How to Select an Appropriate Similarity Measure: Towards a Symmetry-Based Approach

Ildar Batyrshin[1], Thongchai Dumrongpokaphan[2], Vladik Kreinovich[3], and Olga Kosheleva[3]

[1]Centro de Investigaciń en Computación (CIC)
Instituto Politécnico Nacional (IPN), México, D.F.
batyr1@gmail.com
[2]Department of Mathematics, Chiang Mai University
Thailand, tcd43@hotmail.com
[3]University of Texas at El Paso, USA
vladik@utep.edu, olgak@utep.edu

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

1. Outline

- When practitioners analyze the similarity between time series, they often use correlation.
- Sometimes this works.
- However, sometimes, this leads to counter-intuitive results.
- In such cases, other similarity measures are more appropriate.
- An important question is how to select an appropriate similarity measures.
- In this talk, we show, on simple examples, that
  - the use of natural symmetries – scaling and shift
  - can help with such a selection.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

2. Practitioners Routinely Use Correlation to Detect Similarities

- Practitioners are often interested in gauging similarity:
  - between two sets of related data or
  - between two time series.
- A natural idea seems to be to look for (sample) *correlation*: $\rho(a,b) = \dfrac{C_{a,b}}{\sigma_a \cdot \sigma_b}$, where

$$C_{a,b} \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} (a_i - \overline{a}) \cdot (b_i - \overline{b}), \quad \overline{a} \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} a_i, \quad \overline{b} \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{b} b_i,$$

$$\sigma_a \stackrel{\text{def}}{=} \sqrt{V_a}, \ \sigma_b \stackrel{\text{def}}{=} \sqrt{V_b}, \ V_a \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} (a_i - \overline{a})^2, \ V_b \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} (b_i - \overline{b})^2.$$

- Practitioners understand that correlation only detects *linear* dependence.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

3. Limitations of Correlation

- In some cases, the dependence is non-linear.
- In such cases, simple correlation does not work.
- More complex methods are needed to detect dependence.
- Also, correlation assumes that the value b_i is only affected by the value of a_i at the same moment of time i.
- In real life, we may have a delayed effect – and the corresponding delay may depend on time.
- However, in simple linear no-delay cases, practitioners expect correlation to be a perfect measure of similarity.
- And often it is. But sometimes, it is not. Let us give two examples.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

4. First Example

- We ask people to evaluate movies on a scale 0–5.
- Persons a, b, and c gave the following grades:

$$a_1 = 4, \quad a_2 = 5, \quad a_3 = 4, \quad a_4 = 5, \quad a_5 = 4, \quad a_6 = 5;$$
$$b_1 = 5, \quad b_2 = 4, \quad b_3 = 5, \quad b_4 = 4, \quad b_5 = 5, \quad b_6 = 4;$$
$$c_1 = 0, \quad c_2 = 1, \quad c_3 = 0, \quad c_4 = 1, \quad c_5 = 0, \quad c_6 = 1.$$

- From the common sense viewpoint, a and b have similar tastes: they like all the movies.
- However, between a_i and b_i, there is a perfect *anti*-correlation $\rho = -1$.
- The opposite opinion is expressed by c who does not like the movies.
- However, between a_i and c_i, there is a perfect correlation $\rho = 1$; so, correlation is counter-intuitive.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

5. Second Example

- Suppose that the US stock market shows periodic oscillations, with relative values

$$a_1 = 1.0, \quad a_2 = 0.9, \quad a_3 = 1.0, \quad a_4 = 0.9.$$

- Stock market in a small country X shows similar relative changes, but with a much higher amplitude:

$$b_1 = 1.0, \quad b_2 = 0.5, \quad b_3 = 1.0, \quad b_4 = 0.5.$$

- These sequences are somewhat similar, but not the same:
  - while the US stock market has relatively small 10% fluctuations,
  - the stock market of the country X changes by a factor of two.
- However, the two stock markets have a perfect positive correlation $\rho = 1$.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

6. Other Similarity Measures

- The need to go beyond correlation is well known.
- Many effective similarity measures have proposed.
- Most of these measures start:
  - either with correlation,
  - or with the Euclidean distance

$$d(a,b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

  - or with a more general l^p-distance

$$\left(\sum_{i=1}^{n} |a_i - b_i|^p \right)^{1/p}.$$

- Sometimes, a linear or nonlinear transformation is applied to the result, to make it more intuitive.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

