An Energy Effectual Swarm Based Hierarchical Density Peak Clustering (H-DPCS) Method for Multi-sensor Data Fusion Strategy

Easpin Brumancia[1]*   Justin Samuel[2]

[1]*School of Computing, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India*
[2]*Computer Science Department, PSN College of Engineering, Melathediyoor, Tamil Nadu, India*
* Corresponding author's Email: easpinbrumancia@gmail.com

Abstract: Clustering is the most frequently used physical architecture in data fusion, causing a problem as it has random distribution within local clusters, we propose a density peak based clustering algorithm utilizing a hierarchical approach. As a proficient method to joint decision-making of multi sensors, data fusion contains several benefits in data mining. In this work, we use the Salp Swarm Optimizer (SSO) algorithm based Hierarchical Density Peak Clustering (DPC) method to cluster the multiple sensor nodes as per the data similitude for the process of data fusion. Here, the higher density, distance and the cut off values are determined by using the SSO algorithm and then the chosen optimal fixed values were used to further process. Hierarchical Density Peak Clustering estimates the density and distance of each point. It is validated experimentally and compared it with the datasets and the error prediction with the test case numbers. The experimental results of the method are 96% accuracy determined from different datasets.

Keywords: Density peak clustering (DPC), Salp swarm optimizer (SSO), Data fusion, Fuzzy clustering, Clustering fast search (CFS), Euclidean distance.

1. Introduction

Data fusion is a technology that allows combine information from multiple sources to create a unified image or data, and it is extensively utilized in several areas like sensor networks, robotics, and video with image processing. Dempster–Shafer theory is used to many types of the fusion information particularly for some applications are aircraft landing and many computational intelligence systems [1]. The algorithms for clustering the nodes in the data equivalent for the best results the clustering for fusion methods is preferred. The Gaussian space time is used to divide the region as clearly and correlated strongly with the measurements in the spatial coherence region. [2]. Salp Swarm Algorithm (SSA) is an irregular populace based calculations proposed by Mirjalili et al. (2017) in 2017. SSA reproduces amassing salps component while scavenging in seas. Such as other multitude based procedures, salps condition is characterized as s-dimensional inquiry space, where s refers quantity of factors in obstacle.

In this manner, the situation of all salps is put away in a two-dimensional network called z. It is additionally accepted that there is a food source called P in the pursuit space as the multitude objective [3]. The dataset is assigned to the users on the experience and the priority of cut-off distance. The DPC assignment will be assigning the centre points by the local density of neighbour cluster it will be done in the error propagation of the label cluster [4]. The clustering and recombination levels of the network are improved taking into account factors including the remaining energy of the nodes and the distance between the nodes [5].

The contribution of the proposed method is,

- In this manuscript we propose DPC clustering algorithm using hierarchical strategy, to solve the problem of random distribution within local clusters.
- By using the SSO algorithm the higher density, distance and the cut off values are determined and then the selected optimal fixed values are used for further process.

- It consists of two steps. The first step is, DPC approach using hierarchical strategy assesses the density with distance of each point. Points with high density and distance are selected as candidate centres, after that sub clusters centred on them are acquired.
- In second step, we measure the dissimilarity and connectivity between the sub clusters. Such high similar and coupled sub-clusters are merged to maximize the dissimilarity between different clusters and to get the eventual clustering results.

1.1 Organization of the paper

The rest of this manuscript is designed as: About the related works are discussed in comprehensive in section 2. The proposed method is delineated in section 3. Results and discussion are illustrated in section 4. The scopes for future studies are discussed in Section 5. Finally, section 6 concludes the manuscript.

2. Related works

Redhu (2019) [6] was introduced the multiple-sensor data fusion to the cluster the method was used to the different sensor modalities, for developing the network into various clusters using fused resemblance. The drawback of the method was poor cluster description. Shen Xinglin et al. (2019) [7] was found the fast density peak- clustering algorithm for partition and obviously decrease the computation burden without lose the efficiency while the measurements of created cluster. The disadvantage of this method is datasets with altering densities are tricky. The problem of such method was more energy and time needed. Cai (2020) [8] was found new clustering approach depends on DPC to solve effect of selection parameter in the calculating the density and results of clustering. The drawback was complicated to assess the cut-off distance values depend on basis for selection.

Jiang (2019) [9] have found the DPC-LG density peaks clustering in terms of logistic distribution including gravitation to find the count of clusters accurately and detect the outliers. The drawback of such method was processing certain datasets changing the densities and the irregular shape. Li (2020) [10] have found the DPC based in the relative optimization was used to CFS and detect density peaks presents a completely novel clustering frame and re-defined clustering centre type. With DPC, model data density peaks were simply and rapidly seen. DPC also displays higher performance in assignment and noise removal.

Ren (2020) [11] was explained for extracting the data of lower-dimensional feature denotes from higher-dimensional image data, also to decrease the data of 2-dimensional space favouring density-depend clustering algorithms. The main drawback was it needs the number of clusters in advance. Jiande (2019) [12] have presented the clustering algorithm includes calculate the cluster for each data point and determine whether there was an overlapping cluster in dataset according to filter out clutter and obtain for k-means clustering. The disadvantage of such method was local minima and slow convergence. Milad (2019) [13] presented a spectral clustering heuristic approach in particular avoids the required input by evaluating parameters from data, which introduced the heuristic nature of the iterative. The major drawback was network delay and network energy consumption.

3. Proposed method

The local clusters distribution is uneven for the determination of the centres and to overcome this density peak clustering is used with the hierarchy strategy. We use the Salp Swarm Optimizer (SSO) algorithm based Hierarchical Density Peak Clustering (DPC) method for clustering the multiple sensor nodes as per the data similitude for the process of data fusion. Our method consists of two stages are estimating the density and the distance in the point from HCFS (Hierarchical CFS) with the candidate centres for the decision graph and the second stage is high similarity and connected sub clusters are merging with increasing of dissimilarity between the different clusters for obtaining the final clustering.

