalar College for Women (Autonomous), Erode, Tamil Nadu, India.