Constructing non-statistical association measures on the sets with involution and similarity measure

Ildar Batyrshin
Computing Research Center (CIC), National Polytechnic Institute (IPN), México, DF
Correlation coefficient

Correlation coefficient is a fundamental concept in data analysis measuring positive and negative relationships between variables.

Sample Pearson’s correlation coefficient:

\[-1 \leq corr(x,y) \leq 1\]

\[
corr(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}},
\]
Limitations of the correlation coefficient

Correlation coefficient is a measure of linear relationship of variables.

**Example:** For data with perfect nonlinear relationship

\[ y = Ax^2 + Bx + C, \]

\[ \text{corr}(x,y) = 0 \]

Shortcomings of the correlation coefficient

INVESTOPEDIA: http://www.investopedia.com/terms/c/correlation.asp

**Positive Correlation:** A relationship between two variables in which both variables move in tandem. A positive correlation exists when as one variable decreases, the other variable also decreases and vice versa.

For the example above it should be $A(x,y) > 0$ but we have $corr(x,y) = 0$

**Negative Correlation:** A relationship between two variables in which one variable increases as the other decreases, and vice versa.

For the example below it should be $A(x,y) < 0$ but we have $corr(x,-y) = 0$
The magic of statistics

2567 work-daily spot prices (foreign currency in dollars) over the period 10/9/86 - 8/9/96.

Papadimitriou, S., Sun, J., Yu, P. S. Local correlation tracking in time series. In: ICDM'06. Sixth Intern. Conf. Data Mining:
“The global cross-correlation coefficient of 4-FRF and 9-ESP is 0.30, which is statistically significant, exceeding the 95% confidence interval of ±0.04. “
corr(4-FRF, 6-JPY) = 0.62 > 0.30  But it should be inverse relation  !!!
Correlation is not useful for TS shape association analysis.
We need another time series association measure, not the correlation coefficient!
Correlation between membership functions

The measure of correlation of membership functions (in case of finite domain):

\[
c(f_1, f_2) = \begin{cases} 
1 - \frac{4}{H} \sum_{x \in U} [f_1(x) - f_2(x)]^2, & \text{if } H \neq 0 \\
1, & \text{if } H = 0
\end{cases}
\]

\[
H = \sum_{x \in U} [2f_1(x) - 1]^2 + \sum_{x \in U} [2f_2(x) - 1]^2.
\]

Does it has some statistical meaning?

Association measures

- Pearson's correlation coefficient
- Spearman's rank correlation coefficient
- Kendall's rank correlation coefficient (τ)
- Local trend association measure
- Correlation of fuzzy sets
- Association rules

1. Is it possible to introduce and analyze the general class of functions that can serve as non-statistical association measures similar to correlation coefficient?

2. How to generate such functions for different classes of objects: time series, elements of [0,1], fuzzy sets, fuzzy sets of type 2, interval valued fuzzy sets etc?
Properties of Pearson’s correlation coefficient

The properties of the correlation coefficient (defined for \(n\)-tuples):

\[
\text{corr}(x,y) = \text{corr}(y,x), \quad (1)
\]

\[
\text{corr}(x,x) = 1, \quad (2)
\]

\[
\text{corr}(x, -y) = -\text{corr}(x,y), \quad (3)
\]

\[
\text{corr}(-x, -y) = \text{corr}(x,y), \quad (4)
\]

\[
\text{corr}(x, -x) = -1, \quad (5)
\]

\[
\text{corr}(x,x) \geq \text{corr}(x,y). \quad (6)
\]

(4)-(6) follow from (1)-(3).

(5) and (2) are contradictive if \(x_i = 0\) for all \(i = 1, \ldots, n\)

\(\text{corr}(x,y)\) does not defined for \(x = \text{const}\), we have 0 in denominator

- We need to avoid such contradictions between desirable properties of association measure.
- We need to define association measure on different domains and can replace \(-x\) by operation with the similar property
- It is possible to consider additional axioms (requirements) depending on domain
Reflection operation

**Definition 1.** A function \(N:X \rightarrow X\), \(|X|>1\), satisfying the following properties:

\[
N(N(x)) = x \quad \text{for all } x \in X \quad \text{(involutivity)} \tag{2}
\]

\[
N(x) \neq x \text{ for some } x \in X \tag{3}
\]

is called a **reflection** on \(X\). An element \(x \in X\), such that

\[
N(x) = x, \tag{4}
\]

is called a **fixed point** of \(N\) in \(X\).