3.1 Clustering

Clustering [14, 15] is aimed for the classification of elements as their similarity it is based on the assumption of the cluster centres are in lower of local density

It is sensitive for only themagnitude of the cluster centre is found in the point for assigning the density in the neighbours cluster. The point of the cluster density is higher in the cluster core and the count of local clusters is created to acquisition the local cluster structure. DPC will assigned the points in descending order in the density of cluster label error propagation when the point in the incorrect cluster with the nearest points density on cluster. The cut-off distance

Figure. 1 Basic block diagram of proposed method

will be prescribed based on the experience and the knowledge of datasets that can optimized by the actual class labels in the red point.

3.2 Density peak based clustering algorithm

The algorithm is related with density peak clustering as CFS (Clustering by fast search) and new density CFS [16]. CFS is the thickness of based algorithm top based calculation. Clustering by fast search depends upon two suppositions. The first one focuses may have high density then the second one focuses in generally huge good ways from the closest neighbour along high local density. Certain comprehensive descriptions are given below.

1) Distance matrix D: The Gauss distance between any two points at given dataset constitutes a distance matrix $D=duv$. The symbol d_{uv} indicates the distance between the u-th and v-th points and is computed by the given equation:

$$d_{uv} = \left(\sum_{k=1}^{dim} (x_{uk} - x_{vk})^2 \right)^{\frac{1}{2}} \quad (1)$$

Here dim represents count of point features

2) Pdenotes neighbours average percentage, dc indicates cut-off distance; $p$implies neighbours average percentage also it is set about 1% to 2%. The cut-off distance dcimplies nearby radius it must be computed as:

$$dc = sort(X) \quad (2)$$

where,

$$X = \{d_{uv}|u, v \varepsilon S \text{ and } u < v\} \quad (3)$$

and,

$$cut = round(\|S\|^{*}p) \quad (4)$$

The number of points in S is representing as $\|S\|$ and the Eq. (3) and (4) is the function of inconsistent of MATLAB function.

3) The density of points in the two kernels are estimate by ρ in the cut off and kernel function of the Gaussian kernel the density is,

$$\rho_u = \sum_{v \varepsilon S} X(d_{uv} - d_c) \quad (5)$$

4) The distance of the point is defined as nearest point of the neighbour as,

$$\delta_u = \min_{v \varepsilon S, \rho v > \rho u} d_{uv} \quad (6)$$

5) The decision graph is constructing the coordinate points in the candidate centre of the distance

matrix including cut-off distance. Centres are defined using two thresholds, which are central to the CFS algorithm. Nonetheless, there are other candidate centres additionally to the real centres in this small set.

3.3 Salp swarm optimizer (SSO)

The optimizer used to continuous problem [15]. It begins with the random population and develops the solutions by two stages as exploring and exploitation. In the exploitation the specific solutions are searched the neighbourhood solutions from the existing ones of better solutions. The SSO is used to avoid the local optimum value, determines issues in the real world, flexibility and simplicity and inspired by the swallow behaviour as swarms, physics, humans, etc

The salp chain model as the population and categorizing is two types as leader and followers. The salp in front of the chain is consider as leader and remain salp is the followers. The position of salps is updated for the leader is given below:

$$X_d^1 = \begin{cases} F_d + c_1\big((ub_d - lb_d)c_2 + lb_d\big)c_3 \geq 0 \\ F_d - c_1\big((ub_d - lb_d)c_2 + lb_d\big)c_3 < 0 \end{cases} \quad (7)$$

SSO Pseudo code
Initialize the salp population $x_u(u =1,2,3,....m)$
while(end condition is not satisfied)
Compute each search agent fitness
F= better search agent
Update the equationC_1
for every salp(X_d)
if $u = 1$
update the leading salp position through Eq. (7)
else
update the follower salp position through Eq. (9)
end
end
alter the salps depend on upper and lower bounds of variables
end
return F

The leader salp updates the food source position, where, X_d^1 implies leader salp position on dth dimension. F_d represents food source of dth dimension, ub_d represents upper bound of dth dimension and lb_d denotes lower bound of dth dimensions C_1, C_2, C_3 specifies uneven numbers.

$$c_1 = 2e^{-\left(\frac{4l}{lmax}\right)^2} \tag{8}$$

The parameter c_1 is balanced both exploration and exploitation, where l represents the iteration of current value and l_{max} is the maximum iteration value. The time of the optimization is the iteration of discrepancy the iterations between 0 and 1 consider v = 0 the equation as,

$$x_d^u = \frac{1}{2}\left(x_d^u + x_d^{u-1}\right) \tag{9}$$

Where the $u \geq 2$ and x_d^u the show the partition of u-th follower salp in d-th dimension

3.3.1. Steps

Initialization: The values are initialized for x and the salp value is also initialized.

Evaluate initial parameters: The values of x are evaluated and the x_d Evaluate fitness value is also evaluated. Fitness value is evaluated by the search agent.

