Multi-Layer Combinatorial Fusion Using Cognitive Diversity

LANDON HURLEY[1,2], **BRUCE S. KRISTAL**[3,4], **SUMAN SIRIMULLA**[5],
CHRISTINA SCHWEIKERT[6], **(Member, IEEE),**
AND D. FRANK HSU[7], **(Life Senior Member, IEEE)**

[1]Department of Biostatistics, School of Public Health, Yale University, New Haven, CT 06510, USA
[2]VA CSPCC, West Haven, CT 06516, USA
[3]Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
[4]Division of Sleep Medicine, Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[5]Department of Pharmaceutical Sciences, School of Pharmacy, The University of Texas at El Paso, El Paso, TX 79968, USA
[6]Division of Computer Science, Mathematics, and Science, St. John's University, Queens, NY 11439, USA
[7]Laboratory of Informatics and Data Mining, Department of Computer and Information Science, Fordham University, New York, NY 10458, USA

Corresponding authors: D. Frank Hsu (hsu@cis.fordham.edu) and Landon Hurley (landon.hurley@yale.edu)

ABSTRACT Multiple scoring systems (including rank and score functions; MSS) have been widely used in multiple regression, intelligent biometric systems, multiple artificial neural nets, combining pattern classifiers, ensemble methods, machine learning and artificial intelligence (AI), data and information fusion, preference ranking, and deep learning. Combining MSS has achieved numerous successful results in a variety of domain applications. However, the reasons why this happens remains an active area of investigation. Combinatorial fusion analysis (CFA) combines MSS using the rank-score characteristic (RSC) function and cognitive diversity (CD). The RSC function was proposed to characterise the predictive behaviour of a scoring system. It was subsequently used to define the notion of ''cognitive diversity'', which measures the dissimilarity in the representation of information between two scoring systems. In this article, we first examine characterizations of and diversity between scoring systems. Then, we review combinatorial fusion analysis with a variety of domain applications, including biometric systems in cognitive neuroscience, and joint decision making with visual cognitive systems. Finally, we demonstrate that multi-layer combinatorial fusion (MCF) on the Kemeny rank space is a viable machine learning and AI framework for preference ranking and reinforcement learning. This work provides a scientific foundation and technological insights for the use of Combinatorial Fusion in ensemble methods, data and information fusion, preference ranking, and deep reinforcement learning with applications to a variety of domains in data science and informatics for secure and sustainable societies.

INDEX TERMS Bubble-sort Cayley graph B_n, cognitive diversity (CD), combinatorial fusion analysis (CFA), intelligent biometric systems, Kemeny rank space H_n, multiple scoring systems (MSS), rank-score characteristic (RSC) function, symmetric group S_n.

I. INTRODUCTION

According to Jim Gray, in The Fourth Paradigm [2], the scientific discovery process has gone through three phases: (a) empirical (thousands of years ago), (b) theoretical (in the last few hundreds years), and (c) modelling and simulation (beginning about a century ago). Today's scientific inquiry and knowledge discovery is not only data-intensive but also data-centric, which requires the joint effort and fusion of the

The associate editor coordinating the review of this manuscript and approving it for publication was Marina Gavrilova.

three disciplines: mathematics, statistics, and computation, used in the previous three phases.

The current process of scientific discovery can be characterised as an inductive, rather than deductive, problem [3]. Early scientific discovery, which was also deductive empiricism, focused upon describing homogeneous, or universal, processes, which are used to validly predict singular scores in relation to specific characteristics. One early example of this technique was Galileo's experimental construction of the arc-length of a pendulum, which lead to the property of isochronism [4].

When the complexity of a system increases, the number of necessary parameters to describe the system also increases. Each parameter must be differentially weighted by the exhaustive enumeration across the set of possible points. Within the exhaustive space, scientists are able to determine the effect of X, a superset of the possible feature spaces, on Y, the target result of a function that we attempt to approximate. The expectation of performance is realised as a form of a conditional random variable. Hence, the uncertainty in predicting the state of a system is reduced as a consequence of the increased ability to predict how the system is expected to perform with the introduction of new data. Typical examples include scoring systems in biometric systems [5], [6], information retrieval [7], internet search, and search engine optimisation, as well as biomedical informatics, chemoinformatics, virtual screening and drug discovery [8], [9]. Two particular examples are the score function used by Google to determine the relevance of page presentation as preference learning [10], protein structure prediction, and protein-ligand interaction score functions in pharmaceutical settings [11]–[13].

Many problems in the science, technology, engineering and mathematics (STEM) areas and social-ecosystems emphasise the relevance, closeness, or similarity between two data measurement vectors of subjects, objects, or entities. Similarity is also a well-studied problem in the field of functional approximation, allowing us to view a system in a specific, quantitative and convex framework. Methods are not initially constructed to measure the explicit probability of success or the relevance to the target prototype that a score function s_A of a scoring system A was trained upon. The initial desired goal is rather to provide an optimal and/or desired ordering of the candidate data items that can serve as a proxy for preference ranking of the data items. This can be accomplished by sorting the implicit score values in descending (or ascending) order and assigning natural numbers to that rank order [14]. Our perspective depends upon the fusion of the mathematical, statistical and computational approaches, including the 'Two Cultures' as introduced by Breiman [15], wherein we treat the data models themselves as fundamental and necessary. However, each model has its strengths and weaknesses in the process of acquiring generalisable knowledge, as characterised by the maximisation of information by model fusion.

In this article, we focus on the RSC function f_A, which characterises the information-based relation between the scores and their orderings (ranks) of a scoring system A. Here we present a new paradigm that exhibits an analogy between the scoring system A on the data items $\mathcal{D} = \{d_1, d_2, \cdots, d_n\}$ with s_A and r_A as defined, and the variable $X(A)$, on the data set points in $\mathcal{D}$ with Euclidean score values $s_A(d_i)$ and the rank values $r_A(d_i)$ at the data point d_i. These rank values constitute a permutation of the set of natural numbers $\mathbb{N} = \{1, 2, 3, \cdots, n\}$. Current work provides a new paradigm for machine learning and artificial intelligence (ML/AI) with respect to general purpose deep reinforcement learning on the Kemeny rank space, which allows tied rankings. A variety

of diverse application domains are discussed, including data science, informatics, biomedicine and cheminformatics, data and information fusion, virtual screening and drug discovery, ensemble methods, information and cyber security, and model fusion.

II. BACKGROUND AND OUTLINE OF CURRENT WORK

When a score function, s_A, of a scoring system A within the candidate data space $\mathcal{D}$ is converted (sorted) into a rank function r_A, potential loss of information in the reduction from a measurable scoring space to a ranking space has historically been often discussed but little acted upon. When represented as a rectangular matrix, the data space $\mathcal{D}$ is of order $n \times t$, reflecting the row space where each of the n rows is a single data item, and t is the dimension of the system space. The early work of ordered similarity measures in non-parametric space, by [16], resulted in a widely applicable measurement of normed similarity, analogous to the Euclidean Pearson's r. However, the lack of a complete ordinal metric topology, especially in the presence of ties (data items d_i and d_j have the same value $s_A(d_i) = s_A(d_j)$), led to an inability to follow upon the early successes after the 1970's. Recently, rank aggregation, model fusion, or ensembles of model systems, have been applied many practical uses and demonstrations of utility at the cutting edge of mathematical development. These have produced, in a number of data mining competitions, academic disciples, and industrial sectors, strong accolades not as a result of the specific model construction techniques, but the ensembling techniques themselves [17]. When the available data grows in size and complexity, the need to operate upon the expanding domain of X in a computationally meaningful approach has resulted in a black box perspective with respect to the local function approximations. This situation is further complicated when multiple divergent systems are combined to produce singular approximations of Y.

A scoring system A on the data space $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$, consisting of a score function s_A and a derived rank function r_A, was proposed by Hsu *et al.* [1]. By sorting the values in the score function $s_A : \mathcal{D} \to \mathbb{R}$ in descending order, a rank function $r_A : \mathcal{D} \to \mathbb{N}$, where $\mathbb{N} = \{1, 2, 3, \ldots, n\}$, is obtained. The rank-score characteristic (RSC) function $f_A : \mathbb{N} \to \mathbb{R}$ in equation 1 upon scoring system A, was introduced in [1]

$$f_A(i) = s_A(r_A^{-1}(i)) = (s_A \circ r_A^{-1})(i). \tag{1}$$

Combinatorial fusion analysis [18], using multiple scoring systems and the RSC function, was subsequently developed and shown to be useful in a variety of domains [18]–[20]. These include text categorisation [21] and ranked versions thereof, protein structure prediction [13], [22], information retrieval [1], [23], target tracking and robotics [24], [25], motif detection [26], CHIP-seq peak detection [27], visual informatics [28], [29], cognitive neuroscience [30], virtual screening and drug discovery [8], [9], [31], deep learning [3], microarray analysis [32]–[34], data fusion [1], [7], and portfolio management [35], [36].

In what follows, Section III covers the characterisation and properties of the RSC function f_A of scoring system A in informatics and its counterpart, the cumulative distribution function (CDF) in statistics. Section IV covers correlation and diversity between scoring systems, focusing on three estimators of statistical data similarity: the Kolmogorov-Smirnov (KS) statistic, the Kullback-Leibler (KL) divergence, and the Cramér-von Mises statistics in light of the RSC function. Section IV also covers various computational information diversity measures as defined in machine learning and artificial intelligence (AI) [37]. In particular, $CD(A, B)$, which we define to measure the dissimilarity between two scoring inputs A and B, is compared to the notion of correlation (Pearson's correlation between $s_{X(A)}$ and $s_{X(B)}$, both Spearman's ρ and Kendall's τ rank correlation between $r_{X(A)}$ and $r_{X(B)}$) in statistics. In Section V, we review the approach of Combinatorial Fusion Analysis (CFA), which combines multiple scoring systems using the RSC function and CD. Finally, in Section VI, we discuss and demonstrate a recent advance in multi-layer combinational fusion (MCF) on the Kemeny rank space as opposed to the traditional deep learning practices on the Euclidean space.

III. CHARACTERISATION OF THE SCORING SYSTEMS

Given a scoring system A with score function s_A and its rank function r_A, the RSC function f_A, as shown in formula 1, can be derived. In addition, a scoring system A (see III-A) can be characterised w.r.t. statistical distribution and matrix (subsection III-B), permutation group (subsection III-C), and RSC function (subsection III-D).

A. SCORING SYSTEMS

An approximating model or scoring system A on items in a data set $\mathcal{D} = \{d_1, d_2, \cdots, d_n\}$, consists of a score function $s_A : \mathcal{D} \rightarrow \mathbb{R}$ which assigns a value $s_A(d_i)$ to the data item d_i. Given a score function s_A of the scoring system A, a rank function $r_A : \mathcal{D} \rightarrow \mathbb{N}$, where $\mathbb{N} = \{1, 2, 3, \cdots, n\} = [1, n]$, can be derived by sorting the score values $s_A(d_i)$, for all $d_i \in \mathcal{D}$, into descending order. Under this interpretation, a rank-score characteristic (RSC) function $f_A : \mathbb{N} \rightarrow \mathbb{R}$ was defined [1] as a mapping from the ordering of the score values upon the n score values with $f_A(i) = s_A(r^{-1}(i)) = (s_A \circ r_A^{-1})(i)$. The RSC function f_A of a scoring system A has three distinctive characteristics that we formulated as the following three remarks:

Remark 1: For a scoring system A with score function s_A and its derived rank function r_A, the rank-score characteristic (RSC) function $f_A : \mathbb{N} \rightarrow \mathbb{R}$ has either one of two possibilities: (a) f_A is a monotonically decreasing function, or (b) f_A is a non-increasing function. More specifically, case (a) consists of scoring systems without tied score values, while in case (b), tied score values (and hence tied rankings) are allowed.

