Similarity Metrics Based on the Rearrangement Inequality

Su Jianlin

July 16, 2026

Yesterday, I saw a classmate @Ji_Ha_Kim share an interesting paper "Beyond Cosine Similarity", which proposes constructing new and more flexible similarity metrics based on the rearrangement inequality. I found it quite interesting after reading it, so I am writing this brief note.

Cosine Similarity

For two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, a commonly used normalized similarity metric is the cosine similarity:

$$\cos(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|} \in [-1, 1] \tag{1}$$

Setting aside the geometric interpretation and looking from a purely algebraic perspective, cosine similarity is a normalized inner product based on the Cauchy-Schwarz inequality:

$$-\|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \leq \boldsymbol{x} \cdot \boldsymbol{y} \leq \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \tag{2}$$

When $\cos(\boldsymbol{x}, \boldsymbol{y}) = \pm 1$, it means there exists $k > 0$ such that $\boldsymbol{y} = \pm k\boldsymbol{x}$. This shows that cosine similarity is insensitive to magnitude; it measures the degree of similarity in the directions of the two vectors, or in other words, their degree of linear correlation. Specifically, from the AM-GM inequality we have $\|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \leq \frac{1}{2}(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2)$. Substituting this into the Cauchy-Schwarz inequality for further relaxation yields:

$$-\frac{1}{2}(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2) \leq \boldsymbol{x} \cdot \boldsymbol{y} \leq \frac{1}{2}(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2) \tag{3}$$

The condition for the first equality to hold is $\boldsymbol{x} = -\boldsymbol{y}$, and for the second equality it is $\boldsymbol{x} = \boldsymbol{y}$. From this, we can define a new normalized inner product:

$$\text{dcos}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\frac{1}{2}(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2)} \tag{4}$$

In this case, $\text{dcos}(\boldsymbol{x}, \boldsymbol{y}) = \pm 1 \Leftrightarrow \boldsymbol{x} = \pm \boldsymbol{y}$, meaning dcos is a metric that measures the degree of exact equality between two vectors.

Rearrangement Inequality

For the inner product $\boldsymbol{x} \cdot \boldsymbol{y}$, the rearrangement inequality also holds:

$$\boldsymbol{x}^{\downarrow} \cdot \boldsymbol{y}^{\uparrow} = \boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\downarrow} \leq \boldsymbol{x} \cdot \boldsymbol{y} \leq \boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\uparrow} = \boldsymbol{x}^{\downarrow} \cdot \boldsymbol{y}^{\downarrow} \tag{5}$$

Here, $\boldsymbol{x}^{\downarrow}, \boldsymbol{x}^{\uparrow}$ refer to the new vectors obtained by arranging the components of $\boldsymbol{x}$ in descending/ascending order (and similarly for $\boldsymbol{y}$). Simply put, "reverse order inner product $\leq$ unordered inner product $\leq$ same order inner product". The condition for the first equality to hold is that $\boldsymbol{x}, \boldsymbol{y}$ are completely reversely ordered, meaning $(x_i - x_j)(y_i - y_j) \leq 0$ holds for all i, j; the condition for the second equality is that $\boldsymbol{x}, \boldsymbol{y}$ are completely similarly ordered, i.e., $(x_i - x_j)(y_i - y_j) \geq 0$.

Clearly, reordering does not change the magnitude, i.e., $\|\boldsymbol{x}^{\downarrow}\| = \|\boldsymbol{x}^{\uparrow}\| = \|\boldsymbol{x}\|$. Thus, further combining this with the Cauchy-Schwarz inequality, we get:

$$-\|\boldsymbol{x}\|\,\|\boldsymbol{y}\| \leq \boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\downarrow} \leq \boldsymbol{x} \cdot \boldsymbol{y} \leq \boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\uparrow} \leq \|\boldsymbol{x}\|\,\|\boldsymbol{y}\| \tag{6}$$

This shows that for $\boldsymbol{x} \cdot \boldsymbol{y}$, the rearrangement inequality is tighter than the Cauchy-Schwarz inequality. The standard proof of the rearrangement inequality uses the method of local adjustments, starting from the identity:

$$(x_i - x_j)(y_i - y_j) = (x_i y_i + x_j y_j) - (x_i y_j + x_j y_i) \tag{7}$$

The first term on the right is the inner product of (x_i, x_j) and (y_i, y_j), and the second term is the inner product after swapping x_i, x_j or y_i, y_j. If $x_i > x_j$ and $y_i > y_j$ (same order), then the left side is greater than 0, meaning that swapping from same order to reverse order decreases the inner product. In this way, the maximum is achieved when they are completely in the same order, and similarly, the minimum is achieved when they are completely in reverse order.

