CrossMark

The diversity rank-score function for combining human visual perception systems

Christina Schweikert · Darius Mulia · Kilby Sanchez · D. Frank Hsu

Abstract There are many situations in which a joint decision, based on the observations or decisions of multiple individuals, is desired. The challenge is determining when a combined decision is better than each of the individual systems, along with choosing the best way to perform the combination. It has been shown that the diversity between systems plays a role in the performance of their fusion. This study involved several pairs of people, each viewing an event and reporting an observation, along with their confidence level. Each observer is treated as a visual perception system, and hence an associated scoring system is created based on the observer's confidence. A diversity rank-score function on a set of observation pairs is calculated using the notion of cognitive diversity between two scoring systems in the combinatorial fusion analysis framework. The resulting diversity rank-score function graph provides a powerful visualization tool for the diversity variation among a set of system pairs, helping to identify which system pairs are most likely to show improved performance with combination.

Keywords Cognitive diversity · Combinatorial fusion analysis · Diversity rank-score function · Multiple scoring systems · Rank-score characteristic (RSC) function

C. Schweikert (✉)
Division of Computer Science, Mathematics and Science, St. John's University, Queens, NY, USA
e-mail: schweikc@stjohns.edu

D. Mulia · K. Sanchez · D. F. Hsu
Laboratory of Informatics and Data Mining, Department of Computer and Information Science, Fordham University, New York, NY, USA
e-mail: hsu@cis.fordham.edu

1 Introduction

The concept of multiple scoring systems has been applied to a variety of domains [1, 2]. In situations where multiple scoring systems are constructed, we are interested in conducting a meta-analysis to gain an understanding of the relationship between the systems, specifically the diversity between them. It has been shown that the combination of two scoring systems can outperform individual systems when there is some diversity between the systems, and they are of relatively good performance [1, 3]. To this end, quantitative measures of diversity can be used to generate diversity scores for pairs of systems, which can then be analyzed within the combinatorial fusion analysis (CFA) framework [1].

Human beings are constantly and naturally performing *fusion* of information within and among the senses. There is extensive research in this area on the neurological level pertaining to how fusion in the sensory system works [4–6], how visual information is combined with information from other senses [7–11], and how visual systems are combined [12, 9, 13]. In this study, however, we are focused on the inter-human level of information and fusion of the information at the decision level.

There are many situations in which two people's observations are considered for a decision, such as referees in a football or tennis match, physicians examining a patient, co-pilots navigating a plane, and so on. For example, when two physicians are examining a new patient, each may observe different symptoms that can indicate different diseases; interactive consultation may lead to a final diagnosis. When two people are interactively making a decision based on visual input, research by Bahrami et al [12], Ernst and Banks [7], and Kepecs et al [13] suggests that these decisions are improved when two

people are interactively making the decision, rather than an individual. The question then becomes, if we have two people making visual observations of an event, how do we integrate these observations or decisions? Do we choose one of the observer's results, or create a combination of the two? Koriat [14] emphasizes the importance of confidence, and that it may be a good option to take the decision of the more confident person. The approach taken in our study is to combine the observations or decisions made by two people in an attempt to outperform the individual decisions. The visual observations tested in this project involve pairs of volunteers that are asked to give the location of a small object they observe being tossed in a field.

In order to perform the desired combination, by score or by rank, a scoring system must first be constructed for each participant in a trial. Each participant's observation, or perception system, is represented as a scoring system, which is made up of a score function and a rank function. Given this multiple scoring system scenario, we then analyze the cognitive diversity between the scoring systems of a trial. A quantitative diversity measure, the distance between two rank-score functions, is used to represent the cognitive diversity between two scoring systems [1, 2]. Examining the relative diversities between the system pairs, together with the performance of their combinations, can give us insight into how diversity variation may play a role in the performance of system combinations. The diversities between systems are analyzed using the diversity rank-score functions, which are then visualized in diversity rank-score graphs. This visualization of diversity variation is beneficial in situations where there are a large number of scoring system pairs (hundreds or thousands). Interactive data visualization [15–17] is a dynamic field in which data are visualized with the intent to facilitate an end user in a particular task. The diversity rank-score function graph is such a tool that has potential to be integrated into various data analytics and software systems.

Information fusion can be applied to many situations where there are multiple scoring systems, or multiple classifiers. For example, the CFA framework [18, 1, 2] has been applied to information retrieval [19], text categorization [20], target tracking [21], sensor feature selection and combination [22], and image skeleton pruning [23]. Combinatorial fusion has also been used for enhancing the analysis of various biomedical datasets including virtual screening for molecular compounds [3], protein structure prediction [24], and ChIP-seq peak detection [25]. When combining multiple models (performing information fusion), it would be useful to know in advance whether the fusion will outperform the best model. Ng and Kantor [26] identify system features that can help predict whether fusion will be beneficial. Combination of multiple

classifiers has also been shown to improve results in the area of pattern recognition. [27, 28]

The content of this paper is organized as follows: Sect. 2 describes the concept of multiple visual perception systems, along with the corresponding multiple scoring systems, which are considered a generalization of multiple classifier systems. The CFA framework, which establishes each visual perception system as a scoring system and combines two such systems, is also described. The diversity rank-score function can be used as a guiding light to combine pairs of visual perception systems based on the diversity variation across a set of trials. In Sect. 3, we describe the visual perception dataset, present the results of scoring system combinations, and examine the role of the diversity rank-score function graph in the context of diversity variation and visualization. Concluding remarks and discussion are included in Sect. 4.