Remark 2: Since the RSC function f_A for the scoring system A is from $\mathbb{N} \rightarrow \mathbb{R}$, it possesses the following two fundamental properties: (a) The removal of the data items in $\mathcal{D}$ as an
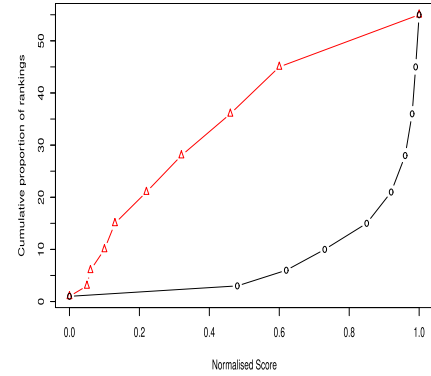

FIGURE 1. CDF function graphs for variables X_A and X_B in Table 2.

explicit middle transition step between the conversion of the scores to rankings leads to a function that is based on $\mathbb{N}$, not on D; and (b) For two scoring systems A and B on the same domain data set $\mathcal{D}$, the relationship between A and B defined using f_A and f_B is independent of the data set items d_i in $\mathcal{D}$.

Remark 3: For two scoring systems A and B on the dataset items $\mathcal{D}$, we define rank combination C as $s_C(d_i) = (r_A(d_i) + r_B(d_i))/2$ and score combination D as $s_D(d_i) = (s_A(d_i) + s_B(d_i))/2$, then under certain conditions including the different functions f_A and f_B, rank combination C is better than score combination D [1], [38].

Here we give one example. Let $D = \{d_1, d_2, \cdots, d_n\}$ be a set of n data items and $\mathbb{N} = [1, n]$ be the set of all integers from 1 to n. Let A and B be two scoring systems with score function s_A and s_B, and rank functions r_A and r_B, respectively. Let f_A and f_B be the rank-score characteristic (RSC) functions of A and B, respectively. Tables 2 and 3 list the normalised scores s_A and s_B, score combination $SC(A, B)$, rank combination $RC(A, B)$, and f_A and f_B, for the case $n = 10$, respectively. Figure 2 depicts the RSC function graph for f_A and f_B.

The RSC function f_A characterises the scoring system A, which is obtained by formula 1. It provides the relationship between the ranks and the score values of the scoring system A (See Figure 2 for RSC function graphs f_A and f_B, respectively). In what follows, we review three other different types of characterisations: the cumulative distribution function (CDF), the score matrix, and the permutation.

B. CDF & SCORE MATRIX

Treating normalised score function s_A of a scoring system A as a random variable x_A, we can construct a cumulative distribution function such that

$$F_A(x) = \int_{i=1}^{n} \mathbb{P}(r_A(d_i) \leq r_A(d_j)) \, di$$
$$s.t., i \leq j \in \{1, \ldots, n\} \quad (2)$$

Figure 1 depicts the CDF function graphs of X_A and X_B. There are several different ways of using matrices to characterise the scoring system A. First, the traditional Kendall's τ
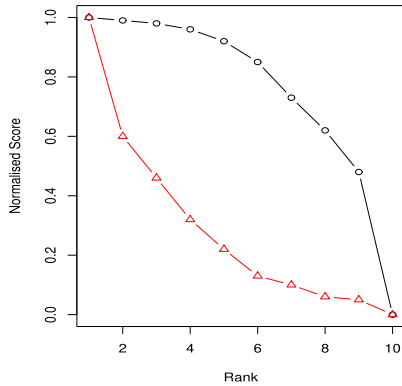
FIGURE 2. RSC function graphs for scoring systems A and B in Table 3.

FIGURE 3. The bubble-sort graph B_4. B_4 has connectivity = 3 [43].

FIGURE 4. There exists 3 disjoint paths between any two nodes, such as 1432 and 4123: P_1 = {1432, 4132, 4123}, P_2 = {1432, 1423, 4123}, and P_3 = {1432, 1342, 3142, 3412, 4312, 4321, 4231, 4213, 4123}, [43].

has the form for scoring system *A* without ties:

$$M_{\tau_a}(A) = M_{\tau_a}(a_{ij}) = \begin{cases} 1 & \text{if } d_i > d_j \\ -1 & \text{if } d_i < d_j \end{cases} \quad (3)$$

Then the extended Kendall's τ_b matrix representation which can handle ties [39]:

$$\tau_b(A, B) = \frac{\sum\sum a_{ij}b_{ij}}{\sqrt{\sum\sum a_{ij}^2 \sum\sum b_{ij}^2}} \quad (4)$$

$$M_{\tau_b}(A) = M_{\tau_b}(a_{ij}) = \begin{cases} 1 & \text{if } d_i > d_j \\ -1 & \text{if } d_i < d_j \\ 0 & \text{if } d_i = d_j \text{ or } i = j \end{cases} \quad (5)$$

In their study of consensus ranking problems, Emond and Mason [40] used the Banach inner-product nature of the Kemeny rank space τ_x, same as d_H in Section VI, to characterise the scoring system *A*:

$$\tau_x(A, B) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}b_{ij} \quad (6)$$

$$M_{\tau_x}(A) = M_{\tau_x}(a_{ij}) = \begin{cases} 1 & \text{if } d_i \geq d_j \\ -1 & \text{if } d_i < d_j \\ 0 & \text{if } i = j \end{cases} \quad (7)$$

This characterisation is similar to the description and proofs of [41], utilising the Borda count for Mixed Group Ranks fusion, with the combination function tending towards the expectation of the group, defined as the median of the set of scoring functions. The Borda count is, however, a non-compliant Condorcet technique, in that it is expected, but not guaranteed, that scoring systems should rank candidates as a linear combination that produces a one-to-one correspondence between the highest composite scores and the highest rankings. However, the Borda count fails to satisfy this criterion in certain specific conditions, wherein a lesser scoring candidate combination may be ranked higher than a higher scoring candidate. Our utilisation of the Kemeny metric (See Section VI), which is linearly invariant under monotonic transformation, allows our combinations to reduce
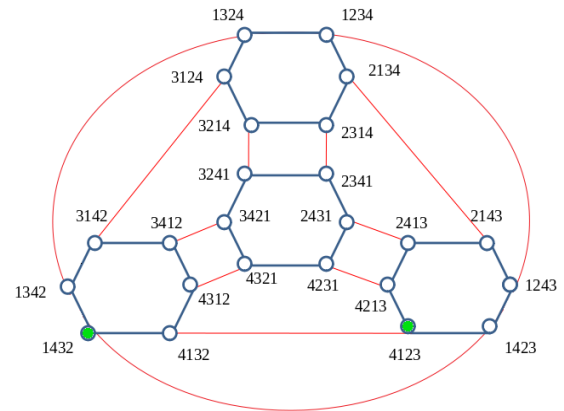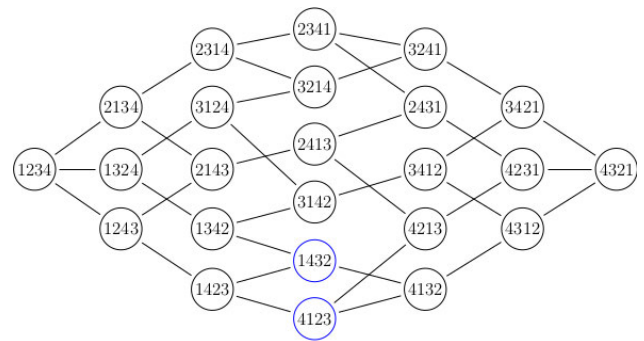
the potential for Condorcet failures occurring in the presence of ties upon scoring systems.

C. PERMUTATIONS

A rank function is really a permutation, with potential ties, of the data. The permutation unit distance under the Kendall metric, defined upon S_n, provides a combinatorial and computational framework in which distances are computed across the comparative number of swaps that are necessary in order to recover the target ordering [14], [42]. As the score matrix (equation 5 and equation 7), rather than the scores themselves, are the comparative basis upon which all unit evaluations (i.e., errors in prediction) are computed, the inclusion of a proper metric analysis, with ties, sub-additively decomposes the bias realised in [42]. This, in turn, enables construction of a linear function of the cumulative distribution function across the ranks of all *m* units that is optimal in the Gauss-Markov sense. Thus, independent of actual construction of the target score approximation, which is orderable, the permutation measure determines a compact and comparable measure space.

For example, consider Figure 3 and Figure 4, which depict the permutation ranking of $n = 4$ elements upon the complete symmetric permutation space. This restricts the score realisations to be monotonically invariant, such that the lowest rank denotes the highest score realisation upon the symmetric group space S_4 with adjacency defined as a swap of two

adjacent elements. In general, S_n with the metric Kendall's τ, is called a bubble-sort graph (B_n) [43]. In said graph (B_n), no tied rankings may occur. Figure 3 shows that B_4 can be constructed using 4 copies of B_3. Figure 4 depicts the layout of the $4! = 24$ nodes w.r.t. the distance from the identity permutation $I_n = 1234$ or its inverse $I'_n = 4321$.

D. RSC FUNCTION

Rank-score characteristic (RSC) function as defined by [1] is the function $f_A : \mathbb{N} \to \mathbb{R}$ that assigns the score value of a data item to the rank of that data item under the scoring system A. In mathematical terms, the composite function of s_A and r_A^{-1}, is constructed in the following way: $f_A(i) = s_A(r_A^{-1}(i)) = (s_A \circ r_A^{-1})(i)$. As discussed earlier in this section, the RSC function of a scoring system A characterises the scoring system A analogous to the role played by the cumulative distribution function (CDF) and scoring matrix ($M_{\tau_a}, M_{\tau_b}, M_{\tau_x}$, in Section III-B) that characterises Euclidean spaces with parametric Euclidean score values and non-parametric Kemeny rank values of the scoring variables x_A, respectively, in computation.

We note that given the score function s_A of a scoring system A, the RSC function f_A can be computed efficiently [38], [44]. More specifically, f_A is obtained by sorting the score values $\{s_A(d_i)|d_i \in \mathcal{D}\}$ using the rank value $[1, n] = \{1, 2, 3, \cdots, n\}$, with $n = |\mathcal{D}|$ as the key.

IV. DIVERSITY BETWEEN SCORING SYSTEMS

Given two scoring systems A and B on the same dataset items $\mathcal{D} = \{d_1, d_2, \cdots, d_n\}$, similarity (or dis-similarity) between A and B has been studied due to its importance in statistics and informatics. In statistics, correlation (Pearson) and rank correlation ([16], [39], [40], [44], [45], [49]) were studied. So were the Kolmogorov [53] and Smirnov [54], the Kullback and Leibler [55], and the Cramér [57] statistics.

In informatics, diversity has been studied in combining artificial neural networks [58], combining pattern classifiers [59], complex systems and ensemble methods [47], [60], [61], and combining multiple scoring systems [1], [18]. We illustrate more details in the following two subsections: statistical data correlation and computational information diversity, in the context of a pair of scoring systems A and B.

A. STATISTICAL DATA CORRELATION

For two scoring systems A and B, statistical correlation focuses on the correlation between data distributions of the random variables x_A and x_B. In this regard, we begin with the empirical Euclidean definition of distance for the Pearson correlation (equation 8) and compare it to the Spearman data correlation (equation 9),

$$\text{Pear}_{s_A, s_B} \propto \left(d = \frac{1 - r}{2} \right),$$

$$s.t., d = \sqrt{\sum_{i=1}^{n} \left(s_A(d_i) \cdot s_B(d_i) \right)^2}. \tag{8}$$

$$r_S(r_A, r_B) = 1 - \frac{6 \sum \left(r_A(d_i) - r_B(d_i) \right)^2}{n(n^2 - 1)}, \tag{9}$$

where in equation 9 the Spearman footrule is defined in which the squared difference between the rankings of the two scoring systems upon each of n observations, in addition to the previous metrics in equations 3,4, and 6.

