Constructing Correlation Coefficients from Similarity and Dissimilarity Functions

Ildar Z. Batyrshin

Centro de Investigacion en Computacion, CIC-IPN, Av. Juan de Dios Bátiz S/N, Nueva Industrial Vallejo, Gustavo A. Madero, 07738 Ciudad de México, CDMX, e-mail: batyr1@cic.ipn.mx

Abstract. Correlation coefficients (association measures) were introduced more than one hundred years ago as measures of relationship between variables that usually belong to one of the following basic types: continuous, ordinal, or categorical. Nowadays, it appears the growing demand for the development of new correlation coefficients for measuring associations between variables or objects with more sophisticated structures. The paper presents a non-statistical, functional approach to the study of correlation coefficients. It discusses the methods of construction of correlation coefficients using similarity and dissimilarity measures. Generally, all these measures are considered as functions defined on the underlying (universal) domain and satisfying some sets of properties. The methods of construction of correlation functions on the universal domain can be easily applied for constructing correlation coefficients for specific types of data. The paper introduces a new class of correlation functions, satisfying a weaker set of properties than the previously considered correlation functions (association measures) defined on a set with involution (negation) operation called here strong correlation functions. The methods of constructing both types of correlation functions are discussed. The one-to-one correspondence between the strong correlation functions and the bipolar similarity and dissimilarity functions is established. The theoretical results illustrated by examples of construction of classical Pearson's product-moment correlation coefficient, Spearman's and Kendall's rank correlation coefficients, etc. from similarity and dissimilarity functions.

Keywords: similarity; dissimilarity; correlation; association; Spearman's and Kendall's rank correlations; Pearson's correlation; cosine similarity, and cosine correlation

1 Introduction

Similarity and association measures actively studied more than one hundred years as measures of relationship between data initially in pattern classification and statistics and later in data mining and machine learning [1, 9-11, 13, 14, 16-20]. Correlation coefficients (association measures) were introduced in statistics as measures of the relationship between variables of one of the following types: continuous, ordinal, or categorical. Nowadays, we see the growing demand for

measuring the relationship between data with diverse structures: sequences and time series, rating profiles, vectors with different attributes, fuzzy sets, matrices, images, texts with syntactic structures, etc. An application of classical correlation coefficients to new types of data often impossible or gives misleading results. For this reason, to have the correlation coefficients specific for the analyzed data type is of great interest.

The paper considers the methods of construction of correlation coefficients using similarity and dissimilarity measures defined on an underlying set referred to as a universal domain. One can easily apply these methods to a specific type of data. The similarity, dissimilarity, and correlation measures are considered as functions satisfying some sets of properties [3, 4, 8]. This paper introduces a new class of correlation functions, satisfying a weaker set of properties than the previously considered correlation functions (association measures) [4, 8] called here strong correlation functions and defined on a set with involution (negation) operation. We discuss the methods of constructing both types of correlation functions using (dis)similarity functions. We establish the one-to-one correspondence between the strong correlation functions and "bipolar" (dis)similarity functions. The examples of the construction of correlation coefficients for specific domains illustrate the theoretical results.

The paper has the following structure. Section 2 gives definitions of similarity and dissimilarity functions. Sections 3 and 4 consider the methods of construction of strong correlation functions (association measures) on the set with involution operation. In Section 5, we introduce generalized correlation functions that can be defined on a set without involution operation and study their relationships with (dis)similarity functions. In Section 6 we establish one-to-ome correspondence between strong correlation functions and bipolar (dis)similarity functions. Section 7 discusses related works and includes the conclusion.

2 Similarity and Dissimilarity Functions

The paper considers similarity, dissimilarity, and correlation measures or coefficients as functions defined on some nonempty underlying set Ω [8]. As such set one can use, for example, any specific domain: the set of all real-valued *n*-tuples, the set of binary vectors, the set of membership values, the set of images, etc. To emphasize that the underlying set Ω is not a specific domain but any domain this set will be referred to as a universal domain.

