Association Measures and Aggregation Functions

Ildar Batyrshin

Research Program of Applied Mathematics and Computations
Mexican Petroleum Institute
batyr1@gmail.com

Abstract. The concept of association measure generalizing the Pearson correlation coefficient is introduced. The methods of generation of association measures by means of pseudo-difference associated to some t-conorm and by similarity measures are proposed. The association measure can be introduced on any set with involutive reflection operation and suitably defined similarity measure. The methods of construction of association measures by Minkowski metric and data standardization using the aggregation functions are considered. The cosine similarity and the Pearson's correlation coefficient are obtained as partial cases of the proposed general methods.

Keywords: association measure, t-conorm, pseudo-difference, similarity measure, Minkowski distance, correlation coefficient, cosine similarity, involutivity, reflection, idempotence, data standardization.

1 Introduction

The Pearson correlation coefficient plays an important role in data analysis giving possibility to measure possible direct and inverse relationships between variables. It is considered as a measure of the strength of linear relationship between variables but it is not always suitable for measuring possible associations between variables in general case [1] and for measuring associations between time series shapes [2]. It arises the problem of creation of association measures suitable for different applications. An axiomatic definition of time series shape association measures generalizing the properties of correlation coefficient has been considered in [4]. In [3], the general methods of construction of association measures satisfying to the axioms of time series shape association measure have been proposed. In the present work the results of [3] are extended in several directions. First, the problem of definition and construction of association measures is considered here from the more general point of view of the theory of aggregation functions [6]. It gives possibility to extend the methods of generation of association measures using the concept of pseudo difference associated with some *t*-conorm. Second, the concept of association measure is extended from the set of time series on a general domain where some involutive mapping together with a similarity measure related with this mapping can be introduced. It gives possibility to extend the class of association measures that can be considered and generated

F. Castro, A. Gelbukh, and M. González (Eds.): MICAI 2013, Part II, LNAI 8266, pp. 194–203, 2013.

on wide class of objects different from time series. The cosine similarity and the Pearson correlation coefficient are obtained as particular cases of the proposed approach.

The paper has the following structure. Section 2 gives definitions of *t*-conorms and pseudo-differences. Section 3 introduces the concept of the association measure and proposes the methods of construction of these measures on the sets with involutive reflection operation and suitably defined similarity measures. Section 4 considers a set of *n*-tuples of real values (vectors, time series or samples) where association measures can be defined and discusses the methods of standardization of *n*-tuples. Section 5 shows how dissimilarity measures and the Minkowski distance together with standardizations can be used for constructing association measures considered in Section 3. The cosine similarity and the Pearson's correlation coefficient are obtained from the general methods of construction of association measures using standardization transformation and Minkowski distance. Conclusions are given in Section 6.

2 Basic Definitions

Consider some definitions from [5-7].

A ***t*-conorm** is a function $S:[0,1]^2 \rightarrow [0,1]$ such that for all $a,b,c \in [0,1]$ the following axioms are satisfied:

$$S(a,b) = S(b,a), \qquad \text{(commutativity)}$$

$$S(a,S(b,c)) = S(S(a,b),c), \qquad \text{(associativity)}$$

$$S(a,b) \leq S(a,c), \text{ whenever } b \leq c, \qquad \text{(monotonicity)}$$

$$S(a,0) = a. \qquad \text{(boundary condition)}$$

From the definition of *t*-conorms it follows for all $a \in [0,1]$:

$$S(1,a) = S(a,1) = 1, \qquad S(0,a) = a.$$

An element $a \in]0,1[$ will be referred to as **a nilpotent element** [5] of S if there exists some $b \in]0,1[$ such that $S(a,b)=1$. A *t*-conorm S has no nilpotent elements if and only if on [0,1] it is fulfilled:

$$\text{from } S(a,b) = 1 \text{ it follows } a = 1 \text{ or } b = 1.$$

Consider simplest *t*-conorms:

$$S_M(a,b) = max\{a,b\}, \qquad \text{(maximum)}$$

$$S_L(a,b) = min\{a+b, 1\}, \qquad \text{(Lukasiewicz } t\text{-conorm)}$$

$$S_P(a,b) = a+b-ab. \qquad \text{(probabilistic sum)}$$

It is clear that the maximum and the probabilistic sum have no nilpotent elements but the Lukasiewicz *t*-conorm has.