When performing statistical tests within the rank space, the Kolmogorov-Smironov (KS) statistic [53], [54] is an easily understood measure. In the following discussion of statistical methods, we note the equivalence between $x = \{x_1, x_2, \ldots, x_n\} = \{s(d_1), s(d_2), \ldots, s(d_n)\}$. In the context of the RSC function for each of the scoring systems A and B [1], the KS statistic computes the supremum of the differences in ranks between two ranks f_A^{-1} and f_B^{-1} for each score value $x \in (0, 1)$:

$$\text{KS}(A, B) = \sup_x |F_A(x) - F_B(x)|, \tag{10}$$

where F_A, F_B, f_A and f_B are all the cumulative distribution functions of X_A and X_B (equation 2 and Figure 1) and RSC functions of A and B (equation 1 and Figure 2), respectively. Importantly, while the score functions may be non-linearly related to the feature spaces of interest, rankings are linearly monotonically invariant function spaces [62], allowing for more fundamental distributional characteristics to be defined without the embedding complexity otherwise necessary upon the Euclidean plane.

The Kullback and Leibler divergence [55] can also be understood in the context of the RSC function. The expression measures the divergence from the function approximation distribution X_B to the target X_A, weighted by the potential score difference in magnitude consequent from the choice of difference in Euclidean parametric formulates from B under A:

$$KL(A, B) = -\sum_X F_A(x) \log \left(\frac{F_B(x)}{F_A(x)} \right) \tag{11}$$

$$= -\sum_X f^{-1}(x) \log \left(\frac{f_B^{-1}}{f_A^{-1}} \right), \tag{12}$$

where F_A and $F_B : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ and $x \in \mathbb{R}_0^+; x \leq 1$. The Cramér-von Mises criterion [57] operates upon an identical domain as the KS-statistic, supplanting the ℓ_∞-norm with the ℓ_2-norm.

$$T = n \cdot \omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i - 1}{2n} - F(x_i) \right]^2 \tag{13}$$

with the same defined domain and range as the RSC function, from which it can be easily seen that the Cramér-von Mises statistic is equivalent to the total marginalised distance between the two scores under both functions over all data items in the sample, thus representing the total dissimilarity between a sequence of n scores for any empirical finite sample. It can be easily understood that the term

$$\frac{2i - 1}{2n}$$

TABLE 1. Statistical and computational diversity.

(a) Statistical Data Correlation	Range	(b) Computational Information Diversity	Range
(1) Pearson's correlation r [45]	$[-1, 1]$	(1) Q-statistics [46, 47]	$[-1, 1]$
(2) Kendall's τ_a [16]	$[-1, 1]$	(2) KW-variance [48, 47]	$[0, 1]$
(3) Spearman's ρ [49]	$[-1, 1]$	(3) Entropy [50, 47]	$[0, 1]$
(4) Kendall's τ_b [39]	$[-1, 1]$	(4) Generalised diversity [51, 47]	$[0, 1]$
(5) τ_x [40]	$[-1, 1]$	(5) Double-fault [52, 47]	$[0, 1]$
(6) KS-statistic [53, 54]	$[0, 1]$	(6) Coincident system failure [52, 47]	$[0, 1]$
(7) KL-divergence [55]	$(-\infty, 0]$	(7) E_{CC} [56, 47]	$[0, 1]$
(8) Cv [57]	$[0, 1]$		

in ascending order with respect to i reflects an intrinsic ordering which is present, as proven before, within any cumulative distribution function. The existence of the ratio $\frac{2i}{2n} - \frac{1}{2n}$ allows for the assigned probability of comparison to be a permanently increasing value whose asymptotic limit is bound to the support $(0, 1]$, purely as a function of n, the sample space. This direct orderability of both allows for a comparison to be made across common points of the area between the two curves, which thereby allows us to determine whether the proportion of the distribution function indexed upon the rank space, that is less than a given threshold, is roughly equivalent to the theoretical Euclidean object of comparison. It also allows for the first term to be treated as the rank score function, upon which the rank with respect to i and the aggregated score $1 - \frac{2i-1}{2n}$ up to point i possess a proportional equivalence. In the left of Table 1(a), we provide a number of different score based statistical data aggregations marginalising over n, which allow for the predictive similarity of bivariate scoring systems to be compared.

B. COMPUTATIONAL INFORMATION DIVERSITY

In informatics and in computational science, scoring systems (with score functions and/or rank functions) have been used extensively. In data mining, machine learning, and data and information fusion, the notion of "diversity" between two scoring systems has been widely discussed and used in the context of artificial neural nets [58], pattern classifiers [59], complex systems [61], ensemble methods [47], and multiple scoring systems [1], [63].

In [59], diversity in classification ensembling was compared to diversity in biology, software engineering, and statistical measures of relationship. Kuncheva reviewed pairwise measures and six non-pairwise measures and discussed the relationship between diversity and accuracy in the context of combining pattern classifiers. Reference [19] showed shared relationships between various diversity measures in multiple classifier systems, and [63], [64] established a diversity-performance relationship for majority voting and plurality voting in classifier ensembles.

Diversity in ensemble methods [47] among the individual learners has been considered as a fundamental issue and is crucial to the accuracy of the ensemble. Although many diversity measures have been proposed and developed in the field

of ensemble methods, the right formulation and measures for diversity have not been resolved yet [37], [47], [59], [65], [66].

In the field of complex systems, measuring diversity exists in three categories: diversity within a type, such as variates, diversity across types, such as entropy and attributes, and diversity of community composition such as population [60], [61]. Complex systems exist in a variety of domains, including ecological systems, economic systems, financial systems, political systems, and biological systems. In the right of Table 1(b), we provide a number of different approaches to computational information diversity, which allow for the predictive similarity of bivariate scoring systems to be compared.

C. COGNITIVE DIVERSITY

CD was proposed to measure the dissimilarity between two scoring systems A and B [1], [18]. More specifically, it was calculated using RSC functions f_A and f_B of A and B [67]. For example, cognitive diversity between A and B, $CD(A, B)$, is computed as in formula 14 [8], and utilised as depicted in Figure 5 to depict the relationships between A and B [67].

$$CD(A, B) = d(f_A, f_B) = \sqrt{\frac{1}{n^2 - n} \sum_{i=1}^{n} (f_A(i) - f_B(i))^2} \quad (14)$$

It was shown [38] that under certain conditions, rank combination performs better than score combination. These conditions for which better performant combinations are obtained in favour of larger cognitive diversity between A and B. More details on the notion of cognitive diversity and its domain applications can be found in [67].

D. EXAMPLES

In information retrieval systems, each of the two search algorithms uses similarity scoring systems A and B, respectively. Let $\mathcal{D}$ be the set of documents $\mathcal{D} = \{d_1, d_2, \cdots, d_n\}$. Tables 2 and 3 illustrate: score functions s_A and s_B; rank functions r_A and r_B; score combinations $SC(A, B)$ and rank combinations $RC(A, B)$, as well as RSC-functions f_A and f_B in the case of $n = 10$.

For the pair of scoring systems A and B as illustrated in Table 2, we list the diversity of A and B, $d(A, B)$ in the context of the following fifteen measurements related to this section.

TABLE 2. Score functions and Rank functions of *A*, *B*; Score combinations *SC*(*A*, *B*); and Rank combination *RC*(*A*, *B*).

		A		B		$SC(A, B)$	$SC(A, B)$	$RC(A, B)$	RC(A,B)
$\mathcal{D}$	$s_A(d_i)$	$r_A(d_i)$	$s_B(d_i)$	$r_B(d_i)$	$s_{SC}(d_i)$	$r_{SC}(d_i)$	$s_{RC}(d_i)$	$r_{RC}(d_i)$
d_1	0.73	7	0.13	6	0.43	7	6.5	5,6
d_2	0.99	2	0.32	4	0.655	3	3	3
d_3	0.98	3	1.00	1	0.49	4,5	2	2
d_4	0.85	6	0.05	9	0.45	6	7.5	7,8,9
d_5	1.00	1	0.60	2	0.80	1	1.5	1
d_6	0.00	10	0.22	5	0.11	10	7.5	7,8,9
d_7	0.96	4	0.46	3	0.71	2	3.5	4
d_8	0.62	8	0.10	7	0.36	8	7.5	7,8,9
d_9	0.92	5	0.06	8	0.49	5,4	6.5	5,6
d_{10}	0.48	9	0.00	10	0.24	9	9.5	10
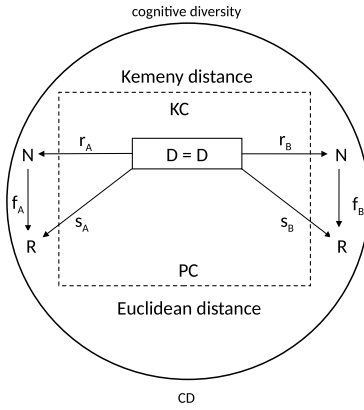
FIGURE 5. Scoring systems *A* and *B*: score functions s_A and s_B with corresponding rank functions r_A and r_B; RSC functions f_A and f_B; Pearson's score correlations (PC), Kemeny rank correlation (KC) and cognitive diversity (CD) (see [67] for ranking without ties).

See V-A and V-B. The first eight in Table 1(a) are in terms of statistical data correlations:

1) Pearson correlation $r = 0.4214$
2) Kendall-$\tau_a = 0.4667$
3) Kendall-$\tau_b = 0.4667$
4) Spearman $\rho = 0.6606$
5) $\tau_x = 0.4666667$
6) KS-statistic $D = 0.7$
7) KL-divergence $(x||y) = -17.0104$
8) Cramér-von Mises statistic $\omega^2 = 0.2803$

The second group, in Table 1(b), consists of seven computational information diversities [40], [46], [48], [50]–[52], [56], [65], [68]–[70]. For each measure, we also indicate the range of measurement and its reference in the table.

V. COMBINATORIAL FUSION ANALYSIS

In this section, we review the field of Combinatorial Fusion Analysis (CFA) proposed by [1], [18] for optimally combining multiple scoring systems. CFA was founded upon the concept of using the "rank-score characteristic (RSC) function," to characterise a predictive engine, and the concept of "cognitive diversity" was introduced to measure the dissimilarity between multiple scoring systems [1], [18], [67]. One of the novelties of the CFA approach is the use of various

combinatorial methods including both rank and score combination [18], [67] to produce predictive model combinations whose predictions are both more accurate and more robust to cross-validation [30].

Combining multiple scoring systems (MSS) has become an emerging field of machine learning, AI, forecasting, and predictive analytics. It has been shown to be useful in a variety of domains, such as combining pattern classifiers [59], weighted score and rank combination [18], [67], ensemble methods [47], data fusion in information retrieval [7], multi-indicator systems [70], and online learning algorithms [71].

Given t scoring systems $A_j, j = 1, 2, \cdots, t$ with score functions s_{A_j} and rank functions r_{A_j} respectively, we have either score combination $SC(A_1, A_2, \cdots, A_t) = SC(A_j^*)$ or rank combination $RC(A_1, A_2, \cdots, A_t) = RC(A_j^*)$. Allowing $SC = SC(A_j^*)$ and $RC = RC(A_j^*)$ to be the score combinations and rank combinations of $A_j^* = \{A_1, A_2, \cdots, A_t\}$, we have

$$s_{SC}(d_i) = \frac{1}{t} \sum_{j=1}^{t} (s(A_j)(d_i)), \quad i \in [1, n] \quad (15)$$

and

$$s_{RC}(d_i) = \frac{1}{t} \sum_{j=1}^{t} (r(A_j)(d_i)), \quad i \in [1, n] \quad (16)$$

as the score functions of $SC = SC(A_j^*)$ and $RC = RC(A_j^*)$, respectively. Performance of said scores are evaluated using the permutation distance between the target permutations and the model combination permutations, by means of the Kemeny metric (see Section VI). The Kemeny metric resolves the biased performance of the Kendall's τ_b distance as determined in [42], [43], allowing for a performance evaluation and probability distribution to be determined for any monotonically non-decreasing cumulative distribution of scores and score combinations.

A. METHOD AND PRACTICE OF COMBINING MULTIPLE SCORING SYSTEMS

Let $CFA(M; n, t)$ be the combinatorial fusion analysis (CFA) algorithm which combines t scoring systems on the dataset items $d_1, d_2, \cdots, d_n$, using the set of combination methods M: (1) average combination (AC), (2) weighted combination

TABLE 3. Rank-score characteristic (RSC) functions f_A and f_B with data items.