The fixed points will be denoted by \(x_{FP}\), hence for any fixed point \(x_{FP}\) it is fulfilled:

\[
N(x_{FP}) = x_{FP}. \tag{5}
\]

Denote \(FP(N,X)\) the set of all fixed points of \(N\) in \(X\). This set can be empty.

**Examples.**

1) \(X = \mathbb{R}^n\), \(x = (x_1,\ldots,x_n)\), \(N(x) = -x = (-x_1,\ldots,-x_n)\)

2) \(X = [0,1]\), \(N(x)\) is a strong negation, e.g. \(N(x) = 1 - x\)
Association measure

**Definition 2.** Let \( V \) be a subset of \( X \), \(|V| > 1\), \( N \) be a reflection on \( X \) and the restriction of \( N \) on \( V \) be a reflection on \( V \). A function \( A : V \times V \rightarrow [-1,1] \) satisfying for all \( x,y \in V \) the properties:

\[
\begin{align*}
A(x,y) &= A(y,x), \quad \text{(symmetry)} \quad (6) \\
A(x,x) &= 1, \quad \text{(reflexivity)} \quad (7) \\
A(x,N(y)) &= -A(x,y), \quad \text{(inverse relationship)} \quad (8)
\end{align*}
\]

is called an *association measure on* \( V \).

**Proposition 1.** If \( A \) is an association measure on \( V \subseteq X \) then \( V \subseteq X \setminus FP(N,X) \).

**Proposition 2.** The association measure \( A \) on \( V \) satisfies for all \( x,y \in V \) the properties:

\[
\begin{align*}
A(x,N(x)) &= -1, \quad (9) \\
A(N(x),N(y)) &= A(x,y), \quad \text{(cancellation of reflections)} \quad (10) \\
A(x,N(y)) &= A(N(x),y). \quad \text{(permutation of reflections)} \quad (11)
\end{align*}
\]

**Definition 4.** A function \( A : X \times X \rightarrow [-1,1] \) will be called *an association measure of type 1 on* \( X \) if (6) and (8) are fulfilled for all \( x,y \in X \) and (7) is fulfilled for all \( x \notin FP(N,X) \):

\[
A(x,x) = 1, \quad \text{for all } x \notin FP(N,X) \quad (7a)
\]
Association measure of type 1 on $X$

**Definition 4.** A function $A:X \times X \rightarrow [-1,1]$ will be called an association measure (of type 1) on $X$ if for all $x,y \in X$ it is fulfilled

\[
A(x,y) = A(y,x), \quad (6)
\]

\[
A(x,x) = 1, \quad \text{for all } x \notin FP(N,X) \quad (7a)
\]

\[
A(x,N(y)) = -A(x,y), \quad (8)
\]

**Proposition 4.** An association measure $A$ (of type 1) on $X$ satisfies for all $x,y \in X$ the properties

\[
A(N(x),N(y)) = A(x,y), \quad \text{(cancellation of reflections)} \quad (10)
\]

\[
A(x,N(y)) = A(N(x),y). \quad \text{(permutation of reflections)} \quad (11)
\]

\[
A(x,N(x)) = -1 \quad \text{if } x \notin FP(N,X), \quad (14)
\]

\[
A(x,x_{FP}) = A(x_{FP},x) = 0, \quad \text{for all } x_{FP} \in FP(X,N), \quad (15)
\]

\[
A(x_{FP},x_{FP}) = 0. \quad (16)
\]
Association measure of type 1 on \( X \)

**Definition 5.** Suppose \( X \) is a subset of real values \( \mathbb{R} \), and \( N \) is a reflection on \( X \) with a unique fixed point \( c = x_{FP} \). An association measure \( A:X \times X \rightarrow [-1,1] \) on \( X \) is called **\( c \)-separable** if the following properties are fulfilled for all \( x,y \in X \):

\[
A(x,y) > 0 \quad \text{if} \quad x, y > c \quad \text{or} \quad x, y < c, \quad (17)
\]

\[
A(x,y) = 0 \quad \text{if} \quad x = c \quad \text{or} \quad y = c, \quad (18)
\]

\[
A(x,y) < 0 \quad \text{if} \quad x < c < y \quad \text{or} \quad y < c < x. \quad (19)
\]

**Example 3.** \( X = [0,1] \)

\( N(x) = 1 - x \). \( c = x_{FP} = 0.5 \)

\( A(x,y) = A(y,x) \)

\( A(x,x) = 1, \quad \text{if} \quad x \neq c \)

\( A(x,N(x)) = -1, \quad \text{if} \quad x \neq c \)

\( A(x,c) = A(c,x) = 0 \)