Find best solution: The best solution is determined and the $c1$ value is determined. Update the food solution as best solution. Update all the values for searching the agents by using the Eq. 8 and Eq. 9
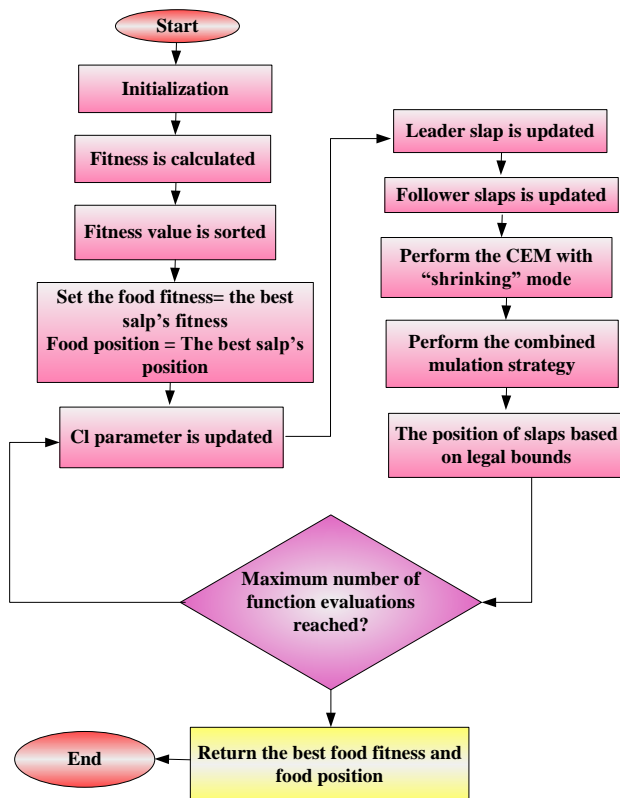

Figure. 2 Block diagram of SSO

3.3 HCFS algorithm

The density peak clustering employs the hierarchical strategy [16], the improvement of CFS is called as HCFS and it represents some features as 1) the reason of the problem is determined in the local clusters, 2) calculate the candidate centres of the sub clusters, 3) merged the estimation in the strategy.

3.3.1. Problem description

The points are misclassified in exists of uneven distribution of the local clusters were shown in the Fig. 3 (a) and (b) the cluster are in ground truth. The analyses are conducted with the misclassification of the procedure. The first 6 points are *A, B, C, D, E,* and *F* for the candidate centres select in the decision graph shows in the Fig. 3 (c). The candidate centres are high density and the respective local sub clusters. After, 6 candidate centres are chosen in cluster results and the marked points of the cluster are star points with similar colour in Fig. 3 (d).

The count of cluster is 2 together with green and brown colour points in Fig. 3 (c) indicate real clusters. So the assignment points are required to the star points *C, D, E* and *F* in the neighbours of the density of the points *I, K, H* and *G* respectively.

The distance of points for *C, D, E* and *F* are calculated as:

$$\delta_C = d_{ci};\ \delta_D = d_{DK}$$
$$\delta_E = d_{EH};\ \delta_F = d_{FG} \tag{10}$$

Where δ is the distance and d is the distance between of two points and the $\delta_C, \delta_D, \delta_E,$ and δ_F are the points of *C, D, E,* and *F*. The point *C* is the nearest neighbour with high density in point *I* the cluster is situated centred at point *A*. The merged cluster of red cluster point C and the brown cluster *A* and then the blue cluster *D*, are merged with brown cluster *A*. The pale blue and the yellow are clustered in the points as *E* with *F* is combining merged with green clustered *B*. The points with the high decision graph are in the point of $\rho;\delta$ with real centres which is due to statistical errors. In the fundamental reason the distribution of local cluster density will occur in the assignment strategy.

3.3.2. Constructing the sub clusters

In this construction of sub clusters the candidate centres find a set of clusters in the neighbours has high density and clustered with the candidate centres. The cluster analysed within the dataset and distributed in order of real centres having high δ with the other density of non-centre points have
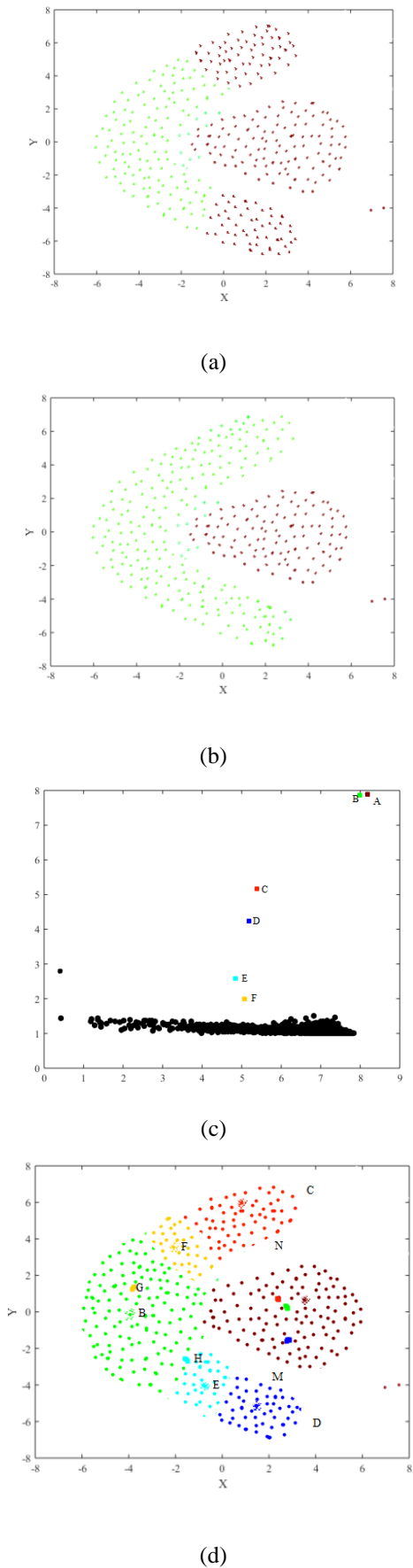
(a)

(b)

(c)

(d)

Figure. 3 Clustering for the CFS algorithm

small δ. The candidate centres are the real centres with the other point that has higher density of δ and ρ be selected candidates. The distribution of cluster is uneven multiple points with high density δ and ρ is besides in the centre. This is the inspiration to identifying the candidate centres. In the decision graph the points are focused on the narrow region depicts in Fig. 3 (c) in this time the set of high density is in above region. The point of the dataset will be fall in the dimensional in less rank space of candidate that are recognized in the outliers thus occurs the datasets in the narrow region with the appropriate candidate centres, candidate is selected by the two thresholds in the decision graph. It will represent the algorithm deals with the uneven distribution within the clusters. The candidates centres is higher than the count of real clusters the candidate centres are assign the similar cluster of the neighbour of higher density. The candidate centres are used to construct the sub clusters in the merging process. The candidate centres will avoid the wrong clusters in the merging time.

