Theoretical Model for a Special Type of Distance Based Clustering of Feature Points Based on Distance to Complement Feature Point or Orthogonal Feature Point of Each Feature Point

Ramesh Chandra Bagadi

Fellow, Institution of Engineers (F124-1135) MIG-905, Mithilapuri Colony, VUDA Layout, Madhurawada, Visakhapatnam 530041, Andhra Pradesh State, India

Abstract- **In this research investigation, the author has detailed a scheme for a special type of distance based clustering of feature points of concern based on distance to complement feature point or orthogonal feature point of each feature point.**

Index terms- **Clustering, Distance Based Clustering**

I.INTRODUCTION

There have been many propositions regarding Clustering Models, the major among them being [1], [2], [3], [4], [5]. Also, there have been a few propositions on Overlapping Clustering Models [6], [7].

II. PROPOSED THEORETICAL MODEL

Notion of the Complement of A Given Vector

For any given Vector $A=[x_1 \quad x_2 \quad x_3 \quad . \quad . \quad x_{n-1} \quad x_n]$ the Complement of this Vector is given by filling it with the Complement of each element w.r.t all other elements of the vector.

That is, the complement of x_i, namely x_i^c is

Case 1: Only Complement the Weighted Average of all other elements of this Vector except x_i

Case 2: Orthogonal Complement the Weighted Average of all other elements of this Vector except x_i, with a Sign to be fixed as follows:

If $\sum_{i=1}^{j} x_i x_i^c$ is Positive, then $x_{(j+1)}^c$ is chosen such that $x_{(j+1)}x_{(j+1)}^c$ is Negative. And if, $\sum_{i=1}^{j} x_i x_i^c$ is Negative, then $x_{(j+1)}^c$ is chosen such that $x_{(j+1)}x_{(j+1)}^c$ is Positive. Also, $1 < j \leq (n-1)$.

That is, $A^c = \left[x_1^c \quad x_2^c \quad x_3^c \quad . \quad . \quad x_{n-1}^c \quad x_n^c\right]$.

The weight is given by

$$w_i = \left\{ \frac{x_i}{\sum_{i=1}^{n} x_i} \right\}$$

$$x_{i=p}^c = \frac{\sum_{\substack{i=1 \\ i \neq p}}^{n} w_i x_i}{(n-1)}$$

Notion of the Complement of A Given n Dimensional Matrix

The Complement of any element $A(p_1, p_2, p_3, ..., p_{n-1}, p_n)$ of an n Dimensional Matrix A with dimension sizes $l_1, l_2, l_3, ..., l_{n-1}, l_n$ is given as follows:

Case 1: Simple Complement

The required value is given by

$$\left\{ \frac{\sum_{\substack{i_n=1 \\ i_n \neq p_n}}^{l_n} \sum_{\substack{i_{n-1}=1 \\ i_{n-1} \neq p_{n-1}}}^{l_{n-1}} ......\sum_{\substack{i_3=1 \\ i_3 \neq p_3}}^{l_3} \sum_{\substack{i_2=1 \\ i_2 \neq p_2}}^{l_2} \sum_{\substack{i_1=1 \\ i_1 \neq p_1}}^{l_1} \left\{ w_{i_1 i_2 i_3 \ldots i_{n-1} i_n} a_{i_1 i_2 i_3 \ldots i_{n-1} i_n} \right\}}{\left(\prod_{i=1}^{n} l_i \right) - 1} \right\}$$

Where the weight term is given by

$$w_{i_1 i_2 i_3 \ldots i_{n-1} i_n} = \left\{ \frac{a_{i_1 i_2 i_3 \ldots i_{n-1} i_n}}{\sum_{i_n=1}^{l_n} \sum_{i_{n-1}=1}^{l_{n-1}} \ldots\ldots \sum_{i_3=1}^{l_3} \sum_{i_2=1}^{l_2} \sum_{i_1=1}^{l_1} \left\{ a_{i_1 i_2 i_3 \ldots i_{n-1} i_n} \right\}} \right\}$$