\( A(x,N(y)) = -A(x,y) \)

**Example.** \( A(x,y) = \text{sign}((x-c) \cdot (y-c)) \)
Association between fuzzy (plausible) predicates $P(x), Q(y) \in [0,1]$

It is reasonable to have association measure such that:

$A(P,Q) > 0$ if $P(x), Q(y) > 0.5$

or $P(x), Q(y) < 0.5$

$A(P,Q) < 0$ if $P(x) < 0.5 < P(y)$

or $P(x) > 0.5 > P(y)$

for all $x, y \in X$
Association measure on the set of time series

$X$ is a set of real $n$-tuples $x = (x_1, \ldots, x_n)$, $(n>1)$, $q_{(n)} = (q, q, \ldots, q)$ is a constant $n$-tuple, $q$ is real. $N(x) = -x = (-x_1, \ldots, -x_n)$ is a reflection with a unique fixed point $0_{(n)} = (0,0,\ldots,0)$. Define $x+y = (x_1+y_1, \ldots, x_n+y_n)$, $py+q = (py_1+q, \ldots, py_n+q)$, $p,q$ are real, $p > 0$.

**Definition 3.** Suppose $V \subseteq X$, $|V| > 1$, such that

- from $x \in V$ it follows $-x \in V$, and $x+q \in V$ for all real $q$.

A function $A: V \times V \to [-1,1]$ satisfying for all $x,y \in V$ and for all real $q$ the properties

\[
A(x,y) = A(y,x),
\]

\[
A(x,x) = 1,
\]

\[
A(x,N(y)) = -A(x,y),
\]

\[
A(x+q,y) = A(x,y),
\]

is called a **time series shape association measure** on $V$.

If for all $x \in V$ and for all $p > 0$ it is fulfilled $px \in V$ and $A$ satisfies on $V$ the property:

\[
A(px,y) = A(x,y),
\]

**scale invariance**

then $A$ is called a **scale invariant time series shape association measure**.
Moving Approximation Transform (MAT) and local trend association measure

Calculate least squares approximations $f_i = a_it + b_i$ of time series $x = (x_1, \ldots, x_n)$ in sliding window of size $k$. Replace $x$ by sequence of local trends:

$$MAT_k(x) = (a_1, \ldots, a_{n-k+1}).$$

Local trends $a_1, \ldots, a_{n-k+1}$ depend on the size of window $k$.

**Example:** $k = 5$

Local trend association measure:

$$lta_k(x, y) = \frac{\sum_{i=1}^{m} a_{xi} \cdot a_{yi}}{\sqrt{\sum_{i=1}^{m} a_{xi}^2 \cdot \sum_{j=1}^{m} a_{yj}^2}}$$

Batyrsbin I., Sheremetov L. …(2004,2007)
Association measure on the set of time series

**Proposition 4.** If $A$ is a time series shape association measure on $V$ then $V \subseteq X \setminus X_C$, where $X_C$ is a set of all constant $n$-tuples.

**Example.** $corr(x,y)$ is a scale invariant association measure on $V = X \setminus X_C$

**Example.** Local trend association measure $lta_k(x,y)$ is a scale invariant association measure on $V = X \setminus X_{Mk0}$ where $X_{Mk0}$ is a set of all time series $x$ such that $MAT_k(x) = 0_{(m)}$, i.e. $a_{xi} = 0$ for all $i=1,\ldots,m$. 
Association measures related with similarity measures

A similarity measure $SIM:X \times X \rightarrow [0,1]$ satisfies for all $x, y$ from $X$:

$$SIM(x,y) = SIM(y,x),$$
$$SIM(x,x) = 1.$$

An association measure $A$ related with $SIM$:

$$A(x,y) > 0, \text{ if } SIM(x,y) > SIM(x,N(y)),$$
$$A(x,y) < 0, \text{ if } SIM(x,y) < SIM(x,N(y)).$$

Can we use (?) the difference between $SIM(x,y)$ and $SIM(x,N(y))$:

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

- **What properties should be fulfilled for $SIM$ to obtain association measure $A$ satisfying the desirable properties?**
- **Can we use here pseudo-difference operation related with t-conorms?**
**t-conorms**

*t-conorm* is a function $S: [0,1]^2 \rightarrow [0,1]$ satisfying commutativity, associativity, monotonicity boundary condition: $S(a,0) = a$.

We have: $S(1,a) = S(a,1) = 1$, $S(0,a) = S(a,0) = a$.

$a \in ]0,1[\text{ is a nilpotent element of } S \text{ if there exists } b \in ]0,1[ \text{ such that } S(a,b) = 1$.

