Computación y Sistemas Vol. 9 Núm. 3, pp. 191-202

Formation of Resemblance Measures Among Sets

Formación de Medidas de Equivalencia entre Conjuntos

Otar Verulava, Ramaz Khurodze and Lasha Verulava.
Artificial Intelligence Departament
University of GTU
Georgia, Tbilisi
verulava@gtu.ge

Article received on February 2, 2005; accepted on June 22,2005

Abstract

In this paper it is described a method to compute the distance between sets, that implies the formation of distance functions different from Hausdorff metric. Two functions with metric properties, which describe quantitatively distances between sets, are formed. First function can be used for sets arbitrary situated from each other. Second distance is more suited for sets clustered by rank links. For reconstructing the metric functions, we define the so-called boundary points between sets. This allows to defining the minimal and the maximal distances between them, which represents the arguments for the formed metric functions. This also allows quantitatively estimate in amore complete way the isolation degree between given sets.
Keywords: Clusterization, Metric Functions, Pattern Recognition.

Resumen

En este artículo se describe un método para el cálculo de la lejanía entre conjuntos. Esto implica la formación de funciones distancia pero diferentes a la métrica de Hausdorff. Se forman dos clases de funciones con propiedades métricas, que describen cualitativamente distancias entre conjuntos. La primera función puede ser usada para conjuntos arbitrariamente situados entre ellos. La segunda función es más aplicable para conjuntos que pueden ser acumulados a través del método de "Rank Links". Para re-construir las funciones métricas, se definen los llamados puntos de la frontera entre conjuntos. Esto permite definir distancias mínimas y máximas entre ellos, lo que representa sus argumentos. También, esto permite estimar cuánticamente de firma más completa el índice de separación entre ellos.
Palabras clave: Clusterización, Funciones de medida, Reconocimiento de patrones.

1 Introduction

The sets, whose elements represent vectors with real-valued components, are considered. The equi-dimensional realizations of training set in pattern recognition are examples of such sets.

Similarity measures among the sets represent significant values for estimation of feature space. It is known, that the more patterns are isolated in future space, the easier the formation of pattern descriptions and highly reliable recognition processes.

Similarity measures among the sets can be metric and non-metric. Hausdorff function belongs to a metric measure [1], which is oriented to the definition of the maximal distances among the elements of sets. It doesn't give perfect representation about isolation of sets. In [2], [3] are presented clustering procedure; one of obtained parameters of this procedure is ranks missing number, which describes clusters isolation degree. Described procedure doesn't satisfy all conditions of a metric. It is also known the Kendall's function, but in this case any metric condition has not been proved.

Below the metric functions formation processes among the sets are described, which are based on the definition of Euclidean distances among the elements of sets. As a result, some parameters are obtained. These parameters, as arguments, are used for the formation of new functions and metrics properties of these functions have been proved.

2 Formation of the Similarity Measures Among The Sets

Hausdorff metric and Kendal function are commonly used as a similarity measure between the sets. Hausdorff metric could not obtain wide spread occurrence in pattern recognition field, while Kendal function uses the relations (more or less, equivalence) for comparing process, which is a cause for inaccuracies and errors during the recognition.

Hence, it is desirable to form some sort of similarity measure, based on which it is possible to perform comparing process and to obtain accurate and correct results.

Let us assume, that for a set of patterns $\{A\}$ a feature set $\{x\}$ is defined, for power of which we have $Card\{x\} = N$, where $N = const.$ It means, that dimension of arbitrary realization of the set $\{A\}$ is equal to N. From the given set of patterns choose the arbitrary patterns A_i and A_j, which are correspondingly presented by the sets of realizations X_i and X_j, for the powers of which we have

$$Card\{X_i\} = M_i; \quad Card\{X_j\} = M_j \tag{1}$$

For simplicity and clarity assume, that N =2. It'll allow us to present realizations of the patterns A_i and A_j as points in two-dimensional Cartesian coordinate system (figure 1).

