(Dis)Similarity-Based Correlation Functions on Polar Scales

Ildar Batyrshin[a], Nailya Kubysheva[b], Raouf Akhtiamov[b]

[a] *Instituto Politécnico Nacional, CIC-IPN, Av. Juan de Dios Bátiz S/N, CDMX, 07738, México*
[b] *Kazan Federal University, 18 Kremlyovskaya str., Kazan, 420008, Russian Federation*

Abstract
We observe the methods of constructing (dis)similarity-based correlation functions defined on scales with polarity given by an involutive operation. We consider basic polar scales: binary scales and bipolar scales with more than two grades. Correlation functions are constructed using similarity and dissimilarity functions defined on such scales and satisfying suitable properties.

Keywords 1
Correlation, similarity, dissimilarity, binary data, bipolar scale

1. Introduction

Correlation and association coefficients were used for more than one hundred years as important tools of knowledge extraction from data in medicine, ecology, signal processing, social and behavioral sciences, etc. [1-10]. In the last decades, they are also considered together with similarity and association measures as measures of relationship and interestingness in data mining, recommender systems, and machine learning [8, 11-14]. Recently, a correlation was defined as a function defined on a set with involution operation and satisfying some properties of Pearson's product-moment correlation coefficient [15-18]. It was shown that the correlation function could be defined using suitable similarity or dissimilarity functions. We will call these functions for short as (dis)similarity functions. This new approach to the construction of correlation functions was used for introducing new correlation coefficients on different domains [17-24]. In this work, we observe the methods of construction of correlation coefficients on polar scales: binary scales and bipolar scales with more than two grades. These scales differ by the methods of definition of involution operation on the scale and by the methods of definition of (dis)similarity functions.

The paper has the following structure. Section 2 describes the definition of the correlation and (dis)similarity functions and considers the methods of construction of correlation functions from (dis)similarity functions. Section 3 considers the methods of constructing correlation functions on a binary scale. Section 4 describes the method of constructing a correlation function on bipolar scales using a utility function defined on the scale. Section 5 contains the conclusion.

2. Correlation and (dis)similarity functions

Let Ω be a nonempty set with involutive operation $N(x)$ called *reflection* or *negation* such that for all x in Ω the reflection $N(x)$ belongs to Ω, and satisfies the property:

$$N\big(N(x)\big) = x. \qquad \text{(involutivity)}$$

The element x in Ω such that $N(x) = x$, is referred to as a fixed point of the negation in Ω. The set of all fixed points of N in Ω denoted by $FP(N,\Omega)$ or FP. The set of fixed points can be empty.

Russian Advances in Fuzzy Systems and Soft Computing: selected contributions to the 8-th International Conference on Fuzzy Systems, Soft Computing and Intelligent Technologies (FSSCIT-2020), June 29 – July 1, 2020, Smolensk, Russia
EMAIL: batyr1@gmail.com

Let Ω be a set with a reflection operation *N*, and *V* be a non-empty subset of $\Omega \setminus FP$ closed under operation *N*. The *correlation function (association measure)* on *V* is a function $A(x, y)$ such that for any *x* and *y* in *V* it takes values in [-1, 1], and satisfies the following properties:

$$A(x,y) = A(y,x), \quad \text{(symmetry)}$$
$$A(x,x) = 1, \quad \text{(reflexivity)}$$
$$A(x,N(y)) = -A(x,y). \quad \text{(inverse relationship)}$$

From the definition of the correlation function, it follows the fulfillment of the following properties fulfilled for all *x* and *y* in *V*:

$$A(x,N(x)) = -1, \quad \text{(opposite elements)}$$
$$A(N(x),N(y)) = A(x,y), \quad \text{(co-symmetry-I)}$$
$$A(x,N(y)) = A(N(x),y). \quad \text{(co-symmetry-II)}$$

The correlation between the elements of Ω and possible fixed points x_{Fp} of *N* is not defined. Depending on the underlying domain, such correlations can be redefined for all *x* in Ω as follows: $A(x, x_{FP}) = 0$ or $A(x, x_{FP}) = 1$.

