An Innovative Drastic Metric for Ranking Similarity in Decision-Making Problems

Wojciech Sałabun
0000-0001-7076-2519
West Pomeranian University of Technology
in Szczecin
ul. Żołnierska 49, 71-210 Szczecin, Poland
Email: wojciech.salabun@zut.edu.pl

Andrii Shekhovtsov
0000-0002-0834-2019
National Telecommunications Institute
ul. Szachowa 1, 04-894 Warsaw, Poland
Email: a.shekhovtsov@il-pib.pl

Abstract—In this paper, we propose a novel approach to distance measurement for rankings, introducing a new metric that exhibits exceptional properties. Our proposed distance metric is defined within the interval of 0 to 1, ensuring a compact and standardized representation. Importantly, we demonstrate that this distance metric satisfies all the essential criteria to be classified as a true metric. By adhering to properties such as non-negativity, identity of indiscernibles, symmetry, and the crucial triangle inequality, our proposed distance metric provides a robust and reliable approach for comparing rankings in a rigorous and mathematically sound manner. Finally, we compare our new metric with distances such as Hamming distance, Canberra distance, Bray-Curtis distance, Euclidean distance, Manhattan distance, and Chebyshev distance. By conducting simple experiments, we assess the performance and advantages of our proposed metric in comparison to these established distance measures. Through these comparisons, we demonstrate the superior properties and capabilities of our new drastic weighted similarity distance for accurately capturing the dissimilarities and similarities between rankings in the decision-making domain.

I. Introduction

DISTANCE measures are fundamental tools in many areas of data analysis, including machine learning, statistics, data mining, and many more [1], [2]. They quantify the difference or dissimilarity between pairs of objects, like vectors, sets, or more complex structures, providing a quantitative basis for their comparison [3].

A key aspect of distance measures is that they must satisfy certain properties, such as non-negativity (distances are always non-negative), identity of indiscernibles (the distance between an object and itself is zero), symmetry (the distance from A to B is the same as from B to A), and the triangle inequality (the direct distance from A to B is always shorter or equal to the distance from A to B via an intermediary point C) [4].

There are various types of distance measures, including Euclidean [5], Manhattan [6], Chebyshev [7], Hamming [6], Canberra, Bray-Curtis, and many others [8], [9], each with their own characteristics and use-cases. Some measures like Euclidean and Manhattan are primarily used for continuous variables [6], while others like Hamming are used for categorical variables [10]. Some measures are sensitive to the scale and distribution of the data, while others are more robust.

The choice of the appropriate distance measure is highly dependent on the nature of the data and the specific objectives of the analysis [11], [12]. For example, in a scenario where extreme values or outliers are important, a measure such as the Chebyshev distance could be useful as it focuses on the maximum difference in any one dimension. On the other hand, for data that represents rankings or preferences, a measure like Spearman's footrule or the Kendall tau distance might be more appropriate [13].

When comparing rankings in decision-making, distance measures play a vital role. To compare rankings, we need a way to quantify how similar or different two rankings are [14], [15]. That's where distance measures come in. They provide a numeric value representing the dissimilarity between two rankings, with lower values typically indicating greater similarity.

The choice of distance measure can have a significant impact on the comparison. Some measures are more sensitive to the exact order of the rankings, while others, like Spearman's footrule [16], are more focused on the overall similarity. Moreover, some measures are more sensitive to differences at the top of the rankings [17], [18], while others treat all positions equally. Overall, comparing rankings using distance measures can provide valuable insights in decision-making, helping decision-makers understand how different choices, evaluations, or scenarios compare to each other, and aiding in making more informed, data-driven decisions [19].

Rankings and comparisons form an integral part of decision-making processes in diverse fields such as information retrieval, sports [20], elections, and more [21]. However, a significant challenge that persists in these scenarios is quantifying the dissimilarity or distance between different rankings effectively and accurately. Traditional distance measures, while useful, can often fail to capture the nuances and subtleties inherent in the comparison of rankings. To address these limitations and introduce a more robust and versatile solution, the motivation behind this paper emerges.

In this paper, the main contribution is to propose a novel distance metric that is particularly well-suited for ranking comparisons. We aspire to create a metric that not only captures the dissimilarity between rankings accurately but also

exhibits essential properties required of a true metric. A key part of our motivation is to ensure that this new measure is defined within the interval of 0 to 1, thus providing a compact and standardized representation that is easy to interpret across diverse scenarios.

The structure of the paper is as follows: In Section II, the necessary groundwork is laid by introducing and defining key distance measures. Section III is dedicated to proposing a novel distance metric, WS_{dra}, along with comprehensive proof of its properties. Section IV then provides a comparative study of this new metric against the traditional measures introduced in Section II. Finally, Section V concludes the paper by summarizing the research findings and their potential implications.

II. PRELIMINARIES

A. Weighed similarity

The Weighted Similarity (WS) measure aims to be sensitive to significant changes in rankings while remaining robust against minor fluctuations. It also offers the advantage of being easy to interpret, with its values falling within a specified range [17].

In designing the WS measure, a key assumption is made that differences in the top rankings are more impactful than those lower down the list. This is intuitive in scenarios where top-ranked items often have more importance, such as in competitive rankings or search results.

