ISSN 1512-1232

Formation of the Algorithmic Similarity Measures for Recognition Processes

Otar Verulava, Taliko Zhvania, Tea Todua

Georgian Technical University
verulava@gtu.ge, talizhvania@gmail.com, tea_todua@gtu.ge

Abstract

In the article there are methods of realization of the recognition process described in the theory of rank of links according to the clustering drawn parameters, where clustering surface points are used for etalon descriptions, which are divided by positive and antipodal surface points. In this work there are methods of drawing etalon descriptions and its appropriate similarity measures offered, which is based on linking rank clustering method. In particular, there are parameters by clustering: rank of cluster drawing, quantity of rank omitting.

Keywords: Pattern Recognition, Clustering, Etalon Descriptions, Rank of Link.

I. INTRODUCTION

In the article there are methods of realization of the recognition process described in the theory of rank of links [1] according to the clustering drawn parameters, where clustering surface points are used for etalon descriptions, which are divided by positive and antipodal surface points.

In the multidimensional spaces it is quite difficult to find cluster surface points; especially in case of clustering complex form and a low-quality of separation. When a cluster is incompact, the surface points might not represent description of incompact patterns combined in a cluster and, hence, the use of positive and antipodal points for etalon descriptions becomes impossible.

In this work there are methods of drawing etalon descriptions and its appropriate similarity measures offered, which is based on linking rank clustering method. In particular, there are parameters by clustering: rank of cluster drawing, quantity of rank omitting.

II. CONSTRUCTING SIMILARITY MEASURES FOR A COMPACT CLUSTER

Let us say, we got a set of clusters $\{CL\}$ as a result of the clustering process realized by rank of linking method, its $CL_i \in \{Cl\}$ element meets the condition of compactness, which by definition given in [1], means that in CL_i cluster, there is only one A_i type of realizations (Fig.1).

In Fig.1, CL_j cluster is shown, for which, let us say, CL_j meets the condition of compactness. In the same figure there is also Cl_i cluster's positive E_i^+ and antipodal E_i^- points, and CL_j for cluster E_j^+ positive surface points.

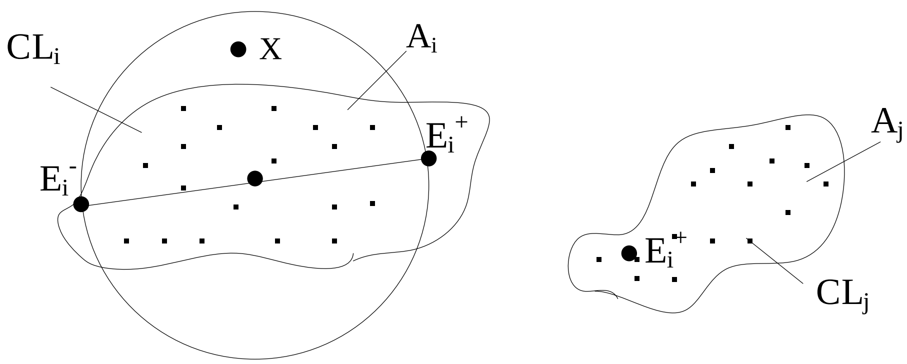

Figure 1. Constructing similarity measures for a compact cluster

ISSN 1512-1232

Definition 1- The point included into a cluster is called "central", if it is located in the midpoint of the connecting line of positive and antipodal points.

In Figure 1, are given E_i^+ and E_i^-, the central point created by etalon descriptions C_i. Let us use C_i point as a center of hyper-sphere, the radius of which is $C_i E_i^+ = C_i E_i^-$ and, define the status relating to any kind of the points' location in the drawn hyper-sphere. Here we might have the following situations:

a) In the hyper-sphere there are only A_i type of realizations; which means that any X point in the hyper-sphere may belong to A_i type, hence for the similarity measures we will have:

$$X \in A_i \qquad \rho(X, C_i) < R_i^0 \tag{1}$$

Where $\rho(X, C_i)$ – is an Euclidean distance, between the points X and C_i and $R_i^0 = \mathrm{C_i E}_i^+ = \frac{\mathrm{E}_i^- E_i^+}{2}$ - is the hyper-sphere radius.

b) There are also different types of realizations in the hyper-sphere. In such case we have to define the minimal rank link C_i between the realizations except the central points and other type (A_i) that are located in the drawn hyper-sphere.

$$rank\,(C_i; X_k) = \min_q \left\{ \mathrm{rank}\left(\mathrm{C_i}; X_q\right)\right\} \tag{2}$$

where $q = \overline{1;Q}$, Q represents the amount of other type of realizations in the drawn hyper-sphere.

Let us compute the value $\rho(C_i; X_k)$ and draw a new hyper-sphere R_i^1 with the radius:

$$R_i^1 < \rho(C_i; X_k) \tag{3}$$

For the more accurate definition of R_i^1 radius, let us compute A_i type X_i^p realization that is related to X_k realization with minimal rank links.

$$rank\,\left(X_k; X_i^p\right) = \min_g \left\{\mathrm{rank}\,\left(X_k; X_i^g\right)\right\} \tag{4}$$

Where $g = \overline{1;G}$, G represents the value of A_i type of realizations in the hyper sphere. According to the results we can define reduced R_i^1 radius of the hyper-sphere.

$$R_i^1 = \rho\left(C_i; X_i^k\right) \tag{5}$$

It is obvious that in the hyper-sphere, there will not be any other type of realizations except A_i type.