Definition 1. A similarity function on a set Ω is a function $S:\Omega\times\Omega\rightarrow[0,1]$ satisfying for all x, y in Ω the properties:

$S(x,y) = S(y,x)$, (symmetry)

$S(x,x) = 1$. (reflexivity)

If for some x, y in Ω it is fulfilled:

$S(x,y) = 0$,

then S is called (0)-*normal* (in (x,y)).

Definition 2. A dissimilarity function on a set Ω is a function $D:\Omega\times\Omega\rightarrow[0,1]$ satisfying for all x, y in Ω the properties:

$D(x,y) = D(y,x)$, (symmetry)

$D(x,x) = 0$. (irreflexivity)

If for some x, y in Ω it is fulfilled:

$D(x,y) = 1$,

then D is called 1-*normal* (in (x,y)).

Dissimilarity functions are *dual* to similarity functions.

Definition 3. Similarity S and dissimilarity D functions are called complementary if for all x, y in Ω it is fulfilled:

$S(x,y)+D(x,y) = 1$.

One can obtain one of these functions from the corresponding complementary function for all x, y in Ω as follows:

$S(x,y) = 1 - D(x,y)$,

$D(x,y) = 1 - S(x,y)$.

It is clear that a similarity function is 0-normal if and only if its complementary dissimilarity function is 1-normal.

Example 1. Let Ω be a set of nonnegative real-valued n-tuples $x = (x_1,\ldots,x_n)$ such that $x \neq (0,\ldots,0)$. Then for any $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ in Ω the following function:

$$S(x,y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}},$$

is the symmetric and reflexive similarity function called a *cosine similarity measure* and denoted as $\cos(x,y)$. It is 0-normal for orthogonal pairs of n-tuples x and y such that $x_i y_i = 0$ for all $i=1,\ldots,n$.

The similarity function $\cos(x,y)$ has the following complementary dissimilarity function [8]:

$$D(x,y) = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i}{\sqrt{\sum_{i=1}^{n} x_i^2}} - \frac{y_i}{\sqrt{\sum_{i=1}^{n} y_i^2}}\right)^2.$$

Due to the duality of similarity and dissimilarity functions, one can consider only one of these functions, but they have different interpretations and methods of construction; hence, we consider them together. These functions studied in [8]. For short, we call them also (dis)similarity functions.

3 Strong Correlation Functions

The correlation functions (association measures) were introduced in [3, 4, 8] on a universal domain Ω with an involutive operation N as functions satisfying several properties of Pearson's product-moment correlation coefficient. Here these correlation functions called strong correlation functions. The methods of construction of correlation functions using similarity functions have been proposed, and it was shown how the Pearson's correlation and Yule's Q association coefficient could be constructed using suitable similarity functions [2, 5, 8]. In the following sections, we will consider correlation functions satisfying a weaker set of properties.

Definition 4. A function $N:\Omega\rightarrow\Omega$ satisfying for all x in Ω the property:

$N(N(x))=x$, (involutivity)

is called a *reflection* or a *negation* on Ω if it is not an identity function, i.e. if for some x in Ω it is fulfilled:

$N(x) \neq x$.

An element x in Ω such that $N(x) = x$ is called a *fixed point* and the set of all fixed points of the negation N in Ω is denoted as $FP(N, \Omega)$ or $FP(\Omega)$.

Below $\Omega\backslash FP(\Omega)$ denotes the set of all elements in Ω, which are not fixed points. The set $FP(\Omega)$ can be empty. It is easy to show that the set $\Omega\backslash FP(\Omega)$ is closed under reflection operation.

Proposition 1 [4]. Let $N:V\rightarrow V$ be a reflection on a set V, and $R:V\times V\rightarrow R$ be a symmetric real-valued function, i.e.:

$R(x,y) = R(y,x)$,

for all x,y in V. The function R satisfies for all x,y in V the property:

$R(N(x),N(y)) = R(x,y)$,

if and only if it is fulfilled:

$R(x,N(y)) = R(N(x),y)$.

Symmetric function R satisfying these properties will be called a *co-symmetric* function [8]. Further, we will consider co-symmetric similarity, dissimilarity, and correlation functions.

A similarity function $S:V\times V\to[0,1]$ satisfying for all x, y in V the property:

$S(x,N(x)) = 0$,

is called a *consistent* similarity function [8].