Case 2: Orthogonal Complement

The required value is given by

$$\left\{ \frac{\sum_{\substack{i_n=1 \\ i_n \neq p_n}}^{l_n} \sum_{\substack{i_{n-1}=1 \\ i_{n-1} \neq p_{n-1}}}^{l_{n-1}} \ldots\ldots \sum_{\substack{i_3=1 \\ i_3 \neq p_3}}^{l_3} \sum_{\substack{i_2=1 \\ i_2 \neq p_2}}^{l_2} \sum_{\substack{i_1=1 \\ i_1 \neq p_1}}^{l_1} \left\{ w_{i_1 i_2 i_3 \ldots i_{n-1} i_n} a_{i_1 i_2 i_3 \ldots i_{n-1} i_n} \right\}}{\left(\prod_{i=1}^{n} l_i \right) - 1} \right\}$$

Where the weight term is given by

$$w_{i_1 i_2 i_3 \ldots i_{n-1} i_n} = \left\{ \frac{a_{i_1 i_2 i_3 \ldots i_{n-1} i_n}}{\sum_{i_n=1}^{l_n} \sum_{i_{n-1}=1}^{l_{n-1}} \ldots\ldots \sum_{i_3=1}^{l_3} \sum_{i_2=1}^{l_2} \sum_{i_1=1}^{l_1} \left\{ a_{i_1 i_2 i_3 \ldots i_{n-1} i_n} \right\}} \right\}$$

And the sign of the term

$$\left\{ \frac{\sum_{\substack{i_n=1 \\ i_n \neq p_n}}^{l_n} \sum_{\substack{i_{n-1}=1 \\ i_{n-1} \neq p_{n-1}}}^{l_{n-1}} \ldots\ldots \sum_{\substack{i_3=1 \\ i_3 \neq p_3}}^{l_3} \sum_{\substack{i_2=1 \\ i_2 \neq p_2}}^{l_2} \sum_{\substack{i_1=1 \\ i_1 \neq p_1}}^{l_1} \left\{ w_{i_1 i_2 i_3 \ldots i_{n-1} i_n} a_{i_1 i_2 i_3 \ldots i_{n-1} i_n} \right\}}{\left(\prod_{i=1}^{n} l_i \right) - 1} \right\}$$

is given as follows:

If $\sum_{i_n=1}^{j_n} \sum_{i_{n-1}=1}^{j_{n-1}} \ldots\ldots \sum_{i_3=1}^{j_3} \sum_{i_2=1}^{j_2} \sum_{i_1=1}^{j_1} \left\{ a_{i_1 i_2 i_3 \ldots i_{n-1} i_n} b_{i_1 i_2 i_3 \ldots i_{n-1} i_n} \right\}$ is Positive, then the sign of $b_{(j_1+1)(j_2+1)(j_3+1)\ldots\ldots(j_{n-1}+1)(j_n+1)}$ is chosen such that $a_{(j_1+1)(j_2+1)(j_3+1)\ldots\ldots(j_{n-1}+1)(j_n+1)} b_{(j_1+1)(j_2+1)(j_3+1)\ldots\ldots(j_{n-1}+1)(j_n+1)}$ is Negative.

And if $\sum_{i_n=1}^{j_n} \sum_{i_{n-1}=1}^{j_{n-1}} \ldots\ldots \sum_{i_3=1}^{j_3} \sum_{i_2=1}^{j_2} \sum_{i_1=1}^{j_1} \left\{ a_{i_1 i_2 i_3 \ldots i_{n-1} i_n} b_{i_1 i_2 i_3 \ldots i_{n-1} i_n} \right\}$ is Negative, then the sign of $b_{(j_1+1)(j_2+1)(j_3+1)\ldots\ldots(j_{n-1}+1)(j_n+1)}$ is chosen such that $a_{(j_1+1)(j_2+1)(j_3+1)\ldots\ldots(j_{n-1}+1)(j_n+1)} b_{(j_1+1)(j_2+1)(j_3+1)\ldots\ldots(j_{n-1}+1)(j_n+1)}$ is Positive. It should be noted that here, $b_{i_1 i_2 i_3 \ldots i_{n-1} i_n}$ represents the Complement of the Matrix Element of A, namely $a_{i_1 i_2 i_3 \ldots i_{n-1} i_n}$ in the Complement Matrix B which is the complement of Matrix A.