In Fig. 3 (c) the coloured non-black points are locate in the narrow region are select as the centre of candidate in the points are in the non-centres. The candidate centres are determined in the remaining point that the clusters are assigned in the nearest neighbours of the higher density. The candidate and the sub clusters are merged with any clusters in the non-classification. The hierarchy based methods, the construction of the sub clusters are the candidate in the centre are determined. It is inconvenient for the method of sub clusters in the complex computation. The fixed number 1.5% for constructing the decision graph resulting will indicates the constructing of sub clusters has the misclassifications.

3.3.3. Merging the sub clusters

Merging the sub cluster is constructed in the algorithm of chameleon in the connectivity and the similarities in between two adjacent sub-clusters are taken in the procedure. The algorithm has its assumption of the two adjacent sub clusters in the same clusters in the intersection region in the dense and distributions for some of the statistical information will be similar. In terms of the inference, the definitions also state that the clusters are determined by a novel strategy that measures the connections and dissimilarity among the sub-clusters and the count of clusters:

1) Intersect between the 2 sub clusters: the intersection set B_{uv} between the sub cluster in B_u and the sub cluster are calculated as,

$$\forall_x \varepsilon B_u, \forall_y \varepsilon B_v, if \ d_{uv} \leq \alpha d_c \qquad (11)$$

Then,

$$x \varepsilon B_{uv}, y \varepsilon B_{uv} \qquad (12)$$

Where d_c represents cut off distance and is computed as the adjacent percentage andα implies adjustable coefficientαd_c represents sub clusters intersection area size

2) Local set of sub cluster for their adjoining sub cluster: the sub clusters$B_u$in the adjacent sub clusters B_v with the related local set B_{uv} are calculated as:

$$\forall_x \varepsilon B_u, \forall_y B_v, if\ d_{xy} \leq d_{mumv} \qquad (13)$$

$$x \varepsilon B_{u-v}, y \varepsilon B_{u-v} \qquad (14)$$

Where m_u is the centre of B_u and m_v indicates centre of B_v, associated local set B_{u-v} is also calculate same time. The algorithm is used to the problem solving of distribution in the local clusters in the hierarchical strategy,
 Compute the distance between total dataset points in distance D.
  • Compute the local distance with density to every point.
  • Make decision graph, for determines the set of centre's candidate based on sub clusters candidate centres.
  • Sub clusters dissimilarity is calculated form lower to higher bounds.

4. Result and discussion

The dataset used in the proposed method is clustering basic benchmark is used to the analysis which was proved in the experimentally by the MATLAB 14.1 platform in the Intel core i3 processor of 3.5 GHZ with computational time of 25 sec to randomly selected data. The performance analysis is done with the different datasets such as Iris, S2, Flame, Star log and Path based [18]. The clustering evaluation is performed with different clustering algorithms and our proposed method select the adjusted mutual information (AMI) and adjusted rand index (ARI). The ARI is used to data clustering resemblance and it

Table 1. ARI performance

Dataset	HCFS	CFS	Proposed
Iris	0.886	0.759	0.789
S2	0.993	0.993	0.993
Flame	1.000	1.000	0.215
Star log	0.998	0.998	0.914
Path Based	0.970	0.453	0.789

Table 2. AMI performance

Dataset	HCFS	CFS	Proposed
Iris	0.871	0.806	0.568
S2	0.994	0.994	0.994
Flame	1.000	1.000	0.362
Star log	0.957	0.957	0.873
Path Based	0.960	0.539	0.772

defined as expected index is reduced from the original index that is divided by the expected index reduces from the maximum index. The equation of the ARI is given below:

$$ARI = \frac{SJ - E(SJ)}{\max(SJ) - E(SJ)} \qquad (16)$$

where SJ denotes the index value$E(SJ)$ is the expected index and $max(SJ)$ is the maximum index value.
AMI is the difference of the clustering comparisons in the theory of information and probabilistic. The effect of the agreement is due to the chance in between of the clusters in the way of ARI related to the information variation.

$$AMI = \frac{J(X,Y) - E\left\{\frac{J(N)}{(c,d)}\right\}}{\sqrt{P(U)P(V)} - E\left\{\frac{J(N)}{(c,d)}\right\}} \qquad (15)$$

Where $\sqrt{P(U)P(V)}$is the maximum value of valid upper bounds of the mutual information and the U and V are the data points in the cluster respectively. The marginal of the contingency table is given as c and d, the $I(M)$ denotes the mutual information in between of the two clustering for consistency of M table.