The formula to calculate the WS measure is:

$$WS = 1 - \sum_{i=1}^{n} 2^{-x_i} \frac{|x_i - y_i|}{\max\{|1 - x_i|, |N - x_i|\}} \quad (1)$$

In this equation: WS represents the similarity coefficient's value, n is the length of the ranking, and x_i and y_i represent the place in the ranking for the ith element in the respective rankings x and y.

This formula implies that WS calculates the absolute differences in the ranks of each element in two rankings, normalizes them by the maximum possible difference for that element, and then sums the results. This total is subtracted from 1 to convert it into a similarity measure. Thus, a larger WS value indicates a higher similarity between the two rankings, making WS an effective tool for comparing and analyzing rankings [17].

B. Hamming distance

Hamming distance is a metric that measures the difference between two strings of equal length. It counts the number of positions at which the corresponding symbols in the strings differ [6]. The formula for calculating Hamming distance is as follows:

$$d(x, y) = \frac{\sum_{i=1}^{n} \delta(x_i, y_i)}{n} \quad (2)$$

where x_i and y_i represent the symbols at position i in the two strings, and $\delta(x_i, y_i)$ is an indicator function that equals 0 if x_i and y_i are equal, and 1 otherwise. The Hamming distance provides a way to quantify the dissimilarity between two vectors by measuring the number of symbol mismatches [22].

C. Canberra distance

The Canberra distance is a metric used to quantify the dissimilarity between two vectors or points in a multidimensional space. It takes into account both the magnitude and direction of differences between corresponding components of the vectors [23]. The formula for calculating the Canberra distance is as follows:

$$d(x, y) = \sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|} \quad (3)$$

where x_i and y_i represent the components at position i in the two vectors. The Canberra distance considers the absolute difference between the components, normalized by the sum of their magnitudes. This normalization accounts for differences in scale and ensures that each component contributes proportionally to the overall distance calculation.

D. Bray-Curtis distance

The Bray-Curtis distance is a metric used to measure the dissimilarity between two vectors or points in a multidimensional space. It considers both the magnitude and direction of differences between corresponding components of the vectors, taking into account their relative proportions [24]. The formula for calculating the Bray-Curtis distance is as follows:

$$d(x, y) = \frac{\sum_{i=1}^{n} |x_i - y_i|}{\sum_{i=1}^{n} |x_i + y_i|} \quad (4)$$

where x_i and y_i represent the components at position i in the two vectors. The Bray-Curtis distance calculates the absolute difference between the components and normalizes it by the sum of their absolute values. This normalization accounts for differences in scale and ensures that each component contributes proportionally to the overall distance calculation.

E. Euclidean distance

The Euclidean distance is a metric used to measure the straight-line distance between two points in a multidimensional space. It calculates the length of the line connecting the two points, taking into account the differences between their corresponding components [6]. The formula for calculating the Euclidean distance is as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (5)$$

where x_i and y_i represent the components at position i in the two points. The Euclidean distance computes the squared differences between the components, sums them up, and takes the square root of the result. This computation ensures that each component's contribution to the distance calculation is positive and reflects the actual geometric distance between the points.

F. Manhattan distance

The Manhattan distance, also known as the city block distance or L1 distance, is a metric used to measure the distance between two points in a multidimensional space. It calculates the sum of the absolute differences between the corresponding components of the two points [25]. The formula for calculating the Manhattan distance is as follows:

$$d(x,y) = \sum_{i=1}^{n} |x_i - y_i| \qquad (6)$$

where x_i and y_i represent the components at position i in the two points. The Manhattan distance measures the distance traveled along the grid-like streets of a city, where movement can only occur in vertical and horizontal directions. It sums up the absolute differences between the components, disregarding their sign.

G. Chebyshev distance

The Chebyshev distance, also known as the maximum value or L_∞ distance, is a metric that measures the dissimilarity between two vectors or points in a multidimensional space. It calculates the maximum difference between the corresponding components of the two vectors [7]. The formula for calculating the Chebyshev distance is as follows:

$$d(x,y) = \max_{i=1}^{n} |x_i - y_i| \qquad (7)$$

where x_i and y_i represent the components at position i in the two vectors. The Chebyshev distance provides a measure of the largest difference between any pair of corresponding components in the vectors, which corresponds to the maximum distance in any dimension.

III. A NEW PROPOSED DRASTIC METRIC

In the realm of data analytics and decision-making, the concept of distance plays a pivotal role, enabling us to evaluate similarities, disparities, and rank variables effectively. However, traditional distance metrics have their inherent strengths and limitations. To address these shortcomings and propel the field forward, we introduce a novel distance measure based on the WS coefficient.

Our proposed distance metric revolutionizes the notion of distance by adopting a drastic approach. Instead of penalizing discrepancies in ranking, we treat each comparison in the ranking position as a binary attribute, representing a significant or non-significant relationship. This novel perspective eliminates the conventional notion of assigning varying degrees of penalty based on the magnitude of ranking differences. In essence, our approach treats all errors equally, as an error is an error regardless of what is given in analysed position. This drastic approach fosters a fairer assessment of rankings.