*t-conorm $S$ has no nilpotent elements* if and only if from $S(a,b) = 1$ it follows $a = 1$ or $b = 1$.

Examples:

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

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

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

$S_M$ and $S_P$ have no nilpotent elements but $S_L$ has.
\textbf{S-difference and pseudo-difference}

Let $S$ be a $t$-conorm. \textbf{S-difference} is defined by:

$$a \underset{S}{-} b = \inf \{c \in [0,1] | S(b, c) \geq a\} \quad \text{for any } a, b \text{ in } [0,1].$$

\textbf{Pseudo-difference} associated (related) to $S$ is defined by:

$$a(-)_{S} b = \begin{cases} 
  a \underset{S}{-} b, & \text{if } a > b \\
  - \left( b \underset{S}{-} a \right), & \text{if } a < b \\
  0, & \text{if } a = b 
\end{cases} \quad \text{for any } a, b \text{ in } [0,1]^2$$

Equivalently

$$a(-)_{S} b = \text{sign}(a - b)(\max(a, b) \underset{S}{-} \min(a, b)).$$
Pseudo-differences related with basic t-conorms

\[ S_M(a,b) = \max\{a, b\} \quad \text{(maximum)} \]
\[ S_L(a,b) = \min\{a+b, 1\} \quad \text{(Lukasiewicz t-conorm)} \]
\[ S_P(a,b) = a+b-ab \quad \text{(probabilistic sum)} \]

\[ a(-)_M b = \begin{cases} 
  a, & \text{if } a > b \\
  -b, & \text{if } a < b, \\
  0, & \text{if } a = b 
\end{cases} \]
\[ a(-)_L b = a - b. \]
\[ a(-)_P b = (a - b)/(1 - \min(a, b)). \]
Similarity measures

SIM: X × X → [0, 1] is a similarity measure on X if it satisfies the properties:

\[ SIM(x, y) = SIM(y, x) \] (symmetry)
\[ SIM(x, x) = 1 \] (reflexivity)

---

\[ SIM(N(x), N(y)) = SIM(x, y) \] (cancellation of reflections)
\[ SIM(x, N(y)) = SIM(N(x), y) \] (permutation of reflections)

\[ SIM(x, y) < SIM(x, x) \] for all \( x \neq y \) (strict reflexivity)
\[ SIM(x, N(x)) < 1 \] (weak similarity of reflections)
\[ SIM(N(x), x) = 0. \] (non-similarity of reflections)
Main result

Theorem 2. Suppose $X$ is a set with a reflection $N$, $V \subseteq X \setminus \text{FP}(N,X)$, $|V| > 1$, $V$ is closed under $N$ and the restriction of $N$ on $V$ is a reflection on $V$, $S$ is a t-conorm and $SIM$ is a similarity measure on $X$ satisfying the properties of cancellation of reflections and weak similarity of reflections then the function

$$A_{SIM,S}: V \times V \rightarrow [-1,1]$$

defined for all $x,y \in V$ by

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

is an association measure on $V$ if one of the following is true

a) the t-conorm $S$ has no nilpotent elements \hfill (29)

b) $SIM(x,N(x)) = 0$, for all $x \in V$ \hfill (non-similarity of reflections)
Association measure on $[0,1]$

**Theorem 3.** Suppose $N$ is a strong negation, $S$ is a t-conorm, $SIM$ is a similarity measure on $[0,1]$ satisfying the properties of cancellation of reflections and weak similarity of reflections then the function $A_{SIM,S} : [0,1] \times [0,1] \to [0,1]$ defined for all $x,y \in [0,1]$ by:

$$A_{SIM,S}(x, y) = SIM(x, y) \ominus_S SIM(x, N(y)), \quad (37)$$

is an association measure on $[0,1]$ if one of the following is fulfilled:

1) $SIM(x,N(x)) = 0$, for all $x \in [0,1]$, \quad (38)
2) the t-conorm $S$ has no nilpotent elements. \quad (39)
Association measure of type 1 on $X$

**Definition.** Suppose $X = [0,1]$, $N$ is a strong negation with a unique fixed point $c$. An association measure $A:X \times X \to [-1,1]$ on $[0,1]$ is called **c-separable** if the following properties are fulfilled for all $x,y \in X$:

- $A(x,y) > 0$ if $x, y > c$ or $x, y < c$,
- $A(x,y) = 0$ if $x = c$ or $y = c$,
- $A(x,y) < 0$ if $x < c < y$ or $y < c < x$.