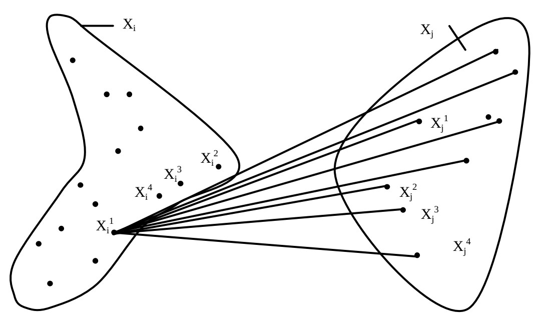

Fig.1

Presented on figure 1 sets X_i and X_j are isolated, that makes easy a geometrical interpretation of further discussions. In general, the sets X_i and X_j can intersect. For the estimation of similarity between the elements of the sets X_i and X_j let us apply some metric. It is quite possible to apply the Euclidean metric $\rho(\cdot)$. Particularly, for the realizations (points) $X_i^{m_i} \in \{X_i\}$ and $X_j^{m_j} \in \{X_j\}$ we'll have:

$$\rho(\cdot) = \rho(X_i^{m_i}; X_j^{m_j}) \tag{2}$$

where, $m_i = \overline{1, M_i}$, $m_j = \overline{1, M_j}$,

Let us define the distance values between point X_i^1 and all points of $\{X_j\}$ by term (2). As a result, we'll obtain a set of distance values:

$$\{\rho(X_i^1; X_j^{m_j})\},$$

an element of which for a pair of points X_i^1 and X_j^1 is denoted by $\rho(X_i^1; X_j^1)$. Based on condition (1), it is evident, that condition

$$Card\{\rho(X_i^1; X_j^{m_j})\} = M_j \tag{3}$$

where $m_j = \overline{1, M_j}$, is satisfied.

Let us define the minimal element of the set of distance values and denote it by

$$\rho^*(X_{ij}^1) = \min_{m_j}\{\rho(X_i^1; X_j^{m_j})\} \tag{4}$$

If we perform computations given by terms (2) and (4) over all points of the set X_i, we'll obtain a set of values of $\rho(\cdot)$, i.e. $\rho^*(X_{ij}^{m_i})$, $m_i = \overline{1, M_i}$.

Let us extract in the set X_j all those elements, for which condition (4) was satisfied and denote the obtained subset by $\{X_{m_j}^*\}$. Therefore, for all set of elements the following condition should be satisfied:

$$X_j^{m_j} \in \{X_{m_j}^*\}, \text{ if } \rho(X_i^{m_i}; \mathrm{X}_\mathrm{j}^{\mathrm{m_j}}) = \rho^*(X_{ij}^{m_i}) \tag{5}$$

Evidently, that subset of points $\{X_{m_j}^*\}$ is included in set X_j and presents closest points to set X_i. Hence, let us call subset $\{X_{m_j}^*\}$ as set X_j boundary points relative to set X_i.

According to terms (2), (4) and (5) define the set of points $\{X_{m_i}^*\}$, that give us boundary points of the set X_i relative to points of the set X_j. On figure 2 boundary points of the sets X_i and X_j are presented, which are surrounded by circles. Let us denote them correspondingly by $X_{m_i}^p \in \{X_{m_i}^*\}$ and $X_{m_j}^q \in \{X_{m_j}^*\}$.

From the boundary points sets $\{X_{m_i}^*\}$ and $\{X_{m_j}^*\}$ define some pairs of points $X^p \in \{X_{m_i}^*\}$ and $X^q \in \{X_{m_j}^*\}$, which will satisfy the following conditions:

$$\rho(X_{m_i}^p; X_{m_j}^q) = \min_{p_1} \rho(X_{m_i}^{p_1}; X_{m_j}^q) \tag{6}$$

$$\rho(X_{m_i}^p; X_{m_j}^q) = \min_{q_1} \rho(X_{m_i}^{p_1}; X_{m_j}^{q_1}) \tag{7}$$

where $p_1 = \overline{1, P}$; $q_1 = \overline{1, Q}$; $P = Card\{X_{m_i}^*\}$; $Q = Card\{X_{m_j}^*\}$;

The terms (6) and (7) describe situations according to which, if point $X_{m_i}^p$ is at the minimal distance from point $X_{m_j}^q$, then point $X_{m_j}^q$ is at the minimal distance from point $X_{m_i}^p$ also. On figure 2 the points obtained by terms (6) and (7) are connected by segments.