Special Type of Distance Based Clustering Using Distance to Complement of the Feature Point

Let there be m number of feature points each of n dimensions. Let them be represented by $\overline{x}_p$, where $p = 1 \ to \ m$. Also, let the elements of the feature points be represented by x_{pq}, where $p = 1 \ to \ m$ and $q = 1 \ to \ n$. We now find the weighted average of all these feature points which is just the feature point gotten by taking the weighted averages element-wise as follows

$${}^{r}x_q = \frac{\sum_{p=1}^{m} w_{pq} x_{pq}}{\sum_{p=1}^{m} w_{pq}}$$

with

$$w_{pq} = \frac{x_{pq}}{\sum_{p=1}^{m} x_{pq}}$$

This weighted average point is represented by ${}^{r}\overline{x}$ indicating that it is the most representative point for all the given feature points. Its elements are represented by ${}^{r}x_q$ where $q = 1 \ to \ n$.

We now find the distances between this most representative point ${}^{r}\overline{x}$ and each of all other feature points. Let these be represented by $d\left(\overline{x}_p, {}^{r}\overline{x}\right)$ for $p = 1 \ to \ m$. We now arrange these distances in increasing order. Let this order be a function f_1 given by a map from the Set $\{p\}_{p=1 \ to \ m}$ to the same Set $\{p\}_{p=1 \ to \ m}$ but with the possibility that the map need not be necessarily congruent but such that the increasing order of distances aspect is satisfactorily met. Let these distances be denoted by

$$d_1\left(\overline{x}_{p=f_1^{-1}(1)}, {}^{r}\overline{x}\right), d_2\left(\overline{x}_{p=f_1^{-1}(2)}, {}^{r}\overline{x}\right), d_3\left(\overline{x}_{p=f_1^{-1}(3)}, {}^{r}\overline{x}\right), \ldots\ldots, d_m\left(\overline{x}_{p=f_1^{-1}(m)}, {}^{r}\overline{x}\right)$$

Now, if we need K number of clusters, we find the first K Number of points that are closest to the most representative point ${}^{r}\bar{x}$. That is, we consider the points $\bar{x}_{p=f_1^{-1}(1)}$, $\bar{x}_{p=f_1^{-1}(2)}$, $\bar{x}_{p=f_1^{-1}(3)},\ldots\ldots,\bar{x}_{p=f_1^{-1}(K-1)},\bar{x}_{p=f_1^{-1}(K)}$. Now, we consider each of these K points and find the distances to their respective *Complement* points. Let these be represented by

$$g_1 = d_1\left(\bar{x}_{p=f_1^{-1}(1)},\bar{x}^c_{p=f_1^{-1}(1)}\right), g_2 = d_2\left(\bar{x}_{p=f_1^{-1}(2)},\bar{x}^c_{p=f_1^{-1}(2)}\right), g_3 = \left(\bar{x}_{p=f_1^{-1}(3)},\bar{x}^c_{p=f_1^{-1}(3)}\right),\ldots$$
$$\ldots, g_{K-1} = d_{K-1}\left(\bar{x}_{p=f_1^{-1}(K-1)},\bar{x}^c_{p=f_1^{-1}(K-1)}\right), g_K = d_K\left(\bar{x}_{p=f_1^{-1}(K)},\bar{x}^c_{p=f_1^{-1}(K)}\right)$$

We now arrange these distances in increasing order. Let this order be a function f_2 given by a map from the Set $\{g_h\}_{h=1\ to\ K}$ to the same Set $\{g_h\}_{h=1\ to\ K}$ but with the possibility that the map need not be necessarily congruent but such that the increasing order of distances aspect is satisfactorily met.

Let these distances be represented by $g_{h=f_2^{-1}(1)}, g_{h=f_2^{-1}(2)}, g_{h=f_2^{-1}(3)},\ldots\ldots, g_{h=f_2^{-1}(K-1)}, g_{h=f_2^{-1}(K)}$

We now consider the distance $g_{h=f_2^{-1}(1)}$ and the point corresponding to it, namely $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(1)\right)}$ and find all points that bear distance less than or equal to the distance $g_{h=f_2^{-1}(1)}$. Now these points along with the point $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(1)\right)}$ comprise the First Cluster.

