A report on fuzzy c-means using Jeffreys-divergence based similarity measure

Mathav Raj - 194102311

***Abstract*— The proposed FCM uses Jeffrey's divergence to find soft assignments of each data point to the clusters and is found to be a better similarity measure than the minkowski distances especially for clusters having different densities and sizes. The algorithm converges to a local minima of the objective function which is based on overall within-group sum of squared errors. Original paper link : Fuzzy c-means clustering using Jeffreys-divergence based similarity measure**

I. INTRODUCTION

Clustering is unsupervised machine learning that finds groups of data that are close together within a group and dissimilar to data in other groups based on a similarity measure. Traditional FCM employing euclidean distance measures the distance between each sample and every group centroid. Intuitively, the Euclidean distance implies that each sample is equally important and independent from others which may not be always true in many real world problems. The FCM does not take the density distribution of a data set into account, since it evaluates only the distance between two individual sample points, and such evaluation may degrade the performance of FCM for the data with uneven densities. Non-linear similarity measures such as Jeffrey's divergence identify more accurate cluster boundaries.

II. TERMINOLOGY AND ALGORITHM

A. *Jeffrey's divergence*

Divergence is a weaker distance measure between two probability distributions. The proposed similarity measure is among the family of f-divergences which means it is characterised by the function $\ln \frac{p(x)}{q(x)}$ where $p(x)$ and $q(x)$ represent two distributions but it is not a Bregman divergence owing to the function's concaveness. Jeffrey's divergence is symmetric which facilitates lesser computations. Jeffrey's divergence between any two data points **a** and **b**, where **a**, **b** $\in \mathbb{R}^{\ltimes}_{+}$ as,

$$S(a,b) = \sum_{i=1}^{n}(a_i - b_i)log(\frac{a_i}{b_i}) \tag{1}$$

Unlike the concentric circle contoured norm ball of euclidean measure, the contour plot of Jeffrey divergence is somewhat like distorted ovals. Jeffrey's divergence between two points is higher when they are near to origin and it decreases while they are away from the origin. On the other hand, Euclidean distance between two points are same irrespective of the position. This property would be useful when clusters are having different densities and sizes.

B. *Algorithm*

Consider a set of data points $O[o_1, o_2, ..., o_m]$. Clustering is splitting of O into c groups of homogeneous data points, $K[(k_1, k_2, ..., k_c)]$ in such a way that there is degree of strong association within group and weak association between different groups and $1 < c < m$. Here, data point, o_i is expressed using d dimensional feature vector, o_i in $\mathbb{R}^{\ltimes}_{+}$, where i varies from 1 to m.
$k_i \neq \emptyset \quad for \quad i = 1, .., c$
$k_i \cap k_j = \emptyset \quad for \quad i, j = 1, .., c \quad and \quad i \neq j$
$\bigcup_{i=1}^{c} k_i = K$
If λ_{xy} represents the membership of o_y to the centroid k_x, the membership of all the data points to different centroids is stored in a matrix $D(O)_{(cxm)}$. f, a real quantity between 1 and ∞ is the fuzziness coefficient. If we increase f then the partition becomes fuzzier. Fuzzy c means partitioning is achieved by optimizing the following objective function,

$$\underset{\substack{K \in \mathbb{R}^{\gtrdot\frown} \\ D \in M}}{min} E_f = \sum_{y=1}^{m}\sum_{x=1}^{c}(\lambda_{xy})^f S(o_y, k_x) \,, 1 \leq S < \infty$$

$$M = \left\{ D = \lambda_{xy} \mid \lambda_{xy} \in [0,1], \sum_{x=1}^{c} \lambda_{xy} = 1, \sum_{y=1}^{m} \lambda_{xy} > 0 \right\}$$

Algorithm 1 Modified Fuzzy C Means

1: *Initialization: Randomly initialize the membership matrix D such that it belongs to M*
2: *Iteration i:*
   a: *membership values:*

$$\tau_y = \{x \mid x \in [1, c], o_y = k_x^{(i)}\}$$

$$\lambda_{xy}^{i+1} = \begin{cases} \left(\sum_{l=1}^{c} \left[\frac{S(o_y, k_x^i)}{S(o_y, k_l^i)}\right]^{\frac{2}{f-1}}\right)^{-1} & , if\ \tau_y = 0 \\ \frac{1}{|\tau_y|}, if\, \tau_y \neq 0\ and\ x \in \tau_y \\ 0, if\ \tau_y \neq 0\ and\ x \notin \tau_y \end{cases}$$

   b: *centroids:*

$$k_x^{i+1} = \frac{\sum_{y=1}^{m} \left[\lambda_{xy}^{i+1}\right]^f .v_y}{\sum_{y=1}^{m} \left[\lambda_{xy}^{i+1}\right]^f}$$

3: *Termination:* $max_{xy}\{\mid \lambda_{xy}^{i+1} - \lambda_{xy}^{i} \mid\} < \epsilon \,, \epsilon \in (0, 1)$

C. Characteristics of the resulting clusters:

- The goodness of resulting cluster is studied using cluster validity indexes, external(Normalized Mutual Information, Adjusted Rand Index) and internal(Silhouette index, Dunn index, Davies Boulden index). The external indices measures the resemblance to ground truth. Internal indices measure how similar a data point is to its own group (cohesion) compared to other groups (separation). A value closer to 1 for NMI, ARI and SI, a higher DI and lower DBI means a good performance.The proposed FCM with Jeffrey's divergence outperforms FCM with euclidean and minkowksi weighted distance measures.
- Assuming a null hypothesis that the there is no statistically significant variation existing between the median values of ARI, NMI, SI, DI and DBI of the proposed FCM and other algorithms taken one at a time, we find that the p values so found are less than 0.05 significance which gives us enough evidence to reject the null hypothesis and conclude that the better median values of the clustering validity indexes generated by proposed FCM is statistically significant and it does not happen by chance
- It gives best results for overlapped data and creating better cluster boundaries that have varying densities better than euclidean measure

D. Drawbacks:

- Though a good clustering technique it does have the common drawback of a general clustering algorithm, requiring to specify the number of clusters before hand
- With lower value of ϵ we get the better result but at the expense of more number of iterations.

III. CONCLUSION AND FURTHER STUDIES

In this study, a new distance metric for FCM on $\mathbb{R}_+^{\ltimes}$ has been addressed using Jeffrey's divergence. The study of data complexity metrics is an promising area of research in the field of clustering. It deserves further study. So, we would focus on the analysis of several data set characteristics to retrieve information from them and this could further be considered to design the proper clustering algorithm.