**Definition.** A similarity measure $SIM$ is **strict monotonic** if for all $x,y,z \in [0,1]$ it is fulfilled:

$$SIM(x,z) < \min(SIM(x,y), SIM(y,z)) \text{ if } x < y < z.$$ 

**Theorem 4.** Suppose in the conditions of Theorem 3 the $t$-conorm $S$ is continuous at the point 0 in both arguments and the similarity measure $SIM$ is strict monotonic then the association measure

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

is c-separable.
Constructing $SIM$ satisfying the cancellation of reflections property $SIM(N(x), N(y)) = SIM(x, y)$

**Proposition 7.** Suppose $f, g: [0,1] \rightarrow [0,1]$ are automorphisms of $[0,1]$ and $g$ defines the strong negation $N$ on $X=[0,1]$ then the function

$$SIM(x,y) = 1 - f(|g(x) - g(y)|),$$

is a similarity measure on $X=[0,1]$ satisfying the properties of the strict monotonicity, the strict reflexivity and hence the weak similarity of reflections, the cancellation of reflections with respect to $N$, but it does not satisfy the non-similarity of reflections property.

$$SIM(x, y) = 1 - \varphi^{-1}(|\varphi(x) - \varphi(y)|),$$

For generator $\varphi(x) = x$ of standard negation $N(x) = 1 - x$, we obtain:

$$SIM(x, y) = 1 - |x - y|.$$
For standard negation $N(x)=1-x$, fixed point $c = 0.5$, pseudo-difference $(-)_S$ related with $S_M(a,b) = \max\{a,b\}$, (maximum), $SIM(x,y) = 1 - |x - y|$: 

$$A_{SIM,S_M}(x,y) = \begin{cases} 
1 - |x - y| & \text{if } x, y > 0.5 \text{ or } x, y < 0.5 \\
|x + y - 1| - 1 & \text{if } x < 0.5 < y \text{ or } y < 0.5 < x \\
0 & \text{if } x = 0.5 \text{ or } y = 0.5 
\end{cases}$$
For standard negation $N(x) = 1 - x$, fixed point $c = 0.5$, pseudo-difference $(-)_S$ related with $S_p(a,b) = a + b - ab$, (probabilistic sum), $SIM(x,y) = 1 - |x - y|$: 

$$A_{SIM,S_P}(x,y) = \begin{cases} 
\frac{|x + y - 1| - |x - y|}{\max\{|x + y - 1|, |x - y|\}} & \text{if } x, y \neq 0.5 \\
0 & \text{if } x = 0.5 \text{ or } y = 0.5
\end{cases}$$
Association measure of type 1 on $X$

**Example 3.** $X = [0,1]$, Yager negation: $N_p(x) = \frac{p}{\sqrt{1 - x^p}}$, $p > 0$, $c = \frac{p}{\sqrt{0.5}}$. 

![Yager Negation](image-url)
Example of distance based association measure on fuzzy sets

\[ \text{SIM}(x,y) = 1 - \frac{1}{n} \sum_{i=1}^{n} |g(x_i) - g(y_i)|^2 \]

where \( g(x) \) is a generator of involutive negation on \([0,1]\).

For \( A(x,y) = A_{SIM,P}(x,y) \) and Zadeh negation \( N(x) = 1 - x \) with generator \( g(x) = x \) we obtain:

\[ A(x,y) = \frac{\sum_{i=1}^{n} (2x_i - 1)(2y_i - 1)}{\max(\sum_{i=1}^{n} |x_i - y_i|^2, \sum_{i=1}^{n} |x_i + y_i - 1|^2)}. \]
Constructing \(SIM\) satisfying the cancellation of reflections property \(SIM(N(x),N(y)) = SIM(x,y)\)

**Definition 14.** A function \(M:[0,1] \times [0,1] \rightarrow [0,1]\) is an aggregation function of two arguments if it is non-decreasing in each arguments and satisfies:

\[
M(x,y) = M(y,x) \quad \text{ (symmetry)}
\]

\[
M(0,0) = 0, \quad M(1,1) = 1, \quad \text{ (boundary conditions)}
\]

**Proposition 6.** If \(M\) is an aggregation function and \(SIM\) is a similarity measure then

\[
SIM_M(x,y) = M(SIM(x,y), SIM(N(x), N(y)))
\]  \hspace{1cm} (32)

is the similarity measure satisfying the cancellation of reflections property (26).

