An improved clustering algorithm for Preshower Detector using fuzzy logic

Susanta Kumar Pal,* Subhasis Chattopadhyay, and Y. P. Viyogi
Variable Energy Cyclotron Centre, 1/AF Bidhan Nagar, Kolkata 700064, INDIA

Introduction

Fuzzy C-Means (FCM) algorithm is already used for finding clusters in a preshower Photon Multiplicity Detector (PMD) in High Energy Heavy Ion Experiments [1]. But clustering precision of this algorithm is affected by its equal partition trend for data sets. The optimum clustering results obtained from the FCM algorithm may not give correct partition for data sets being large discrepency of every class samples. A density function which calculates density of each sample by Gaussian function is introduced in the objective function of FCM and an improved algorithm, weighted Fuzzy C-Means (WFCM) is proposed to solve the problem of equal partion trend retaining favourable convergence and stability of the FCM algorithm. The results obtained from this new algorithm have been compared with the results shown in [1], where FCM is used. This algorithm has also been applied on PMD which is designed to take data at $\sqrt{s_{NN}} = 5.5 TeV$ per nucleon with minimum bias trigger configuration. This data contain larger number of track densities with many overlapping clusters. The performance of this algorithm has been studied in detail with the variation of track densities.

Weighted Fuzzy c-mean Algorithm

A. Weighted matrix

It is common that if one sample is surrounded by other more samples or overlapped with other more samples, the density value of that sample is higher and the sample has a great influence in classification process. The weighted information of the sample in the WFCM alogorithm is defined by a gaussian function z_i in (1).

$$z_i = I_i \sum_{j=1}^{n} e^{-(||x_i - x_j||_A)^2} \qquad (1)$$

$||x_i - x_j||_A \leq r$, i=1,...,n. Where I_i is the intensity value of the pixel i and r is a bound of neighbourhood sample point x_i. Normalzing the density funcltion z_i, the weight matrix W = { w_i, i=1,...,n} becomes

$$w_i = \frac{z_i}{\sum_{j=1}^{n} z_i} \quad i = 1, ..., n. \qquad (2)$$

The optimal fuzzy partition matrix U_{cxn} ={μ_{ik}} and the optimal cluster centre matrix $V_{cxp} = \{v_i\}$ are obtained through (3) and (4), respectively.

$$\mu_{ik} = \left[\sum_{j=1}^{c} \left[\frac{||x_k - v_i||_A}{||x_k - v_j||_A} \right]^{\frac{2}{m-1}} \right]^{-1} \qquad (3)$$

for $1 \leq i \leq c, \quad 1 \leq k \leq n$

$$v_i = \frac{\sum_{k=1}^{n} w_k \mu_{ik}.x_k}{\sum_{k=1}^{n} \mu_{ik}} \quad for \quad i = 1, 2, ..., c \qquad (4)$$

In the WFCM algorithm, the cluster centres v_i in (4) are not the same with the FCM algorithm's in [2]. Fuzzy partion matrix μ_{ik} is also influenced by w_k in the iteration process as it includes $||x_k - v_i||$.

B. Finding Optimum Clusters

To realize FCM or WFCM, fixed number of clusters are to be supplied that will be grouped

*Electronic address: sushant@vecc.gov.in

in the correct manner. In our earlier work [1], the fixed number of clusters that are the optimum clusters have been found out by optimizing the different validity indexes. This is a time comsuming procedure. In the present work, we have used two dimensional high resolution peak searching alogithm to determine the optimum clusters [3]. The merit of the results are related to Raw Efficiency and shown in Fig.1.

I. IMPLEMENTATION OF WFCM ALGORITHM

We apply WFCM algorithm to get the clusters from the data points using the following iterative steps:

1) Fix the degree of fuzziness m=1.8, the convergent threshold *e=0.001*, maximum iteration number T_{max} and generate an initial fuzzy partition matrix U^0.

2) For Event Numbers ev=0 to evMax

3) Get the the number of clusters c from Peak Searching Method described in earlier section.

4) Calcuate weighted matrix w_j from (1) and (2).

5) for iteration number t=0, .. T_{max}

6) Calculate fuzzy partition matrix $\mu_{ij}(t)$.

7) Calculate the cluster centres $v_{i(t)}$ from (4) by $\mu_{ij}(t)$.

8) Update fuzzy partition matrix $\mu_{ij}(t+1)$ from (3).

9) If $||\mu_{ij}(t+1) - \mu_{ij}(t)|| \leq e$, then go to step 2; otherwise go to step 5.

Results and Discussion

A. Clustering Efficiency and Ghost clusters

To justify the merit of WFCM algorithm Raw efficiency, Absolute efficiency and Ghost clusters are considered from [1], and the results are plotted in Fig.1. Raw efficiency ($\epsilon_{\rm raw}$) = $N_{\rm raw}/N_{\rm inp}$, Absolute efficiency ($\epsilon_{\rm abs}$)

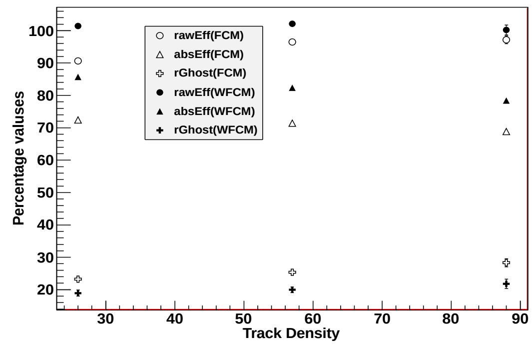

FIG. 1: Comparison of the clustering efficiencies and Ghost clusters obtained from FCM and WFCM for different track densities

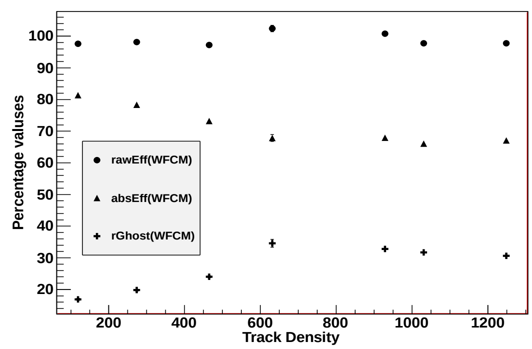

FIG. 2: Clustering efficiency and Ghost clusters obtained from WFCM for different track densities

= $N_{\rm abs}/N_{\rm inp}$. The result shows that the performance of WFCM is 10-15 % better than FCM in all track dencities. Fig.2 justifies the WFCM result for much higher track densities.

References

[1] Susanta Kumar Pal et al., Nucl. Inst. Meth. Phys Res.**A 626-627 (2011) 105-113**.

[2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithm, Plenum, New York **(1981);**.

[3] M.Morhac et al.,Nucl. Inst. Meth. Phys Res.**443(2000), 108-125**.