Let *S* be a *t*-conorm. The ***S*-difference** is defined by [6]:

$$a \overset{S}{-} b = inf\{c \in [0,1] | S(b,c) \geq a\}$$

for any *a,b* in [0,1].

From the properties of *t*-conorms it follows:

$$1 \overset{S}{-} 0 = 1,$$

$1 \overset{S}{-} b = 1$, if $b < 1$ and *t*-conorm *S* has no nilpotent elements.

Let *S* be a *t*-conorm. The **pseudo-difference** associated to *S* is defined by [6]:

$$a(-)_S b = \left\{ \begin{array}{ll} a \overset{S}{-} b, & if\ a > b \\ -\left(b \overset{S}{-} a\right), & if\ a < b \\ 0, & if\ a = b \end{array} \right\}$$

for any *a,b* in $[0,1]^2$. Equivalently

$$a(-)_S b = sign(a-b)(max\ (a,b) \overset{S}{-} min\ (a,b)).$$

The following pseudo-differences are associated with *t*-conorms S_M, S_L and S_P respectively:

$$a(-)_M b = \left\{ \begin{array}{ll} a, & if\ a > b \\ -b, & if\ a < b \\ 0, & if\ a = b \end{array} \right\},$$

$$a(-)_L b = a - b,$$

$$a(-)_P b = (a-b)/(1 - \min\ (a,b)).$$

3 Association Measures

Suppose *X* is a set with a mapping $N: X \rightarrow X$ satisfying for all elements *x* from *X* the property:

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

This mapping will be called **a reflection operation**.

As an example of a set with a reflection operation one can consider the set $X = [0,1]$ with an involutive negation *N*, defined, e.g. by [7]: $N(x) = 1\text{-}x$, the set of fuzzy sets *X* with an involutive negation of fuzzy sets, the set of vectors or time series of the

length n with real valued elements $x= (x_1,\ldots,x_n)$ and reflection operation $N(x)=(-x_1, \ldots, -x_n)$ etc.

Suppose A is a function $A:X\times X\rightarrow$ [-1,1] satisfying for all x and y from X the properties:

$$A(x,y) = A(y,x), \qquad \text{(symmetry)}$$

$$A(x,x) = 1, \qquad \text{(reflexivity)}$$

and N is a reflection operation on X. The function A will be called an **association measure** (with respect to N) if for all x from X such that $A(N(x),x) \neq 1$, it is fulfilled:

$$A(N(x),x) = -1, \qquad \text{(inverse reflexivity)}$$

$$A(N(x),y) = -A(x,y). \qquad \text{(inverse relationship)}$$

Generally, a function $SIM:X\times X \rightarrow [0,1]$ satisfying for all x and y from X the properties:

$$SIM(x,y) = SIM(y,x), \qquad \text{(symmetry)}$$

$$SIM(x,x) = 1, \qquad \text{(reflexivity)}$$

will be referred to as a **similarity measure**.

Suppose SIM for all x, y satisfies some of the following properties:

$$SIM(N(x),y) = SIM(x,N(y)), \qquad \text{(permutation of reflections)}$$

$$SIM(N(x),x) < 1, \qquad \text{(weak similarity of reflections)}$$

$$SIM(N(x),x) = 0. \qquad \text{(non-similarity of reflections)}$$

It is clear that from the non-similarity of reflections it follows the weak similarity of reflections. Below it is a generalization of the result from [3] on pseudo-differences and reflection operation N.

Theorem 1. Suppose SIM is a similarity measure satisfying the property of permutation of reflections and S is a t-conorm. Then the function:

$$A_{SIM}(x,y) = SIM(x,y)(-)_S SIM(x,N(y))$$

defined for all y such that $SIM(N(y),y)\neq 1$ is an association measure if one of the following is fulfilled:

1. SIM satisfies the non-similarity of reflections;
2. SIM satisfies the weak similarity of reflections and t-conorm S has no nilpotent elements.

Since maximum S_M and probabilistic S_P t-conorms has no nilpotent elements but Lukasiewicz t-conorm S_L has, from the Theorem 1 the following specific methods for construction of association measures can be obtained.