4.1 Clustering results

The decision graph is plot for the clusters 39. It produces the fuzzy models and has better value for the continuity data when compared to theδ and ρ.. The results give better value for the fuzzy system. The count of points along high$\delta$and ρvalues are taken as the cluster and it is chosen appropriate in number of clusters. The points of same colour are from the same cluster. The decision graph is plot for the clusters 19. It produces the fuzzy models and has better value for the continuity data when compared to the δ and ρ. The results give better value for the fuzzy system. The clustering are the methods for the analysing of data with the proximity data consisting of objects in the same kind either individuals, among variables in the objects for any of single cohesive type. The models in the structure of set of data are graphically represented in the set of points at space.
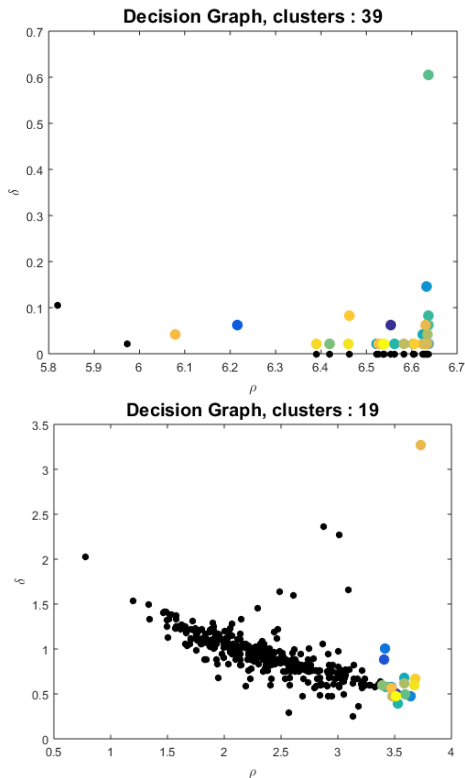
Figure. 5 Non classical multidimensional scaling

Figure. 4 Decision graph

The non-classical multidimensional explore the clusters of spatial resolutions in largest events for the accumulation of the cluster in the feature space for combining the events in precociously and unravel vital for the information in the cluster. The decision graph is plot for the clusters 19 and the graph is given in 2D non-classical scaling of multidimensional value. It produces the fuzzy models and has better value for the continuity data when compared to δ and ρ. The results give better value for the fuzzy system.

The decision is Shown Fig. 5 plot for clusters 16. It produces the fuzzy models and has better value for the continuity data when compared to the δ and ρ. The results give better value for the fuzzy system. The decision is Shown Fig. 6 plot for the clusters 16 and the graph is given in 2D non-classical scaling of multidimensional value. It produces the fuzzy models and has better value for the continuity data when compared to the δ and ρ. The results give better value for the fuzzy system.

The decision is Shown Fig. 7 plot for the clusters 16. It produces the fuzzy models and has better value for the continuity data when compared to the $\boldsymbol{\delta}$ and ρ. The results give better value for the fuzzy system. The first points are selected as candidate features δ and ρ remaining are sub clusters to same cluster in the nearest neighbour of the higher density. The 2D directly plot for the clusters value 10. It produces the fuzzy models and has better value for
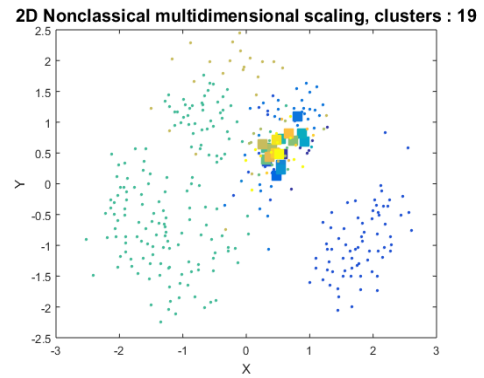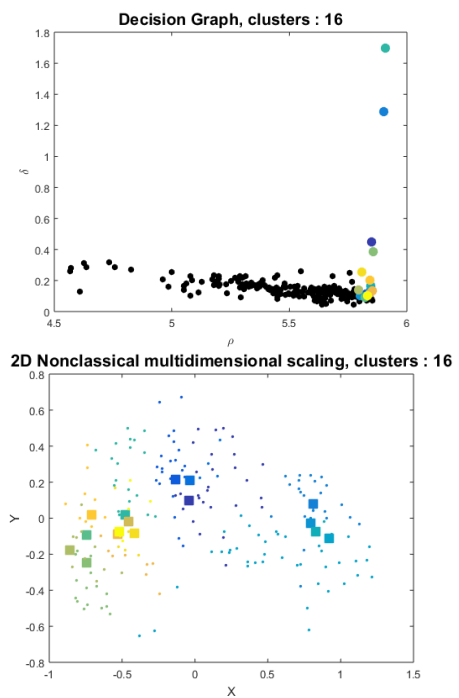