Let us denote by $Z = \{Z_1, Z_2, ..., Z_s\}$ a set of pairs of connected points and distance values between points of each pair by $\rho(Z_s)$.

Define values:

$$\rho_{\min}(Z) = \min_s \{\rho(Z_s)\} \tag{8}$$

$$\rho_{\max}(Z) = \max_s \{\rho(Z_s)\} \tag{9}$$

It is necessary to note, that result obtained by term (8) is identical to result obtained by Hausdorff metric. Hence, it is evident, that only the minimal distance between the sets is taken into consideration by Hausdorff metric. On figure 2 the corresponding segment is surrounded by circles. Based on our approach, we take into consideration a distance (a segment marked by rectangles) also obtained by term (9).

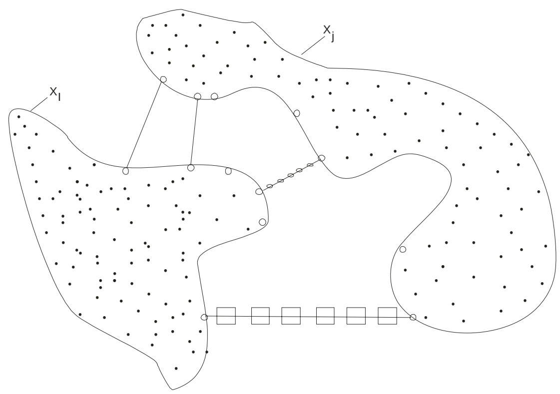

Fig. 2

Theorem 1. The set Z corresponding to the formation process or the subsets X^*_{ij} and X^*_{ji} corresponding to the obtaining process are invariant with respect to a process of the definition of the metric measures between the elements of the sets X_i and X_j.

The statement in Theorem 1 means that if we initially execute operations given by terms (2), (4), (5) over the set X_j elements, and then realize the operations given by terms (6), (7), (8), and (9) over the set X_i elements, we will obtain correspondingly the subsets $\{X^*_{m_j}\}$ and $\{X^*_{m_i}\}$. This means, that by displacement of the sets X_i and X_j over operations given by above mentioned terms, list of the boundary points will keep unchanged.

Let us take into account, that for implementation of the operations (2), (3), (4), and (5) it is necessary to compute the distances (metric measures of similarity) for all elements of the sets X_i and X_j. As the metric functions are symmetrical, we can admit that the distances from all elements of the set X_j to all elements of the set X_i also are computed.

The relation of symmetry will be satisfied by the defining procedures of the minimal and the maximal elements, given by terms (6), (7), (8) and (9), which means that by replacing of the sets X_i and X_j in computing procedures we'll obtain set Z. Thus Theorem 1 is proved.

It is clear that set Z elements represents minimal separated points of the sets X_i and X_j.

Let us construct the function $d(\cdot)$ which arguments are the values given by terms (8) and (9) - $\rho_{\min}(Z)$ and $\rho_{\max}(Z)$, that is,

$$d(Z) = \rho_{\min}(Z)\left(\rho_{\min}(Z) + \rho_{\max}(Z)\right) \quad (10)$$

Prove now that the function given by (10) is a metric, i.e. it satisfies non-negation, reflexive, symmetry and triangle conditions.