Corollary 2. Suppose *SIM* is a similarity measure satisfying the property of permutation of reflections. For all y such that $SIM(N(y),y) \neq 1$ the association measure can be defined as follows. If *SIM* satisfies the non-similarity of reflections then the function:

$$A_{SIM,L}(x,y) = SIM(x,y) - SIM(x,N(y))$$

is an association measure. If *SIM* satisfies the weak similarity of reflections then the following functions are association measures:

$$A_{SIM,M}(x,y) = \begin{cases} SIM(x,y), & if\ SIM(x,y) > SIM(x,N(y)) \\ -SIM(x,N(y)), & if\ SIM(x,y) < SIM(x,N(y)) \\ 0, & if\ SIM(x,y) = SIM(x,N(y)) \end{cases},$$

$$A_{SIM,P}(x,y) = (SIM(x,y) - SIM(x,N(y)))/(1 - \min(SIM(x,y), SIM(x,N(y)))).$$

In the following section, we will consider the set X of n-tuples of real values $x = (x_1,\ldots,x_n)$ of the length n with the reflection operation $N(x) = -x = (-x_1,\ldots,-x_n)$. In this case a symmetric and reflexive function A will be an association measure if for all x from X such that $A(-x,x) \neq 1$, it is fulfilled:

$$A(-x,x) = -1, \qquad \text{(inverse reflexivity)}$$

$$A(-x,y) = -A(x,y). \qquad \text{(inverse relationship)}$$

The corresponding properties of similarity measures related with reflection operation will have the following notations:

$$SIM(-x,y) = SIM(x,-y), \qquad \text{(permutation of reflections)}$$

$$SIM(-x,x) < 1, \qquad \text{(weak similarity of reflections)}$$

$$SIM(-x,x) = 0. \qquad \text{(non-similarity of reflections)}$$

Generally we do not require as in [3] that association measure satisfies for any real value q the following property:

$$A(x+q,y) = A(x,y). \qquad \text{(translation invariance)}$$

But this property will be considered as necessary if X is a set of time series $x = (x_1,\ldots,x_n)$ [3]. The association measure will be referred to as scale invariant if for all positive real values p it is fulfilled [3]:

$$A(px,y) = A(x,y). \qquad \text{(scale invariance)}$$

It is clear that A_{SIM} is translation or scale invariant if *SIM* satisfies the corresponding properties.

4 Standardization

For any n-tuples x, y and real values p,q define $x+y = (x_1+y_1, \ldots, x_n+y_n)$, $px+q= (px_1+q, \ldots,px_n+q)$. Denote $q_{(n)}$ a constant n-tuple with all elements equal to q. We will write $x= const$ if $x = q_{(n)}$ for some q, and $x \neq const$ if $x_i \neq x_j$ for some $i \neq j$ from $\{1,\ldots,n\}$. From definitions above it follows: $px+q = px+q_{(n)}$.

A transformation $F:R^n \to R^n$ is said to be a **standardization** if for all $x \in R^n$ it is fulfilled $F(x) \neq const$ if $x \neq const$:

$$F(F(x)) = F(x), \qquad \text{(idempotence)}$$

$$F(q_{(n)}) = 0_{(n)}, \qquad \text{for any real value } q.$$

A n-tuple x is said to be in a **standard form wrt a standardization** F if $F(x) = x$.

As it follows from the definition, a standardization F transforms any x into a standard form $F(x)$. We will say that $F(x)$ satisfies ***r*-normality** for some $r = 1,2,\ldots$ if:

$$\sum_{i=1}^{n} \left|F(x)_i\right|^r = 1 \quad .$$

A transformation $E:R^n \to R$ is said to be an **estimate** if $E(q_{(n)})=q$ for any real value q.

It is clear that any aggregation function [6] is an estimate.

We will use the following terminology, if for all n-tuples x,y, for any real value q and for any positive value $p > 0$, F satisfies the properties:

$$F(x+q) = F(x)+q, \qquad \text{(translation additivity)}$$

$$F(x+q) = F(x), \qquad \text{(translation invariance)}$$

$$F(x+y) = F(x)+F(y), \qquad \text{(additivity)}$$

$$F(px) = pF(x), \quad p> 0, \qquad \text{(scale proportionality)}$$

$$F(px) = F(x). \qquad \text{(scale invariance)}$$

Note that in literature the translation additivity is often referred to as shift invariance or translation invariance, the scale proportionality is referred to as scale invariance or homogeneity of degree 1. It is clear that from the additivity of F it follows its translation additivity. The same terminology will be used for E.

Proposition 3. The following transformations are standardizations:

1. $F_1(x)= x-E_1(x)$, if E_1 is a translation additive estimate.
   F_1 is translation invariant and $E_1(F_1(x))= 0$.