\(SIM_M\) is strict reflexive if \(SIM\) is strict reflexive and \(M\) satisfies:

\[
M(a,b) < 1 \quad \text{if} \quad \text{min}(a,b) < 1.
\]  \hspace{1cm} (33)

**Examples:**

\[
SIM_M(x,y) = \text{min}(SIM(x,y), SIM(N(x), N(y))), \quad (34)
\]

\[
SIM_M(x,y) = \frac{SIM(x,y) + SIM(N(x), N(y))}{2}. \quad (35)
\]
Similarity and dissimilarity measures

**Definition 15.** A function $D: X \times X \rightarrow [0,1]$ is a **dissimilarity measure** on $X$ if it satisfies:

\[
D(x,y) = D(y,x),
\]

\[
D(x,x) = 0.
\]

If $D$ is a dissimilarity measure and $U: \mathbb{R} \rightarrow \mathbb{R}$ is a strictly decreasing nonnegative real function such that $U(0) = 1$, then the function

\[
SIMD(x,y) = U(D(x,y))
\]

is a similarity measure.

If it exists some positive constant $H$ such that $D(x,y) \leq H$ for all $x,y$ and $W$ is a strictly increasing function such that $W(0) = 0$, $W(H) \leq 1$, then a similarity function can be obtained as follows:

\[
S_D(x,y) = 1 - W(D(x,y)).
\]
Simmilarity and dissimilarity measures

Examples of \( S_D(x, y) = U(D(x, y)) \)

\[
S_D(x, y) = \frac{K}{D(x, y) + K}, \quad K > 0,
\]

\[
S_D(x, y) = \frac{1}{e^{D(x, y)}}.
\]

\[
D_{\gamma,F}(x, y) = \left( \sum_{i=1}^{n} |F(x_i) - F(y_i)|^\gamma \right)^{1/\gamma}
\]
Standardization of time series values

A transformation $F$ of time series $x$ of length $n$ into time series $F(x)$ of the same length is said to be a **standardization** if for all non-constant time series $x$ it is fulfilled:

$$F(F(x)) = F(x).$$  \hspace{1cm} \text{(idempotency)}

Two additional requirements on standardization transformation can be considered:

$$F(x) \neq \text{const} \quad \text{if } x \neq \text{const},$$

$$F(q_{(n)}) = 0_{(n)}, \quad \text{for any real value } q.$$
Standardization of time series values

A time series $x$ is said to be in a standard form wrt a standardization $F$ if $F(x) = x$. As it follows from the definitions, a standardization transforms any time series $x$ into a standard form $F(x)$.

A transformation $E$ of time series $x$ of length $n$ into real value $E(x)$ is said to be an estimate of $x$. 
Proposition. Suppose $E_1(x)$ is a translation additive estimate such that $E(q_{(n)})=q$, then the transformation

$$F(x) = x - E(x),$$

is a translation invariant standardization such that

$$E(F(x)) = 0.$$

If $E(x)$ is an odd function, then $F(x)$ is an odd function. If $E(x)$ is scale proportional then $F(x)$ is scale proportional.
Proposition 3. Suppose $E_1(x)$ is a translation additive and scale proportional estimate such that $E(q_{(n)})=q$, and $E_2(x) \neq 0$ is a translation invariant and scale proportional estimate then the transformation $F(x) = (x - E_1(x))/E_2(x)$ is a translation invariant and scale invariant standardization such that $E_1(F(x)) = 0$.

If $E_1(x)$ is an odd function and $E_2(x)$ is an even function, then $F(x)$ is an odd function.

If then $F(x)$ satisfies $r$-normality.
Constructing $SIM$ satisfying the cancellation of reflections property $SIM(N(x), N(y)) = SIM(x, y)$

**Proposition 8.** Let $X$ be the set of time series of the length $n$ and $D(x, y) = D_{r,F}(x, y)$:

$$D_{r,F}(x, y) = \left( \sum_{i=1}^{n} |F(x_i) - F(y_i)|^r \right)^{1/r} \quad (39)$$

where $F: \mathbb{R} \rightarrow \mathbb{R}$. A dissimilarity measure $D$ satisfies on $V \subseteq X \setminus \mathcal{C}$ the property

$$D(N(x), N(y)) = D(x, y), \quad \text{ (cancellation of reflections)} \quad (42)$$

if for all $x \in V$ the function $F$ satisfies the following condition:

$$F(N(x)) + F(x) = Q, \quad (43)$$

where $Q = \text{const}$.

$D(N(x), x) > 0$ if in (43) $F(x_i) \neq Q/2$ for some $x \in V$.