Dually, a dissimilarity function $D:V\times V\to[0,1]$ will be called a *consistent* dissimilarity function if for all x, y in V it is fulfilled:

$D(x,N(x)) = 1$.

A similarity function S is consistent or co-symmetric if and only if its complementary dissimilarity function is consistent or co-symmetric, respectively.

Definition 5 [4]. Let N be a reflection on Ω and V be a subset of $\Omega\backslash FP(\Omega)$ closed under N. A *strong correlation function* (association measure) on V is a function $A:V\times V\to[-1,1]$ satisfying for all x, y in V the properties:

$A(x,y) = A(y,x)$, (symmetry)

$A(x,x) = 1$, (reflexivity)

$A(x,N(y)) = -A(x,y)$. (inverse relationship) (1)

A strong correlation function will be also referred to as an *invertible correlation function.*

Proposition 2 [4]. A strong correlation function A on V satisfies for all x, y in V the following properties:

$A(x,N(x)) = -1$,

$A(N(x),N(y)) = A(x,y)$. (co-symmetry)

Example 2. Let Ω be the set of all real-valued n-tuples $x = (x_1,\ldots,x_n)$ with the reflection operation $N(x) = -x = (-x_1,\ldots, -x_n)$. Let V be a set of all non-constant n-tuples from Ω such that $x\neq(q,\ldots,q)$ for any real value q. It is clear that V is closed under N, and it has no fixed elements. It is easy to check that the Pearson's product-moment correlation coefficient:

$$A(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}, \quad (2)$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the strong correlation function on V.

Consider the methods of construction of strong correlation functions (association measures) [3, 4, 8].

Theorem 1. Let N be a reflection on Ω and V be a nonempty subset of $\Omega\backslash FP(\Omega)$ closed under N. Let $S:V\times V\to[0,1]$ be a co-symmetric and consistent similarity function, then the function $A:V\times V\to[-1,1]$ defined for all x, y in V by

$A(x,y)=S(x,y) - S(x,N(y))$, (3)

is a strong correlation function on V.

The formula (3) has a simple interpretation: *the correlation between x and y is positive if x is more similar to y than to its negation, and the correlation is negative in the opposite case.*

Replacing in (3) the similarity function S by the complementary dissimilarity function: $D(x,y) = 1 - S(x,y)$ obtain the following formula for constructing a strong correlation function from a co-symmetric and consistent dissimilarity function:

$$A(x,y) = D(x,N(y)) - D(x,y). \quad (4)$$

This formula can be more convenient than (3) for constructing strong correlation functions when we use distance-based dissimilarity functions [8].

Example 3 [2, 5, 8]. Consider the dissimilarity function on the set V of non-constant n-tuples (see Example 2) that generates by (4) the Pearson's product-moment correlation coefficient (2):

$$D(x,y) = \frac{1}{4}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} - \frac{y_i-\bar{y}}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}\right)^2.$$

One can check that it is co-symmetric and consistent.

Example 4. If the cosine function from Example 1 is defined on the set of all nonzero real-valued n-tuples with the negation operation $N(x) = -x = (-x_1,\ldots, -x_n)$, then it will satisfy the properties of the strong correlation function and can be generated by (4) using the following dissimilarity function [8]:

$$D(x,y) = \frac{1}{4}\sum_{i=1}^{n}\left(\frac{x_i}{\sqrt{\sum_{i=1}^{n}x_i^2}} - \frac{y_i}{\sqrt{\sum_{i=1}^{n}y_i^2}}\right)^2.$$

This dissimilarity function is co-symmetric and consistent.

As it follows from Examples 1 and 4, the cosine function will be the similarity function if it is defined on the set of nonzero, nonnegative real-valued n-tuples, and it will be the strong correlation function if it is defined on the set of nonzero real-valued n-tuples [8]. The cosine is often used as a similarity measure between the texts represented by vectors of attributes where x_i equals to the frequency of appearing of the ith attribute in the text [18]. In this case, all vector components take nonnegative values. As a correlation function, the cosine used, for example, in measuring associations between time series presented by sequences of local trends that can have positive and negative values [2, 6]. Generally, the cosine correlation can be used instead of Pearson's correlation on the set of real-valued n-tuples when the calculation of means used in Pearson's correlation has not much sense, for example, when the signs of the elements of n-tuples are important, or these elements contain the values of attributes measured in different scales.