A function $S(x,y)$ is called a *similarity function* on Ω if for all *x*,*y* in Ω it takes values in [0, 1] and satisfies the following properties:

$$S(x,y) = S(y,x), \quad \text{(symmetry)}$$
$$S(x,x) = 1. \quad \text{(reflexivity)}$$

Dually, a function $D(x,y)$ is called a *dissimilarity function* on Ω if for all *x*,*y* in Ω it takes values in [0, 1] and satisfies the following properties:

$$D(x,y) = D(y,x), \quad \text{(symmetry)}$$
$$D(x,x) = 0. \quad \text{(irreflexivity)}$$

Similarity and dissimilarity functions are called *complementary* if, for all *x*, y in Ω it is fulfilled:

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

A non-negative real-valued function $d(x,y)$ of elements of Ω will be referred to as a *distance* if for all *x*, y in Ω it fulfills:

$$d(x,y) = d(y,x), \quad \text{(symmetry)}$$
$$d(x,x) = 0. \quad \text{(irreflexivity)}$$

If for some positive real value M for all *x*, y in Ω it is fulfilled $d(x,y) \leq M$, then the function

$$D(x,y) = \frac{d(x,y)}{M},$$

will be a dissimilarity function taking values in the interval [0, 1].

Let *V* be a subset of $\Omega \setminus FP(N)$ closed under operation *N*. A similarity and dissimilarity functions *S* and *D* are called *consistent* on *V* if for *x* in *V* it is fulfilled, respectively:

$$S(x,N(x)) = 0. \quad \text{(consistency)}$$
$$D(x,N(x)) = 1. \quad \text{(consistency)}$$

(Dis)similarity functions S and D are called *co-symmetric* on *V* if for all *x*,*y* in *V* it is fulfilled, respectively:

$$S(N(x),N(y)) = S(x,y). \quad \text{(co-symmetry-I)}$$
$$D(N(x),N(y)) = D(x,y). \quad \text{(co-symmetry-I)}$$

It was shown that co-symmetry-I is equivalent to co-symmetry-II:

$$S(x,N(y)) = S(N(x),y). \quad \text{(co-symmetry-II)}$$
$$D(x,N(y)) = D(N(x),y). \quad \text{(co-symmetry-II)}$$

Theorem 1 [15-17]. Let Ω be a set with a reflection operation N, and V be a subset of $\Omega \setminus FP$ closed under N. If S is a co-symmetric and consistent similarity function on V, then the function:

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

defined for all x,y in V is a correlation function on V.

Dually we obtain for co-symmetric and consistent dissimilarity function:

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

3. Correlation functions for binary scales

Let $X = \{0,1\}$ be the binary scale. X can correspond to the measurements of a binary attribute or variable. Consider the Cartesian product of n binary scales: $\Omega = X_1 \times \ldots \times X_n = \{(x_1, \ldots, x_n) | x_i \in X_i, i = 1, \ldots, n\}$. If the binary scales correspond to different attributes, then n-tuples $(x_1, \ldots, x_n)$ correspond to measurements of the attribute values of some object x. If all binary scales correspond to the same attribute X, i.e. $X_1 = X_2 = \cdots X_n = X$, then n-tuples $(x_1, \ldots, x_n)$ usually give the values of the attribute X for n different objects or measurements. If it is not confusing, X denotes an attribute and the set of its possible values.

Similarity functions between two binary n-tuples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ use the following parameters. Denote:

- a, the number of measurements when $x_i = y_i = 1$;
- b, the number of measurements when $x_i = 1, y_i = 0$;
- c, the number of measurements when $x_i = 0, y_i = 1$;
- d, the number of measurements s when $x_i = y_i = 0$,

where $i = 1, \ldots, n$. The numbers a and d are called the numbers of *positive and negative matches*, respectively. For numbers a, b, c, d, we have $a + b + c + d = n$.