1. Non-negativity:

$$d(Z) \geq 0 \quad (11)$$

Inequality (11) is true because the following conditions are satisfied:
$\rho_{\min}(Z) \geq 0;\ \rho_{\max}(Z) \geq 0 \Rightarrow d(Z) \geq 0$.

2. Symmetry condition:

An element (point) of the set Z, for which the conditions (8) or (9) are satisfied, presents a pair of points such that by changing of their places does not change the function $\rho(\cdot)$ and, particularly, $\rho_{\min}(Z)$ and $\rho_{\max}(Z)$, i.e., this will not change function $d(Z)$; it means, that function $d(Z)$ is symmetrical.

3. Reflectivity condition:

It is clear, that $\rho_{\min}(Z_s) = 0$ if and only if $X_{ij}^* = \mathrm{X}_{\mathrm{ji}}^* = \mathrm{X}$. If we substitute the given values in the term (10), we'll obtain

$$d(Z) = d(X;X) = 0 \quad (12)$$

4. The triangle condition: it defines the existence of a certain relation between the distances of three objects (points) situated in the given space. Particularly, for arbitrary three points, for example, X_i, X_j, X_k non-strict inequality must be satisfied:

$$\rho(X_i, X_j) \leq \rho(X_i, X_k) + \rho(X_j, X_k) \quad (13)$$

It is clear that the uses of term (13) as the triangle condition, when in space three points are not given but sets consist of many points, are impossible. That's why in the sets case it is necessary to form a new condition, similar to the triangle condition, for example, based on term (13).

To do that let us consider the sets X_i and X_j, whose boundary points X_{ij}^* and $\mathrm{X}_{\mathrm{ji}}^*$ form set Z.

Definition 1. The triangle condition for the sets X_i, X_j and X_k holds, if for arbitrary pair of points from subsets X_{ij}^* and $\mathrm{X}_{\mathrm{ji}}^*$, and also for arbitrary point $X_k^{m_k} \in X_k$ the triangle condition (13) is satisfied.

According to the definition 1 we have:

$$\rho(X_{ij}^*; X_{ji}^*) \leq \rho(X_{ij}^*; X_k^{m_k}) + \rho(X_k^{m_k}; X_{ji}^*) \quad (14)$$

The geometry interpretation of the condition (14) is given on Figure 3.

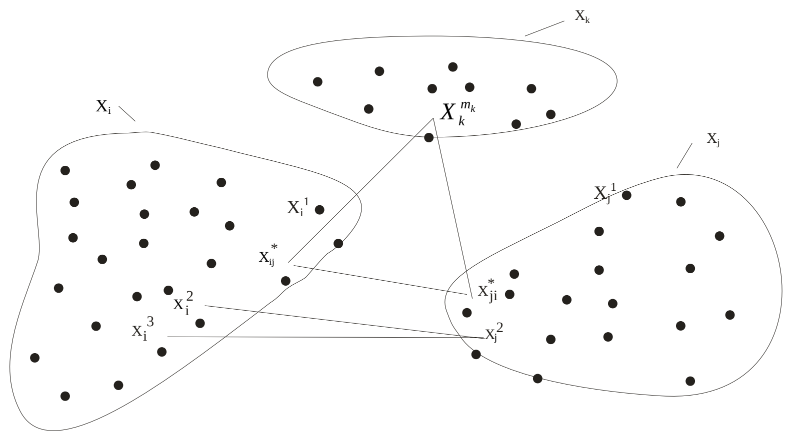

Fig.3

Let us call the condition presented by term (14) the triangle condition for sets.

Definition 2. In the given space arbitrary function $\rho(\cdot)$ is a metric one for the sets X_i, X_j and X_k, if the pairs formed by a triplet of arbitrary points of those sets satisfy conditions given by terms (11), (12) and fulfill the symmetry property.

The sets X_i, X_j and X_k in the given space can be considered as one set $X = \{X_i, X_j, X_k\}$, which is formed by union of the points three of sets.