N	f_A	D	f_B	D
1	1.00	d_5	1.00	d_3
2	0.99	d_2	0.60	d_5
3	0.98	d_3	0.46	d_7
4	0.96	d_7	0.32	d_2
5	0.92	d_9	0.22	d_6
6	0.85	d_4	0.13	d_1
7	0.73	d_1	0.10	d_8
8	0.62	d_8	0.06	d_9
9	0.48	d_{10}	0.05	d_4
10	0	d_6	0	d_{10}
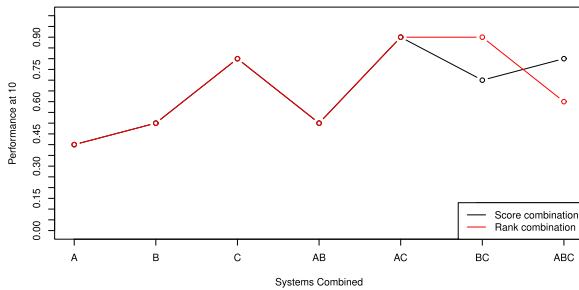
FIGURE 6. Performance of combinatorial fusion.

FIGURE 7. RSC function graphs for f_A, f_B, f_C, Table 4.

TABLE 4. Score function and rank function for scoring systems A,B, and C.

	Score function			Rank function		
	s_A	s_B	s_C	r_A	r_B	r_C
d_1	15.17	15.02	14.35	8	7	14
d_2	14.36	12.80	17.12	14	18	8
d_3	16.22	14.36	17.13	3	13	7
d_4	15.49	12.58	11.29	5	19	19
d_5	13.70	15.37	22.25	19	4	1
d_6	14.47	14.86	16.39	13	8	10
d_7	15.08	13.36	16.81	10	17	9
d_8	16.84	15.82	10.70	2	3	20
d_9	14.75	14.42	18.70	11	12	2
d_{10}	17.19	16.25	13.89	1	2	15
d_{11}	13.86	12.47	13.71	18	20	16
d_{12}	13.99	14.65	17.42	17	10	6
d_{13}	13.17	15.10	15.96	20	6	11
d_{14}	14.31	13.84	15.46	15	16	13
d_{15}	15.12	14.64	17.55	9	11	3
d_{16}	15.42	13.90	11.46	6	15	18
d_{17}	15.20	14.21	17.52	7	14	4
d_{18}	14.69	15.22	15.73	12	5	12
d_{19}	14.25	14.80	17.49	16	9	5
d_{20}	15.95	16.56	12.24	4	1	17

by performance (WCP) (3) weighted combination by diversity strength(WCDS), (4) geometric mean, (5) mixed group rank, and/or (6) other combination method, on the bubble-sort Cayley graph B_n and the Euclidean space $E_n = \mathbb{R}^n$ for rank combination and score combination, respectively. Each of these methods (1)–(4) has both rank and score combinations while method (5) has rank combination only.

Combinatorial fusion consists of a scoring system A on the domain set $\mathcal{D} = \{d_1, d_2, \cdots, d_n\}$, consists of a score function $s_A : \mathcal{D} \to \mathbb{R}$ which assigns a score (real numbers in $\mathbb{R}$) to each item d_i of the domain set $\mathcal{D}$ and a rank function $r_A : \mathcal{D} \to \mathbb{N}$ which assigns a rank (natural numbers in $\mathbb{N} : [1, 2, \cdots, n], n = |\mathcal{D}|$). The rank function $r_A(d_i)$ is the result of sorting the score values $s_A(d_i)$ in descending order and assigning rank to each data item accordingly. In Subsection V-B, we illustrate more details using two domain examples from cognitive neuroscience: preference detection based on eye movement in V-B1 [30], and joint decision making using visual cognitive systems, found in V-B2 [28].

B. DOMAIN EXAMPLES

We first illustrate the above discussion with an example in the information retrieval domain before we get to the two examples in V-B1 and V-B2. Let $\mathcal{D} = \{d_i\}$ be the domain space where $i \in [1, n]$, $n = 20$. Let A, B, C be three retrieval systems with score functions s_A, s_B, and s_C and rank functions r_A, r_B and r_C, respectively. Table 4 lists the score functions and rank functions. Table 5 gives the list of normalised score functions s'_A, s'_B, s'_C, along with respective ranks, while Table 6 lists the RSC functions f_A, f_B and f_C, respectively. Figure 7 plots the RSC-function graphs for f_A, f_B, f_C on the

same space. Figure 6 illustrates the performance of all combinations by rank or by score.

1) PREFERENCE DETECTION BASED ON EYE MOVEMENT

Cognitive neuroscience researchers are similarly often interested in modelling and understanding decision processes, which are reflected in neurological electrical activation and eye movement capture, for contexts such as text comprehension. In eye movement tracking [30], the 'gaze cascade' effect was observed, in which a subject is given two images and prompted to select their preference. The subject's variance in focus is initially evenly distributed, but the gaze focus is found to follow a Pareto distribution and converges as time increases, for which the focus modality corresponds to the preferred image. In the eye tracking trials, the x,y coordinates of the subjects' gaze locations, along with the duration of the gaze at the point, are collected. Based on this data, the following features were constructed: duration of focus in the last 200ms of a trial (A), total duration (B), gaze point count (C), interest sustainability (D), and count of edges between face regions (E).

Preliminary analysis found that, with 90% accuracy, the selection of the individual face deemed most attractive by the subject was determined by the visual focus in the

TABLE 5. Normalised scores for scoring systems *A*, *B*, and *C*.

	Normalised score functions					
	$s_{A'}$	$s_{B'}$	$s_{C'}$			
d_1	0.50	0.62	0.32	8	7	14
d_2	0.30	0.08	0.56	14	18	8
d_3	0.76	0.46	0.56	3	13	7
d_4	0.58	0.03	0.05	5	19	19
d_5	0.13	0.71	1.00	19	4	1
d_6	0.32	0.58	0.49	13	8	10
d_7	0.47	0.22	0.53	10	17	9
d_8	0.91	0.82	0.00	2	3	20
d_9	0.39	0.48	0.69	11	12	2
d_{10}	1.00	0.92	0.28	1	2	15
d_{11}	0.17	0.00	0.26	18	20	16
d_{12}	0.20	0.53	0.58	17	10	6
d_{13}	0.00	0.64	0.46	20	6	11
d_{14}	0.28	0.33	0.41	15	16	13
d_{15}	0.49	0.53	0.59	9	11	3
d_{16}	0.56	0.35	0.07	6	15	18
d_{17}	0.50	0.42	0.59	7	14	4
d_{18}	0.38	0.67	0.44	12	5	12
d_{19}	0.27	0.57	0.59	16	9	5
d_{20}	0.69	1.00	0.13	4	1	17

TABLE 6. RSC functions for scoring systems *A*, *B*, and *C* with data items $\in D$ from Table 5.

	RSC functions					
N	f_A		f_B		f_C	
	s_A	D	s_B	D	s_C	D
1	1.00	d_{10}	1.00	d_{20}	1.00	d_5
2	0.91	d_8	0.92	d_{10}	0.69	d_9
3	0.76	d_3	0.82	d_8	0.59	d_{15}
4	0.69	d_{20}	0.71	d_5	0.59	d_{17}
5	0.58	d_4	0.67	d_{18}	0.59	d_{19}
6	0.56	d_{16}	0.64	d_{13}	0.58	d_{12}
7	0.50	d_1	0.62	d_1	0.56	d_2
8	0.50	d_{17}	0.58	d_6	0.56	d_2
9	0.49	d_{15}	0.57	d_{19}	0.53	d_7
10	0.47	d_7	0.53	d_{12}	0.49	d_6
11	0.39	d_9	0.53	d_{15}	0.46	d_{13}
12	0.38	d_{18}	0.48	d_9	0.44	d_{18}
13	0.32	d_6	0.46	d_3	0.41	d_{14}
14	0.30	d_2	0.42	d_{17}	0.32	d_1
15	0.28	d_{14}	0.35	d_{16}	0.28	d_{10}
16	0.27	d_{19}	0.33	d_{14}	0.26	d_{11}
17	0.20	d_{12}	0.22	d_7	0.13	d_{20}
18	0.17	d_{18}	0.08	d_2	0.07	d_{16}
19	0.13	d_5	0.03	d_4	0.05	d_4
20	0.00	d_{13}	0.00	d_{11}	0.00	d_8

FIGURE 8. Rank-Score characteristic graph for gaze attributes [30].

intelligence as a consequence of CD. It has been observed that non-redundancy (i.e., dissimilarity) in well-performing scoring systems tends to improve overall system performance. This can be mathematically expressed by an algebraic distance formula. For example, the information, presented as a posterior probability distribution of the scores, and the corresponding integration over said space, thereby defining both the probability distribution function (PDF) and cumulative distribution function (CDF) of the two systems *A* and *B* are compared and combined. Then the redundant learned structure, or overlapping information, is removed from each system, producing a measure of the distribution of distinct information contained in these two systems, as reflected in the differences in respective predictions.

In joint decision making problems, a decision maker is often presented with various expert decisions, for which they must choose to combine together a subset of the recommendations to make a decision. In a real case experiment, the authors [28] selected two candidates, who observed a small projectile being tossed onto an open grassy spot. The observers were asked to indicate the location at which the token landed, and to provide their confidence (as measured by the size of radius of about their chosen location in which they were most confident the token lay) in their decision. For 96 distinct combinations of design factors, 34 cases were equal or better performant than the best decision of the two subjects. The results of this experiment are provided in Table 8, with the different weighting systems assigned to the confidence radius referred to as M_0, M_1 and M_2.

VI. MULTI-LAYER COMBINATORIAL FUSION ON THE KEMENY RANK SPACE

A. RANKING UPON THE BUBBLE-SORT CAYLEY GRAPH B_n

Each of the scoring systems *A* and *B* on data items $d_i \in \mathcal{D}$ represents a respective score function s_A and s_B and rank functions r_A and r_B. The rank function r_A(or r_B) : $\mathcal{D} \rightarrow \mathbb{N}$ with $r_A(d_i) \in [1, n]$ is the rank order of the data items d_i in $\mathcal{D}$. In other words, the rank function r_A, $r_A = [r_A(d_1), r_A(d_2), \cdots, r_A(d_n)]$, can be considered as a

final 200 milliseconds, which is represented by feature A. From this observation followed the construction of the five features recorded for each of 720 trials $T = \{t_1, t_2, \ldots, t_{720}\}$ (see features A, B, C, D, and E in Figure 8). Univariate and pairwise score combinations for $\binom{5}{2}$ of the features are recorded in Table 7, with precision recorded on the left-most column for each approach with the corresponding RSC function found in Figure 8.

2) JOINT DECISION MAKING USING VISUAL COGNITIVE SYSTEMS

In this section, we revisit the task of combining multiple scoring systems in exploring the generalisable computational

TABLE 7. Performance of feature scoring system combinations [30].

precision	A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
@100	.90	.83	.80	.51	.73	.95	.94	.98	.95	.82	.71	.82	.76	.82	.67
@200	.89	.78	.735	.535	.695	.945	.915	.965	.94	.78	.715	.765	.67	.745	.66
@300	.91	.747	.707	.503	.68	.913	.917	.913	.913	.743	.673	.717	.693	.697	.63

TABLE 8. Experimental results of 16 trials [28].