For each n-tuple $x = (x_1, \ldots, x_n)$, its negation $N(x) = \bar{x}$ defined as follows: $\bar{x} = (1 - x_1, \ldots, 1 - x_n)$ This negation is involutive, i.e., for any binary n-tuple x it is fulfilled: $\bar{\bar{x}} = x$.

Consider examples of similarity functions [8, 17]:

Simple Matching:

$$S_{SM}(x,y) = \frac{a+d}{a+b+c+d},$$

YQ similarity:

$$S_{YQ}(x,y) = \frac{ad}{ad+bc}.$$

These similarity functions symmetric, reflexive, co-symmetric, and consistent, hence satisfy the conditions of Theorem 1.

We obtain:

$$S_{SM}(x,N(y)) = \frac{b+c}{a+b+c+d},$$
$$S_{YQ}(x,N(y)) = \frac{bc}{ad+bc}.$$

and from Theorem 1 we obtain, respectively, the well-known association coefficients:

Hamann's association coefficient:

$$A_H = S_{SM}(x,y) - S_{SM}(x,N(y)) = \frac{(a+d)-(b+c)}{a+b+c+d},$$

Yule's Q association coefficient:

$$A_{YQ} = S_{YQ}(x,y) - S_{YQ}(x,N(y)) = \frac{ad-bc}{ad+bc}.$$

Consider Yule's W association coefficient [8, 10, 17]:

$$A_{YW}(x,y) = \frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}.$$

One can check that it takes values in $[-1,1]$ and satisfies the properties (1)-(3). Hence it is a correlation function. The function

$$S_{YW}(x,y)=\frac{\sqrt{ad}}{\sqrt{ad}+\sqrt{bc}},$$

satisfies the properties of co-symmetric and consistent similarity functions. Calculate:

$$S_{YW}(x,\bar{y})=\frac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}},$$

and from Theorem 1 obtain Yule's W:

$$A(x,y)=S_{YW}(x,y)-S_{YW}(x,\bar{y})=\frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}=A_{YW}(x,y).$$

Consider phi coefficient that is Pearson's product-moment correlation coefficient applied to binary variables [8]:

$$A_{\rho}(x,y)=\frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}.$$

One can check that it is a correlation function. Using *Sokal and Sneath* similarity measure [8]:

$$S_{\rho}(x,y)=\frac{ad}{\sqrt{(a+b)(a+c)(b+d)(c+d)}},$$

that is a co-symmetric and consistent similarity function, we obtain:

$$S_{\rho}(x,\bar{y})=\frac{bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}.$$

Applying Theorem 1, we obtain phi coefficient.

From Theorem 1, it follows that these association coefficients satisfy the properties of correlation functions. In [24], using Theorem 1, a new parametric correlation function for binary data was proposed.

4. Bipolar scales

Many linguistic scales, like Likert scales, etc. [25-29], are linear ordered and have symmetric structure such that the grades of one half of the scale can be mapped onto another half of the scale. For example, the grades of the scale (*never, seldom, sometimes, often, always*) are ordered from the left to the right and symmetric with respect to the center of the scale *C*= *sometimes*. This symmetry can be represented by an involutive operation *N* as follows: *N*(*never*)= *always*, *N*(*seldom*)= *often*, *N*(*sometimes*)= *sometimes*, *N*(*often*)= *seldom*, *N*(*always*)= *never*. For example, we have: *N*(*N*(*never*))= *N*(*always*) = *never*, etc. A *bipolar* scale *L* with *n* ordered categories $c_1<\ldots<c_n$ can be represented by an ordered set of indexes of these categories $J=\{1,\ldots,n\}$, $n>1$, with the *negation* operation *N(j)* defined by:

$$N(j) = n+1-j \text{ for all } j\in J.$$

This negation is involutive:

$$N(N(j))=j,\quad \text{for all } j\in J.$$

It is supposed that the bipolar scale has an odd number of elements, i.e., $n=2m+1$ for some positive integer *m*. In this case, it has a *fixed point* of the negation *N*: $x_{FP}=m+1$, coinciding with the center of the scale: $C=m+1$. One can check that $C=m+1$ is the fixed point:

$$N(C) = n+1-m-1 = 2m+1-m = m+1 = C.$$

The fixed point is called the *center* of the bipolar scale.

The values $P_1 = min(J) = 1$ and $P_2 = max(J) = n$ are called the *negative* and the *positive* poles. From $N(j)+j = 1+n$, for all *j* in *J*, it follows that the formula of the negation can be presented in the form of *bipolarity*:

$$N(j)+j = P_1+P_2.$$

From $C = \frac{1}{2}(P_1 + P_2)$ it follows also:

$$N(j) + j = 2C, \text{ for all } j \in J.$$

Also it is fulfilled: $N(P_1) = P_2$ and $N(P_2) = P_1$.

A bipolar scale can be represented by another ordered set of indexes: $K = \{-m, \dots, -1, 0, 1, \dots, m\}$ that is the *centered form* of the bipolar scale $J = \{1, \dots, 2m + 1\}$, $m > 0$. The negation operation on the scale K is defined by:

$$N(k) = -k, \qquad \text{for all } k \in K.$$

Unless it can cause confusion, we will use the same letter N for the negation on J and on K, using the arguments j or k, respectively. It is clear that N on K is a strictly decreasing and involutive function, i.e., $N(N(k)) = k$, for all $k \in K$. This scale has the center $C = 0$ with $N(0) = -0 = 0$, and poles $P_1 = min(K) = -m$ and $P_2 = max(K) = m$. The bipolarity properties also fulfilled for the scale K:

$$N(k) + k = 2C = P_1 + P_2 = 0, \quad \text{for all } k \in K.$$

The scale $K = \{-m, \dots, 0, \dots, m\}$, $m > 0$, with the corresponding negation is called a *centered bipolar scale.*

For example, the 5-point bipolar scale (*never*, *seldom, sometimes*, *often*, *always*) can be given by two sets *of* indexes. The set $J = \{1, 2, 3, 4, 5\}$ has the negation $N(j) = 6–j$, poles $P_1 = 1$, $P_2 = 5$, and the center $C = 3$. The set $K = \{-2, -1, 0, 1, 2\}$ has the negation $N(k) = -k$, poles $P_1 = -2$, $P_2 = 2$, and the center $C = 0$.

The bipolar scales $J = \{1, \dots, 2m+1\}$, $m > 0$, and $K = \{-m, \dots, 0, \dots, m\}$ can be transformed one into the another as follows:

$$k = j - m - 1, \; j = k + m + 1, \quad \text{for all } j \in J \text{ and } k \in K.$$

Let I denotes the scale J or K. A strictly increasing real function U defined on I is called a *scoring* or *utility function* on I. This function is called a *bipolar scoring function* (BSF) if it satisfies the condition:

$$U(N(i)) + U(i) = U(P_1) + U(P_2), \quad \text{for all } i \in I. \qquad \text{(bipolarity)}$$

For the scales with the center C from $N(C) = C$, we have $U(P_1) + U(P_2) = 2U(C)$, and the bipolarity property can be given by:

$$U(N(i)) + U(i) = 2U(C), \qquad \text{for all } i \in I. \qquad \text{(bipolarity)}$$