Moreover, our new distance measure recognizes the inherent significance disparity across different ranking positions. It assigns greater consequence to the head of the ranking, acknowledging the top positions as more crucial than the lower ones. This acknowledgment aligns with the understanding that

errors at the top of the ranking can have more significant implications than errors further down the list. By considering this significance disparity, our distance measure offers a more nuanced and accurate evaluation of rankings.

Crucially, our proposed measure is normalized within the interval from 0 to 1, enabling straightforward interpretation and comparison across diverse contexts. This normalization facilitates intuitive understanding and ensures that the distance measure remains consistent and interpretable regardless of the specific data or application domain.

By embodying these innovative characteristics, our proposed distance measure qualifies as a true metric in the rigorous mathematical sense. Its drastic approach, significance-awareness, and normalized range combine to offer a comprehensive and reliable framework for comparing rankings in various decision-making scenarios. Through empirical evaluations and theoretical analyses, we demonstrate the superiority and practical utility of our proposed distance measure, paving the way for enhanced ranking analysis and informed decision-making in diverse domains.

A. Definition

The new metric, denoted as $WS_{dra}(x,y)$, is defined as follows:

$$WS_{dra}(x,y) = \frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}} \qquad (8)$$

The metric operates on two rankings, denoted as x and y, with each ranking consisting of N elements. The key element of this metric is the function $f(x_i, y_i)$, which compares the elements at corresponding positions in the two rankings.

The function $f(x_i, y_i)$ is defined as follows:

$$f(x_i, y_i) = \begin{cases} 0, & \text{if } x_i = y_i \\ 1, & \text{if } x_i \neq y_i \end{cases} \qquad (9)$$

In other words, if the elements at position i in the rankings x and y are the same, $f(x_i, y_i)$ is assigned a value of 0. Conversely, if the elements are different, $f(x_i, y_i)$ takes the value of 1.

The $WS_{dra}(x,y)$ metric computes the weighted sum of $f(x_i, y_i)$ values for each position i, using the weights given by the geometric series 2^{-i}. The weights decrease exponentially as i increases, reflecting a decreasing level of importance for elements further down the rankings. The summation of the weighted $f(x_i, y_i)$ values is then divided by the factor $1 - 2^{-N}$ to ensure normalization within the range of 0 to 1.

Overall, this new metric captures the dissimilarities between two rankings by assigning a weight to each pairwise comparison based on the function $f(x_i, y_i)$. It combines these weighted comparisons to provide a comprehensive measure of dissimilarity between the rankings x and y, where a higher value indicates greater dissimilarity. The normalization factor ensures that the metric remains consistent and interpretable across different ranking sizes.

TABLE I
A SIMPLE EXAMPLE OF TWO RANKINGS, I.E., x_i AND y_i.

i	x_i	y_i	$f(x_i, y_i)$	2^{-i}
1	1	3	1	$\frac{1}{2}$
2	2	2	0	$\frac{1}{4}$
3	3	1	1	$\frac{1}{8}$

We demonstrate a short computational example of the newly proposed metric. Consider the example shown in Table I, which illustrates two rankings, denoted as x_i and y_i. Each row in the table corresponds to a position i in the rankings, and we calculate the associated values of $f(x_i, y_i)$ and 2^{-i}. Thus, ranking x_i means the order of alternatives in the form $A_1 > A_2 > A_3$, and ranking y_i in the form $A_3 > A_2 > A_1$. To calculate the $WS_{dra}(x, y)$ value for these rankings, we use the formula (8) and we get the following result:

$$WS_{dra}(x, y) = \frac{\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 0 + \frac{1}{8} \cdot 1}{1 - \frac{1}{8}} = \frac{\frac{5}{8}}{\frac{7}{8}} = \frac{5}{7}$$

A true metric, also known as a metric space or distance metric, is a mathematical concept used to quantify the distance or similarity between objects within a set. It defines a set of rules or properties that a distance function must satisfy to be considered a true metric [2].

In a true metric, the following properties should hold:

1) Non-negativity: The distance between any two objects is non-negative. It is always equal to or greater than zero.
2) Identity of indiscernibles: The distance between two objects is zero if and only if the objects are identical.
3) Symmetry: The distance between object A and object B is the same as the distance between object B and object A.
4) Triangle inequality: The distance from object A to object B, added to the distance from object B to object C, is always greater than or equal to the distance from object A to object C.

In the following subsections, we will explore each of the presented properties to demonstrate the validity of the proposed measure as a true metric. Our objective is to carefully analyze and evaluate these properties, providing a solid foundation for the metric's credibility. Through a systematic examination, we will investigate the non-negativity, identity of indiscernibles, symmetry, and triangle inequality properties. By establishing the fulfillment of these properties, we aim to establish the proposed metric as a reliable tool for comparing rankings. The goal is to offer a well-founded framework that promotes accurate assessments and meaningful insights for decision-making.

B. Non-negativity

To prove the inequality $\frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}} \geq 0$ when $f(x_i, y_i)$ can take the values 0, 1, or a combination of 0 and 1, we will consider three different cases.