We now consider the distance $g_{h=f_2^{-1}(2)}$ and the point corresponding to it, namely $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(2)\right)}$ and find all points that bear distance greater than or equal to the distance $g_{h=f_2^{-1}(1)}$ and less than or equal to the distance $g_{h=f_2^{-1}(2)}$. Now these points along with the point $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(2)\right)}$ comprise the Second Cluster.

We now consider the distance $g_{h=f_2^{-1}(3)}$ and the point corresponding to it, namely $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(3)\right)}$ and find all points that bear distance greater than or equal to the distance $g_{h=f_2^{-1}(2)}$ and less than or equal to the distance $g_{h=f_2^{-1}(3)}$. Now these points along with the point $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(3)\right)}$ comprise the Third Cluster. In this fashion, we find all K number of Clusters. It should be noted that these Clusters may be Overlapping in nature.

In this fashion, we can even find *m* number of Clusters for the given *m* number of feature points.

Special Type of Distance Based Clustering Using Distance to Orthogonal Complement of the Feature Point

Let there be *m* number of feature points each of *n* dimensions. Let them be represented by $\bar{x}_p$, where $p = 1\ to\ m$. Also, let the elements of the feature points be represented by x_{pq}, where $p = 1\ to\ m$ and $q = 1\ to\ n$. We now find the weighted average of all these feature points which is just the feature point gotten by taking the weighted averages element-wise as follows

$${}^{r}x_q = \frac{\sum_{p=1}^{m} w_{pq} x_{pq}}{\sum_{p=1}^{m} w_{pq}}$$

With
$$w_{pq} = \frac{x_{pq}}{\sum_{p=1}^{m} x_{pq}}$$

This weighted average point is represented by ${}^{r}\bar{x}$ indicating that it is the most representative point for all the given feature points. Its elements are represented by ${}^{r}x_q$ where $q = 1\ to\ n$.

We now find the distances between this most representative point ${}^{r}\bar{x}$ and each of all other feature points. Let these be represented by $d\left(\bar{x}_p, {}^{r}\bar{x}\right)$ for $p = 1\ to\ m$. We now arrange these distances in increasing order. Let this order be a function f_1 given by a map from the Set $\{p\}_{p=1\ to\ m}$ to the same

Set $\{p\}_{p=1\ to\ m}$ but with the possibility that the map need not be necessarily congruent but such that the increasing order of distances aspect is satisfactorily met. Let these distances be denoted by $d_1\left(\bar{x}_{p=f_1^{-1}(1)}, {}^r\bar{x}\right), d_2\left(\bar{x}_{p=f_1^{-1}(2)}, {}^r\bar{x}\right), d_3\left(\bar{x}_{p=f_1^{-1}(3)}, {}^r\bar{x}\right), \ldots\ldots, d_m\left(\bar{x}_{p=f_1^{-1}(m)}, {}^r\bar{x}\right)$

Now, if we need K number of clusters, we find the first K Number of points that are closest to the most representative point ${}^r\bar{x}$. That is, we consider the points $\bar{x}_{p=f_1^{-1}(1)}, \bar{x}_{p=f_1^{-1}(2)}, \bar{x}_{p=f_1^{-1}(3)}, \ldots\ldots, \bar{x}_{p=f_1^{-1}(K-1)}, \bar{x}_{p=f_1^{-1}(K)}$.

Now, we consider each of these K points and find the distances to their respective *Orthogonal Complement* points. Let these be represented by $g_1 = d_1\left(\bar{x}_{p=f_1^{-1}(1)}, \bar{x}^{oc}_{p=f_1^{-1}(1)}\right), g_2 = d_2\left(\bar{x}_{p=f_1^{-1}(2)}, \bar{x}^{oc}_{p=f_1^{-1}(2)}\right), g_3 = \left(\bar{x}_{p=f_1^{-1}(3)}, \bar{x}^{oc}_{p=f_1^{-1}(3)}\right), \ldots$
$\ldots, g_{K-1} = d_{K-1}\left(\bar{x}_{p=f_1^{-1}(K-1)}, \bar{x}^{oc}_{p=f_1^{-1}(K-1)}\right), g_K = d_K\left(\bar{x}_{p=f_1^{-1}(K)}, \bar{x}^{oc}_{p=f_1^{-1}(K)}\right)$

We now arrange these distances in increasing order. Let this order be a function f_2 given by a map from the Set $\{g_h\}_{h=1\ to\ K}$ to the same Set $\{g_h\}_{h=1\ to\ K}$ but with the possibility that the map need not be necessarily congruent but such that the increasing order of distances aspect is satisfactorily met.