A BSF is called a *centered bipolar scoring function* (CBSF) if

$$U(C) = 0.$$

The definition of the centered bipolar scoring function implies

$$U(N(i)) = -U(i), \qquad \text{for all } i \in I.$$

For a CBSF defined on a centered bipolar scale $K = (-m, \dots, 0, \dots, m)$, $m>0$, we have for all $k \in K$:

$$U(k) > 0 \; if \; k > 0,$$
$$U(0) = 0,$$
$$U(k) < 0 \; if \; k < 0,$$
$$U(-k) = -U(k).$$

In [16], it was proposed several methods of generation of a bipolar utility functions. Consider the method of construction of generator-based bipolar utility function on the bipolar scale

$$K = \{-m, \dots, -1, 0, 1, \dots, m\}, \; m > 0.$$

Let G be a positive real value and g:$\{0, \dots, m\} \rightarrow [0, G]$ be a strictly increasing function such that $g(0) = 0$, $g(m) = G$. This function *generates* the function W: $K \rightarrow [-G, G]$ defined by:

$$W(k) = g(k) \quad \text{for all } k \in \{0, \dots, m\},$$
$$W(k) = -g(-k) \quad \text{for all } k \in \{-m, \dots, -1\}.$$

Here is an example of a parametric generator of bipolar utility function inspired by the parametric Sugeno negation used in fuzzy logic, with the parameter $p > -1$:

$$g_1(k) = \frac{Gk(1+p)}{m+pk},\ k = 0, \dots, m.$$

In [20-21] it is proposed the following correlation function on rating profiles with a bipolar utility function. Let I be a bipolar scale ($I = J$ or $I = K$) with the center C, X be a set of profiles $x = (x_1, \dots, x_M)$ of the length M, $x_s \in I$, $s = 1, \dots, M$, with the central profile $C_X = (C, \dots, C)$, and U be a bipolar utility function on I. The following function is a correlation function on $X \backslash \{C_X\}$:

$$A_U(x, y) = \frac{\sum_{S=1}^{M}(U(x_S)-U(C))(U(y_S)-U(C))}{\sqrt{\sum_{S=1}^{M}(U(x_S)-U(C))^2}\sqrt{\sum_{S=1}^{M}(U(y_S)-U(C))^2}}.$$

5. Conclusion

The theory of correlation functions gives the possibility to introduce correlation functions on polar scales. For binary scales, it can be determined which of existing association coefficients are correlation functions. These association coefficients satisfy the properties, similar to Pearson's product-moment correlation coefficient. Also, new correlation functions for binary data can be introduced [24]. For bipolar scales, the general methods for construction correlation coefficients have been considered [20, 21]. Binary scales can often be obtained from Likert scales. Bipolarity usually does not consider for Likert scales, although bipolarity in the form of symmetry is presented in such scales. Bipolar correlation functions give the possibility to take into account this bipolarity of scales. Considered correlation coefficients can be used to measure the correlation between data from polar scales. The dissimilarity-based correlations on circular scales are considered in [23]. One of the promising areas of application of correlation on polar scales is the sentiment analysis.

6. Acknowledgements

This work is partially supported by projects IPN SIP 20200853 and RFBR No 20-07-00770.

7. References

[1] P. Y. Chen, P. M. Popovich, Correlation: Parametric and Nonparametric Measures, Sage, Thousand Oaks, CA, 2002.

[2] J. D. Gibbons, S. Chakraborti, Nonparametric Statistical Inference. 4th ed., Dekker, New York, 2003.

[3] M. G. Kendall, Rank Correlation Methods, Griffin, London, 4 (1970).

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

[5] B. Mirkin, Core Concepts in Data Analysis: Summarization, Correlation and Visualization, Springer Science & Business Media, 2011.

[6] J. L. Rodgers, W. A. Nicewander, Thirteen ways to look at the correlation coefficient, The American Statistician, 42 (1988) 59-66.

[7] M. E. Rodriguez-Salazar, S. Alvarez-Hernandez, E. Bravo-Nunez, Coeficientes de Asociación, Plaza y Valdes Editores, Mexico, 2001.