Trial	(a) Per(p,q)	(b) Conf.(p,q)	(c) Rank M0	M1	M2	p	q	M0	C	D	p	q	M1	C	D	p	q	M2	C	D
1	(28.1,17.6)	(12,6)	2	1	3	4	2	1	2	4	4	3	1	4	2	4	3	1	4	2
2	(16,5.7)	(8,6)	3	2	1	4	1	3	1	4	4	3	1	4	2	4	3	1	4	2
3	(101.7,47.4)	(18,10)	3	2	1	4	1	3	1	4	4	2	1	3	4	4	2	1	3	4
4	(21.6,34.5)	(10,16)	3	1	2	2	5	1	2	2	3	4	1	4	2	3	4	1	4	2
5	(27.4,23.1)	(14,16)	1	2	3	3	1	2	3	3	5	1	2	3	3	5	1	2	3	3
6	(133.3,169.1)	(16,12)	1	2	3	1	4	3	3	3	5	1	2	3	3	5	1	2	3	3
7	(18,25.2)	(16,8)	1	2	3	1	4	3	4	1	1	5	3	1	4	1	5	3	1	4
8	(42.8,20)	(12,6)	3	1	2	4	2	1	2	4	4	3	1	4	2	4	3	1	4	2
9	(5.1,16.1)	(12,12)	1	1	1	1	4	3	4	1	1	4	3	4	1	1	5	4	1	1
10	(242.2,267.1)	(8,10)	3	2	1	1	5	4	1	1	2	4	3	4	1	1	4	3	4	2
11	(96,26.5)	(20,16)	3	2	1	4	1	3	1	4	4	1	3	4	2	4	1	3	4	2
12	(7.2,9)	(10,8)	1	2	3	2	4	1	4	2	2	4	1	5	3	2	5	1	3	3
13	(9.2,15.2)	(10,8)	1	2	3	2	4	1	4	2	2	5	1	2	4	2	5	1	2	4
14	(23.8,32.6)	(10,8)	1	2	3	2	4	1	4	2	2	5	1	2	4	2	5	1	2	4
15	(21.1,15.8)	(18,12)	3	2	1	4	1	3	1	4	5	1	4	1	1	5	1	4	2	2
16	(10,19)	(10,10)	1	1	1	1	2	3	1	3	2	3	1	3	3	2	3	1	3	3

permutation of the data items $[i = 1, 2, \cdots, n]$. As such, r_A is an element of the symmetric group of order n (i.e., S_n) which consists of all permutations of the elements $[1, 2, \cdots, n]$. Let T_i be the operation which consists of any of the adjacent swaps $(i, i + 1)$, for $i = 1, \cdots, n - 1$. The symmetric group S_n together with the operations $T_i, i = 1, 2, \cdots, n - 1$, form a graph called the *bubble-sort Cayley graph*, $(S_n; T_i, i = 1, 2, \cdots, n - 1)$ where two nodes (permutations) A and B are adjacent to each other if and only if $A \circ T_i = B$ or $T_i \circ B = A$ for any $i \in [1, n - 1]$. It is straightforward to show that the distance between two nodes A and B is the number of adjacent interchanges between A and B.

The RSC function of a scoring system A builds upon the algebraic representation of the CDF, providing a comparative basis across different fields reflecting the range of the learning functions in which we are interested. RSC functions that are both ordered and scored identically are, unsurprisingly, identical functions. For two rank functions r_A and r_B as two permutation arrays of scoring systems A and B of the n elements $[1, n] = \{1, 2, 3, \cdots, n\}$, the following are equivalent:

1) distance $dist(A, B) =$ is the number of adjacency swaps between systems A and B;
2) Kendall's τ_a defined in Section III-B, formula 3;
3) the number of interchanges between adjacent elements from array A to array B, by using bubble-sort algorithm.

Although statements (1) and (2) are equivalent, they have different ranges of values. (1) is defined upon the support 0 to $(n(n - 1)/2)$, while (2) maps the distance of (1) onto $[-1, 1]$. If $dist(A, B) = x$, then $\tau_a(A, B) = y = 1 - \frac{2x}{n(n-1)}$. Likewise if $\tau_a(A, B) = y$, then $dist(A, B) = x = n(n - 1)/4(1 - y)$. Statement (3) indicates that the distance between node (permutation) A and node (permutation) B can be computed by sorting the array A to become B and counting the number of interchanges between adjacent elements.

The embedding of a metric space upon a graph allows for the visual depiction of relations between measurements and predictions independent of the specific nature of the relations. This graphical structure allows for a vast array of computational applications to be undertaken [72]–[74]. The domain of knowledge and information representation [75]–[77] allows for the network structure of the relations between nodes in the graph to exist without a necessary linear parametric score relation.

When performing either rank combination or score combination, the resulting scoring system may have tied scores among different data items. Hence the resulting rank function of the combinations are not permutations and hence not in the bubble-sort Cayley graph space $G = (S_n; T_i, i \in [1, n])$. In this regard, Kemeny and Snell [78] formalised the metric d_K as the distance between rank functions A and B using the τ_b characteristic score matrix (formula 4 in Section III-B) as follows:

$$\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij} - b_{ij}|, \qquad (17)$$

where the characteristic score matrix $M_{\tau_b}(A)$ is as defined in formula 4. It was shown by [78] that this metric in formula 17 satisfies a list of axioms which lead to a metric space H_n that includes rank functions of the scoring systems with ties. However, since this computation involves absolute values of differences, the application of this metric is hindered in contexts for which large data sets are conventional [40].

B. THE KEMENY RANK SPACE H_N

Emond and Mason [40] proposed formula 6 to characterise a scoring system A and then constructed an alternative representation of the correlation $\tau_x(A, B)$, as found in formula 6 with the score matrix in formula 7. They also showed by defining the concept of 'half-flip' that the Kemeny rank space H_n with $\tau_x(A, B)$ as the metric satisfies these axioms defined by [78], by leveraging the Banach inner-product nature of the ultrametric space. Figure 9 depicts the Kemeny rank space

TABLE 9. The number of nodes in H_n, $p(H_n)$, vs $n! = p(S_n)$ [80].

n	2	3	4	5	6	7	8	9	10
$p(H_n)$	3	13	75	541	4,883	47,293	545,835	7,087,261	102,247,563
$p(S_n)$	2	6	24	120	720	5,040	40,320	362,880	3,028,800

H_3 with half-flip as an unit of distance [78], [79]. We note that H_3, with 13 nodes and 18 edges, is an extension or a sup-imposed version of the graph S_3 which has 6 nodes and 6 edges. We further note that the number of nodes p in H_n is as follows [80]:

$$p(H_n) = \sum_{b=1}^{n} b! \binom{n}{b}, \tag{18}$$

where $\binom{n}{b}$ is the Stirling number of the second kind, which is the same as the number of ways to partition a set of n objects into b non-empty subsets. Table 9 lists the numbers $p(H_n)$ for $n = 2, \cdots, 10$.

C. MULTI-LAYER COMBINATORIAL FUSION (MCF) – THE ER-ALGORITHM

With the Kemeny rank space H_n and the matrix $d_H = \tau_x(A, B)$ in place, we aim to provide an evolutionary algorithm on H_n using multi-layer combinatorial fusion. We focus first on an expansion-reduction (ER) computational algorithm, which consists of three steps.

The CLEAR$(M; D; n, t, l)$ Framework: The framework uses the multi-layer combinatorial fusion (MCF) to conduct the expansion-reduction (ER)-algorithm on the Kemeny rank space H_n:

1) Expansion:
   a) Generate all the $2^t - t - 1$ combinations (2-com, 3-com,..., t-com) using a variety of combination methods and diversity measures (e.g., methods of combination $M = \{1, 2, 3, 4\}$ and diversity measurement $D = \{d_H, CD\}$ in Section V-A).
   b) Generate the mixed group rank combination [41].
   c) Each of the four methods of combination uses both weighted combination by performance and geometric mean combination [81]. In total, $8(2^t - t - 1)$ new scoring systems are obtained.

2) Reduction: Pick top q rank orders, $0 \leq q \leq 2t$, which are better than the t rank orders in previous step, from the $8(2^t - t - 1) + 1$ rank order obtained in Step 1.

   a) If q = 0, stop;
      If $0 < q < t$, go to Step 1;
      If $t \leq q$, go to Step 2.b;
   b) Calculate diversity, using either d_H or CD, between each pair of the $q \cdot (q - 1)/2$ pairs of rank orders;
   c) Calculate performance(using d_H and diversity strength (using either d_H or CD) of each of the q rank orders;
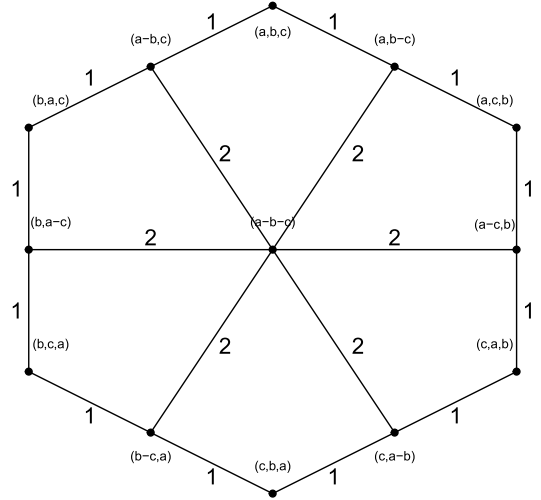
FIGURE 9. The rank space $H_3 = \mathrm{Cay}(H_3, \tau_x)$.

   d) Using the sliding rule to pick the top t rank orders that has the highest performance and diversity strength.
3) Go to Step 1.
4) l is the number of iterations of the expansion process in Step 1.

The CLEAR$(M; D; n, t, l)$ framework works as follows. Starting with t rank orders (with ties allowed) in H_n, we begin with expansion, generating $e(t) = 2 \times (2^t - 1 - t) + 1 = 2^{t+1} - 2t - 1$ rank orders, by three combination methods:

1) $2^t - 1 - t$ weighted combinations using performance as weights [63]
2) $2^t - 1 - t$ combinations by geometric mean [81]
3) combination by mixed group rank [41].

The second step in the CLEAR framework, reduction, first selects the top $q(\sim 2t)$ performing subset from the $e(t)$ rank orders, computes the "diversity" between every pair of these $(q(q-1))/2$ elements, and calculates the "diversity strength" for each of the q rank orders by the average of the CD between this rank order and other $q-1$ rank orders. Step 2 produces the top t rank orders from the two q rank orders using the "sliding rule" upon the joint performance and diversity strength. The reinforcement ER-algorithm continues as a multi-layer combinatorial fusion process in l number of times (layers) until one of the following stopping rules is met:

1) none of the top $q \sim 2t$ rank orders in the expanded set of rank orders is better than the best of the initial t rank orders, or
2) the expanded set $e(t)$ rank orders converges to less than t rank orders.

TABLE 10. H_{10} simulated Case with reported permutation performance [43].

Group	Index	Input	Input	Input	Input	Input	Input	Output	Layer	Time (sec.)
G_1	73,74,75,76,77,78	0.489	0.444	0.489	0.444	0.467	0.489	0.0889	5	5.93
G_2	7,32,57,82,107,132	0.489	0.489	0.467	0.467	0.444	0.444	0.0889	4	7.68
G_3	16,41,66,91,116,141	0.489	0.489	0.467	0.467	0.489	0.444	0.1333	4	6.90
G_4	49,50,51,52,53,54	0.444	0.467	0.467	0.467	0.444	0.489	0.2000	4	4.40
G_5	25,26,27,28,29,30	0.467	0.489	0.467	0.444	0.444	0.467	0.2000, 0.2222, 0.2444	3	5.50
G_6	15,40,65,90,115,140	0.444	0.467	0.489	0.467	0.467	0.467	0.2667, 0.2889, 0.3111	2	2.61
G_7	133,134,135,136,137,138	0.444	0.489	0.489	0.489	0.444	0.489	0.3111,0.3333	2	2.30
G_8	8,33,58,83,108,133	0.489	0.489	0.489	0.489	0.444	0.444	0.4000, 0.4222	1	1.47

Diversity between a pair of two rank orders A and B, $d(A, B)$ can be calculated either by the distance between A and B on the Kemeny rank space, $d(A, B) = d_H(A, B)$, or by the CD between the RSC functions, $d(A, B) = d(f_A, f_B)$. For the set of top q rank orders from the $e(t)$ expanded rank orders, diversity for each of the $q(q-1)$ pairs of rank orders can be obtained. Diversity strength of each of the q rank orders $ds(A)$ is calculated as the number of pairs on which A appears from the relatively top (meaning there is a sharp drop, or scree) pairs with high diversity. The q rank orders can be listed in two columns using performance and diversity strength in descending order. A sliding rule is placed on each row until t rank orders are obtained. In the ER-algorithm process, performance is meant to be how close the rank order is to the identity permutation by calculating the d_H distance. In case of a tie, the median rank is assigned to all three rank positions.