$D(N(x), x) = 2$ if in (43) $Q = 0$ and for all $x \in V$ it is fulfilled $r$-normality:

$$\sum_{i=1}^{n} |F(x_i)|^r = 1$$
Minkowski distance $\rightarrow$ Association

\[ \text{SIMD}(x,y) = U(D(x,y)) \]

\[ D_{r,F}(x,y) = (\sum_{i=1}^{n} |F(x_i) - F(y_i)|^r)^{1/r} \]

If it exists some positive constant $H$ such that $D(x,y) \leq H$ for all $x,y$ and $W$ is a strictly increasing function such that $W(0) = 0$, $W(H) \leq 1$, then a similarity function can be obtained as follows:

\[ S_D(x,y) = 1 - W(D(x,y)) \]

**Corollary.** A shape association measure defined in Proposition coincides with a cosine similarity measure:

\[ A_{\cos,F}(x,y) = \cos(F(x),F(y)) \]

if $r=2$, i.e. $D(x,y) = D_{2,F}(x,y)$ is Euclidean distance, and if $W(D)$ is defined as follows:

\[ W(D(x,y)) = \frac{(D(x,y))^2}{4} \]
Minkowski distance → Association

\[ D \rightarrow S \rightarrow A, \quad F \rightarrow D \rightarrow S \rightarrow A \]

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

where \( F(x) \) is some standardization of time series \( x \).

\( F \) is an **odd function**, i.e. it satisfies: \( F(-x) = -F(x) \).
Constructing Pearson’s correlation coefficient

**Example 5.** Let $X$ be a set of real $n$-tuples $x = (x_1, \ldots, x_n)$, $(n>1)$, with the reflection operation $N(x) = -x = (-x_1, \ldots, -x_n)$. Define dissimilarity measure by

$$D(x, y) = \sqrt{\sum_{i=1}^{n}|F(x_i) - F(y_i)|^2},$$

where $F(x_i)$ is given by:

$$F(x_i) = \frac{x_i - \bar{x}}{\sqrt{\sum_{j=1}^{n}(x_j - \bar{x})^2}}, \quad \bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j,$$

and $SIM(x, y) = 1 - \frac{1}{4}D(x, y)^2 = \frac{1}{4} \sum_{i=1}^{n}|F(x_i) - F(y_i)|^2$. Then

$$corr(x, y) = SIM(x, y) (-)_s SIM(x, N(y)),$$

where the pseudo-difference operation $a(-)_s b = a - b$, i.e. associated to Lukasiewicz t-conorm.
Example of normalized Google Finance data after smoothing by moving average ($w=5$).
Positively (two charts at the top) and negatively (two charts on the bottom) associated moving approximations of BBRY and AAPL data in sliding window of size $k = 30$. 
Google Finance data, win=30, positive associations, Npos=10

Google Finance data, win=30, negative associations, Nneg=79
Moving Approximation Transform (MAT) and local trend association measure

Calculate least squares approximations $f_i = a_i t + b_i$ of time series $x = (x_1, \ldots, x_n)$ in sliding window of size $k$. Replace $x$ by sequence of local trends:

$$MAT_k(x) = (a_1, \ldots, a_{n-k+1}).$$

Local trends $a_1, \ldots, a_{n-k+1}$ depend on the size of window $k$.

**Example:** $k = 5$

Local trend association measure:

$$lta_k(x, y) = \frac{\sum_{i=1}^{m} a_{xi} \cdot a_{yi}}{\sqrt{\sum_{i=1}^{m} a_{xi}^2 \cdot \sum_{j=1}^{m} a_{yj}^2}}$$

Batyrsihn I., Sheremetov L. …(2004,2007)
Foreign Exchange Rates

Time series of Foreign Exchange Rates (money of different countries to one U.S. Dollar) measured daily since 2004-09-02 to 2004-10-15.
Local trend association network of foreign exchange rates obtained for small windows. Only links with high associations are shown.
Global trend association network of foreign exchange rates obtained for large windows. Only high associations are shown.
Gross internal product of Mexico checked quarterly over the period 1980 – 2003 (96 data).