[8] M. J. Warrens, Similarity measures for 2×2 tables. Journal of Intelligent & Fuzzy Systems, 36, (2019) 3005-3018.

[9] G. U. Yule, On the association of attributes in statistics. Philosophical Transactions of the Royal Society, A 194 (1900) 257–319.

[10] G. U. Yule, On the methods of measuring the association between two attributes. Journal of the Royal Statistical Society, 75 (1912) 579–652,

[11] P. N. Tan, V. Kumar, J. Srivastava, Selecting the right interestingness measure for association patterns, Proc. Eighth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 8 (2002) 32-41.
[12] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17 (2005) 734-749.
[13] U. Shardanand, P. Maes, Social information filtering: algorithms for automating "word of mouth". in: Proceedings of the SIGCHI conference on Human factors in computing systems, 1995, 210-217.
[14] F. Ricci, L. Rokach, B. Shapira, P. B. Kantor, Recommender Systems Handbook, Springer US, 2011.
[15] I. Batyrshin, Association measures and aggregation functions. in: Advances in Soft Computing and Its Applications. Lecture Notes in Computer Science, 8266 (2013) 194-203.
[16] I. Z. Batyrshin, On definition and construction of association measures, Journal of Intelligent & Fuzzy Systems, 29 (2015) 2319-2326.
[17] I. Batyrshin, Towards a general theory of similarity and association measures: similarity, dissimilarity and correlation functions. Journal of Intelligent & Fuzzy Systems, 36 (2019) 2977-3004.
[18] I. Z. Batyrshin, Constructing correlation coefficients from similarity and dissimilarity functions. Acta Polytechnica Hungarica, 16 (2019) 191-204.
[19] I. Z. Batyrshin, Association measures on [0,1], Journal of Intelligent & Fuzzy Systems, 29 (2015) 1011-1020.
[20] 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, 14 (2017) 33-57.
[21] F. Monroy-Tenorio, I. Batyrshin, A. Gelbukh, V. Solovyev, N. Kubysheva, I. Rudas, Correlation measures for bipolar rating profiles, in: Fuzzy Logic in Intelligent System Design. NAFIPS 2017. Advances in Intelligent Systems and Computing, 648 (2018) 22-32. Springer, doi: 10.1007/978-3-319-67137-6_3.
[22] F. J. Garcia-Lopez, I. Batyrshin, A. Gelbukh, Analysis of relationships between tweets and stock market trends. Journal of Intelligent and Fuzzy Systems, 34 (2018) 3337-3347.
[23] I. Batyrshin, N. Kubysheva, V. Tarassov, Dissimilarity-Based Correlation of Movements and Events on Circular Scales of Space and Time, in: Mexican International Conference on Artificial Intelligence, Springer, Cham, 2020, 237-246.
[24] I. Z. Batyrshin, I. Ramirez-Mejia, I. I. Batyrshin, V. Solovyev, Similarity-based correlation functions for binary data, in: Mexican International Conference on Artificial Intelligence, Springer, Cham, 2020, 224-233.
[25] R. Likert, A technique for the measurement of attitudes. Archives of Psychology, New York, 22 (1932) 55.
[26] R. B. Gunderman, S. Chan, The 13-point Likert scale: a breakthrough in educational assessment, Academic radiology, 20 (2013) 1466-1467.
[27] D. A. Pospelov, Models of Reasoning. Essay in the Analysis of Mental Acts. Radio y Svyaz. Moscow, 1989.
[28] I. Z. Batyrshin, On the structure of verbal scales, in: Proceedings of the Second All-Union Conference on Artificial intelligence, Minsk, 1 (1990) 37-40.
[29] V. B. Tarassov, Analysis and modeling of NOT-factors on polar scales, in: Integrated Models and Soft Computing in Artificial Intelligence, Moscow, Fismatlit, 2001, 65-71.
[30] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61 (2010) 2544-2558.