Figure. 6 Decision graph and multidimensional

the continuity data when compared to the X and Y. The results give better value for the fuzzy system. The decision is Shown Fig. 8 plot for the clusters 15. It produces the fuzzy models and has better value for the continuity data when compared to the δ and ρ. The results give better value for the fuzzy system. The 2D directly plot for the clusters value 15. It produces the fuzzy models and has better value for the continuity data when compared to the X and Y. The results give better value for the fuzzy system. The decision graph is plot for the clusters 11. It produces the fuzzy models and has better value for the continuity data when compared to the δ and ρ. The results give better value for the fuzzy system. The 2D directly plot for the clusters value 11. It produces the fuzzy models and has better value for the continuity data when compared to the X and Y. The results give better value for the fuzzy system.
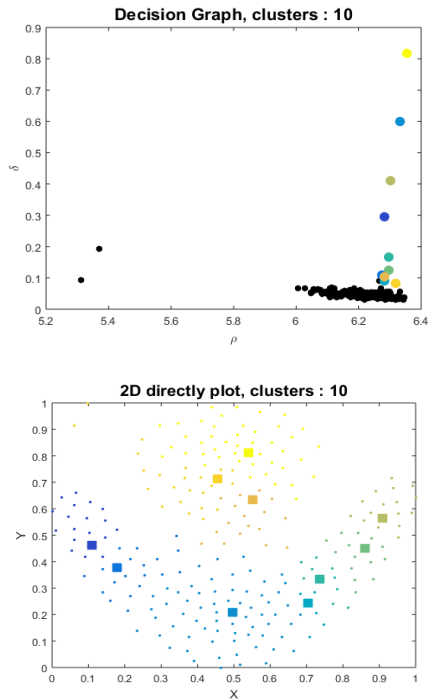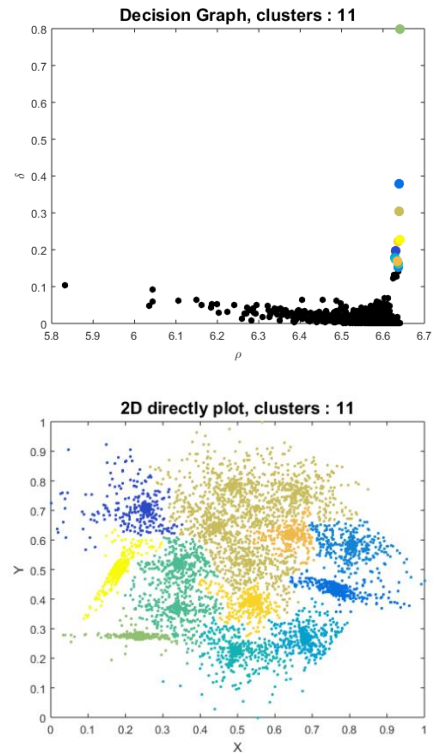
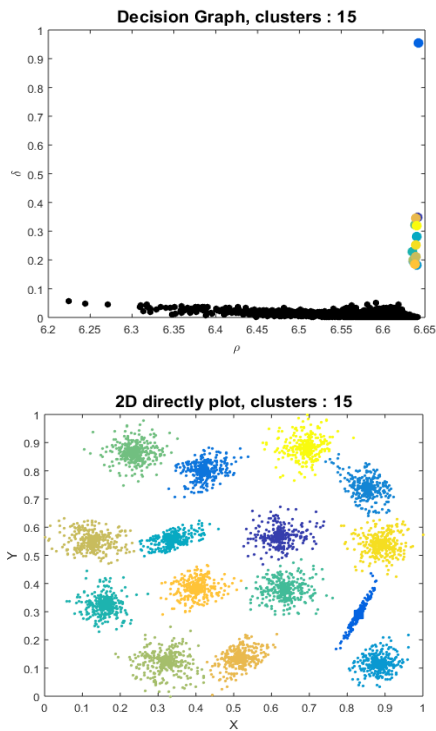Figure. 7 Decision graph and direct plot

Figure. 8 Decision graph and direct plot

4.2 Parameters evaluations

The Figs. 10-15 depicts the accuracy, AMI also ARI values, which are steady to dataset without abruptions changes and with fluctuations. In the figure every line denotes one experiment and single dataset. So the result of the figure gives as the cluster is not randomness and it gives order sensitivity. The
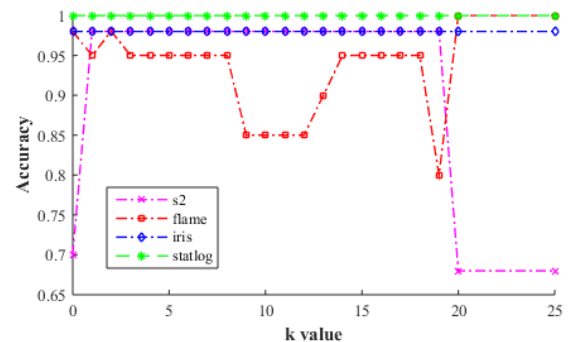

Figure. 9 Decision graph and direct plot

Figure. 10 Comparison between accuracy and k value

dataset chosen is iris, $s2$, star log and flame for the normal ACC (Accuracy), AMI, and ARI. For random order the path based dataset is also include with other dataset

The Fig. 10 shows the accuracy value for the parameters as s2, flame, iris, star log with the k value. The comparison of the data description with k value the star log is 50% higher when compared to each data. The accuracy of clustering can be used to reordering the confusion matrix rows of sum in the diagonal values that is maximum linear problem that implement in the accuracy for the clusters. The count of data points is general in cluster. The sum over all the clusters and divides the total number of data points.

The Fig. 11 shows the AMI value for the parameters as s2, flame, iris, star log with the k value.
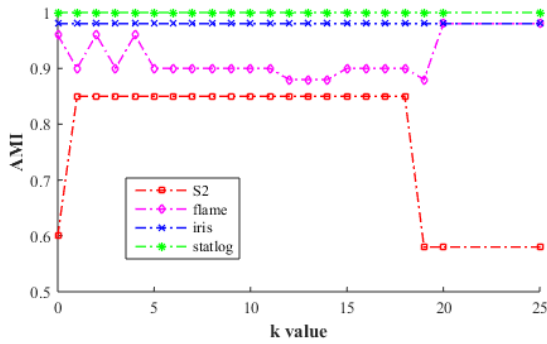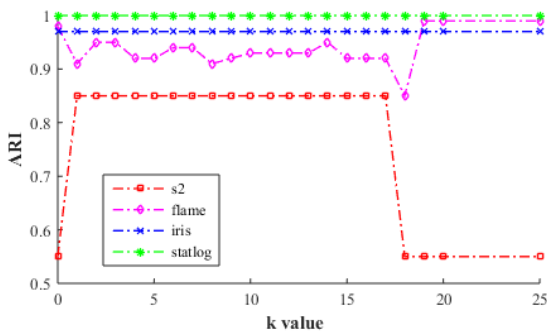
Figure. 11 AMI and k value comparison

Figure. 13 AMI results of different datasets in random order

Figure. 12 Comparison of ARI and k value
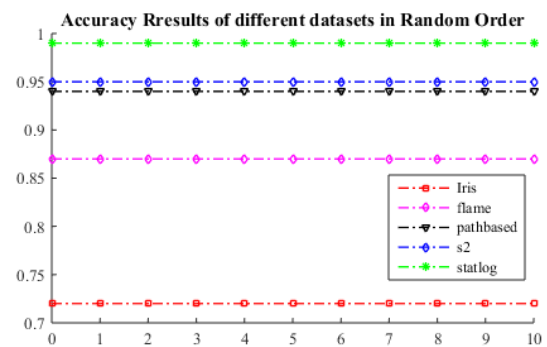
Figure. 14 Accuracy results of different datasets in random order

The comparison of the data description with k value the star log is 50% higher when compared to each data. The mutual description is in the amount clusters between the two clustering's on account of the mutual information fact has high clustering value with the large number of clusters. The metric is not dependent on absolute values of labels in the permutation for cluster label in the score value. The quality of clustering in the average in the coefficient value of all objects in the data set. The count of data points is general in cluster the sum over all the clusters and divides the total numbers of data points.