Case 1: $f(x_i, y_i) = 0$ for all i: When all terms in the summation are multiplied by 0, the numerator becomes zero. The denominator, $1 - 2^{-N}$, is positive since $2^{-N} < 1$ for all positive N. Thus, the inequality holds trivially: $0 \geq 0$.

Case 2: $f(x_i, y_i) = 1$ for all i: In this case, each term in the summation will be equal to 2^{-i} since $f(x_i, y_i)$ is always 1. The numerator then becomes:

$$\sum_{i=1}^{N} 2^{-i} = 2^{-1} + 2^{-2} + \ldots + 2^{-N}$$

This sum is a finite geometric series, and its sum can be calculated as follows:

$$\sum_{i=1}^{N} 2^{-i} = \frac{2^{-1}(1 - 2^{-N})}{1 - 2^{-1}} = \frac{1 - 2^{-N}}{2 - 1} = 1 - 2^{-N}$$

Since $1 - 2^{-N}$ is positive, the numerator is non-negative. The denominator, $1 - 2^{-N}$, is also positive and is equal to the nominative. Therefore, when $f(x_i, y_i) = 1$ for all i, the inequality holds:

$$\frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}} = 1 \geq 0$$

Case 3: $f(x_i, y_i)$ is equal 0 or 1: In this case, the numerator of the expression is a sum of terms, each multiplied by $2^{-i} f(x_i, y_i)$. Since $f(x_i, y_i)$ can be 0 or 1, the product $2^{-i} f(x_i, y_i)$ will be either 0 or 2^{-i}. It means that nominative will be limited to interval:

$$0 \leq \sum_{i=1}^{N} 2^{-i} f(x_i, y_i) < 1$$

Since both 0 and 2^{-i} are non-negative, the numerator is a non-negative number. The denominator, $1 - 2^{-N}$, is positive and it is the biggest possible value of nominative, therefore WS_{dra} will be limited to:

$$0 \leq \frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}} \leq 1$$

Hence, when $f(x_i, y_i)$ takes 0 or 1, the inequality holds:

$$\frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}} \geq 0$$

In all cases, we have shown that the inequality holds. Therefore, we can conclude that $\frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}} \geq 0$ when $f(x_i, y_i)$ can be equal to 0, 1, or a combination of 0 and 1.

C. Identity of indiscernibles

To prove that only $WS_{dra}(x, x) = 0$ for the given expression $WS_{dra}(x, y) = \frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}}$, we can substitute x for y in the expression:

$$WS_{dra}(x, x) = \frac{\sum_{i=1}^{N} 2^{-i} f(x_i, x_i)}{1 - 2^{-N}}$$

Now, let's focus on the numerator of the expression. Since $f(x_i, x_i)$ represents the function f evaluated at the same

element x_i for both arguments, it will always yield the same result. Therefore, $f(x_i, x_i)$ is a constant for all i. Let's denote this constant as c, such that $f(x_i, x_i) = c$ for all i. Substituting c into the numerator, we have:

$$\sum_{i=1}^{N} 2^{-i} f(x_i, x_i) = \sum_{i=1}^{N} 2^{-i} c = c \sum_{i=1}^{N} 2^{-i}$$

The sum $\sum_{i=1}^{N} 2^{-i}$ is a finite geometric series and can be computed as:

$$\sum_{i=1}^{N} 2^{-i} = \frac{2^{-1}(1 - 2^{-N})}{1 - 2^{-1}} = \frac{1 - 2^{-N}}{2 - 1} = 1 - 2^{-N}$$

Now, substituting this value back into the expression, we get:

$$WS(x, x) = \frac{c \sum_{i=1}^{N} 2^{-i}}{1 - 2^{-N}} = \frac{c(1 - 2^{-N})}{1 - 2^{-N}} = c$$

Since c is a constant, it does not depend on the choice of x, and therefore, c is equal to $f(x_i, x_i)$ for any x_i. Since $f(x_i, x_i)$ can take the values of 0 or 1 (according to the given property), we can notice that in this case c is equal 0. Therefore, based on the given expression, we can universally prove that only $WS(x, x) = 0$ and it depends on the specific value of c (i.e., the constant $f(x_i, x_i)$).

D. Symmetry

To prove that $WS_{dra}(x, y) = WS_{dra}(y, x)$ for the expression $WS_{dra}(x, y) = \frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}}$, we need to show that the weighted sum is symmetric with respect to its arguments. Let's consider the left-hand side $WS_{dra}(x, y)$ and the right-hand side $WS(y, x)$ of the equation separately and compare them.

$$WS_{dra}(x, y) = \frac{\sum_{i=1}^{N} 2^{-i} f(x_i, y_i)}{1 - 2^{-N}}$$

$$WS_{dra}(y, x) = \frac{\sum_{i=1}^{N} 2^{-i} f(y_i, x_i)}{1 - 2^{-N}}$$

To show that $WS_{dra}(x, y) = WS_{dra}(y, x)$, we need to demonstrate that the numerator and denominator of both expressions are equal. For each term in the numerator, we have: $f(x_i, y_i)$ in the expression for $WS_{dra}(x, y)$ and $f(y_i, x_i)$ in the expression for $WS_{dra}(y, x)$. Since the order of the arguments is switched between the two expressions, we can see that $f(x_i, y_i) = f(y_i, x_i)$ for each i. Therefore, the numerator of both expressions is identical. The denominator of both expressions is the same: $1 - 2^{-N}$. Since the numerator and denominator of both $WS_{dra}(x, y)$ and $WS_{dra}(y, x)$ are equal, we can conclude that $WS_{dra}(x, y) = WS_{dra}(y, x)$. Hence, we have proven that the weighted sum expression $WS_{dra}(x, y)$ is symmetric with respect to its arguments, satisfying the property $WS_{dra}(x, y) = WS_{dra}(y, x)$.