For a set of points in a space, where some metric is defined, the triangle condition is true for arbitrary triplet of points, for instance, in our case for boundary points Z and some set $\{X_k^{m_k}\}$. Hence, the inequality (14) is true for the arbitrary points X_{ij}^*, X_{ji}^* and $X_k^{m_k}$. It means, that the function $d(Z)$ according to definitions 1 and 2 is a metric function.

3 Formation of Similarity Measure By Rank of Links

This process in Euclidean space sets, particularly, the formation of a similarity measure between realizations of pattern and clusters, obtained from them, is discussed in this section. The similarity measure is based on well-known notions of theory rank of links: a compactness of patterns, an isolation measure – ranks, a missing number and cluster construction ranks [1], [2], [3].

Let us suppose that for a set of patterns A we have a set of realizations X. By the rank of links method, the clusters set *CL* is obtained in the set X, where each pattern $A_i \in \{A\}$ is presented by one or several clusters CL_i: $CL_i = \{X_i\}$, where set $X_i \in \{X\}$ represents the realizations of learning set of pattern A_i. Each cluster CL_i is represented by the parameters obtained after clustering [1], [2]:

1. Let us denote the cluster construction rank by r_i;
2. Let us denote the ranks of missing number by l_i;
3. the list of the realizations united in cluster CL_i is defined as follows: If patterns A_i and A_j intersect, then according to [1], $l_i = 0$ and $l_j = 0$.

Let us denote the cluster obtained after intersection by CL_{ij}, the cluster construction rank by r_{ij}, while the ranks of missing number, i.e. isolation degree from set $X \setminus CL_{ij}$, by l_{ij}.

Suppose that $Card\{A\} = 2,\ Card\{CL\} = 2.$ The dimension of the feature set is equal to two, i.e., we have points on the plane. This allows us to present the process of similarity measure formation geometrically. On Figure 4 the clusters CL_i and CL_j corresponding to the patterns A_i and A_j are shown as sets of points on the plane.

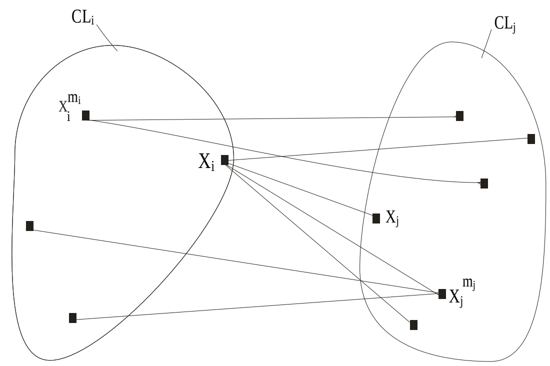

Fig. 4

Let us define the open rank links from each point of cluster CL_i to all points of cluster CL_j and choose a minimal one between them. For instance, for point $X_i^{m_i} \in CL_i$ we'll have (Figure 4) :

$$rank(X_i^{m_i}; X_j^*) = \min_{m_j} rank(X_i^{m_i}; X_j^{mj}) \tag{15}$$

where $rank(X_i^{m_i}; X_j^{m_j})$ represents the open rank links from point $X_i^{m_i}$ to point $X_j^{m_j}$, $m_j = \overline{1, M_j}$ [1]. If we implement the procedure given by term (15) to all points of cluster CL_i, we will get boundary points of cluster CL_j to the points of cluster CL_i [2]. Denote the boundary points set by S_{ji}, which is a subset of set X_j, $S_{ji} \subset X_j$.

Let us implement the same procedure from points of the cluster CL_j to points of the cluster CL_i, then instead of term (15) we'll obtain :

$$rank(X_j^{m_j}; X_i^*) = \min_{m_i} rank(X_j^{m_j}; X_i^{m_i}) \tag{16}$$

where $m_i = \overline{1, M_i}$. As a result of implementation of term (16), we get boundary points of the cluster CL_i to the points of cluster CL_j. Let us denote the boundary points set by $S_{ij} \subset X_i$ (Figure 5).