4 Pseudo-Differences in Constructing Strong Correlation Functions

Let TC:[0,1]×[0,1]→[0,1] be a *t*-conorm [15], i.e. commutative, associative, monotonic function satisfying boundary condition: $TC(a,0) = 0$, for all a in [0,1]. Usually, *t*-conorm is denoted by the letter *S*, but in this paper, the symbol *S* is used for similarity functions and similarity measures, for this reason, we denote *t*-conorm as *TC*.

We will say that a *t*-conorm *TC* has no nilpotent elements if for all a, b in [0,1] it is fulfilled: $TC(a,b) = 1$ if and only if $a = 1$ or $b = 1$.

Consider the examples of basic *t*-conorms defined for all a, b in [0,1] as follows [15]:

$TC_M(a,b) = \max(a,b)$, (maximum)

$TC_P(a,b) = a+b-ab$, (probabilistic sum)

$TC_L(a,b) = \min(a+b, 1)$. (Lukasiewicz *t*-conorm)

t-conorms TC_M and TC_P have no nilpotent elements but TC_L has.

Definition 6 [12]. Let *TC* be a *t*-conorm. A pseudo-difference operation $\ominus_{TC}$ associated to *TC* is defined for all a, b in [0,1] as follows:

$$a \ominus_{TC} b = \begin{cases} a \underline{^{TC}} b & if\ a > b \\ -(b \underline{^{TC}} a) & if\ a < b, \\ 0 & if\ a = b \end{cases}$$

where $a \underline{^{TC}} b$ is the *TC-difference* defined by:

$$a \underline{^{TC}} b = \inf\{c \in [0,1] | TC(b,c) \geq a\}.$$

Consider the pseudo-difference operations associated to basic *t*-conorms [12]:

$$a \ominus_M b = \begin{cases} a & if\ a > b \\ -b & if\ a < b, \\ 0 & if\ a = b \end{cases}$$

$$a \ominus_P b = \begin{cases} \frac{a-b}{1-\min(a,b)}, & if\ a \neq b \\ 0, & if\ a = b \end{cases},$$

$$a \ominus_L b = a - b.$$

Theorem 2 [3,4]. Let N be a reflection on Ω and V be a nonempty subset of $\Omega \backslash FP(\Omega)$ closed under N. Let $S:V\times V\rightarrow[0,1]$ be a co-symmetric and consistent similarity function then the function

$A(x,y) = S(x,y) \ominus_{TC} S(x,N(y))$.

is a strong correlation function on V.

Theorem 1 is a particular case of Theorem 2 when it is used the pseudo-difference operation $\ominus_{TC} = \ominus_L$ associated to Lukasiewicz *t*-conorm.

In some domains, it is difficult to construct consistent similarity functions. In such cases, the property of consistency *S* can be replaced by the property of *weak consistency* (*weak similarity of reflections*) of *S* defined for all *x,y* in *V* by:

$S(x,N(x)) < 1$.

Theorem 3 [3,4]. Let *N* be a refkection on Ω and *V* be a nonempty subset of $\Omega \backslash FP(\Omega)$ closed under *N*. Let $S: V \times V \rightarrow [0,1]$ be a co-symmetric similarity function satisfying weak consistency, then the function

$$A(x,y) = S(x,y) \ominus_{TC} S(x,N(y)), \tag{5}$$

is a correlation function on *V* if *t*-conorm *TC* has no nilpotent elements.

From Theorem 3, it follows that if similarity function *S* is co-symmetric but only weakly consistent, then there is no reason to use in (5) pseudo-difference operation $a \ominus_{TC} b = a - b$ but one can use pseudo-difference operations associated to maximum TC_M and product TC_P t-norms. Some examples one can find in [3, 4].

5 Generalized Correlation Functions

Here we introduce the correlation functions that can be not strong. In the definition of such correlation functions, we do not require that the underlying set Ω equipped with some negation operation. We will consider the methods of construction of such correlation functions and, further, we will show when these methods will define strong correlation functions. Finally, we establish the one-to-one correspondence between "bipolar" similarity functions and strong correlation functions.