TABLE II
ALL POSSIBLE BINARY COMBINATIONS FOR $f(a_i, b_i)$, $f(b_i, c_i)$, $f(a_i, c_i)$ AND THE VALUE X_i OBTAINED AS $f(a_i, b_i) + f(b_i, c_i) - f(a_i, c_i)$.

$f(a_i, b_i)$	$f(b_i, c_i)$	$f(a_i, c_i)$	X_i
0	0	0	0
0	1	1	0
1	0	1	0
1	1	1	1
1	1	0	2

E. Triangle inequality

We'll work with the formula (8) to prove the triangle inequality property:

$$WS_{dra}(a, b) + WS_{dra}(b, c) \geq WS_{dra}(a, c)$$

Let's denote $A = WS_{dra}(a, b)$, $B = WS_{dra}(b, c)$, and $C = WS_{dra}(a, c)$. Substituting these values into the inequality, we have $A + B \geq C$. Now, let's consider the individual terms in the numerator of each expression.

$$WS_{dra}(a, b) : \sum_{i=1}^{N} 2^{-i} f(a_i, b_i)$$

$$WS_{dra}(b, c) : \sum_{i=1}^{N} 2^{-i} f(b_i, c_i)$$

$$WS_{dra}(a, c) : \sum_{i=1}^{N} 2^{-i} f(a_i, c_i)$$

Now, let's examine the numerator term-wise for the three expressions: for each i, 2^{-i} is a non-negative constant; for each term $f(a_i, b_i)$ in $WS_{dra}(a, b)$, $f(b_i, c_i)$ in $WS_{dra}(b, c)$, and $f(a_i, c_i)$ in $WS_{dra}(a, c)$, they can take the values of 0 or 1 according to the formula (9). Now, we can compare the terms between the expressions, where for each term i, we have:

$$2^{-i} f(a_i, b_i) + 2^{-i} f(b_i, c_i) \geq 2^{-i} f(a_i, c_i)$$

This inequality holds because for any given term, either $f(a_i, b_i) = f(a_i, c_i) = 1$ or $f(b_i, c_i) = 0$ (which makes the left-hand side greater than or equal to the right-hand side) or $f(a_i, b_i) = f(b_i, c_i) = 0$ (which makes the left-hand side equal to the right-hand side). The all possible cases are presents in Table II, where $X_i = f(a_i, b_i) + f(b_i, c_i) - f(a_i, c_i)$. Now, summing up these inequalities over all i from 1 to N, we have:

$$^{dra}WS(a, b) + {}^{dra}WS(b, c) \geq {}^{dra}WS(a, c)$$

$$\frac{\sum_{i=1}^{N} 2^{-i} f(a_i, b_i)}{1 - 2^{-N}} + \frac{\sum_{i=1}^{N} 2^{-i} f(b_i, c_i)}{1 - 2^{-N}} \geq \frac{\sum_{i=1}^{N} 2^{-i} f(a_i, c_i)}{1 - 2^{-N}}$$

$$\frac{\sum_{i=1}^{N} (2^{-i} f(a_i, b_i) + 2^{-i} f(b_i, c_i) - 2^{-i} f(a_i, c_i))}{1 - 2^{-N}} \geq 0$$

TABLE III
THE COMPARED RANKINGS, I.E., x_i AND $y_i^{(j)}$ FOR $j = 1, 2, ..., 7$, WHERE RED COLOR INDICATES THE DIFFERENCES WITH THE ORIGINAL RANKING.

i	x_i	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(3)}$	$y_i^{(4)}$	$y_i^{(5)}$	$y_i^{(6)}$	$y_i^{(7)}$
1	1	2	1	1	1	3	4	5
2	2	1	3	2	2	2	2	2
3	3	3	2	4	3	1	3	3
4	4	4	4	3	5	4	1	4
5	5	5	5	5	4	5	5	1

$$\sum_{i=1}^{N} 2^{-i} \left(f(a_i, b_i) + f(b_i, c_i) - f(a_i, c_i) \right) \geq 0$$

$$\sum_{i=1}^{N} 2^{-i} X_i \geq 0$$

Therefore, X_i is always non-negativity, we can conclude that we have proven triangle inequality for WS_{dra} distance.

IV. COMPARISON AND DISCUSSION

Table III provides a visual representation of the initial ranking x_i and sample rankings $y_i^{(j)}$ for $j = 1, 2, ..., 7$. The table is designed to compare the considered distance measures with the proposed new measure. In this table, the red color highlights the differences between each example ranking and the initial ranking.