2 Multiple visual perception systems

2.1 From multiple classifier systems to multiple scoring systems

In many domains, such as biomedical informatics, finance, security, information retrieval, among others, classification models are created in order to generate class predictions for new data. Binary classifiers attempt to categorize items into one of two classes (or labels). For example, determining whether a webpage is relevant to a search term or not, or whether a patient tests positive or negative for a disease. Some binary classification problems are asymmetric, meaning one class occurs much less frequently than the other. Multiclass classifiers involve more than two classes.

The output of a classification system includes a class prediction, along with an associated probability. Treating these probabilities as scores, and sorting the results by score to generate rankings, enables us to consider classification systems as a scoring system that have a score function and a rank function.

In an effort to improve classification accuracy, it is often desired to incorporate the results from multiple classifiers that are varied in terms of their approach or algorithm. The element of variety, or diversity, is essential since different classifiers may contribute various perspectives, results, or predictions, on the data. Generally, the results from multiple classifier systems are combined using ensemble methods such as majority voting (bagging) or weighted voting (boosting). Table 1a, b contains a snapshot from a classification example in which the class label of a sample document is predicted in each of the following two cases: (a) 3 class labels, and (b) 6 class labels (Table 1a, b). The document is analyzed by 4 different classifiers, each of

Table 1 Combination of multiple classifier systems with (a) 3 class labels and (b) 6 class labels using majority voting (C_{MAJ}), score combination (C_{SC}), and rank combination (C_{RC})

a

Classifier	C_1	C_2	C_3	C_4	C_{MAJ}	C_{SC}	C_{RC}
Class label							
A(score, rank)	(0.74, 1)	(0.05, 3)	(0.55, 1)	(0.31, 3)	2	0.41	2
B(score, rank)	(0.14, 2)	(0.48, 1)	(0.25, 2)	(0.33, 2)	1	0.30	1.75
C(score, rank)	(0.12, 3)	(0.47, 2)	(0.20, 3)	(0.36, 1)	1	0.29	2.25
Class label	A	B	A	C	A	A	B

b

Classifier	C_1		C_2		C_3		C_4		C_{MAJ}	C_{SC}	C_{RC}
Class label	s	r	s	r	s	r	s	r			
A	0.32	1	0.05	3	0.03	6	0.30	1	2	0.18	2.75
B	0.11	4	0.04	4	0.25	2	0.11	4	0	0.13	3.5
C	0.01	6	0.03	5	0.45	1	0.08	5	1	0.14	4.25
D	0.27	2	0.81	1	0.04	5	0.22	3	1	0.34	2.75
E	0.05	5	0.01	6	0.11	4	0.06	6	0	0.06	5.25
F	0.24	3	0.06	2	0.12	3	0.23	2	0	0.16	2.5
Class label	A		D		C		A		A	D	F

Table 2 Combining multiple scoring systems (with 3 scoring systems) to rank a set of items (with 8 items)

	J_1		J_2		J_3		s(SC)	r(SC)	s(RC)	r(RC)
	s	r	s	r	s	r				
d_1	8.5	4	7	5	9.7	4	25.2	4	13	4.5
d_2	7.6	7	8.4	3	9.6	6	25.6	3	16	7
d_3	8.3	5	5.6	7	9.75	3	23.65	7	15	6
d_4	6.4	8	7.4	8	9.81	2	21.61	8	18	8
d_5	9.4	3	7.8	4	9.68	5	26.88	2	12	3
d_6	9.5	2	8.5	2	9.2	7	27.2	1	11	2
d_7	7.9	6	6.3	6	10	1	24.2	6	13	4.5
d_8	10	1	10	1	5.1	8	25.1	5	10	1

which output the probability that the document belongs to class A, B, or C, in the case of 3 class labels. In Table 1b, each document belongs to one of the 6 class labels: A, B, C, D, E, or F. For each classifier, the class label with the highest probability is considered the predicated class label and is assigned rank 1. Likewise, the next highest probability is assigned rank 2, and so on. The ensemble approach of majority voting is used to combine the results of the individual classifiers. For each class label, we count the number of times that class is ranked 1 (has the highest probability) by a classifier. Then, the class label with the highest number of votes is considered the predicted class for the document.

If we consider the classifiers as scoring systems (see Table 2), we can apply score and rank combinations as an alternative *ensemble* approach. Here, the probabilities are treated as scores, which are then ranked. Score combination (SC), in this example, is the average of the scores for a class label across the 4 classifiers. The class label with the highest average score is chosen as the result. The rank combination (RC) is computed as the average rank for a class label for all classifiers. The class label with the lowest average rank is then selected. Weighted averages can be used if the past performance of the classifiers is known. In this example, we can see that combining by score or rank may produce different results. Table 1b is a classification

problem that involves more possible class labels. In this example, we see that classifiers can be viewed as *scoring systems*, where the scores are the class label probabilities. The concept of multiple classifier systems with multiple class labels (the case in Table 1b) is then generalized to multiple scoring systems with multiple choices (items or options) (as is the case in Table 2).