Rearrangement Similarity

If it is known that $l(\boldsymbol{x}, \boldsymbol{y}) \leq \boldsymbol{x} \cdot \boldsymbol{y} \leq u(\boldsymbol{x}, \boldsymbol{y})$, and the equalities on both sides are achievable, then a similarity metric can generally be constructed as:

$$2 \cdot \frac{\boldsymbol{x} \cdot \boldsymbol{y} - l(\boldsymbol{x}, \boldsymbol{y})}{u(\boldsymbol{x}, \boldsymbol{y}) - l(\boldsymbol{x}, \boldsymbol{y})} - 1 = \frac{2 \cdot \boldsymbol{x} \cdot \boldsymbol{y} - l(\boldsymbol{x}, \boldsymbol{y}) - u(\boldsymbol{x}, \boldsymbol{y})}{u(\boldsymbol{x}, \boldsymbol{y}) - l(\boldsymbol{x}, \boldsymbol{y})} \in [-1, 1] \tag{8}$$

Substituting the rearrangement inequality, we get:

$$\mathrm{rcos}(\boldsymbol{x}, \boldsymbol{y}) = \frac{2 \cdot \boldsymbol{x} \cdot \boldsymbol{y} - \boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\downarrow} - \boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\uparrow}}{\boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\uparrow} - \boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\downarrow}} \tag{9}$$

Then $\mathrm{rcos}(\boldsymbol{x}, \boldsymbol{y}) = 1$ means $\boldsymbol{x}, \boldsymbol{y}$ are completely in the same order, and $\mathrm{rcos}(\boldsymbol{x}, \boldsymbol{y}) = -1$ means $\boldsymbol{x}, \boldsymbol{y}$ are completely in reverse order. If cosine similarity describes the degree of linear correlation, then rcos describes some nonlinear correlations to a certain extent. rcos remains invariant under translation and positive scaling, i.e.,

$$\text{rcos}(a\boldsymbol{x} + b\mathbf{1}, c\boldsymbol{y} + d\mathbf{1}) = \text{rcos}(\boldsymbol{x}, \boldsymbol{y}) \tag{10}$$

However, this is not the case for general strictly monotonic component-wise transformations (such as component-wise cubing), meaning rcos preserves and utilizes the actual numerical values, rather than relying solely on the ranking.

Connection and Difference

It should be pointed out that although this post was inspired by "Beyond Cosine Similarity", the rearrangement similarity constructed here is slightly different from the original paper. The construction method in the original paper is (recos is also the notation used in the original paper):

$$\text{recos}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{|\boldsymbol{x}^{\uparrow} \cdot \boldsymbol{y}^{\updownarrow}|}, \qquad \boldsymbol{y}^{\updownarrow} = \begin{cases} \boldsymbol{y}^{\uparrow}, & \boldsymbol{x} \cdot \boldsymbol{y} > 0 \\ \boldsymbol{y}^{\downarrow}, & \boldsymbol{x} \cdot \boldsymbol{y} < 0 \end{cases} \tag{11}$$

That is, it selects the same-order or reverse-order bound based on the sign of the inner product. Although defining it this way can also restrict the result to $[-1, 1]$, the sign of the inner product can no longer be used to distinguish between same order and reverse order. Personally, I think it is not as reasonable as the definition in this post.

For example, let $\boldsymbol{x} = (1, 2, 3), \boldsymbol{y} = (6, 5, 4)$. These two vectors are completely in reverse order, and we would expect a rearrangement similarity metric to at least give a negative result. But obviously, $\text{recos}(\boldsymbol{x}, \boldsymbol{y})$ gives a positive result, whereas $\text{rcos}(\boldsymbol{x}, \boldsymbol{y})$ gives a negative result. The definition in this post directly uses the double bounds of the rearrangement inequality for normalization, anchoring ± 1 to completely same order and completely reverse order respectively. I humbly believe this is more reasonable semantically.

Of course, whether rcos or recos is more reasonable in practical applications depends on the specific scenario. The claimed superiority here is only with respect to the meaning of the term "rearrangement similarity".

Rank Correlation Coefficient

If we treat $\boldsymbol{x}, \boldsymbol{y}$ as individual 2D data points $(x_1, y_1), (x_2, y_2), \cdots$, then cosine similarity is related to their linear correlation coefficient (Pearson correlation coefficient):

$$\text{pearson}(\boldsymbol{x}, \boldsymbol{y}) = \cos(\boldsymbol{x} - \bar{\boldsymbol{x}}, \boldsymbol{y} - \bar{\boldsymbol{y}}) \tag{12}$$

where $\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}$ are the means of the components of $\boldsymbol{x}, \boldsymbol{y}$. Based on the linear correlation coefficient, the rank correlation coefficient (Spearman's rank correlation coefficient) is derived:

$$\text{spearman}(\boldsymbol{x}, \boldsymbol{y}) = \text{pearson}(\boldsymbol{r}_x, \boldsymbol{r}_y) \tag{13}$$

where $\boldsymbol{r}_x, \boldsymbol{r}_y$ are the rank vectors of $\boldsymbol{x}, \boldsymbol{y}$. That is, if x_i is the k-th component of $\boldsymbol{x}$ ordered from smallest to largest, then $(\boldsymbol{r}_x)_i = k$. By definition, the rank correlation coefficient completely ignores the actual values of the data and only cares about their order. We previously used this metric when evaluating semantic similarity. Interestingly, the rank correlation coefficient can also be viewed as a special case of rcos:

$$\text{spearman}(\boldsymbol{x}, \boldsymbol{y}) = \text{rcos}(\boldsymbol{r}_x, \boldsymbol{r}_y) \tag{14}$$

The reader is asked to verify this equation independently; I will not expand on it here. Thus, $\text{rcos}(\boldsymbol{x}, \boldsymbol{y})$ can be regarded as a generalized rearrangement similarity that retains the original numerical information.

In a Word

This post introduced the idea of constructing similarity metrics from the perspective of "similarity = normalized inner product" based on double-sided inequalities of the inner product, focusing on the rearrangement similarity metric constructed from the rearrangement inequality.