Definition 7. A function $A: \Omega \times \Omega \rightarrow [-1,1]$ on a nonempty set Ω is a *correlation function* if it is symmetric, reflexive and has negative value for some *x, y* in Ω: $A(x,y) < 0$. A correlation function *A* is called (–1)-normal (in (*x,y*)) if $A(x,y) = -1$ for some *x, y* in Ω.

A non-strong correlation function will be called a *weak* o *semi-correlation function*.

Proposition 3. Suppose *S* and *D* are similarity and dissimilarity functions on Ω such that for some *x, y* in Ω it is fulfilled: $S(x,y) < D(x,y)$, then the function defined for all *x,y* in Ω by:

$$A(x,y) = S(x,y) - D(x,y), \tag{6}$$

is a correlation function.

Proof. The similarity of *A* follows from the similarity *S* and *D*. The reflexivity of *A* follows from the reflexivity of *S* and irreflexivity of *D*. When for some *x, y* in Ω it is fulfilled $S(x,y) < D(x,y)$, the value of *A* in (6) is negative. ■

If *S* is 0-normal and *D* is 1-normal in the same pair of elements (*x,y*) then *A* is (-1)-normal. If similarity and dissimilarity functions in (6) are complementary, then the function (6) will be a correlation function if for some *x, y* in Ω it fulfills $S(x,y) < 0.5$.

The formula (6) has a reasonable interpretation: *the correlation between x and y is positive if the similarity between them is greater than the dissimilarity, and the correlation is negative in the opposite case*.

Definition 8. If the similarity *S* and dissimilarity *D* functions in (6) are complementary, then the correlation function *A* defined by (6) is called complementary to *S* and *D*, and (*S,D,A*) for such functions is called a *complementary (*or *correlation) triplet*.

From Definitions 3 and 8 and from Proposition 3 it follows that the similarity, dissimilarity and correlation functions from the complementary triplet (*S,D,A*), can be obtained one from another for all *x, y* in Ω as follows:

$$S(x,y) = 1 - D(x,y), \qquad D(x,y) = 1 - S(x,y), \tag{7}$$

$$A(x,y) = 2S(x,y) - 1, \qquad S(x,y) = 0.5(A(x,y)+1), \tag{8}$$

$$A(x,y) = 1 - 2D(x,y), \qquad D(x,y) = 0.5(1 - A(x,y)). \tag{9}$$

Example 5. The Spearman's rank correlation coefficient is equivalent to the Pearson's product-moment correlation coefficient applied to rankings of *n* objects [9, 14]. When each of rankings $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ contains *n* different integer ranks, $1 \le x_i, y_i \le n$, i.e. there are no ties, the Spearman's rank correlation coefficient is calculated as follows:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{10}$$

where $d_i = x_i - y_i$. Consider the function:

$$D(x,y) = \frac{3\sum_{i=1}^{n}(x_i-y_i)^2}{n(n^2-1)}.$$

It is irreflexive, symmetric and takes values in the interval [0,1], hence it is the dissimilarity function and Spearman's rank correlation coefficient defined by $r_S(x,y) = 1 - 2D(x,y)$, is a correlation function, compare with (9).

Example 6. The Kendall rank correlation coefficient is defined for measurements without ties $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ of two variables for *n* objects as follows [11,14,17]:

$$\tau = \frac{number\ of\ concordant\ pairs - number\ of\ disconcordant\ pairs}{n(n-1)/2},$$

that can be represented as the difference between similarity S and dissimilarity D functions (6):

$\tau(x,y) = S(x,y) - D(x,y)$,

defined as follows:

$$S(x,y) = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} s_{ij}}{n(n-1)/2}, \qquad D(x,y) = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{ij}}{n(n-1)/2},$$

where concordant and disconcordant pairs (i,j), $i<j$, defined by:

$$s_{ij} = \begin{cases} 1, & if\ (x_i - x_j)(y_i - y_j) > 0 \\ 0, & otherwise \end{cases},$$

$$d_{ij} = \begin{cases} 1, & if\ (x_i - x_j)(y_i - y_j) < 0 \\ 0, & otherwise \end{cases}.$$