This reinforcement learning algorithm is similar to an evolutionary computational algorithms. An initial population of scoring functions is generated, upon which the predictive performance is then evaluated. The top performant rank orders are selected.

Optimisation under any complete metric sub-space enables linear transitive (or sub-additive) convexity to hold. This allows for convex optimisation upon the entire space to be established for any well-posed problem space with a unique solution, with all requisite maximum likelihood properties (e.g., minimum error and maximum information; [14]). As long as the variance-covariance matrix (reflective of the statistical correlational matrix) is positive-definite, it follows that the optimal combination of the learning systems, as proposed in this system, is expected to be linearly convex and well-posed, and is therefore resistant to over-fitting in the mathematical sense, regardless of the parametric realisations of the score distributions themselves in the Euclidean sub-space.

D. SIMULATED AND EMPIRICAL EXAMPLES

An exploration of the CLEAR model was reported in Zhong *et al.* (2019) [43]. It presented two simulations of MCF in H_{10} and H_{300} and an empirical application developed upon a protein-ligand virtual screening paper using deep-learning neural networks [12] to demonstrate the conceptual framework developed and expanded in the CLEAR framework. These two simulated cases H_{10} and H_{300} and the empirical example in virtual screening and drug discovery are included in Tables 10,11, and 12 [43]. Specifically, we demonstrated how the recursive ultrametric structure of the graph depicted in Figure 4, providing a convex and relatively quickly solved framework from which results a marked improvement of performance which surpasses the initial input models. Here, Index denotes the 300 unique identifiers corresponding to the neural-network systems provided by [12], and the Input denotes the system error for each neural-network system, or their aggregate performance under linear combination. These conventions hold for Tables 10,11, and 12, with each numerical value denoting the total error variance, with improved system performance reflected as the value approaches 0 from the right.

For both H_{10} and H_{300}, 125 rank orders upon the bubble-sort Cayley graph were selected with rank distance d_H greater than 40% and less than 50%, upon which 150 cases were selected. These performance bounds were selected to ensure that the worst case performance be explored, bounded by error rates of between 40% and 50%. They would allow us to characterise the worst-case performance improvement observed within our Fusion framework, relative to I_n. The empirical example utilised the 12 top performant functions reported from the 100 in [12], which are all no less than 87% accurate in performance, as they are the results of a deep learning computing experiment.

In the rank space H_{10}, the 150 rank orders were partitioned into 28 groups of 6 models each, with few duplications of the models across all 28 groups. The initial distances from the input models were found to be 44.4% – 48.9% accurate, while producing results in [8.89%, 42.22%] achieved in between 1 and 5 layers with computational time for all cases (sec) in [1.47, 7.68]. These results for H_{10} are found in Table 10. Upon the rank space H_{300} the 150 rank orders were composed into 50 groups of 6 each, with some duplication. From this set of 50 groups, 5 were subset as given in Table 11, for which the distances of the initial 6 were within [49.949%, 49.996%] distance performance, resulting in an output distance of [46.253%, 48.495%] for either two or three layers with computational time spent (min) in [20.29, 31.64], as provided in Table 11.

TABLE 11. H_{300} simulated Case with reported permutation performance [43].

Group	Index	Input	Input	Input	Input	Input	Input	Output		Layer	Time (sec.)
G_1	19,20,21,22,23,24	0.49996	0.49993	0.49971	0.49973	0.49953	0.49953	0.46253, 0.46462, 0.46537,	0.46446, 0.46524, 0.46540	3	22.90
G_2	13,14,15,16,17,18	0.49996	0.49996	0.499973	0.49971	0.49953	0.49953	0.47061, 0.47374, 0.47737,	0.47166, 0.47670, 0.47819	2	25.14
G_3	25,26,27,28,29,30	0.49993	0.49996	0.49971	0.49976	0.49953	0.49949	0.47458, 0.48495	0.48486,	2	20.29
G_4	1,2,3,4,5,6	0.49993	0.49993	0.49973	0.49973	0.49949	0.49951	0.47594, 0.47690	0.47679,	3	20.98
G_5	7,8,9,10,11,12	0.49996	0.49993	0.49971	0.49971	0.49951	0.49949	0.48370, 0.48375		3	31.64

TABLE 12. Empirical case of virtual screening [43].

Group	Input	Input	Input	Input	Input	Input	Output
$G_1(B6)$	0.126466	0.129164	0.130903	0.131037	0.131371	0.132308	0.117659 0.118149, 0.118172, 0.118239, 0.118305
G_2 (W6)	0.132486	0.132642	0.133289	0.133846	0.133846	0.134448	0.118707
G_3 (R6A)	0.132308	0.131371	0.133846	0.132486	0.131037	0.130903	0.117793, 0.117837, 0.119576, 0.119755, 0.119777, 0.119866
G_4(R6B)	0.131037	0.132642	0.134448	0.132486	0.130903	0.133846	0.119086, 0.119420, 0.119955, 0.120045, 0.120201, 0.120223
G_5 (R6C)	0.131037	0.133846	0.131371	0.129164	0.132642	0.134448	0.11922

The empirical case containing the top 12 performant models were found to possess a distance d_H from the target in [12.6466%, 13.4448%], from which 5 groups of 6 models were chosen to reflect the best 6, worst 6, and a random subset of six. These produced results of combination which were of distance to the target d_H in [11.766%, 12.022%], with only two or three necessary layers which were found to converge (in mins.) in [39.64, 47.74] as reported in Table 12.

It is important to note from these three examples that even though the decrease in performance percentage from 40.01% in H_{10} to 3.74% in H_{300} is substantial, the number of improved steps is quite high as the number of nodes in H_{300} is much larger than those in H_{10}. On the other hand, we also note that systems which have better performance converge to the target much faster because they possess less cumulative redundancy in the convex cone which represents the performance for each approximation. As the performance increases, the relative diversity between said functions must necessary decrease, thereby limiting the available uniqueness that results upon the combinations.

VII. CONCLUDING REMARKS

Combinatorial fusion analysis (CFA) [18], uses the rank-score characteristic (RSC) function [1] and cognitive diversity (CD) [1], [18], [67] to combine MSS with both score combinations (in Euclidean space) and rank combinations (in the bubble-sort Cayley graph based on the symmetric group S_n). By extending the B_n to the Kemeny rank space H_n which allows tied rankings and using the computationally efficient τ_x distance metric [39], [40], [78], [79], multi-layer combinatorial fusion (MCF) provides a robust computational and combinatorial framework on the metric space H_n

using RSC-function based cognitive diversity. Section III contrasts the RSC function in data science and informatics with the empirical CDF in statistics. Section IV compares CD in data science and informatics with other data correlations in statistics and information diversity in computation. Section V reviews the field of CFA and illustrates two intelligent biometric systems among a variety of domain applications. Section VI provides the multi-layer combinatorial fusion (MCF) framework CLEAR$(M; D; n, t, l)$ on H_n with two simulated examples on H_{10} and H_{300} and an empirical example on protein-ligand virtual screening and drug discovery in H_{300}. In the following, we summarise three distinctive characteristics of the MCF approach (A) and suggest several directions for future work (B).

A. MCF ON THE KEMENY RANK SPACE

Multi-layer combinatorial fusion (MCF) using cognitive diversity has the following distinct features worthy of special attention:

1) MCF uses the combinatorial fusion analysis (CFA) [1], [18], [43], which has the following characteristics:

    a) It considers a scoring system A as both a score function s_A in the Euclidean space $\mathbb{R}^n$ and the derived rank function r_A in the bubble-sort Cayley graph space B_n. RSC function

$$f_A(i) = (s_A \circ r_A^{-1})(i) = s_A(r^{-1}(i))$$

    was defined to characterise the scoring system A.

    b) It combines scoring systems A and B in both score and rank combinations. It was shown that under certain conditions (involving cognitive diversity)

rank combination can perform better than score combination [38].

c) Cognitive diversity between scoring systems A and B is defined using RSC functions f_A and f_B. Empirical results on consensus scoring were very useful for improving performance in virtual screening and drug discovery [8].

d) CFA provides combinatorial fusion with ($2^t - t - 1$ combinations for t original scoring systems) for any combination methods, which are efficient and effective in different domain applications in protein-structure predictions [13], stress identification [20], text categorisation [21], target tracking, robot mapping, and localisation [24], [25], identification of degenerated motif [26], CHIP-Seq peak detection [27], combining visual cognitive systems [28], [29], preference detection using eye movement [30], virtual screening [8], microarray gene expression [32]–[34], portfolio management [35], [36], classifier ensemble [63], [64], and online learning [71].

2) MCF is based on the Kemeny rank space H_n which has the following characteristics:

a) It is a natural extension from the symmetric group S_n and the bubble-sort Cayley graph B_n [43], [73].

b) The nodes in H_n include rank orders with ties. This property facilitates advances in the MCF process when ties are the result of combinations.

c) It provides a distance metric d_H which uses the computationally efficient τ_x function.

d) The structure of a network can dictate and affect function and application of the network. H_n has S_n as its sub-network, and B_n as its functional subspace. The structure of S_n has been well studied [82]–[84]. The bubble-sort Cayley graph B_n has many combinatorial properties including combination of mutually independent Hamiltonian cycles [84]. The symmetric group of order n, S_n, has many useful combinatorial structures. For example, the connectivity of S_3 is 3 and there are 3 disjoint paths between any pair of nodes in the graph, as shown in Figure 4. See [83] for general study of container width and length in a variety of graphs and groups where a container between two nodes A and B is a set of disjoint paths between A and B.

3) The expansion property of B_n:

a) Since the Kemeny rank space H_n is a metric space which embeds the bubble-sort Cayley B_n as a subspace, structure and properties of the graph B_n can facilitate the ER-algorithm in the CLEAR(M; D; n, t, l) framework. In fact, the graph B_n has some good expansion properties (see [85] for expander graphs). For example, it was shown [86] that the second eigenvalue of B_n,

$\lambda_2(B_n)$, is at most 1 which is related to certain expansion coefficients.

B. FUTURE WORK

A direction of our future work will focus on applying the CLEAR framework to a variety of domain applications on the Kemeny rank space H_n with distance metric d_H. In this regard, Heiser and D'Ambrosio obtained results on clustering and prediction of ranking using d_H [87]. Beyond improvement in the simulated cases of H_{10} and H_{300}, Tables 10 and 11, Table 12 exhibits a successful case, in the empirical domain of virtual screening and drug discovery, of the multi-layer combinatorial fusion (MCF) and deep reinforcement learning [12], [43].

Other than the work on various domain applications, we will also work on problems which require multidisciplinary approaches to fundamental methods and intelligent systems, including intelligent biometric systems and computational intelligent systems, blending together the three foundational fields for data science and informatics: statistics, mathematics (including combinatorics and graph theory), as well as computing and informatics (including machine learning and AI) [15], [18], [22], [40], [43], [50], [63], [65], [67], [70], [73], [76], [78], [86]. Recent results by experts from diverse fields have made significant contributions, for example, in the following areas: multi-modal biometric systems using rank level fusion for security systems [5], [6], fusion of deep learning and combinatorics [88], emphasising the fusion of computer hardware and software, global network architecture and web systems, proactive and reactive investigation, and public-private collaboration in mitigating cyber attacks and cyber exploitations [89], calculating the thermodynamic limit for the Mallows model on the bubble-sort Cayley graph B_n [90], harnessing fuzzy logic and combinatorial fusion to make network selection simple and effective in the heterogeneous mobile user environment [91], and using model fusion algorithms for neural networks with optimal transport [92]. Equipped with a robust framework such as CLEAR, using multi-layer combinatorial fusion (MCF) and cognitive diversity (CD), on a fundamental metric space such as the Kemeny rank space H_n using the distance metric d_H [1], [18], [38], [40], [67], [78], [82], more exciting results will be forthcoming to the benefit of secure, healthy, and sustainable societies.