The table consists of nine columns. The first column, labeled i, denotes the position in the rankings. The second column represents the initial ranking, denoted as x_i. The remaining columns, labeled $y_i^{(j)}$, correspond to the examplary rankings for $j = 1, 2, ..., 7$.

Each cell in the table represents the element at position i in the corresponding ranking. The red color is used to indicate any differences between the element in the examplary ranking and the initial ranking. By visually highlighting these differences, the table facilitates a clear comparison between the rankings and serves as a reference for evaluating the performance of different distance measures. Table III provides a useful reference point for understanding the subsequent analyses and discussions related to the comparisons between the considered distance measures and the proposed new measure.

Table IV provides a comprehensive summary of the similarity and distance measures for the previously presented rankings, offering valuable insights for analyzing and comparing the relationships between x_i and each $y_i^{(j)}$. The measures included in the table enable a thorough assessment of the similarities and differences among the rankings.

The table introduces a new proposed distance metric denoted as WS_{dra}, represented in blue font. This novel metric returns distances ranging from 0.5484 to 0.7742. It was developed as an enhancement of the WS coefficient. To ensure consistent information direction, the $1 - WS$ coefficient was introduced, yielding values ranging from 0.2083 to 0.5313.

This modification was incorporated into the analysis, as the new distance metric was built upon this coefficient.

In addition to the newly proposed metric, Table IV includes well-known distance measures commonly used for comparison purposes. These measures, namely Hamming, Canberra, Bray-Curtis, Euclidean, Manhattan, and Chebyshev, offer additional perspectives on the dissimilarity between x_i and each $y_i^{(j)}$ ranking.

The Hamming measure, typically employed for comparing categorical data, consistently yields a distance value of 0.4000 for all comparisons. This suggests that all rankings $y_i^{(j)}$ exhibit the same level of distance and similarity with respect to x_i. However, from a decision-making standpoint, it becomes evident that this statement does not hold true, as, for example, ranking $y_i^{(1)}$ is closer to x_i than ranking $y_i^{(2)}$.

The Canberra measure calculates distances ranging from 0.6667 to 1.3333, providing insights into the relative dissimilarity between x_i and the different $y_i^{(j)}$ rankings. This measure considers both the magnitude and direction of differences between the rankings, offering a comprehensive assessment of their dissimilarity.

The Bray-Curtis measure, which evaluates dissimilarity based on the proportions of shared and unique elements, yields distances ranging from 0.0667 to 0.2667. This measure takes into account the presence and absence of specific elements, providing valuable information regarding the relative dissimilarity between the rankings.

Let's delve deeper into the two distinct sets of rankings: $y_i^{(1)}$ to $y_i^{(4)}$, and $y_i^{(1)}$ alongside $y_i^{(5)}$ to $y_i^{(7)}$. What sets these rankings apart is the presence of a singular swap between alternatives, occurring either in adjacent positions or non-adjacent ones.

In the first set of rankings, we begin by examining modifications at the top (or head) of the ranking and gradually proceed towards the bottom (or tail). Ranking tasks inherently pose a significant challenge, as they tend to assign more weight or significance to changes at the beginning of the ranking sequence rather than towards the end. For instance, let's consider a scenario where a company not placed first in the ranking wins a tender—such an event is, of course, wrong, as that company would either won or be removed from the ranking (e.g., due to withdrawal).

By comparing these rankings with the x_i ranking and employing five different distance measurement methods—Hamming, Bray-Curtis, Euclidean, Manhattan, and Chebyshev—we observe that the comparison values for all paired rankings remain constant. The respective constant values assigned to these methods are 0.4, 0.0667, 1.4142, 2.0000, and 1.0000. This indicates that these five measurements may not adequately capture the variability required for decision-making processes, as they remain insensitive to changes in ranking positions, regardless of where those changes occur. Furthermore, when considering the $1 - WS$ ratio and the Canberra distance, both measurements consistently exhibit a decreasing trend in values with each subsequent ranking, $y_i^{(j)}$,

TABLE IV
SUMMARY OF SIMILARITY AND DISTANCE MEASURES FOR x_i AND $y_i^{(j)}$ FOR $j = 1, 2, ..., 7$ RANKINGS.

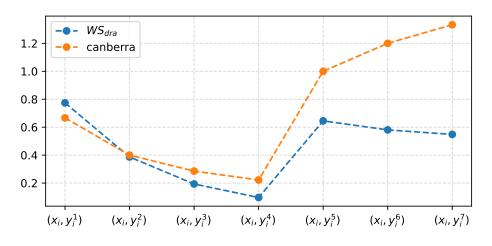| Measures	$d(x_i, x_i)$	$d(x_i, y_i^{(1)})$	$d(x_i, y_i^{(2)})$	$d(x_i, y_i^{(3)})$	$d(x_i, y_i^{(4)})$	$d(x_i, y_i^{(5)})$	$d(x_i, y_i^{(6)})$	$d(x_i, y_i^{(7)})$
WS_{dra}	0.0000	0.7742	0.3871	0.1935	0.0968	0.6452	0.5806	0.5484
$1 - WS$	0.0000	0.2083	0.1458	0.0833	0.0286	0.3750	0.4375	0.5313
Hamming	0.0000	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000
Canberra	0.0000	0.6667	0.4000	0.2857	0.2222	1.0000	1.2000	1.3333
Bray-Curtis	0.0000	0.0667	0.0667	0.0667	0.0667	0.1333	0.2000	0.2667
Euclidean	0.0000	1.4142	1.4142	1.4142	1.4142	2.8284	4.2426	5.6569
Manhattan	0.0000	2.0000	2.0000	2.0000	2.0000	4.0000	6.0000	8.0000
Chebyshev	0.0000	1.0000	1.0000	1.0000	1.0000	2.0000	3.0000	4.0000

Fig. 1. Comparison of the proposed distance WS_{dra} with the Canberra distance.

where j = 1, 2, 3, 4.