When constructing an ensemble, it is desired to have diversity among the component classifiers or scoring systems. Several techniques for measuring diversity have been proposed for regression and classification [29, 30]. It is more challenging to measure diversity between classifiers if we just consider the output class labels, without their associated probabilities [29].

Viewing classification systems as scoring systems enables us to apply the concept of diversity that has been defined for multiple scoring systems [1, 2, 26, 3].

2.2 The combinatorial fusion framework

2.2.1 Establishing each visual perception system as a scoring system

In situations where we have a set of documents (webpages, genes, customers, etc.) that are assigned scores or probabilities by an algorithm or classifier, creating a scoring system is straightforward. However, in cases where we do not have a set of scores to work with, a score function needs to be generated based on the value(s) given. In this experiment, when an observer is deciding on the proposed landing point of the object based on the visual input, he/she is selecting from several locations within a range. Intervals within this visual range will be considered as the items (or options) that will be scored and ranked. Since there are two subjects within each trial, the corresponding score functions must score the same set of intervals. To this end, a common visual space is created, as described in previous work [18]. First, the mean of the decisions (points) for the two observers P and Q is computed in three different versions, varying the weight given to the confidence radius σ. M_0, M_1, and M_2, are computed as

$$M_i = \frac{\frac{P}{\sigma_1^i} + \frac{Q}{\sigma_2^i}}{\frac{1}{\sigma_1^i} + \frac{1}{\sigma_2^i}} \tag{1}$$

The scoring system analysis is performed for each version of M_i. Specifically, the M_i values are used as a foundation point from which to create a common visual space. The M_i points are always located between the P and Q original points. The visual space is also extended on both sides of P and Q. The common visual space is divided into 63 intervals. The interval scores are computed using a normal

Fig. 1 Scoring system P with: (a) score function s_P, (b) rank function r_P, and (c) rank-score characteristic (RSC) function f_P
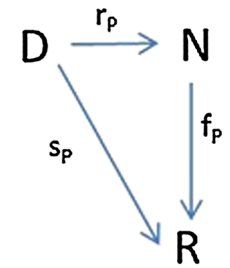
distribution around M_i, using the confidence radius (0.5r) for the standard deviation. The performance of each M_i is measured as the distance from M_i to the actual location of the object [31]. The scores, created for the intervals for P and Q, give us the score functions s_P and s_Q. Given a set of intervals $d_1, d_2, \ldots, d_n$, the scoring system P consists of a score function s_P, rank function r_P, and rank-score characteristic (RSC) function f_P (see Fig. 1). The rank function for the scoring systems P and Q are obtained by sorting s_P and s_Q and assigning ranks to create the rank functions r_P and r_Q. The Rank-Score Characteristic (RSC) function, as defined by Hsu et al [1, 2], is the composite function of s_p and the inverse of r_P. Rank-score functions map ranks to scores, and are independent of the data items. Here, the rank-score characteristic (RSC) function for the scoring system P, $f_P : N \rightarrow R$, is computed as

$$f_P(i) = (s_P \circ r_P^{-1})(i) = s_P(r_P^{-1}(i)) \tag{2}$$

Similarly, f_Q is computed for scoring system Q.

2.2.2 Combining two visual perception systems

Within the CFA framework [1, 2], system combination is performed either by score or rank combination. A score combination is computed as the average of the score functions, s_p and s_Q for each interval, d_i, giving us the score function of the score combination s_{SC}. The rank function of the score combination, r_{SC}, is achieved by sorting s_{SC} in descending order and obtaining ranks for each d_i. In addition, we compute the rank combination by averaging the rank functions r_P and r_Q, to give us the score function of the rank combination, s_{RC}. We sort this function in ascending order and assign ranks to get its associated rank function, r_{RC} (see the example in Table 2). The performance of these combined results is measured by the distance of the newly computed points to the actual x,y coordinates where the object landed in the field.

2.2.3 Cognitive diversity between two scoring systems

In cases where multiple scoring systems, algorithms, or approaches exist, it is beneficial to know under what circumstances combining pairs of these systems could result
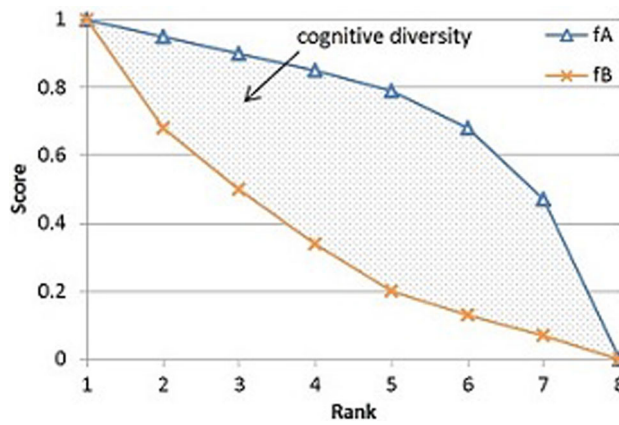
Fig. 2 Rank-score characteristic function graph of two scoring systems, *A* and *B*