REFERENCES

[1] D. F. Hsu, J. Shapiro, and I. Taksa, "Methods of data fusion in information retrieval: Rank vs. Score combination," DIMACS, Rutgers Univ., Piscataway, NJ, USA, Tech. Rep. DIMACS TR 2002-58, 2002.

[2] T. Hey, S. Tansley, and K. Tolle, "Jim Gray on e-science: A transformed scientific method," in *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, 2009, pp. xvii–xxxi.

[3] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 3rd ed. New York, NY, USA: Springer, 2013.

[4] P. Palmieri, "A phenomenology of Galileo's experiments with pendulums," *Brit. J. Hist. Sci.*, vol. 42, no. 4, pp. 479–513, Dec. 2009.

[5] M. M. Monwar and M. L. Gavrilova, "Multimodal biometric system using rank-level fusion approach," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 39, no. 4, pp. 867–878, Aug. 2009.

[6] M. L. Gavrilova and M. Monwar, *Multimodal Biometrics and Intelligent Image Processing for Security Systems*. Information Science Reference, 2013.

[7] S. Wu, *Data Fusion in Information Retrieval*. Berlin, Germany: Springer-Verlag, 2012.

[8] J.-M. Yang, Y.-F. Chen, T.-W. Shen, B. S. Kristal, and D. F. Hsu, "Consensus scoring criteria for improving enrichment in virtual screening," *J. Chem. Inf. Model.*, vol. 45, no. 4, pp. 1134–1146, Jul. 2005.

[9] C. M. Ginn, P. Willett, and J. Bradshaw, "Combination of molecular similarity measures using data fusion," in *Virtual Screening: An Alternative or Complement to High Throughput Screening?* Norwell, MA, USA: Kluwer Academic, 2000, pp. 1–16.

[10] J. Fürnkranz and E. Hüllermeier, Eds., *Preference Learning*. Berlin, Germany: Springer, 2011.

[11] M. Shokouhi, "Segmentation of search engine results for effective data-fusion," in *Advances in Information Retrieval* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2007, pp. 185–197.

[12] M. Hassan, D. C. Mogollon, O. Fuentes, and S. Sirimulla, "DLSCORE: A deep learning model for predicting protein-ligand binding affinities," *ChemRxiv*, 2018, doi: 10.26434/chemrxiv.6159143.v1.

[13] K.-L. Lin, C.-Y. Lin, C.-D. Huang, H.-M. Chang, C.-Y. Yang, C.-T. Lin, C. Y. Tang, and D. F. Hsu, "Feature selection and combination criteria for improving accuracy in protein structure prediction," *IEEE Trans. Nanobiosci.*, vol. 6, no. 2, pp. 186–196, Jun. 2007.

[14] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York, NY, USA: Springer-Verlag, 1998.

[15] L. Breiman, "Statistical modeling: The two cultures (with comments and a rejoinder by the author)," *Statist. Sci.*, vol. 16, no. 3, pp. 199–231, Aug. 2001.

[16] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, nos. 1–2, pp. 81–93, Jun. 1938.

[17] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*, vol. 2, no. 1. New York, NY, USA: Springer, 2013.

[18] D. F. Hsu, Y.-S. Chung, and B. S. Kristal, "Combinatorial fusion analysis: Methods and practice of combining multiple scoring systems," in *Advanced Data Mining Technologies in Bioinformatics*. Hershey, PA, USA: Idea Group Inc, 2006, pp. 32–62.

[19] Y.-S. Chung, D. F. Hsu, and C. Y. Tang, "On the relationships among various diversity measures in multiple classifier systems," in *Proc. Int. Symp. Parallel Archit., Algorithms, Netw. (i-span)*, May 2008, pp. 184–190.

[20] Y. Deng, D. F. Hsu, Z. Wu, and C.-H. Chu, "Feature selection and combination for stress identification using correlation and diversity," in *Proc. 12th Int. Symp. Pervasive Syst., Algorithms Netw.*, Dec. 2012, pp. 37–43.

[21] Y. Li, D. F. Hsu, and S. M. Chung, "Combination of multiple feature selection methods for text categorization by using combinatorial fusion analysis and rank-score characteristic," *Int. J. Artif. Intell. Tools*, vol. 22, no. 2, Apr. 2013, Art. no. 1350001.

[22] T. K. Ho, *Multiple Classifier Combination: Lessons and Next Steps* (Series in Machine Perception and Artificial Intelligence). Singapore: World Scientific, 2002, pp. 171–198.

[23] Y. Li, N. Shi, and D. F. Hsu, "Fusion analysis of information retrieval models on biomedical collections," in *Proc. 14th Int. Conf. Inf. Fusion (FUSION)*, Chicago, IL, USA, Jul. 2011, pp. 1–8.

[24] D. M. Lyons, D. F. Hsu, Q. Ma, and L. Wang, "Selection of fusion operations using rank-score diversity for robot mapping and localization," in *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2007*, B. V. Dasarathy, Ed. Bellingham, WA, USA: SPIE, 2007.

[25] D. M. Lyons and D. F. Hsu, "Combining multiple scoring systems for target tracking using rank–score characteristics," *Inf. Fusion*, vol. 10, no. 2, pp. 124–136, Apr. 2009.

[26] C.-H. Peng, J.-T. Hsu, Y.-S. Chung, Y.-J. Lin, W.-Y. Chow, D. F. Hsu, and C. Y. Tang, "Identification of degenerate motifs using position restricted selection and hybrid ranking combination," *Nucleic Acids Res.*, vol. 34, no. 22, pp. 6379–6391, Dec. 2006.

[27] S. Brown, D. F. Hsu, C. Schweikert, and Z. Tang, "ChIP-seq analytics: Methods and systems to improve ChIP-seq peak identification," in *Systems Biology*. Singapore: World Scientific, 2012, pp. 87–112.

[28] A. Batallones, K. Sanchez, B. Mott, C. Coffran, and D. Frank Hsu, "On the combination of two visual cognition systems using combinatorial fusion," *Brain Informat.*, vol. 2, no. 1, pp. 21–32, Mar. 2015.

[29] C. Schweikert, D. Mulia, K. Sanchez, and D. F. Hsu, "The diversity rank-score function for combining human visual perception systems," *Brain Informat.*, vol. 3, no. 1, pp. 63–72, Mar. 2016.

[30] C. Schweikert, S. Shimojo, and D. F. Hsu, "Detecting preferences based on eye movement using combinatorial fusion," in *Proc. IEEE 15th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI CC)*, Aug. 2016, pp. 336–343.

[31] C. Bissantz, P. Bernard, M. Hibert, and D. Rognan, "Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets?" *Proteins, Struct., Function, Bioinf.*, vol. 50, no. 1, pp. 5–25, Nov. 2002.

[32] C. McMunn-Coffran, C. Schweikert, and D. F. Hsu, "Microarray gene expression analysis using combinatorial fusion," in *Proc. 9th IEEE Int. Conf. Bioinf. BioEng.*, J. J. P. Tsai, P. C. Sheu, and H. C. W. Hsiao, Eds. Taichung, Taiwan: IEEE Computer Society, Jun. 2009, pp. 410–414, doi: 10.1109/BIBE.2009.70.

[33] H.-Y. Chuang, H. Liu, S. Brown, C. McMunn-Coffran, C.-Y. Kao, and D. F. Hsu, "Identifying significant genes from microarray data," in *Proc. 4th IEEE Symp. Bioinf. Bioeng.* Taichung, Taiwan: IEEE Computer Society, Mar. 2004, pp. 358–365, doi: 10.1109/BIBE.2004.1317366.

[34] H.-Y. Chuang, H. Liu, F.-A. Chen, C.-Y. Kao, and D. F. Hsu, "Combination methods in microarray analysis," in *Proc. 7th Int. Symp. Parallel Archit., Algorithms Netw.* Hong Kong, SAR, China: IEEE Computer Society, 2004, pp. 625–630, doi: 10.1109/ISPAN.2004.1300548.

[35] H. D. Vinod, D. F. Hsu, and Y. Tian, "Combining multiple criterion systems for improving portfolio performance," Dept. Econ., Fordham Univ., New York, NY, USA, Fordham Econ. Discuss. Paper Ser. dp2008-07, 2008.

[36] X. Wang, J. Ho-Shek, D. Ondusko, and D. F. Hsu, "Improving portfolio performance using attribute selection and combination," in *Pervasive Systems, Algorithms and Networks*. Naples, Italy: Springer, 2019, pp. 58–70.

[37] E. K. Tang, P. N. Suganthan, and X. Yao, "An analysis of diversity measures," *Mach. Learn.*, vol. 65, no. 1, pp. 247–271, 2006.

[38] D. Frank Hsu and I. Taksa, "Comparing rank and score combination methods for data fusion in information retrieval," *Inf. Retr.*, vol. 8, no. 3, pp. 449–480, Jan. 2005.

[39] M. G. Kendall, "The treatment of ties in ranking problems," *Biometrika*, vol. 33, no. 3, pp. 239–251, 1945.

[40] E. J. Emond and D. W. Mason, "A new rank correlation coefficient with application to the consensus ranking problem," *J. Multi-Criteria Decis. Anal.*, vol. 11, no. 1, pp. 17–28, 2002.

[41] O. Melnik, Y. Vardi, and C.-H. Zhang, "Mixed group ranks: Preference and confidence in classifier combination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 973–981, Aug. 2004.

[42] W. Tang, "Mallows ranking models: Maximum likelihood estimate and regeneration," in *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds. Long Beach, CA, USA: PMLR, 2019, pp. 6125–6134.

[43] X. Zhong, L. Hurley, S. Sirimulla, C. Schweikert, and D. F. Hsu, "Combining multiple ranking systems on the generalized permutation rank space," in *Proc. IEEE 5th Int. Conf. Big Data Intell. Comput. (DATACOM)*, Nov. 2019, pp. 123–129.

[44] D. F. Hsu, B. S. Kristal, and C. Schweikert, "Rank-score characteristics (RSC) function and cognitive diversity," in *Brain Informatics*. Berlin, Germany: Springer-Verlag, 2010, pp. 42–54.

[45] K. Pearson, "Contributions to the mathematical theory of evolution," *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 185, pp. 71–110, 1894. [Online]. Available: https://www.jstor.org/stable/90667

[46] G. U. Yule, "VII. On the association of attributes in statistics: With illustrations from the material of the childhood society," *Phil. Trans. Roy. Soc. London. Ser. A, Math. Phys. Character*, vol. 194, nos. 252–261, pp. 257–319, 1900.

[47] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, 1st ed. London, U.K.: Chapman & Hall CRC, 2012.

[48] R. Kohavi and D. Wolpert, "Bias plus variance decomposition for zero-one loss functions," in *Proc. 13th Int. Conf. Int. Conf. Mach. Learn.* San Francisco, CA, USA: Morgan Kaufmann, 1996, pp. 275–283.

[49] C. Spearman, "Footrule for measuring correlation," *Brit. J. Psychol.*, vol. 2, no. 1, pp. 89–108, Jul. 1906.

[50] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 5, no. 1, pp. 3–55, 1948.

[51] R. A. Kempton and L. R. Taylor, "The Q-statistic and the diversity of floras," *Nature*, vol. 275, no. 5677, pp. 252–253, Sep. 1978.

[52] D. Partridge and W. Krzanowski, "Software diversity: Practical statistics for its measurement and exploitation," *Inf. Softw. Technol.*, vol. 39, no. 10, pp. 707–717, Jan. 1997.

[53] A. Kolmogorov, "On the empirical determination of a distribution function," in *Breakthroughs in Statistics* (Springer Series in Statistics). New York, NY, USA: Springer, 1992, pp. 106–113.

[54] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Ann. Math. Statist.*, vol. 19, no. 2, pp. 279–281, Jun. 1948.