Shifting focus to the second set of rankings, we encounter a scenario where the order of alternatives is swapped, albeit not in adjacent positions. This gives rise to a peculiar situation where the comparison of x_i with $y_i^{(1)}$ results in erroneous rankings at the first and second positions. Similarly, in the comparison of x_i with $y_i^{(5)}$, erroneous rankings occur at the first and third positions, and so on. Intuitively, we would expect a larger distance in the first case, with decreasing values in subsequent comparisons, as errors are initially observed at the first position and then propagate to the further subsequent positions.

However, both the Canberra distance and the $1 - WS$ ratio exhibit counter-intuitive behavior, as their values increase rather than decrease. This contradicts the initial assumption. Moreover, the $1-WS$ ratio cannot be considered a true metric due to its lack of symmetry. The comparison between the results obtained using the WS_{dra} method and the Canberra method is depicted in Fig. 1. The analysis reveals a significant correlation for the first set of alternatives, with a Pearson correlation coefficient of 0.9995. This high correlation indicates a strong agreement between the values obtained from the WS_{dra} method and the Canberra method for this set of rankings.

In the case of the second set of rankings, a similarly strong correlation is observed; however, it possesses a negative nature with a correlation coefficient of -0.9965. This negative correlation suggests an inverse relationship between the values obtained by the WS_{dra} method and the Canberra method for this particular set of rankings.

When considering both sets of rankings collectively, a moderate positive correlation of 0.6995 is observed. This indicates that, overall, there is a consistent relationship between the rankings obtained from the WS_{dra} method and the Canberra method, albeit with a moderate strength of association.

Fig. 1 visually illustrates these correlations, providing a clear understanding of the magnitude and characteristics of the relationship between the WS_{dra} method and the Canberra method for the various sets of rankings. It is important to note that the Canberra distance exceeded the value of 1, which represents the upper limit in the WS_{dra} distance metric.

The WS_{dra} distance metric is based on certain assumptions that contribute to its enhanced reliability and applicability across diverse contexts. Firstly, this metric takes into account the weighted differences between rankings, recognizing that not all changes in rankings hold equal significance. By assigning appropriate weights to these differences, the WS_{dra} metric captures the varying impact of alterations in ranking positions, providing a more accurate assessment of dissimilarity.

Additionally, the WS_{dra} metric adheres to the fundamental

principle that "an error is an error." It acknowledges that any deviation or discrepancy between rankings, regardless of its location or magnitude, should be considered as an error. By treating all errors equally, the WS_{dra} metric ensures a fair and unbiased evaluation of dissimilarity, promoting a more reliable comparison between rankings.

Consequently, these examples demonstrate that the distance metric WS_{dra} provides greater reliability, as it considers the weighted differences between rankings, adhering to the principle that "an error is an error." It is important to note that the value of this metric is normalized within the range of 0 to 1, ensuring its applicability across diverse contexts.

V. Conclusion

In this paper, we introduced a new distance metric for the comparison of rankings, demonstrating its effectiveness and advantages over well-established distance measures. This novel measure, which we term the drastic WS distance, conforms to all necessary properties of a true metric and exhibits a unique capability to capture nuances in the ranking structure.

The drastic distance metric provides a compact, standardized representation within the 0 to 1 interval, making it easy to interpret across a broad spectrum of applications. Importantly, it holds the crucial properties of non-negativity, identity of indiscernibles, symmetry, and the triangle inequality. This compliance ensures that our proposed distance metric offers a mathematically sound and reliable framework for comparing rankings, further enhancing its credibility.

By conducting comparative experiments, we illustrated the superior performance of our new drastic distance metric against established measures such as Hamming, Canberra, Bray-Curtis, Euclidean, Manhattan, and Chebyshev distances. The results showcased the new metric's enhanced sensitivity and ability to accurately quantify dissimilarities between rankings, making it a potent tool in the decision-making domain.

In conclusion, the drastic distance metric proposed in this work represents a significant advancement in the area of distance measurement for rankings. With its proven mathematical robustness and practical effectiveness, it has the potential to contribute significantly to decision-making processes across various fields. Future research directions could explore more extensive applications of this metric and further refine its potential through diverse real-world use cases.

Acknowledgment

References

[1] S.-S. Choi, S.-H. Cha, C. C. Tappert, *et al.*, "A survey of binary similarity and distance measures," *Journal of systemics, cybernetics and informatics*, vol. 8, no. 1, pp. 43–48, 2010.

[2] E. Deza, M. M. Deza, M. M. Deza, and E. Deza, *Encyclopedia of distances*. Springer, 2009.