7. Other Similarity Measures (cont-d)

- In other situations, modifications take care of the possible time lag in describing the dependence.
- For example, we may look for a correlation between b_i and the delayed series a_{i+c} for an appropriate c.
- More generally, we can look for delay $c(i)$ that changes with time, i.e., for correlation between b_i and $a_{i+c(i)}$.
- An example of such a similarity measure is the move-split-merge metric.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

8. How to Select a Similarity Measure

- In different practical situations, different similarity measures are appropriate.
- It is therefore important to be able to select the most appropriate similarity measure for each given situation.
- There have been several papers comparing the effectiveness of different similarity measures in *clustering*.
- Another important practical case is when we simply have two time series.
- In this talk, we show that natural symmetries – shifts and scalings – can help.
- We only consider no-time-lag linear case.
- We hope that symmetries will help in general case as well.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

9. Compared Values Come from Measurements

- We want to understand a discrepancy between commonsense meaning of similarity and correlation.
- For this, let us recall how we get the values a_i and b_i.
- Usually, we get these values from measurements.
- Sometimes, they come from expert estimates:
  - they can also be considered as measurements
  - performed by a human being as a measuring instrument.
- To perform a measuremnt, we need to select a starting point and a measuring unit.
- For example, we can measure temperature in the Fahrenheit (F) scale or in the Celsius (C) scale.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

10. Values Come from Measurements (cont-d)

- F and C scales have:
  - different starting points: $0°\text{C} = 32°\text{F}$, and
  - different units: a difference of 1 degree C is equal to the difference of 1.8 degrees Fahrenheit.
- If we change a measuring unit to a u times smaller one, then all numerical values get multiplied by u: $x' = u \cdot x$.
- For example, a height of $x = 2$ m becomes $x' = 100 \cdot 2 = 200$ cm in the new units.
- If we use a new starting point which is s units earlier, then we get $x' = x + s$.
- If we change both the measuring unit and the starting point, we get new valued $x' = u \cdot x + s$.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

11. Is Correlation Appropriate

- A perfect correlation $\rho = 1$ means that after an appropriate linear transformation, we have $b_i = u \cdot a_i + s$.
- In other words, if we select an appropriate measuring unit and an appropriate starting point for a, then:
  - the values $a'_i = u \cdot a_i + s$ of the quantity a described in the new units
  - will be identical to the values of the quantity b.
- In such cases, correlation is indeed a perfect measure of similarity.
- However, some quantities only allow some of the above symmetries – or none at all.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

12. Is Correlation Appropriate (cont-d)

- For example, for stock markets, 0 is a natural starting point, so:
  - while scalings $x \to u \cdot x$ make sense,
  - shifts $x \to x + s$ change the situation – often drastically.
- For movie evaluations, the results are even less flexible: here:
  - both the measuring unit and the starting point are fixed,
  - so no symmetries are allowed.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

13. Case When No Scaling Is Possible

- Then, a natural measure of dissimilarity is the distance $d(a, b)$ between these tuples: $d(a, b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$.
- The above formula is reasonable.
- However, from the practical viewpoint, the simpler the computations, the better.
- From this viewpoint, the distance is not perfect:
  - we need to compute the square root,
  - which is not easy to perform by hand.
- Good news is that the main purpose of gauging similarity is to compare the degrees.
- Thus, we can use easier-to-compute squares

$$d^2(a, b) = \sum_{i=1}^{n} (a_i - b_i)^2.$$

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

14. Case When All Scalings Are Allowed

- Let us consider the case when all scalings are applicable: $a \to u \cdot a + s$, $b \to u' \cdot b + s'$.
- Then, instead of $d^2(a, b)$, it makes sense to use

$$D_g(a, b) = \min_{u,s} d^2(u \cdot a + s, b) = \min_{u,s} \sum_{i=1}^{n} (b_i - (u \cdot a_i + s))^2.$$

- This formula takes care of re-scaling a_i, but it may change if we re-scale b_i.
- min over all such re-scalings is 0.
- To avoid 0, we can consider *relative* distance:

$$\frac{\min\limits_{u,s} \sum\limits_{i=1}^{n} (b_i - (u \cdot a_i + s))^2}{\sum\limits_{i=1}^{n} b_i^2} = \min_{u,s} \frac{\sum\limits_{i=1}^{n} (b_i - (u \cdot a_i + s))^2}{\sum\limits_{i=1}^{n} b_i^2}.$$