[55] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[56] P. Cunningham and J. Carney, "Diversity versus quality in classification ensembles based on feature selection," in *Machine Learning: ECML 2000*. Berlin, Germany: Springer, 2000, pp. 109–116.

[57] H. Cramér, "On the composition of elementary errors: 2. Statistical applications," *Scandin. Actuarial J.*, vol. 1928, no. 1, pp. 13–74, 1928.

[58] A. J. C. Sharkey, Ed., *Combining Artificial Neural Nets*. London, U.K.: Springer, 1999.

[59] L. I. Kuncheva, *Combining Pattern Classifiers*. Hoboken, NJ, USA: Wiley, 2004.

[60] S. E. Page, *The Difference*. Princeton, NJ, USA: Princeton Univ. Press, 2007.

[61] S. E. Page, *Diversity and Complexity*. Princeton, NJ, USA: Princeton Univ. Press, 2010.

[62] H. P. Martinez, G. N. Yannakakis, and J. Hallam, "Don't classify ratings of affect; rank them!" *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 314–326, Jul. 2014.

[63] Y.-S. Chung, D. F. Hsu, and C. Y. Tang, "On the diversity-performance relationship for majority voting in classifier ensembles," in *Multiple Classifier Systems*. Berlin, Germany: Springer, 2007, pp. 407–420.

[64] Y. Chung, D. F. Hsu, C. Liu, and C. Tang, "Performance evaluation of classifier ensembles in terms of diversity and performance of individual systems," *Int. J. Pervasive Comput. Commun.*, vol. 6, no. 4, pp. 373–403, Nov. 2010.

[65] G. Brown, J. Wyatt, R. Harris, and X. Yao, "Diversity creation methods: A survey and categorisation," *Inf. Fusion*, vol. 6, no. 1, pp. 5–20, Mar. 2005.

[66] S.-S. Choi, S.-H. Cha, and C. C. Tappert, "A survey of binary similarity and distance measures," *J. Syst., Inform.*, vol. 8, no. 1, pp. 43–48, 2010.

[67] D. F. Hsu, B. S. Kristal, Y. Hao, and C. Schweikert, "Cognitive diversity: A measurement of dissimilarity between multiple scoring systems," *J. Interconnection Netw.*, vol. 19, no. 1, Mar. 2019, Art. no. 1940001.

[68] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[69] R. Brüggemann, L. Carlsen, and J. Wittmann, Eds., *Multi-Indicator Systems and Modelling in Partial Order*. New York, NY, USA: Springer, 2014.

[70] R. Brüggemann and G. P. Patil, "Ranking and prioritization with partial order for multi-indicator systems–an integrative view with a look forward," in *Ranking and Prioritization for Multi-indicator Systems*. New York, NY, USA: Springer, 2014, pp. 291–303.

[71] C. Mesterharm and D. F. Hsu, "Combinatorial fusion with on-line learning algorithms," in *Proc. 11th Int. Conf. Inf. Fusion (FUSION)*, Cologne, Germany, Jun./Jul. 2008, pp. 1–8.

[72] C. Schweikert, L. Gobin, S. Xie, S. Shimojo, and D. F. Hsu, "Preference prediction based on eye movement using multi-layer combinatorial fusion," in *Brain Informatics*. Cham, Switzerland: Springer, 2018, pp. 282–293.

[73] J. C. Bermond, F. Comellas, and D. F. Hsu, "Distributed loop computer networks: A survey," *J. Parallel Distrib. Comput.*, vol. 24, no. 1, pp. 2–10, Jan. 1995.

[74] P. Erdös and D. F. Hsu, "Distributed loop network with minimum transmission delay," *Theor. Comput. Sci.*, vol. 100, no. 1, pp. 223–241, Jun. 1992.

[75] K. Mohan, J. Pearl, and J. Tian, "Graphical models for inference with missing data," in *Proc. Neural Inf. Process. Syst. Conf. (NIPS)*, 2013, pp. 1277–1285.

[76] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA, USA: Morgan Kaufmann, 2014.

[77] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*. New York, NY, USA: Basic Books, 2018.

[78] J. G. Kemeny and J. L. Snell, "Preference rankings: An axiomatic approach," in *Mathematical Models in the Social Sciences*. Cambridge, MA, USA: Ginn Company; Blaisdell Publishing Company, 1962, pp. 9–23.

[79] J. Kemeny, "Generalized random variables," *Pacific J. Math.*, vol. 9, no. 4, pp. 1179–1189, Dec. 1959.

[80] I. J. Good, "The number of orderings of n candidates when ties are permitted," *Fibonacci Quart.*, vol. 13, no. 1, pp. 11–18, 1975.

[81] J. Bedö and C. S. Ong, "Multivariate spearman's ρ for aggregating ranks using copulas," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 7072–7101, 2016.

[82] P. Diaconis, *Group Representations in Probability and Statistics*, vol. 11. Hayward, CA, USA: Institute Mathematical Statistics, 1988.

[83] D. F. Hsu, "On container width and length in graphs, groups, and networks–dedicated to professor Paul Erdös on the occasion of his 80th birthday–," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. 77, no. 4, pp. 668–680, 1994.

[84] Y.-K. Shih, C.-K. Lin, D. F. Hsu, J. J. M. Tan, and L.-H. Hsu, "The construction of mutually independent Hamiltonian cycles in bubble-sort graphs," *Int. J. Comput. Math.*, vol. 87, no. 10, pp. 2212–2225, Aug. 2010.

[85] E. Kowalski, *An Introduction to Expander Graphs*. Paris, France: Société Mathématique de France, 2019.

[86] J. Friedman, "On Cayley graphs on the symmetric group generated by transpositions," *Combinatorica*, vol. 20, no. 4, pp. 505–519, Apr. 2000.

[87] W. J. Heiser and A. D'Ambrosio, "Clustering and prediction of rankings within a Kemeny distance framework," in *Algorithms From and for Nature and Life*. Cham, Switzerland: Springer, 2013, pp. 19–31.

[88] M. V. Pogančić, A. Paulus, V. Musil, G. Martius, and M. Rolinek, "Differentiation of blackbox combinatorial solvers," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–19.

[89] D. F. Hsu, D. Marinucci, and J. M. Voas, "Cybersecurity: Toward a secure and sustainable cyber ecosystem," *Computer*, vol. 48, no. 4, pp. 12–14, Apr. 2015.

[90] S. Starr, "Thermodynamic limit for the mallows model on SN," *J. Math. Phys.*, vol. 50, no. 9, Sep. 2009, Art. no. 095208.

[91] I. Kustiawan, C.-Y. Liu, and D. F. Hsu, "Vertical handoff decision using fuzzification and combinatorial fusion," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2089–2092, Sep. 2017.

[92] S. Pal Singh and M. Jaggi, "Model fusion via optimal transport," 2019, *arXiv:1910.05653*. [Online]. Available: http://arxiv.org/abs/1910.05653

LANDON HURLEY received the M.A., M.S., and Ph.D. degrees in psychometrics and statistics from Fordham University. He was a member with the Laboratory of Informatics and Data Mining, Fordham University, and a Consulting Statistician with the Edward Zigler Center in Child Development and Social Policy, Yale University. He is currently a Postdoctoral Fellow with a joint appointment with the Yale University School of Public Health, and the West Haven Veterans' Health Administration Cooperative Studies Program. His research interests include the exploration of dual characterization functionals for non-parametric statistical models and their topology and latent variable measurement models in non-parametric empirical and latent spaces. This work has been extended to applications in early prognostic cancer diagnostics and clinical research trials, as well as child development assessments and non-randomized causal inference.
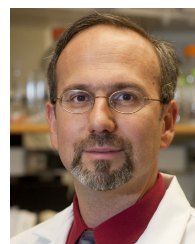
BRUCE S. KRISTAL received the B.S. degree in life sciences from the Massachusetts Institute of Technology, in 1986, and the Ph.D. degree from the Division of Medical Sciences, Graduate School of Arts and Sciences, Committee on Virology, Harvard University, in 1991. He rose from a Postdoctoral position to a Research Assistant Professor at The University of Texas Health Science Center, San Antonio, from 1991 to 1996, then moved to the Burke Medical Research Institute, in 1996, and the Department of Biochemistry and the Department of Neuroscience, Weill Medical College, Cornell University, in 1997 and 1998, respectively, where he became an Associate Professor in neuroscience in 2004. He joined the Brigham and Women's Hospital's Department of Neurosurgery, in April 2007, and the Department of Surgery, Harvard Medical School, in 2008, transitioning to the Department of Neurosurgery when it split off in 2013. He moved to the Division of Sleep and Circadian Disorders, Department of Medicine, in 2016. His research has been recognized by awards from the American Federation for Aging Research, the American Institute for Cancer Research, the Hereditary Disease Foundation, and others. He has served as a Contributing Editor for Science's Science of Aging Knowledge Environment. He was the Founding Secretary from 2004 to 2008 and a member of the Board of Directors of the Metabolomics Society from 2004 to 2011.

SUMAN SIRIMULLA received the M.S. and Ph.D. degrees in computational chemistry from The University of Texas at El Paso (UTEP), in 2010, and the B.Pharm. degree from the JSS College of Pharmacy, India. He is currently working as an Assistant Professor (tenure track) with the Department of Pharmaceutical Sciences with adjunct appointments in computer science, chemistry and biochemistry, computational science, bioinformatics, and data science programs. Prior to joining UTEP as a Faculty Member in 2016, he has worked as a Faculty Member with Northern Arizona University, from 2010 to 2014, and the St. Louis College of Pharmacy, from 2014 to 2016. He also leads the Computer Aided Drug Discovery Laboratory, UTEP, and his laboratory is involved in developing and using computational tools for drug discovery research. His work is highly interdisciplinary. His research interests include drug discovery, genomics, and personalized medicine to high-performance computing, machine learning, model fusion, and artificial intelligence.
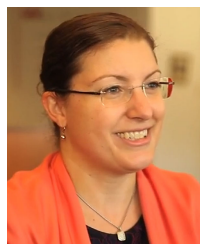
CHRISTINA SCHWEIKERT (Member, IEEE) received the Ph.D. degree. She is currently an Associate Professor of computer science with St. John's University. During her time at the University, she was granted the Clare Boothe Luce Professorship. She is also the Program Director for the M.S. in Data Science Program. She also serves as the Secretary for the IEEE Computational Intelligence Society, NY Section Chapter, and the Assistant Vice President for the Global Business and Technology Association. Her research interests include programming languages, data mining, information fusion, biomedical, and healthcare informatics.

D. FRANK HSU (Life Senior Member, IEEE) received the B.S. degree from Cheng Kung University, the M.S. degree from The University of Texas at El Paso, and the Ph.D. degree from the University of Michigan.

He was a Visiting Professor with Keio University (the IBM Chair in Fall 1991); JAIST (as the Komatsu Chair in Fall 1993); Taiwan University in 1998; and DIMACS Special Years on Information Fusion at Rutgers University, from 2013 to 2015. He is currently the Clavius Distinguished Professor of science and a Professor of computer and information science with Fordham University, New York, NY, USA.

Hsu's research areas include interconnection networks, data science and informatics, and combinatorial fusion analysis (CFA) and multi-layer combinatorial fusion (MCF). He has served on many editorial boards, including IEEE TRANSACTIONS ON COMPUTERS (1992–96), IEEE TRANSACTIONS ON RELIABILITY (2015–16), and IEEE SYSTEMS JOURNAL (2018–***). Among the honors and awards he received are: Fellow of the ICA (1990), Fellow of the New York Academy of Sciences (NYAS) (1999), Fellow of ICIC (2012), IBM Faculty Award (2012), and IEEE Region I Outstanding Teaching Award (2019). He was the Co-Founder and the EIC of the *Journal of Interconnection Networks* in 2000, the Co-Organizer of six DIMACS workshops (Rutgers University, from 1999 to 2013), and the Co-Founder and the Conference Co-Chair of ICCS, a public-private-academia partnership founded at Fordham University, in 2009. He has given more than 400 invited talks, including recent keynote online presentation at the 15th Joint Asia Conference on Information Security (August 2020).

• • •