2. $F_2(x)= x/E_2(x)$, for $x \neq const$, if E_2 is a scale proportional estimate and $E_2(x) > 0$ for all x.
   F_2 is scale invariant and $E_2(F_2(x))= 1$.

If $E_2(x) = \sqrt[r]{\sum_{i=1}^{n}|x_i|^r}$, then $F_2(x)$ satisfies the r-normality property.
If $E_2(x) = \sum_{i=1}^{n} x_i$, then $F_2(x)$ satisfies the normality property: $\sum_{i=1}^{n} F(x)_i = 1$.

3. $F_3(x) = (x - E_{13}(x))/E_{23}(x)$, if E_{13} is a translation additive and scale proportional estimate, E_{23} is a translation invariant and a scale proportional estimate, and $E_{23}(x) > 0$, for all x.
F_3 is translation and scale invariant, $E_{13}(F_3(x)) = 0$.
If $E_{23}(x) = (\sum_{i=1}^{n}|x_i - E_{13}(x)_i|^r)^{1/r}$ then $F_3(x)$ satisfies the r-normality property.

An estimate E is said to be a **mean** if it satisfies the condition [6]:

$$\min\{x_{1,\ldots,} x_n\} \le E(x) \le \max\{x_{1,\ldots,} x_n\}.$$

Most of the means [6] are translation additive and scale proportional estimates and they can be used for generation standardizations considered above. Below are examples of standardizations $F(x) = f(x)$, where the arithmetic mean is denoted by $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$:

$$f_1(x)_i = x_i - \bar{x},$$

$$f_2(x)_i = x_i - MIN(x),$$

$$f_3(x)_i = \frac{x_i - \bar{x}}{\sqrt{\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2}}.$$

5 Dissimilarity Measures

A **dissimilarity measure** D(x,y) is a real valued function satisfying for all n-tuples x and y the properties:

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

$$D(x,y) \ge D(x,x) = 0.$$

D will be called **normalized** if it takes values in [0,1].

Define dissimilarity measure by Minkowski metric and a standardization F:

$$D_{r,F}(x, y) = (\sum_{i=1}^{n}\left|F(x)_i - F(y)_i\right|^r)^{1/r}.$$

$D_{r,F}$ satisfies permutation of reflections property $D_{r,F}(-x,y) = D_{r,F}(x,-y)$ if standardization F used in Minkowski distance is an **odd function,** i.e. it satisfies: $F(-x)= -F(x)$. Standardization F_2 defined in Proposition 3 is an odd function. Standardizations F_1 and F_3 from Proposition 3 will be odd functions if the estimates E_1 and E_{13} are odd functions [3].

If U is a strictly decreasing nonnegative function such that $U(0) = 1$ then the function $SIM_D(x,y)= U(D_{r,F}(x,y))$ with odd standardizations F will be a similarity measure satisfying permutation of reflections property. The property of a weak similarity of reflections $SIM_D(-x,x) < 1$, will be fulfilled because $D_{r,F}(x,-x) > 0$ for odd standardizations F. Such $SIM_D(x,y)= U(D_{r,F}(x,y))$ can be used for generating association measures $A_{SIM,M}$ and $A_{SIM,P}$ considered in Corollary 2 and generally for A_{SIM} from Theorem 1 when t-conorm S has no nilpotent elements. For example, we can use one of the following definitions of SIM, where $D= D_{r,F}$ and C is a positive constant:

$$SIM_D(x,y) = \frac{C}{D(x,y)+C},$$

$$SIM_D(x,y) = \frac{1}{e^{D(x,y)}}.$$

Consider the method of construction of association measure $A_{SIM,L}$ from Corollary 2 by means of standardizations F_2 or F_3 from Proposition 3. If it exists some positive constant H such that $H \geq D(x,y)$ for all x,y, and W is a strictly increasing function such that $W(0) = 0$, $W(H) \leq 1$, then a similarity measure can be defined as follows:

$$SIM_D(x,y)= 1- W(D(x,y)).$$

Such similarity measure will satisfy non-similarity of reflections property if for all n-tuples x,y the following will be fulfilled: $D(-x,x) = H \geq D(x,y)$, $H > 0$, and $W(H) = 1$. If D is normalized then one can define similarity measure by:

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

Such similarity measure satisfies non-similarity of reflections property if $D(-x,x)=1$.