in improved performance. Diversity between two scoring systems A and B can be measured in a few different ways, such as the distance between score or rank functions using covariance (between s_A and s_B) or Kendalls tau (using r_A and r_B), respectively. Another method to measure the diversity between two scoring systems, which is used here and called cognitive diversity, is to measure the *distance* between the rank-score functions (f_A and f_B) of the two systems [1, 2] (see formula (2) and Fig. 1). Figure 2 illustrates two RSC functions, f_A and f_B, for two arbitrary scoring systems A and B. One distance measurement is the area between the two RSC functions. We note that the cognitive diversity between scoring systems A and B, as seen in Fig. 2, provides a powerful visualization tool on the *similarity* or *dissimilarity* between these two visual perception systems, A and B, in the context of the current study.

In this analysis, the concept of cognitive diversity is applied to the trials and scoring systems P and Q, which represent the 2 participants in a given trial pair. Therefore, the cognitive diversity of the two observers P and Q, d(P,Q), defined as the distance between the rank-score functions of two systems P and Q, f_P and f_Q, is computed as follows:

$$d(P,Q) = d(f_P, f_Q) = \sqrt{\sum_{i=1}^{m} \frac{(f_P(i) - f_Q(i))^2}{m}} \qquad (3)$$

2.3 Diversity rank-score function across a set of trials

Let $T = \{(p_1, q_1), (p_2, q_2), \ldots, (p_n, q_n)\}$ represent a set of n trials, each consisting of an ordered pair of participants and let $R = \{d(p_1, q_1), d(p_2, q_2), \ldots, d(p_n, q_n)\}$ represent the diversity scores for each pair in T, where $N = 1, 2, \ldots, n$. The cognitive diversity between each pair of scoring systems, P and Q, is measured by the diversity function d(P,Q), as shown in equation (3), where m is the number of

items (intervals) to be scored; in this case m is 63, indicating the number of intervals in the common visual space. The set of diversity values itself can be treated as a scoring system, making the diversity function into a diversity score function. For this purpose, the number d(P,Q), which is the diversity between scoring systems P and Q, is considered as the diversity score function value of the trial (p,q) and is denoted as $s_{(p,q)}$. The diversity rank function is attained by sorting the score function and generating ranks, giving $r_{(p,q)}$. A diversity rank-score function, $f_{(p,q)}$, is computed as

$$f_{(p,q)}(i) = (s_{(p,q)} \circ r_{(p,q)}^{-1})(i) = s_{(p,q)}(r_{(p,q)}^{-1}(i)) \qquad (4)$$

The diversity rank-score function is a mapping from diversity ranks to diversity scores. The relationship between $s_{(p,q)}$, $r_{(p,q)}$, and $f_{(p,q)}$ is shown in Fig. 3.

3 Case analysis using diversity rank-score graph

3.1 Visual perception dataset

The setting for the data collection was in a grassy field in NYC's scenic Central Park. A lab member was tasked with recruiting pairs of participants for the experiment. The pairs of subjects varied in terms of gender and relationship between the individuals. The subject pairs were randomly chosen and could be friends, siblings, husband and wife, colleagues, or acquaintances. A small metal object that was made of metal plates, nuts, and a bolt, and of size 1.5 by 1.5 inches was used for the experiment, since it was possible to throw it far distances, small enough to be hidden in the grass, and would not roll from its position once landed. The subject pairs stood 40 feet from a marked square of size 250 by 250 inches, and the individuals stood a distance of 10 feet away from each other. A member of our group tossed the metal object into the designated square. Each participant is asked individually to walk and point to where he/she believed the object landed. A marker is placed at these locations. The participants are also asked to give a measure of their confidence of his or her guess in the form of a confidence radius around the specified mark. Lab members helped the participants gauge their confidence

Fig. 3 Diversity scoring system (p,q) with: (a) diversity score function $s_{(p,q)}$, (b) diversity rank function $r_{(p,q)}$, and (c) diversity RSC function $f_{(p,q)}$
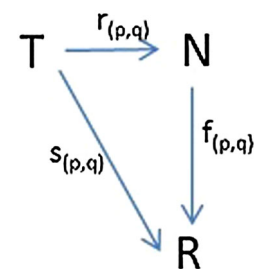
radius by using tool consisting of 2 poles of length 36 by 36 inches to represent the x and y coordinates. Smaller radius values indicate higher confidence of the subject. A lab member measures the distance from the actual position where the object landed and the guess positions of the subjects.

The subjects are given feedback as to how far off their guess is from the actual landing point of the object. The values collected are: x,y coordinates for subject P and Q from each experiment, a confidence radius for each participant, along with the actual landing x,y coordinate of the object. All measurements are in inches. The values for the trials in this most recent experiment are shown in Table 3. Our group has conducted previous data collection activities of this type, the data of which can be found in [18].