Let these be represented by $g_{h=f_2^{-1}(1)}, g_{h=f_2^{-1}(2)}, g_{h=f_2^{-1}(3)}, \ldots\ldots\ldots, g_{h=f_2^{-1}(K-1)}, g_{h=f_2^{-1}(K)}$

We now consider the distance $g_{h=f_2^{-1}(1)}$ and the point corresponding to it, namely $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(1)\right)}$ and find all points that bear distance less than or equal to the distance $g_{h=f_2^{-1}(1)}$. Now these points along with the point $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(1)\right)}$ comprise the First Cluster.

We now consider the distance $g_{h=f_2^{-1}(2)}$ and the point corresponding to it, namely $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(2)\right)}$ and find all points that bear distance greater than or equal to the distance $g_{h=f_2^{-1}(1)}$ and less than or equal to the distance $g_{h=f_2^{-1}(2)}$. Now these points along with the point $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(2)\right)}$ comprise the Second Cluster.

We now consider the distance $g_{h=f_2^{-1}(3)}$ and the point corresponding to it, namely $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(3)\right)}$ and find all points that bear distance greater than or equal to the distance $g_{h=f_2^{-1}(2)}$ and less than or equal to the distance $g_{h=f_2^{-1}(3)}$. Now these points along with the point $\bar{x}_{p=f_1^{-1}\left(h=f_2^{-1}(3)\right)}$ comprise the Third Cluster. In this fashion, we find all K number of Clusters. It should be noted that these Clusters may be Overlapping in nature.

In this fashion, we can even find *m* number of Clusters for the given *m* number of feature points.

III. ACKNOWLEDGMENT

The author would like to express his deepest gratitude to all the members of his Loving Family, Respectable Teachers, En-Dear-Able Friends, Inspiring Social Figures, Highly Esteemed Professors, Reverence Deserving Deities that have deeply contributed in the formation of the necessary scientific temperament and the social and personal outlook of the author that has resulted in the conception, preparation and authoring of this research manuscript document. The author pays his sincere tribute to all those dedicated and sincere folk of academia, industry and elsewhere who have sacrificed a lot of their structured leisure time and have painstakingly authored treatises on Science, Engineering, Mathematics, Art and Philosophy covering all the developments from time immemorial until then, in their supreme works. It is standing on such treasure of foundation of knowledge, aided with an iota of personal god-gifted creativity that the author bases his foray of wild excursions into the understanding of natural phenomenon and forms new premises and scientifically surmises plausible laws. The author strongly reiterates his sense of gratitude and infinite indebtedness to all such 'Philosophical Statesmen' that are evergreen personal librarians of Science, Art, Mathematics and Philosophy.

REFERENCES

[1] Rui Xu, 2005. Survey of Clustering Algorithms. IEEE Transactions On Neural Networks, Vol. 16, No. 3, May 2005

[2] Dongkuan Xu, et. al., 2015. A Comprehensive Survey of Clustering Algorithms, Ann. Data. Sci.

(2015) 2(2):165–193 DOI 10.1007/s40745-015-0040-1

[3] Jain A, Dubes R (1988) Algorithms for clustering data. Prentice-Hall, Inc, Upper Saddle River

[4] Everitt B, Landau S, Leese M (2001) Clustering analysis, 4th edn. Arnold, London

[5] MacQueen J (1967) some methods for classification and analysis of multivariate observations. Proc Fifth Berkeley Symp Math Stat Probab 1:281–297

[6] Said Baadel, et. al. (2016), Overlapping Clustering: A Review, SAI Computing Conference 2016 July 13-15, 2016 | London, UK

[7] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition, Wiley, 1999.