Proposition 4. Suppose $D_{r,F}(x,y)$ is a dissimilarity measure defined by Minkowski distance, F is an odd standardization satisfying r-normality and W is a strictly increasing function such that $W(0)= 0$, $W(2) = 1$, then the function:

$$A_{SIM,L}(x,y) = W(D_{r,F}(x,-y))-W(D_{r,F}(x,y)), \tag{1}$$

defined for all $x,y \neq const,$ is an association measure.

The simplest functions $W(D_{r,F}(x,y))$ considered in Proposition 4 have the form:

$$W(D_{r,F}(x,y)) = \left(\frac{D_{r,F}(x,y)}{2}\right)^p, \tag{2}$$

where p is a positive constant. For $p = 1$ we have:

$$A_{SIM,L}(x,y) = 0.5(D_{r,F}(x,-y)-D_{r,F}(x,y)).$$

For $p = r$ the association measure defined by (1), (2) has the form:

$$A_{SIM,L}(x,y) = \frac{1}{2^r}\left(\sum_{i=1}^{n}\left(\left|F(x)_i + F(y)_i\right|^r - \left|F(x)_i - F(y)_i\right|^r\right)\right).$$

Corollary 5. A shape association measure defined by (1), (2) with parameters $p=r= 2$ coincides with a cosine similarity measure:

$$A_{cos,F}(x,y) = cos(F(x),F(y)).$$

Corollary 6. The shape association measure $A_{cos,F}(x,y) = cos(F(x),F(y))$ with standardization

$$f_3(x)_i = \frac{x_i - \bar{x}}{\sqrt{\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2}}$$

coincides with the sample Pearson's correlation coefficient:

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

6 Conclusions

The paper introduces the concept of association measure in the rapidly developed area of aggregation functions. The operation of pseudo-difference associated to *t*-conorm *S* considered in the theory of aggregation functions [6] gives possibility to generalize the methods of construction of association measures considered in [3] and to propose new methods of construction of such measures. The pseudo-differences associated to *t*-conorms without nilpotent elements play an important part in these methods. Such *t*-conorms are dual to t-norms without zero devisors have been considered in the theory of *t*-norms [5,7]. The main results are given for a wide class of sets with a reflection operation and a suitably defined similarity measure. It gives possibility to introduce association measures on feature spaces, in fuzzy logic, on the set of fuzzy sets, etc. The obtained results can be used for generation of association measures in various application areas, for example, is time series data mining [3]. Possible extensions of considered results can be based on the methods of definition of similarity measures used for generation of association measures. These similarity measures can be given by indistinguishability operators [9], by metrics related with the Archimedian norms [10], by some shape function [8] or kernel function etc.

Acknowledgements. The work presented in the paper has been partially supported by the projects D.00507 of IMP and by PIMAyC research Program of IMP. The author is thankful to reviewers of WCSC 2013 for their comments where the first version of this paper has been accepted.

References

1. Anscombe, F.J.: Graphs in Statistical Analysis. American Statistician 27, 17–21 (1973)
2. Batyrshin, I.: Up and Down Trend Associations in Analysis of Time Series Shape Association Patterns. In: Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Olvera López, J.A., Boyer, K.L. (eds.) MCPR 2012. LNCS, vol. 7329, pp. 246–254. Springer, Heidelberg (2012)
3. Batyrshin, I.: Constructing Time Series Shape Association Measures: Minkowski Distance and Data Standardization. In: BRICS CCI 2013, Brasil, Porto de Galhinas. IEEE Computer Society, Conference Publishing Services (CPS) (in print, 2013)
4. Batyrshin, I., Sheremetov, L., Velasco-Hernandez, J.X.: On Axiomatic Definition of Time Series Shape Association Measures. In: ORADM 2012, Workshop on Operations Research and Data Mining, Cancun, pp. 117–127 (2012)
5. Fodor, J., Roubens, M.: Fuzzy Preference Modelling and Multi-Criteria Decision Support. Kluwer, Dordrecht (1994)
6. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation Functions. Cambridge University Press, Cambridge (2009)
7. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer, Dordrecht (2000)
8. Mesiar, R., Spirková, J., Vavríková, L.: Weighted aggregation operators based on minimization. Information Sciences 178, 1133–1140 (2008)
9. Recasens, J.: Indistinguishability Operators. In: Recasens, J. (ed.) Indistinguishability Operators. STUDFUZZ, vol. 260, pp. 189–199. Springer, Heidelberg (2010)
10. Wagenknecht, M.: On some Relations Between Fuzzy Similarities and Metrics under Archimedian t-norms. The Journal of Fuzzy Mathematics 3, 563–572 (1995)