The distribution functions for P and Q for a sample trial are shown in Fig. 4a. Sample rank-score functions for a trial are shown in Fig. 4b.

3.2 Analysis results of combinations

The experimental results are presented in Fig. 5. The performances of P and Q, shown in column (a), are the distances to the actual landing point of the object. The confidence radii are included in column (b), in which a shaded cell indicates that the more confident participant leads to the best performance. The performance of the weighted means M_0, M_1, and M_2 is listed in column (c). C represents the score combination and D represents the rank combination. The last column, (d), presents information for

the results using each of the weighted means, along with the score and rank combinations (C and D). For each $i = \{0, 1, \text{and } 2\}$, P, Q, M_i, C, and D are ranked in descending order of performance; repeated ranks indicate tied performance. Rank 1 showed the best performance, meaning the closest interval to the actual location of the object. Cases where the score (C) or rank (D) combinations either outperformed or tied the best individual system are highlighted.

3.3 The role of diversity rank-score graphs

After performing the score and rank combinations for the three different computations of M (M_0, M_1, and M_2), we can summarize the results as follows: Using M_0, the score and/or rank combination for 14/16 trials showed either tied or improved performance compared to the best individual system; using M_1, 9/16 trials; and using M_2, 7/16 trials. The diversity rank-score functions for the scoring systems created according to the three different computations of the mean: M_0, M_1, and M_2, are depicted in Fig. 6. Examination of these graphs, along with the performance of the corresponding system pair combinations, can help us understand the role of cognitive diversity in system combinations by score and rank. To make the connection with the trials, Table 4 is included to show the ranking of trials according to the diversity of their component scoring systems, for each case of M_0, M_1, and M_2. When comparing with the performances of the system combinations, we detect a tendency for pairs of systems with relatively high diversity to have more improved performance. In this

Table 3 Data collected for the observed points and confidence radii for trials, along with the actual x,y coordinates

Trial	P		Confidence radius	Q		Confidence radius	Actual	
	x	y		x	y		x	y
1	126	243	12	114	287	6	120	270.5
2	69	362	8	89	358	6	85	362
3	105	220	18	60	287	10	93	321
4	93	336	10	91	285	16	81	318
5	152	170	14	141	162	16	126.5	180
6	66	250.5	16	81	288	12	88	119
7	24	314	16	31	310	8	6	313
8	94	278	12	98	220	6	86	236
9	24	235	12	23	256	12	25	240
10	96	95	8	131	71	10	107	337
11	52	187	20	97	243	16	102	269
12	107	246	10	113	233	8	113	242
13	121	191.5	10	141.5	191	8	127.5	185
14	46	277	10	73	229	8	52	254
15	73	264	18	79	267	12	84	282
16	24	442	10	23	413	10	23	432

Fig. 4 Distribution functions and Rank-score characteristic functions for P and Q in Trial 1 based on M_2. **a** Distribution functions for Trial 1, based on M_2. **b** Rank-score functions for Trial 1, based on M_2
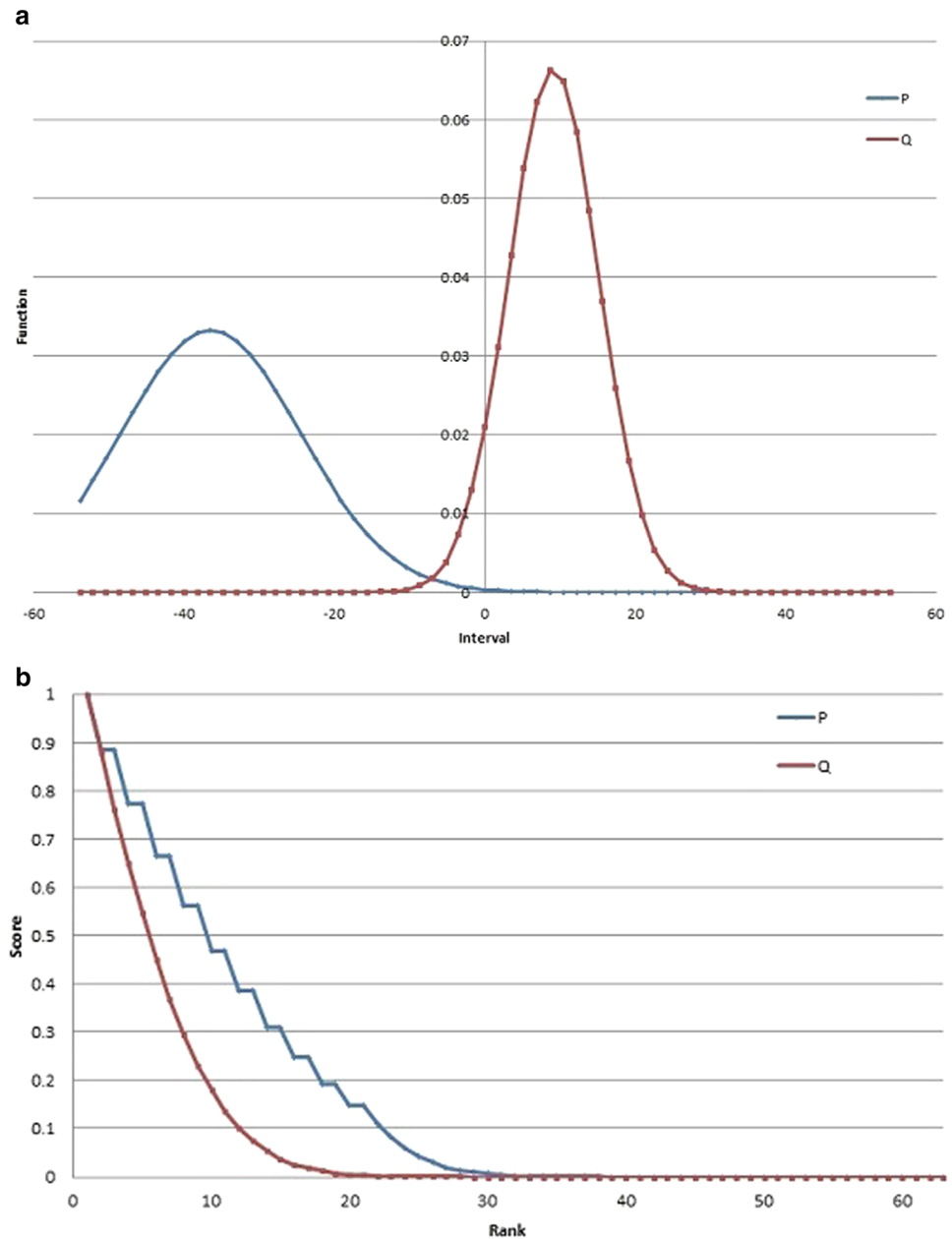
study, this observation is most strongly supported by the data in the M_1 scenario. In new situations, where we may not be able to predict the performance, analyzing the relative diversities between scoring systems may give us insight into which pairs of systems are most likely to show improvement with combination.