The Fig. 12 shows the ARI value for the parameters as s2, flame, iris, star log with the k value. The comparison of the data description with k value the star log is 50% higher when compared to each data.

The ARI is limitary by 1 and 0 when the anticipated value is equal to generalized hyper geometric distribution for uneven in the popular measures. There are many popular measures in the class choosing the measures are appropriate with difficult and confusing. It is very near to 0.0 to the independent random value in count of clusters together with the models will be equal to the clustering that are be identical up to permutation.

The Fig. 13 shows the AMI value of the parameters as s2, flame, iris, path based, star log with the k value. The comparison of the data description
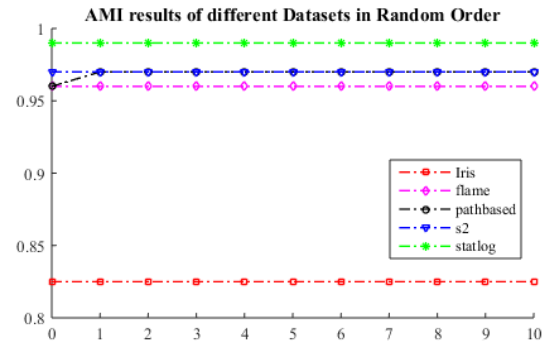
with k value the star log is 65% higher when compared to the s2 data, the path based and the s2 value gets same value. This behaviour is due to the non-standardization variance of the AMI that will need the standardization values for the selection basis.

The Fig. 14 shows the accuracy value of the parameters as s2, flame, iris, path based, star log with the k value. The comparison of the data description with k value the star log is 70% higher when compared to each data. In the path based and the s2 value 25% value closer.

The Fig. 15 shows the ARI value for the parameters as s2, flame, iris, path based, star log with the k value. The comparison of the data description with k value the star log is 60% higher when compared to each data. The random index will be computes the similarity measure in between of two clustering's in the pair of samples and the pairs of counting in the assigned same value of the different clusters in predicted and clustering. ARI needs the ground truth classes accessible in the supervised learning setting, in the unsupervised setting a building block to a consensus index could be utilized to the clustering model selection.

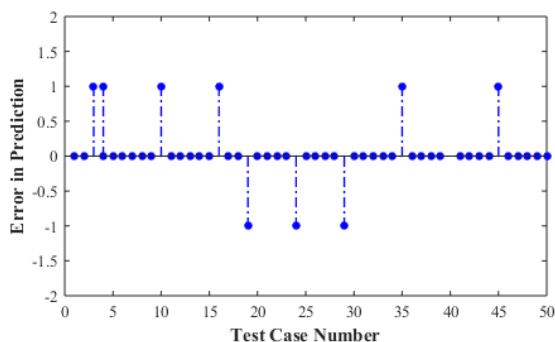Figure. 15 ARI results of different datasets in random order

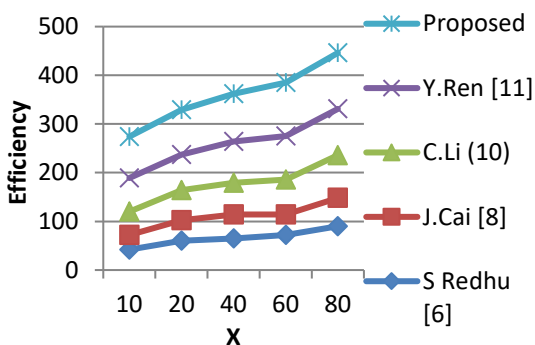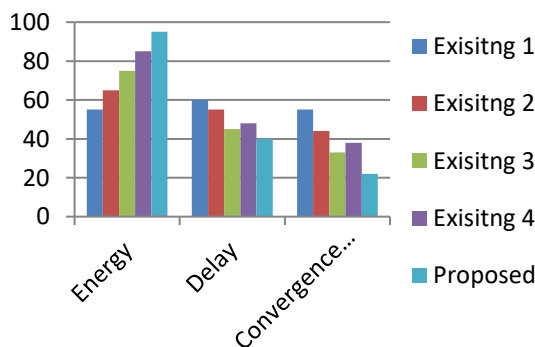Figure. 16 Errors in prediction with test case

Figure. 17 Efficiency

Figure. 18 Comparison of existing methods

The Fig. 16 shows the error prediction of the test case number in the detection for all value the error is positive and for the case 20, 25, 30 the error gets

negative. The RMSE (root mean square error) is predicted in the concrete dataset of test cases as 63 test cases which is equal to 6.32Mpa.

The error prediction in the test case is shown in the Fig. 15 and is computed the deviation value for every predicted value from the strength value of data set.

The Fig. 17 shows the efficiency comparison for the proposed and the existing methods. In this figure the proposed method gets better efficiency when compared to the existing methods as Ren, C. Li, J. Cai, and Redhu. The proposed method gets 96%, S. Redhu gets the value as 17.5% over the proposed methods, Cai gets the value of 35%, C. Li gets the value of 55% and the Ren gets the value of 79% over the proposed methods.