15. When All Scalings Are Allowed (cont-d)

- We can then take max over possible shifts of b:

$$d_g(a,b) = \max_{u',s'} \min_{u,s} \frac{\sum_{i=1}^{n}((u' \cdot b_i + s') - (u \cdot a_i + s))^2}{\sum_{i=1}^{n}(u' \cdot b_i + s')^2}.$$

- **Proposition.** $d_g(a,b) = 1 - \rho^2(a,b)$**.**
- The proof is straightforward: equate derivatives to 0.
- This result explains why correlation is often adequate.
- Moreover, we get a non-statistical explanation of correlation.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

16. Case When Only Scaling Makes Sense

- In this case, we only have transformations

$$a_i \to a_i' = u \cdot a_i \text{ and } b_i \to b_i' = u' \cdot b_i.$$

- In this case, instead of $d^2(a, b)$, we consider

$$D_u(a, b) = \min_u d^2(u \cdot a, b) = \min_u \sum_{i=1}^{n} (b_i - u \cdot a_i)^2.$$

- To take case of re-scalings of b_i, we consider the ratio

$$d_u(a, b) = \frac{\min\limits_u \sum\limits_{i=1}^{n} (b_i - u \cdot a_i)^2}{\sum\limits_{i=1}^{n} b_i^2} = \min_u \frac{\sum\limits_{i=1}^{n} (b_i - u \cdot a_i)^2}{\sum\limits_{i=1}^{n} b_i^2}.$$

- **Proposition.** $d_u(a, b) = 1 - \dfrac{\left(\overline{a \cdot b}\right)^2}{\overline{a^2} \cdot \overline{b^2}}$.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

17. When Only Scaling Makes Sense (cont-d)

- In particular, when $\overline{a} = \overline{b} = 0$, we get a correlation-related formula

$$d_u = 1 - \rho^2(a, b).$$

- In this case, correlation can be reconstructed as

$$\rho(a, b) = \sqrt{1 - d_u(a, b)}.$$

- In general, we can therefore view the expression $\sqrt{1 - d_u(a, b)}$ as an analogue of correlation.
- In the above example of two stock markets, when $\rho(a, b) = 1$, we have:
  - $d_u(a, b) = 0.071$ and
  - $\sqrt{1 - d_u} \approx 0.96 < 1$.

18. Case When Only Shift Makes Sense

- In this case, we can only have transformations

$$a_i \to a_i' = a_i + s \text{ and } b_i \to b_i' = b_i + s'.$$

- In this case, instead of $d^2(a, b)$, we consider:

$$D_s(a, b) = \min_s d^2(a + s, b) = \min_s \sum_{i=1}^{n} (b_i - (a_i + s))^2.$$

- It turns out that this value does not change if we shift b_i as well.
- **Proposition.** $D_s(a, b) = n \cdot (V_a + V_b - 2C_{a,b})$.
- To make sure that the value of dissimilarity does not depend on the sample size n, we divide $D_s(a, b)$ by n:

$$d_s(a, b) \stackrel{\text{def}}{=} \frac{D_s(a, b)}{n} = V_a + V_b - 2C_{a,b}.$$

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

19. Conclusions

- When we ignore time lag and non-linearities, we should select a similarity measure as follows.
- When both a measuring unit and a starting point are fixed, use the distance $\sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$.
- *Example:* movie evaluations.
- When neither a measuring unit nor a starting point are fixed, use correlation $\rho = \dfrac{C_{a,b}}{\sigma_a \cdot \sigma_b}$.
- *Examples:* there are many practical applications of this similarity measure.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

20. Conclusions (cont-d)

- When a starting point is fixed, but we can choose an arbitrary measuring unit, use $\frac{\left(\overline{a \cdot b}\right)^2}{\overline{a^2} \cdot \overline{b^2}}$.
- *Example:* comparing the fluctuations of two stock markets.
- When a measuring unit is fixed, but we can choose an arbitrary starting point, use $\sigma_a^2 + \sigma_b^2 - 2C_{a,b}$.
- *Example:* comparing two sequences of events from different time periods.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .

21. Acknowledgments

- We acknowledge the support of the Center of Excellence in Econometrics, Chiang Mai Univ., Thailand.
- This work was also supported in part:
  - by the National Science Foundation grants
    - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
    - DUE-0926721, and
  - by an award from Prudential Foundation.

Practitioners . . .
Limitations of Correlation
Other Similarity Measures
How to Select a . . .
Compared Values . . .
Case When No Scaling . . .
Case When All . . .
Case When Only . . .
Case When Only Shift . . .