We observe that the diversity rank-score graphs are good indicators for the combination outcome. For example, trials d_5 and d_{16} appear at the very end of the graph in M_0, M_1, and M_2 (see Figure 6 and Table 4). In these two trials, neither rank nor score combination helps improve the outcome. However, even though trial d_9 has a very low diversity (Table 4), its combination of scoring systems P and Q is better than or equal to the best of P and Q since P has a relatively high performance.

4 Conclusion and further work

In this paper, we studied the combination of multiple visual perception systems using the CFA framework and the diversity rank-score function. By establishing each visual perception system as a scoring system on a set of options (possible locations, in our context) in a common visual space, the problem of combining multiple visual perception systems is treated as a problem of combining multiple

Trial	(a): Per (P,Q)	(b): Conf.Rad(P,Q)	M0	M1	M2	P	Q	Mo	C	D	P	Q	M1	C	D	P	Q	M2	C	D
Trial 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
Trial 2	(16,5.7)	(8,6)	3	2	1	4	1	3	1	4	4	3	1	4	2	4	3	1	4	2
Trial 3	(101.7,47.4)	(18,10)	3	2	1	4	1	3	1	4	4	1	3	4	2	4	1	3	4	2
Trial 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
Trial 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
Trial 6	(133.3,169.1)	(16,12)	1	2	3	1	4	3	4	1	1	5	3	1	4	1	5	3	1	4
Trial 7	(18,25.2)	(16,8)	1	2	3	1	4	3	4	1	1	5	2	3	3	1	5	2	3	3
Trial 8	(42.8,20)	(12,6)	3	1	2	4	2	1	2	4	4	3	1	4	2	4	3	1	4	2
Trial 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
Trial 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
Trial 11	(96,26.5)	(20,16)	3	2	1	4	1	3	1	4	4	1	3	4	2	4	1	3	4	2
Trial 12	(7.2,9)	(10,8)	1	2	3	2	4	1	4	2	2	4	1	5	3	2	5	1	3	3
Trial 13	(9.2,15.2)	(10,8)	1	2	3	2	4	1	4	2	2	4	1	5	3	2	5	1	3	3
Trial 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
Trial 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
Trial 16	(10,19)	(10,10)	1	1	1	2	3	1	3	3	2	3	1	3	3	2	3	1	3	3