The Kendall rank correlation coefficient uses the signs of differences between measurement values: $(x_i - x_j)$ and $(y_i - y_j)$, hence takes into account only the ordering (or ranking) of these values. Note that S is reflexive, D is irreflexive, and both functions are symmetric, take values in [0,1]; hence they are similarity and dissimilarity functions, respectively. Due to the measurements are without ties, i.e., all have different values, from $s_{ij} + d_{ij} = 1$ for all $1 \le i < j \le n$, it follows $S(x,y)+D(x,y)= 1$, i.e., S and D are complementary functions, and together with Kendall rank correlation coefficient τ compose complementary triplet (S,D,τ), hence for S, D, and $A= \tau$ all relations (7)-(9) are fulfilled.

6 Relationship between Strong Correlation Functions and Similarity Functions

The following theorem answers the question: when the correlation function from a complementary triplet will be strong.

Theorem 4. Let N be a reflection on Ω and V be a nonempty subset of $\Omega \setminus FP(\Omega)$ closed under N. The formulas (8) establish the one-to-one correspondence between the strong correlation functions and the similarity functions satisfying for all x, y in V the following property:

$$S(x,y) + S(x,N(y)) = 1. \tag{11}$$

Proof. Suppose (1) is fulfilled, then

$$A(x,y) + A(x,N(y)) = 0, \tag{12}$$

applying (8) we obtain from (12): $2S(x,y) - 1 + 2S(x,N(y)) - 1 = 0$, and $2S(x,y) + 2S(x,N(y)) = 2$, i.e. (11) is fulfilled. Similarly, from (11), (8) we can obtain (1) ■

A similarity function satisfying (11) for all x, y in V will be referred to as a *bipolar similarity function*, see [7], because the right side of (11) one can consider as a sum: $1 = 0 + 1$, of "poles" 0 and 1 of the interval of similarity values [0,1]. Similarly, we can present the property (12) of the inverse relationship of correlation function (1) as *bipolarity* condition, where the right side equal to the sum: $0 = -1 + 1$, of the "poles" of the interval of correlation function values [-1,1]. For this reason, a strong correlation function one can consider also as a *bipolar correlation function*, and Theorem 4 one can interpret as follows: *there exists a one-to-one correspondence between bipolar correlation functions and bipolar similarity functions*.

From Theorem 4 it follows also that the formulas (9) establish a one-to-one correspondence between strong correlation functions and *bipolar dissimilarity functions* satisfying for all x, y in V the bipolarity condition:

$$D(x,y) + D(x,N(y)) = 1. \tag{13}$$

It is easy to see from (7) that for complementary similarity and dissimilarity functions the bipolarity relations (11) and (13) are equivalent to:

$$D(x,y) = S(x,N(y)), \quad S(x,y) = D(x,N(y)). \tag{14}$$

Proposition 4. Let N be a reflection on Ω and V be a nonempty subset of $\Omega \setminus FP(\Omega)$ closed under negation N. Let S be a bipolar similarity function on V then it is consistent and co-symmetric.

Proof. From (11) replacing y by x and from the reflexivity of S obtain the consistency of S: $S(x,N(x)) = 1 - S(x,x) = 1 - 1 = 0$.

Replacing in (11) x by $N(x)$ and y by $N(y)$ obtain: $S(N(x),N(y))+S(N(x),N(N(y))) = 1$. Further, applying involutivity of N: $N(N(y)) = y$, and symmetry of S obtain: $S(N(x),N(y)) = 1 - S(N(x),N(N(y))) = 1 - S(N(x),y) = 1 - S(y,N(x))$. From (11) we have $S(y,x) + S(y,N(x)) = 1$, that together with symmetry of S gives: $1 - S(y,N(x)) = S(y,x) = S(x,y)$, and finally we obtain: $S(N(x),N(y)) = S(x,y)$, i.e. co-symmetry of S ■

From Proposition 4 and Theorem 1 it follows that the bipolar similarity function can be used for constructing a strong correlation function by (3): $A(x,y) = S(x,y) - S(x,N(y))$, and from (14) it follows that this strong correlation function is complementary to S, i.e. (6) is fulfilled: $A(x,y) = S(x,y) - D(x,y)$, hence the simplified formulas (8): $A(x,y) = 2S(x,y) - 1$, and (9): $A(x,y) = 1 - 2D(x,y)$, can be used. From (8) and (9) it follows that the strong correlation function is the rescaling of bipolar similarity and dissimilarity functions.