[3] S. Chen, B. Ma, and K. Zhang, "On the similarity metric and the distance metric," *Theoretical Computer Science*, vol. 410, no. 24-25, pp. 2365–2376, 2009.

[4] M. Zhu, V. Lakshmanan, P. Zhang, Y. Hong, K. Cheng, and S. Chen, "Spatial verification using a true metric," *Atmospheric research*, vol. 102, no. 4, pp. 408–419, 2011.

[5] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean distance geometry and applications," *SIAM review*, vol. 56, no. 1, pp. 3–69, 2014.

[6] S. Pandit, S. Gupta, *et al.*, "A comparative study on distance measuring approaches for clustering," *International journal of research in computer science*, vol. 2, no. 1, pp. 29–31, 2011.

[7] R. Coghetto, "Chebyshev distance," *Formalized Mathematics*, vol. 24, no. 2, pp. 121–141, 2016.

[8] M. M. Deza, E. Deza, M. M. Deza, and E. Deza, "Distances and similarities in data analysis," *Encyclopedia of distances*, pp. 291–305, 2013.

[9] A. Bączkiewicz, J. Wątróbski, and W. Sałabun, "Distance metrics library for mcda methods," in *R. A. Buchmann, G. C. Silaghi, D. Bufnea, V. Niculescu, G. Czibula, C. Barry, M. Lang, H. Linger, C. Schneider (Eds.), Information Systems Development: Artificial Intelligence for Information Systems Development and Operations (ISD2022 Proceedings). Cluj-Napoca, Romania: Babeș-Bolyai University.*, pp. 1–8, 2022.

[10] P. Zhang, X. Wang, and P. X.-K. Song, "Clustering categorical data based on distance vectors," *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 355–367, 2006.

[11] H. A. Abu Alfeilat, A. B. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, H. S. Eyal Salman, and V. S. Prasath, "Effects of distance measure choice on k-nearest neighbor classifier performance: a review," *Big data*, vol. 7, no. 4, pp. 221–248, 2019.

[12] G. Glazko, A. Gordon, and A. Mushegian, "The choice of optimal distance measure in genome-wide datasets," *Bioinformatics*, vol. 21, no. Suppl_3, pp. iii3–iii11, 2005.

[13] R. Kumar and S. Vassilvitskii, "Generalized distances between rankings," in *Proceedings of the 19th international conference on World wide web*, pp. 571–580, 2010.

[14] W. Sałabun, J. Wątróbski, and A. Shekhovtsov, "Are mcda methods benchmarkable? a comparative study of topsis, vikor, copras, and promethee ii methods," *Symmetry*, vol. 12, no. 9, p. 1549, 2020.

[15] A. Karczmarczyk, J. Wątróbski, G. Ladorucki, and J. Jankowski, "Mcda-based approach to sustainable supplier selection," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 769–778, IEEE, 2018.

[16] D. K. Bukovšek and B. Mojškerc, "On the exact region determined by spearman's footrule and gini's gamma," *Journal of Computational and Applied Mathematics*, vol. 410, p. 114212, 2022.

[17] W. Sałabun and K. Urbaniak, "A new coefficient of rankings similarity in decision-making problems," in *Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part II 20*, pp. 632–645, Springer, 2020.

[18] C. Genest and J.-F. Plante, "On blest's measure of rank correlation," *Canadian Journal of Statistics*, vol. 31, no. 1, pp. 35–52, 2003.

[19] A. Shekhovtsov, V. Kozlov, V. Nosov, and W. Sałabun, "Efficiency of methods for determining the relevance of criteria in sustainable transport problems: A comparative case study," *Sustainability*, vol. 12, no. 19, p. 7915, 2020.

[20] W. Sałabun, A. Shekhovtsov, D. Pamučar, J. Wątróbski, B. Kizielewicz, J. Więckowski, D. Bozanić, K. Urbaniak, and B. Nyczaj, "A fuzzy inference system for players evaluation in multi-player sports: The football study case," *Symmetry*, vol. 12, no. 12, p. 2029, 2020.

[21] B. Bera, P. K. Shit, N. Sengupta, S. Saha, and S. Bhattacharjee, "Susceptibility of deforestation hotspots in terai-dooars belt of himalayan foothills: A comparative analysis of vikor and topsis models," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 8794–8806, 2022.

[22] M. Norouzi, D. J. Fleet, and R. R. Salakhutdinov, "Hamming distance metric learning," *Advances in neural information processing systems*, vol. 25, 2012.

[23] G. Jurman, S. Riccadonna, R. Visintainer, and C. Furlanello, "Canberra distance on ranked lists," in *Proceedings of advances in ranking NIPS 09 workshop*, pp. 22–27, Citeseer, 2009.

[24] E. W. Beals, "Bray-curtis ordination: an effective strategy for analysis of multivariate ecological data," in *Advances in ecological research*, vol. 14, pp. 1–55, Elsevier, 1984.

[25] M. Malkauthekar, "Analysis of euclidean distance and manhattan distance measure in face recognition," in *Third International Conference on Computational Intelligence and Information Technology (CIIT 2013)*, pp. 503–507, IET, 2013.