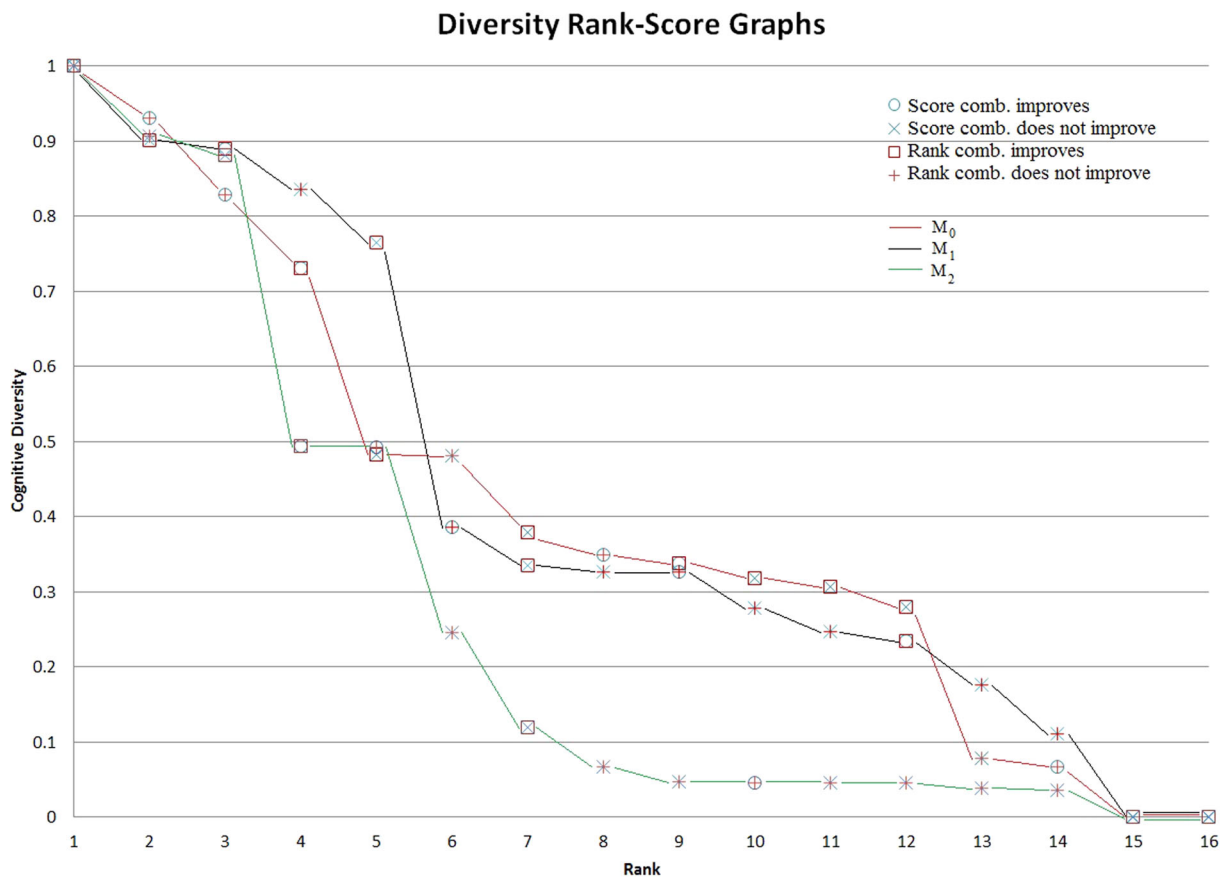Fig. 5 Analysis results for 16 trials [31]

Fig. 6 Diversity rank-score graphs based on M_0, M_1, and M_2, respectively

Table 4 Trials ranked with respect to M_0, M_1, and M_2

Rank	Trials (M_0)	Trials (M_1)	Trials (M_2)
1	d_1	d_1	d_1
2	d_8	d_8	d_3
3	d_3	d_2	d_4
4	d_4	d_3	d_2
5	d_6	d_4	d_6
6	d_2	d_6	$\mathbf{d_5}$
7	d_{13}	d_{10}	d_8
8	d_{11}	d_{11}	d_7
9	$\mathbf{d_{10}}$	d_{14}	d_{10}
10	d_4	d_7	d_{14}
11	d_{12}	d_{13}	d_{11}
12	d_7	d_{15}	d_{15}
13	$\mathbf{d_5}$	d_{12}	d_{13}
14	d_{15}	$\mathbf{d_5}$	d_{12}
15	d_9	$\mathbf{d_{16}}$	$\mathbf{d_{16}}$
16	$\mathbf{d_{16}}$	d_9	d_9

scoring systems. Using a dataset of an experiment with sixteen trials where each trial consists of a pair of two observers, we studied various issues as to how the diversity between these two observers (and their individual perception systems) affects the performance of the combined system.

At the individual trial level, we illustrated that the rank-score characteristic (RSC) function graphs of the two scoring systems (perception systems) can provide a useful visualization tool on the *similarity* or *dissimilarity* between these two visual perception systems (see Fig. 2 and Sect. 2.2.3). At the population level, the diversity rank-score graphs on three common visual space definitions, M_0, M_1, and M_2, respectively provide a powerful visualization comparison, not only among all (sixteen) trials in an experiment, but also among all (three) analytic methods based on M_0, M_1, and M_2, respectively (see Fig. 5 and Sect. 2.3). Our current study suggests a few issues which are worthy of further investigation. We list three here:

1. With the diversity rank-score function defined in formula (4) and the diversity rank-score graphs based on M_0, M_1, and M_2, extend the study to include higher order of M_i, i = 4, 5, and so on (refer to formula 1).
2. Establish a CFA framework to study the combination of more than two visual perception systems. In this regard, the notion of *diversity* among more than two systems would have to be defined differently.
3. Apply the visualization tool illustrated in current work to combination of multiple sensing systems, multiple

robotics systems, and multi-modal physiological imaging systems such as MRI, EEG, and EKG.