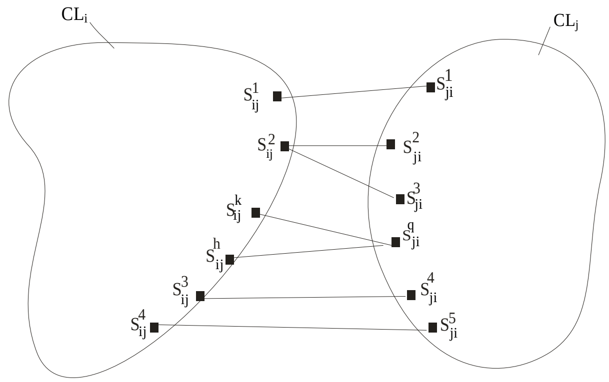

Fig. 5

Define the minimal rank link for each point $S_{ij}^{k} \in S_{ij}$ of cluster CL_i to points of S_{ji} as follows:

$$rank(S_{ij}^{k}; S_{ji}^{q}) = \min_{q_1} rank(S_{ij}^{k}; S_{ji}^{q_1}) \tag{17}$$

where $q_1 = \overline{1, Card\{S_{ji}\}}$, $k = \overline{1, Card\{S_{ij}\}}$.

The same procedure implemented for points of set S_{ji} to points of S_{ij} leads to the relation

$$rank(S_{ji}^{q}; S_{ij}^{k}) = \min_{k_1} rank(S_{ji}^{q}; S_{ij}^{k_1}) \tag{18}$$

where $k_1 = \overline{1, Card\{S_{ij}\}}$, $q = \overline{1, Card\{S_{ji}\}}$. After the implementation of (17) and (18) we will get pairs of points from the sets S_{ij} and S_{ji} which are connected to each other by open rank links. It is well-known that the open rank links, as relations, are not one-to-one, because of which they do not satisfy requirements to metric functions. The illustration of this fact is shown on Figure 5, where points S_{ij}^{k} and S_{ij}^{h} are connected by the minimal rank links to the point S_{ji}^{q} of the cluster CL_j and vice versa, points S_{ji}^{2} and S_{ji}^{3} to S_{ij}^{2} dot. Let us compose close rank links for the paired points of clusters CL_i and CL_j, we'll have[1]:

$$Rank\{S_{ij}^{k}, S_{ji}^{q}\} = \max\{rank(S_{ij}^{k}; S_{ji}^{q}); Rank(S_{ji}^{q}; S_{ij}^{k})\} \tag{19}$$

where $k = \overline{1, Card\{S_{ij}\}}$; $q = \overline{1, Card\{S_{ji}\}}$.

It is known, that for close rank links the condition of metric is satisfied. It means that one closed rank link is satisfied only for one pair of points.

Let us assume that the number of boundary points of clusters CL_i and CL_j satisfied condition (19) is equal to M. It means that set of values of the close rank links obtained by (19) is also equal to M. Let us define the minimal and the maximal elements of that set :

$$\bar{Rank}(i; j) = \min_{m} Rank\{S_{ij}^{m}, S_{ji}^{m}\} \tag{20}$$

$$\overset{+}{Rank}(i, j) = \max_{m} Rank\{S_{ij}^{m}, S_{ji}^{m}\} \qquad (21)$$

where $m = \overline{1, M}$.

Accordingly, for the estimation of the similarity measure between the clusters we obtain two values given by terms (20) and (21). Based on the latter values and data of cluster construction ranks let us compose the following similarity measure function :

$$d(\cdot) = (r_i + r_j)(\overset{+}{Rank}(i; j) + \overset{-}{Rank}(i; j)) \qquad (22)$$

where $d(\cdot)$ is the similarity measure between clusters CL_i and CL_j . Let us prove that function $d(\cdot)$ is a metric.

1. Non-negativity: $d(\cdot) \geq 0$.

If we take into account that all components of sum (22) are non-negative, then we can conclude that the non-negativity condition is true.