In A_i type $\{X_i\}$ realizations let us define realizations - the new positive and antipodal points left outside of the hyper-sphere, and, carry out procedures described in paragraphs A) and B). We will continue procedures until all the realizations of learning sequence of A_i type is covered by hyper-spheres.

III. CONSTRUCTING SIMILARITY MEASURES FOR INCOMPACT CLUSTERS

Let us take incompact cluster CL_{ij} where the learning sequence of A_i type realizations $\{X_i\}$ and A_j type $\{X_j\}$ realizations of learning sequence are united. (Fig.2).

In Figure 2 there are A_j type of realizations given in circles, and A_i type of realizations in crosses. Dashed line shows the area of overlapping the types.

Let us note area of overlapping of types A_i and A_j with $TAN(i,j)$ (Figure 2). We can define $TAN(i,j)$ according to the rule given in the work [2]:

$$X \in TAN(i,j) \qquad Rank\{X \forall X_h(TAN)\} \le r_{ij} \tag{6}$$

ISSN 1512-1232

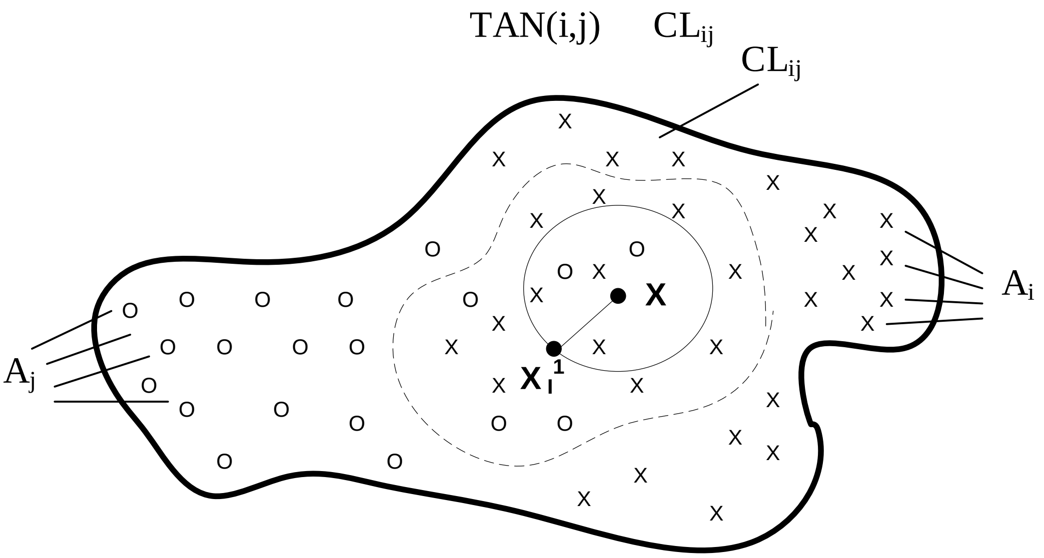

Figure 2. Constructing similarity measures for a incompact cluster

Where $X_h(TAN)$ in the area of overlapping, the existing X_h realization $h=\overline{1;H}$, H represents the quantity of united realizations in $TAN(i,j)$ area; r_{ij} represents CL_{ij} cluster constructing rank. For the realization (point) of unknown X let us select r_{ij} quantity of the neighboring points [1] – realizations, and, define the distance from X point to the farthest neighboring point. For example, in Figure 2, the furthest point is X_i^1 point, and for hyper-sphere radius we have:

$$R_{ij} = \rho\left(X; X_i^1\right) \qquad (7)$$

In the drawn hyper-sphere located a number of A_i type of realizations we mark with m_i^0, And A_j type of realizations with m_j^0. For decision-making, we select parameter -a about X- unknown realization's belonging, which informs us about the X point's minimal Rank link in the R_{ij} radius hyper sphere with the points of cluster. Here we meet the following situations:

a) The newest point with closed ranking link is of type A_j; parameter $a_j = 1$. in opposite case $a_j = 0$.

b) The newest point with closed ranking link is of type A_i, parameter $a_i = 1$. In opposite case. $a_i = 0$.

Decision is made according to the following rule:

$X \in A_i$, if $a_i = 1 \cap \left(m_i^0 > m_j^0\right)$;

$X \in A_j$, if $a_j = 1 \cap \left(m_j^0 > m_i^0\right)$.

Decision is not made if the following terms are met:

$$X \notin A_i \cap X \notin A_j, \text{ if } \left(a_i = 1 \cap a_j = 1\right) \cap \left(m_i^0 = m_j^0\right)$$

Not making decision means that A_i and A_j realizations from the unknown realizations are equally separated.

ISSN 1512-1232

IV. CONCLUSIONS

There is considered in this article the clustering process by rank links and the similarity measure connected with it, which is obtained by using the Euclidean distances. Have been introduced also the notions of cluster featuring, positive, antipodal and central points; After determining these points it is possible to find radiuses of hyperspheres in which realizations of only one pattern are placed. Finding of radiuses is considered for compact and incompact clusters. Effectiveness of this algorithm is proved practically.

REFERENCES

1. O. Verulava and R. Khurodze, *Rank of Link Theory – Recognition Process Modelling*, Technical University, Tbilisi, 2004;.
2. O. Verulava and R. Khurodze, “Clustering Analysis and Decision making by “Rank Links”,” *Mathematical Problems in Engineering, Taylor and Francis Group,* vol. 8 (4-5), 2002.

Article received: 2010-05-10