1. Farming, Forestry and Fishes;
2. Mining
3. Manufacturing Industry
4. Construction
5. Electricity, Gas and Water
6. Commerce, Restaurants and Hotels
7. Transport, Storage and Communications
8. Financial Services, Insurance, Real Estate Activities and Rent
9. Social and Personal Communal Services
10. Liability to Banking Services

Allocate
Association network of Mexican indexes of economics

8: Financial Services, Insurance, Real Estate Activities and Rent

9: Social and Personal Communal Services

5: Electricity, Gas and Water

6: Commerce, Restaurants and Hotels

2: Mining

3: Manufacturing Industry

4: Construction

7: Transport, Storage and Communications

1: Farming, Forestry and Fishes

10: Liability to Banking Services

Relationships:
- 8: Financial Services, Insurance, Real Estate Activities and Rent
  - 9: Social and Personal Communal Services: -0.99
  - 5: Electricity, Gas and Water: -0.81
  - 6: Commerce, Restaurants and Hotels: 0.50
- 9: Social and Personal Communal Services
  - 1: Farming, Forestry and Fishes: 0.72
  - 6: Commerce, Restaurants and Hotels: 0.52
- 6: Commerce, Restaurants and Hotels
  - 2: Mining: 0.50
  - 3: Manufacturing Industry: 0.57
  - 4: Construction: 0.73
How to define association measure on the set of subsets of $[0,1]$, on the set of interval valued truth or membership values ???
Association Measures on the Set of Subintervals of $[0,1]$

Denote $D[0,1]$ the set of all closed subintervals of $[0,1]$, i.e. $D[0,1] = \{a^* = [a^-, a^+] \subseteq [0,1] \}$, where $a^-, a^+ \in [0,1]$, $a^- \leq a^+$.

**Definition.** Suppose $\text{neg}$ is an involutive negation of $[0,1]$. An involution (reflection) $N:D[0,1] \rightarrow D[0,1]$ on $D[0,1]$ is defined for all $a^* \subseteq [0,1]$ as follows:

$$N(a^*) = N([a^-, a^+]) = [\text{neg}(a^+), \text{neg}(a^-)]$$

It is fulfilled: $N(N(a^*)) = a^*$
Set of fixed points of $N$ on $D[0,1]$

**Proposition.** The set of fixed points of $N$ on $D[0,1]$ is the following:

$$FP= \{ [a, \text{neg}(a)] \mid a \in [0,1], a \leq a_{FP} \},$$

where $a_{FP}$ is the fixed point of a negation $\text{neg}$ on $[0,1]$.

**Example.** For the negation of Zadeh $\text{neg}(a)=1-a$, we have

$$FP= \{ [a, 1-a] \mid a \in [0,0.5] \},$$

e.g. $[0,1]$, $[0.1,0.9]$, $[0.4,0.6]$, $[0.5,0.5]$ etc.
Association measure on $D[0,1]\setminus FP$ related by $S_M$

Define the similarity measure on $D[0,1]$: 

$SIM(a^*, b^*) = 1 - 0.5(|a^- - b^-| + |a^+ - b^+|)$.

Association measure on $D[0,1]\setminus FP$ related by $S_M$: 

$A_M(a^*, b^*) = 1 - 0.5P, \quad \text{if } Q > P; \\
A_M(a^*, b^*) = 0.5Q - 1, \quad \text{if } Q < P; \\
A_M(a^*, b^*) = 0, \quad \text{if } Q = P.$

where 

$P = |a^- - b^-| + |a^+ - b^+| \\
Q = |a^- + b^+ - 1| + |a^+ + b^- - 1|.$
Association measure on $D[0,1]\setminus FP$ related by $S_P$

$$A_P(a^*,b^*) = (Q - P)/\max(Q,P)$$

where

$$P = |a^- - b^-| + |a^+ - b^+|$$

$$Q = |a^- + b^+ - 1| + |a^+ + b^- - 1|.$$ 

where $Q$ and $P$ are defined in (29), (30).
Examples

\(a^* = [0.1, 0.3], \quad b^* = [0.2, 0.4], \quad c^* = [0.8, 0.9], \quad d^* = [0.6, 0.8], \quad e^* = [0.2, 0.7]\)

\[A_M(a^*, b^*) = 0.9, \quad A_M(a^*, c^*) = -0.95, \quad A_M(b^*, d^*) = -1, \quad A_M(a^*, e^*) = 0.75.\]

\[A_P(a^*, b^*) = 0.8, \quad A_P(a^*, c^*) = -0.92, \quad A_P(b^*, d^*) = -1, \quad A_P(a^*, e^*) = 0.29.\]
Conclusions

The general approach to definition and construction of association measures is developed. It is an interesting and promising task to extend these results on various types of sets with reflection operations, on other types of similarity measures and to use them in various tasks of data analysis and data mining.
Thank you very much!

Ildar Batyrshin
CIC IPN, Mexico
batyr1@gmail.com
References


Batyrsin I., Solovyev V., Ivanov V. **Time series shape association measures and local trend association patterns**. Neurocomputing, 2015, doi:10.1016/j.neucom.2015.05.127.