Example 7 (see Example 6). Let Ω be the set of real-valued n-tuples $x = (x_1,\ldots,x_n)$. Define the reflection on Ω as follows: $N(x) = (M - x_1,\ldots, M - x_n)$, where M is some constant, for example, $M = 0$ or $M = \max\{x_1,\ldots,x_n\}$. For similarity function S from Example 6 denote s_{ij} the addends that will be used for calculating $S(x,N(y))$:

$$s_{ij} = \begin{Bmatrix} 1, & if & (x_i - x_j)\left(N(y_i) - N\left(y_j\right)\right) > 0 \\ 0, & & otherwise \end{Bmatrix}$$

$$= \begin{Bmatrix} 1, & if & (x_i - x_j)\left((M - y_i) - (M - y_j)\right) > 0 \\ 0, & & otherwise \end{Bmatrix}$$

$$= \begin{cases} 1, & if\ (x_i - x_j)(y_j - y_i) > 0 \\ 0, & otherwise \end{cases}$$

$$= \begin{cases} 1, & if\ (x_i - x_j)(y_i - y_j) < 0 \\ 0, & otherwise \end{cases}$$

$$= d_{ij}.$$

Hence, $D(x,y) = S(x,N(y))$, similarity function S is bipolar and the Kendall rank correlation coefficient is the strong correlation function.

Example 8. It can be shown also that the dissimilarity functions considered in Examples 3, 4 and 5 are bipolar and give by $A(x,y) = 1 - 2D(x,y)$, the following strong correlation functions, respectively: the Pearson's product-moment correlation coefficient, the cosine correlation coefficient, and the Spearman's rank correlation coefficient. In the last case, the reflection operation on the set of rankings reverses the rankings by $N(x_i)=n+1-x_i$.

7 Related Works and Conclusion

Some of the relationships between similarity, dissimilarity, and correlation coefficients considered in Sections 5 and 6 have been mentioned in several works. The formulas like (6), (9) appear in [14] in the calculation of Kendall rank correlation coefficient, where instead of similarity and dissimilarity functions, the positive and negative scores are used. The formula (10) of Spearman rank correlation also given in this book. Kendall [14] proposed the "general correlation coefficient" using the values a_{ij} and b_{ij} defined for Spearman, Kendall and Pearson correlation as functions of differences $x_i - x_j$ and $y_i - y_j$ for all $i,j = 1,\ldots,n$. How to define a_{ij} and b_{ij} in general case, it is not clear. It is only required that $a_{ij} = -a_{ji}$ and $b_{ij} = -b_{ji}$. Also, the generalization of formulas like (6) and (9) on the universal domain not considered. The half of formulas from (7)-(9) used for constructing similarity and dissimilarity measures from correlation coefficients were considered in [1, 16]. The problem of the construction of correlation coefficients from similarity and dissimilarity measures not considered in these works. The properties like symmetry, inverse relationship and co-symmetry of a "good" relative measure of the association were also considered in [11] but the negation has been considered only for real numbers and the general methods of construction of such measures on the universal domain with involution were not

proposed. Some formulas like (6), and from (7)-(9) for the probabilities of concordance and discordance are considered in [11]. Theorems 1-3 are considered in [3, 4, 8].

The methods of construction of correlation functions on universal domain Ω as difference between similarity and dissimilarity functions proposed in this paper in Section 5 and one-to-one correspondence between strong correlation functions and bipolar similarity functions formulated in Section 6 together with the methods of construction of strong correlation functions (association measures) proposed in [2, 3, 4, 8] have more straightforward interpretation of the correlations in terms of similarities and dissimilarities. These methods give a general and regular methodology for constructing correlation functions on different domains where one can introduce a reflection (negation) operation.

