ference on Physics
ark Gluon Plasma

ICPAQGP-2010

Contribution ID : 138

Fuzzy Clustering for High Energy Physics Data

Content :

In high energy physics experiments, various types of detectors are used to measure the characteristics of a large number of particles emitted in collisions. In order to extract information and reconstruct the characteristics of the various incoming particles, clustering is required to be performed on the detector's plane.

Hard clustering techniques such as Local Maxima Search, Connected-cell Search and K-means Clustering simply assign a data point to a cluster. A data point either lies in a cluster or it does not, and so, overlapping clusters are hardly distinguishable.

Fuzzy c-Means (FCM) clustering [1] is a version of the k-means algorithm that incorporates fuzzy logic so that each point has a weak or strong association to the cluster, determined by the inverse distance to the centre of the cluster. The term fuzzy is used because an observation may in fact lie in more than one cluster simultaneously, though to different degrees called 'memberships', as is the case with many high energy physics applications. The centres obtained using the FCM algorithm are based on the geometric locations of the data points. In order to determine how well a set of cluster centres fits the data, a number of validity indices have been introduced in the literature [2].

Clustering performed on simulated clusters using the FCM algorithm and a validity index succeeded in identifying clusters that were uniformly spread, irrespective of whether they were overlapping.

However, for non-uniformly spread data, FCM could not easily identify the clusters. A dynamic version of the algorithm, called dynamic Fuzzy c-Means (dFCM) clustering [3], was studied. It allows data points to stream in one at a time and generates or eliminates clusters as required, thereby adapting to the pattern of the data. The dFCM algorithm could easily identify clusters that both overlapped and were non-uniformly spread.

The dFCM algorithm has been applied to high energy physics data, performing better than current clustering methods used in this field. The details of the dFCM clustering technique along with the results will be presented.

References:
[1] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum, 1981.
[2] N.R. Pal and J. C. Bezdek,"On cluster validity for the fuzzy c-means model" IEEE Trans. Fuzzy Systems, vol. 3, no. 3, 1995.
[3] R.P.Sandhir and S. Kumar,"Dynamic Fuzzy c-Means (dFCM) Clustering for Continuously Varying Data Environments" Proc. IEEE World Congress on Computational Intelligence, Barcelona, July 18-23, 2010.

Primary authors : Mrs. SANDHIR, Radha Pyari (VECC)

Co-authors : Mr. MUHURI, Sanjib (VECC) ; Dr. NAYAK, Tapan K. (VECC)

Presenter : Mrs. SANDHIR, Radha Pyari (VECC) ; Mr. MUHURI, Sanjib (VECC)

Session classification : --not yet classified--

Track classification : --not yet classified--

Type : --not specified--