5. Conclusions & future work

Therefore, we have used salp swarm optimization technique with DPC method to cluster the multiple sensor nodes as per the data similitude for data fusion. It consists of two steps. The first step, HCFS assesses the density with distance of each point. Points with high density and distance were selected as candidate centres, after that sub clusters centred on them were acquired. The second step, assuming that nearby sub clusters in terms of some candidate centre were high similar and coupled within the similar cluster, we measure the dissimilarity and connectivity between the sub clusters. The first step was selected as candidate features like δ and ρ the remaining sub clusters to same cluster in the nearest neighbour of the higher density. Very similar and coupled sub-clusters were consolidated to maximize the dissimilarity between different clusters and to get the eventual clustering results. The experimental results show the accuracy as 96% and also the ARI and AMI are compared value for different clusters. The theoretical and experimental results show the better value of dataset comparisons. Finally, since our proposed method could effectively identify random distributed clusters and gives better performance when compared to other different datasets. The results are obtained for the current sensor data fusion approach and were compared with the other dataset count of trial flows to obtain the values of parameter estimation prevailing the two stage fusion system and optimization was important as well as extensive. The problem relevant to study is that the relationship between the various parameters that affect the process of data fusion can be detected and made dependent on them, so that the accuracy of final results can be controlled by manipulating a small number of such parameters. The peak clustering is

combined with other optimization technique for elaborating the accuracy of datasets.

Conflicts of Interest

The authors declare no conflict of interest

Author Contributions

Conceptualization, methodology, writing original draft, preparation: Easpin Brumancia; supervision: Justin Samuel.

References

[1] E. Brumancia, S. Justin Samuel, L. Gladence, and K. Rathan, "Hybrid data fusion model for restricted information using Dempster–Shafer and adaptiveneuro-fuzzy inference (DSANFI) system", *Soft Computing*, Vol. 23, No. 8, pp. 2637-2644, 2019.

[2] A. D. Ribas, J. G. Colonna, C. M. Figueiredo, and E. F. Nakamura, "Similarity clustering for data fusion in wireless sensor networks using k-means", In: *Proc. of the 2012 International Joint Conf.on Neural Networks (IJCNN)*, Brisbane, QLD, Australia, pp. 1-7, 2012.

[3] S. Mirjalili, A. Gandomi, S. Mirjalili, S. Saremi, H. Faris, and S. Mirjalili, "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems", *Advances in Engineering Software*, Vol.114, No. 12, pp. 163-191, 2017.

[4] L. Zhuo, K. Li, B. Liao, H. Li, X. Wei, and K. Li, "HCFS: A Density Peak Based Clustering Algorithm Employing A Hierarchical Strategy", *IEEE Access*, Vol. 7, No. 5, pp. 74612-74624, 2019.

[5] X. Liu, R. Zhu, A. Anjum, J. Wang, H. Zhang, and M. Ma, "Intelligent data fusion algorithm based on hybrid delay-aware adaptive clustering in wireless sensor networks", *Future Generation Computer Systems*, Vol. 104, No. 3, pp. 1-14, 2020.

[6] S. Redhu and R. M. Hegde, "Multi-Sensor Data Fusion for Cluster-based Data Aggregation in IoT Applications", In: *Proc. of IEEE International Conf. On Advanced Networks and Telecommunications Systems (ANTS)*, GOA, India, pp. 1-6, 2019.

[7] X. Shen, Z. Song, H. Fan, and Q. Fu, "Fast density peak-based clustering algorithm for multiple extended target tracking", *Journal of Systems Engineering and Electronics*, Vol. 30, No. 3, pp. 435-447, 2019.

[8] J. Cai, H. Wei, H. Yang, and X. Zhao, "A Novel Clustering Algorithm Based on DPC and PSO", *IEEE Access*, Vol. 8, No. 5, pp. 88200-88214, 2020.

[9] J. Jiang, Y. Chen, D. Hao, and K. Li, "DPC-LG: Density peaks clustering based on logistic distribution and gravitation", *Physica A: Statistical Mechanics and its Applications*, Vol. 514, No. 1, pp. 25-35, 2019.

[10] C. Li and Y. Zhang, "Density Peak Clustering Based on Relative Density Optimization", *Mathematical Problems in Engineering*, Vol. 2020, No. 6, pp. 1-8, 2020.

[11] F. Barranco, C. Fermuller, and E Ros, "Real-time clustering and multi-target tracking using event-based sensors", In: *Proc. of IEEE/RSJ International Conf. on Intelligent Robots and Systems (IROS)*, Madrid, Spain IEEE, pp. 5764-5769, 2018.

[12] J. Fan, W. Xie, and H. Du, "A Robust Multi-Sensor Data Fusion Clustering Algorithm Based on Density Peaks", *Sensors*, Vol. 20, No. 1, pp.238-239, 2019.

[13] P. Zhou, Y. Shen, L. Du, F. Ye, and X. Li, "Incremental multi-view spectral clustering", *Knowledge-Based Systems*, Vol. 174, No. 6, pp. 73-86, 2019.

[14] D. Huang, C. Wang, J. Wu, J. Lai, and C. Kwoh, "Ultra-Scalable Spectral Clustering and Ensemble Clustering", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 32, No. 6, pp. 1212-1226, 2020.

[15] Y. Ren, N. Wang, M. Li, and Z. Xu "Deep density-based image clustering", *Knowledge-Based Systems*, Vol. 197, No. 6, pp. 105841, 2020.

[16] A. Hegazy, M. Makhlouf, and G. El-Tawel, "Improved salp swarm algorithm for feature selection", *Journal of King Saud University - Computer and Information Sciences*, Vol. 32, No. 3, pp. 335-344, 2020.

[17] L. Zhuo, K. Li, B. Liao, H. Li, X. Wei, and K. Li, "HCFS: A Density Peak Based Clustering Algorithm Employing A Hierarchical Strategy", *IEEE Access*, Vol. 7, No. 5, pp. 74612-74624, 2019.

[18] P. Fränti and S. Sieranoja, "K-means properties on six clustering benchmark datasets", *Applied Intelligence*, Vol. 48, No. 1, pp. 4743–4759, 2018.