The very surprising result has been obtained here in Theorem 4. It establishes deep relationships between correlation coefficients and similarity (dissimilarity) measures. The one-to-one correspondence between strong correlations and bipolar similarity functions together with (8) shows that there is no much difference between these two concepts. This result paves the way for the construction of new strong correlation functions on almost any domain where negation (reflection) operation and similarity or dissimilarity functions satisfying suitable properties can be defined. The methods of construction of similarity and dissimilarity functions suitable for the generation of correlation functions (association measures) can be based on the results obtained in [2, 8]. For example, one can construct dissimilarity functions using Minkowski distance and *p*-transformation of data, using the methods of co-symmetrization of similarity functions, etc.

Acknowledgment

This work was partially supported by project IPN SIP 20196374. The author also thanks Dr. Imre Rudas for his help in the publication of this work.

References

[1] H. T. Clifford, W. Stephenson: An introduction to numerical classification, Academic Press, New York, 1975

[2] I. Batyrshin: Constructing time series shape association measures: Minkowski distance and data standardization, BRICS CCI–2013, IEEE, 2013, pp. 204-212, https://arxiv.org/ftp/arxiv/ papers/1311/1311.1958.pdf

[3] I. Batyrshin: Association measures and aggregation functions, Advances in soft computing and its applications. Lecture Notes in Computer Science, Vol. 8266, Springer, 2013, pp. 194-203

[4] I. Z. Batyrshin: On definition and construction of association measures, Journal of Intelligent & Fuzzy Systems, Vol. 29, 6, 2015, pp. 2319-2326

[5] I. Batyrshin, V. Kreinovich: One more geometric interpretation of Pearson's correlation, Thailand Statistician, Vol. 13, 2015, pp.125-126

[6] I. Batyrshin, V. Solovyev, V. Ivanov: Time series shape association measures and local trend association patterns, Neurocomputing, Vol. 175, 2016, pp. 924-934

[7] I. Batyrshin, F. Monroy-Tenorio, A. Gelbukh, L.A. Villa-Vargas, V. Solovyev, N. Kubysheva: Bipolar rating scales: a survey and novel correlation measures based on nonlinear bipolar scoring functions, Acta Polytechnica Hungarica, Vol. 14, 3, 2017, pp. 33-57

[8] I. Batyrshin: Towards a general theory of similarity and association measures: similarity, dissimilarity and correlation functions, Journal of Intelligent and Fuzzy Systems, Vol. 36, 4, 2019, pp. 2977-3004

[9] P. Y. Chen, P. M. Popovich: Correlation: Parametric and nonparametric measures, Sage, Thousand Oaks, CA, 2002

[10] J. D. Gibbons: Nonparametric measures of association, Sage Publications, Iowa, 1993

[11] J. D. Gibbons, S. Chakraborti: Nonparametric statistical inference, Dekker, New York, 2003, 4th ed.

[12] M. Grabisch, J. L. Marichal, R. Mesiar, E. Pap: Aggregation Functions, Cambridge Univ. Press, Cambridge, UK, 2009

[13] P. Jaccard: Nouvelles recherches sur la distribution florale, Bull. Soc. Vaud. Sci. Nat., Vol. 44, 1908, pp. 223-270

[14] M. G. Kendall: Rank correlation methods, Griffin, London, 1970, 4th ed.

[15] E. P. Klement, R. Mesiar, E. Pap: Triangular norms, Springer Science & Business Media, 2013

[16] M-J. Lesot, M. Rifqi, H. Benhadda: Similarity measures for binary and numerical data: a survey, Int. J. Knowledge Engineering and Soft Data Paradigms, Vol. 1, 2009, pp. 63-84

[17] A. M. Liebetrau: Measures of Association, Sage Publications, Iowa, 1983

[18] G. Salton: Automatic text processing: the transformation, analysis, and retrieval of information by computer, Addison-Wesley, Boston, MA, 1989

[19] P. N. Tan, V. Kumar, J. Srivastava: Selecting the right interestingness measure for association patterns, 8th Proc. Eighth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2002, pp. 32-41

[20] G. U. Yule: On the association of attributes in statistics: with illustrations from the material of the childhood society, &c., Phil. Trans. Royal Soc. of London. Series A, Vol. 194, 1900, pp. 257-319