2. Reflectivity: $d(\cdot) = 0$ (23)

Let us substitute in (22) the condition, which reflects metric of close rank links. If $i = j$, then we will get :

$$d(\cdot) = (r_i + r_i)(\overset{+}{Rank}(i;i) + \overset{-}{Rank}(i;i)) \qquad (24),$$

$Rank(i;i) = 0 \Rightarrow \overset{+}{Rank}(i;i) = 0$ and $\overset{-}{Rank}(i;i) = 0 \Rightarrow d(\cdot) = 0$

3. Symmetry:

$$d(S_{ij}^{m}, S_{ji}^{m}) = d(S_{ji}^{m}, S_{ij}^{m}) \qquad (25)$$

From term (22) we have that the replacement of arguments in the function $d(\cdot)$ implies the rearrangement of components. It is evident, that it doesn't change summation result. Hence, term (25) is true.

4. The Triangle Condition for Sets.

The well-known triangle condition for metric functions presents non strict inequality, where three objects are connected by mathematical symbols, particularly, three points. Let us consider the triangle condition for any three points of given space, for instance, for boundary points S_{ij} , S_{jk} , S_{ki} of the clusters CL_j , CL_i CL_k . Then we'll have:

$$d(S_{ij}; S_{ji}) \leq d(S_{ik}; S_{ki}) + d(S_{jk}; S_{kj}) \qquad (26)$$

where similarity measure function for sets is denoted by $d(\cdot)$. Let us prove that inequality (26) takes place.

Definition 1. The triangle condition for sets in accordance with (31) is satisfied, if for arbitrary triple of points from the given sets triangle condition holds.

If $S_i \in CL_i$, $S_j \in CL_j$, $S_k \in CL_k$, then condition given by term (31) is as follows :

$$d(S_i; S_j) \leq d(S_i; S_k) + d(S_k; S_j) \qquad (27)$$

where $i = \overline{1, Card\{X_i\}}$, $j = \overline{1, Card\{X_j\}}$, $k = \overline{1, Card\{X_k\}}$

For the function $d(\cdot)$ given by (22) we have the maximal and the minimal values of close rank links between boundary points. Therefore, let us assume in (27) that $R^{+}(i;j)$, $R^{+}(i;k)$ and $R^{+}(j;k)$ rank links are realized for points S_i, S_j, S_k. Since the close rank links are metric functions, therefore according to (27), we'll have:

$$\overset{+}{Rank}(i;j) \le \overset{+}{Rank}(i;k) + \overset{+}{Rank}(j;k) \tag{28}$$

Let us suppose that for points $S_{j_1} \in CL_j$, $S_{i_1} \in CL_i$, $S_{k_1} \in CL_k$ are realized rank links $\overset{-}{Rank}(i;j)$, $\overset{-}{Rank}(i;k)$, $\overset{-}{Rank}$ $\overset{-}{Rank}(j;k)$.

Analogously to (28) we'll have: $\overset{-}{Rank}(i;j) \le \overset{-}{Rank}(i;k) + \overset{-}{Rank}(j;k)$ (29)

For terms (28) and (29) by summation of both sides, accordingly, one gets:

$$\overset{+}{Rank}(i;j) + \overset{-}{Rank}(i;j) \le \overset{+}{Rank}(i;k) + \overset{-}{Rank}(i;k) + \overset{+}{Rank}(j;k) + \overset{-}{Rank}(j;k) \tag{30}$$

If we take into account that $r_i + r_j \le (r_i + r_k) + (r_j + r_k)$, then we obtain

$$(r_i + r_j)(\overset{+}{Rank}(i;j) + \overset{-}{Rank}(i;j)) \le (r_i + r_k)(\overset{+}{Rank}(i;k) + \overset{-}{Rank}(i;k)) + (r_j + r_k)(\overset{+}{Rank}(j;k) + \overset{-}{Rank}(j;k)) \tag{31}$$

Therefore, for each component of the inequality (14) we'll have :

$$d(CL_i;CL_j) \Leftrightarrow d(S_{ij};S_{ji}) = (r_i + r_j)(\overset{+}{Rank}(i;j) + \overset{-}{Rank}(i;j)) \tag{32}$$

$$d(CL_i;CL_k) \Leftrightarrow d(S_{ik};S_{ki}) = (r_i + r_k)(\overset{+}{Rank}(i;k) + \overset{-}{Rank}(i;k)) \tag{33}$$

$$d(CL_j;CL_k) \Leftrightarrow d(S_{jk};S_{kj}) = (r_j + r_k)(\overset{+}{Rank}(j;k) + \overset{-}{Rank}(j;k)) \tag{34}$$

Take into account terms (32), (33) and (34), one may conclude, that the triangle condition for sets given by term (26) holds that proves the result.

4 Conclusions

For the similarity measures formed for sets it is proven, that they are metric functions. Therefore, using the constructed in that way functions it is possible to define an isolation degree (distance) among two arbitrary sets.

It is shown, that unlike Hausdorff metric, which determines isolation degree by only one parameter, particularly by computing the minimal Euclidean distance between sets elements, the suggested function has two parameters which allow to define the distance among sets using the constructed metric.

In the first (Hausdorff) case, the sets are given in the absence of clustering process. The similarity measure, formed for such sets, is a metric, but it does not reflect completely an isolation degree when the intersection of sets takes place.

In the second (the suggested approach) case, the results of clustering, obtained by rank links, allow to estimate quantitatively sets isolation degree. In fact, another similarity measure function is constructed, which is formed by two arguments, particularly, the minimal and the maximal distances between boundary points obtained by rank links. This construction allows to estimate isolation degree between given sets quantitatively more completely.

References

1. **Verulava O**., Clustering analysis by “Rank of Links’’. *Transactions of Georgian Technical University, Tbilisi*, #3 (414), 288-296, 1997.
2. **Verulava O. and R. Khurodze**, Clustering Analysis and Decision-making by “Rank of Links’’. *Mathematical Problems in Enginering*, vol. 8(4-5), 475-492, 2002.
3. **Verulava O. and R. Khurodze,** “The Principles of Pattern Recognition Theory’’. Georgian Technical University, Tbilisi, 2001. (in Georgian).
4. **Faure A.,** Perception and Pattern Recognition. Editests, 1985.(in French).
5. **Jain A. K. and R. C. Dubes**, Algorithms for clustering data. Prentice Hall, 1988.
6. Classificat ion and Clustering. Edited by J. Van Risin, Academic Press, 1977.
7. **Verulava O. and R. Khurodze,** Theory of Rank of Links – Modeling and Recognition Process. Georgian Technical University, Tbilisi, 2004. (in Georgian).

Otar Verulava, Doctor of Technical Science, Professor, Head of Artificial Intelligence Department of Georgian Technical University. In the 1967 he graduated from Georgian Technical Institute and got the Bachelors Degree. From 1969 he continued his post-graduated study in the institute of Control Problems at the Russian Academy of sciences, where in 1975 he defenced his Candidate Dissertation. From 1977 he has been working in Georgian Technical University. In 1993 he was awarded scientific degree of Doctor of Technical Science.

Ramaz Khurodze, Doctor of Technical Science, Professor, Corresponding member of Georgian Academy of Science. In the 1967 he graduated from Georgian Technical University and got the bachelor's degree. From 1979 he worked in the Institute of Control Problems at the Russian Academy of Sciences, where in 1974 he defenced his Candidate Dissertation. In 1995 he was awarded scientific degree of Doctor of Technical Science. Since 1994 he has been rector of Georgian Technical University.

Lasha Verulava graduated From Georgian Technical University and got Bachelor's degree in 2002. In 2005 he received Master's degree. Since 2005 he is a lecturer of the faculty of Informatics and Control systems of GTU. Now he is working for doctor's degree