References

1. Hsu DF, Chung YS, Kristal BS (2006) Combinatorial fusion analysis: methods and practice of combining multiple scoring systems. In: Hsu HH (ed) Advanced data mining technologies in bioinformatics. Idea Group Inc., Calgary, pp 1157–1181
2. Hsu DF, Kristal BS, Schweikert C (2010) Rank-score characteristics (RSC) function and cognitive diversity. Brain Inform 8211:42–54
3. Yang JM et al (2005) Consensus scoring for improving enrichment in virtual screening. J Chem Inform Model 45:1134–1146
4. Gold JI, Shadlen N (2007) The neural basis of decision making. Ann Rev Neurosci 30:535–574
5. Hillis JM et al (2002) Combining sensory information: mandatory fusion within, but not between, senses. Science 298(5598):1627–1630
6. Tong F, Meng M, Blake R (2006) Neural basis of binocular rivalry. Trends Cognit Sci 10(11):502–511
7. Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433
8. Ernst MO (2007) Learning to integrate arbitrary signals from vision and touch. J Vis 7(5):1–14
9. Ernst MO (2010) Decisions made better. Science 329(5995):1022–1023
10. Gepshtein S et al (2009) The combination of vision and touch depends on spatial proximity. J Vis 5(11):1013–1023
11. Lunghi C, Binda P, Morrone C (2010) Touch disambiguates rivalrous perception at early stages of visual analysis. Curr Biol 20(4):R143–R144
12. Bahrami B et al (2010) Optimally interacting minds. Science 329(5995):1081–1085
13. Kepecs A et al (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455:227–231
14. Koriat A (2012) When are two heads better than one. Science 336:360–362
15. Holzinger A, Bruschi M, Eder W (2013) On interactive data visualization of physiological low-cost-sensor data with focus on mental stress. In: Cuzzocrea A et al (eds) Multidisciplinary research and practice for information systems, springer lecture notes in computer science LNCS 8127: 469480. Springer, Heidelberg
16. Turkay C et al (2014) On computationally-enhanced visual analysis of heterogeneous data and its application in biomedical informatics. In: Holzinger A, Jurisica I (eds) Interactive knowledge discovery and data mining: state-of-the-art and future challenges in biomedical informatics., Lecture notes in computer scienceSpringer, Berlin, pp 117–140
17. Wong BLW, Xu K, Holzinger A (2011) Interactive visualization for information analysis in medical diagnosis. In: Holzinger A, Simonic KM (eds) information quality in ehealth., Lecture notes in computer scienceSpringer, Berlin, pp 109–120

18. Batallones A et al (2015) On the combination of two visual cognition systems using combinatorial fusion. Brain Inform 2:2132

19. Hsu DF, Taksa I (2005) Comparing rank and score combination methods for data fusion in information retrieval. Inform Retr 8(3):449–480

20. Li Y, Hsu DF, Chung SM (2013) Combination of multiple feature selection methods for text categorization by using combinatorial fusion analysis and rank-score characteristic. Int J Artif Intell Tools 22(2):1350001

21. Lyons DM, Hsu DF (2009) Combining multiple scoring systems for target tracking using rankscore characteristics. Inform Fus 10(2):124–136

22. Deng Y et al (2013) Sensor feature selection and combination for stress identification using combinatorial fusion. Int J Adv Robot Syst 10:306–313

23. Liu H et al (2013) A skeleton pruning algorithm based on information fusion. Pattern Recognit Lett 34(10):1138–1145

24. Lin KL et al (2007) Feature selection and combination criteria for improving accuracy in protein structure prediction. IEEE Trans NanoBiosci 6(2):186–196

25. Schweikert C et al (2012) Combining multiple ChIP-seq peak detection systems using combinatorial fusion. BMC Genomics 13(Suppl 8):S12

26. Ng KB, Kantor PB (2000) Predicting the effectiveness of naive data fusion on the basis of system characteristics. J Am Soc Inform Sci 51(12):1177–1189

27. Ho TK et al (1994) Decision combination in multiple classifier systems. IEEE Trans PAMI 16(1):66–75

28. Ho TK (1995) Random decision forests, Proceedings of the 3rd international conference on document analysis and recognition, Montreal, pp 278–282

29. Brown G, Wyatt J, Harris R, Yao X (2005) Diversity creation method: a survey and categorization. Inform Fusion 6:5–20

30. Kuncheva LI (2004) Combining pattern classifiers: methods and algorithms. Wiley-Interscience, Hoboken

31. Mulia D et al. (2015) Joint decision making on two perception systems using diversity rank-score function graph. Brain informatics and health, Guo Y et al. (eds): BIH 2015, LNAI 9250:337346

Christina Schweikert Ph.D., is an Assistant Professor in the Division of Computer Science, Mathematics and Science at St. John's University in New York, NY. She has been granted a Clare Boothe Luce professorship in computer science. She completed her Ph.D. degree in Computer Science from the City University of New York, Graduate Center and has previously taught at Fordham University. Her research interests include programming languages, data mining, biomedical, and healthcare informatics.

Darius Mulia received a B.S. in mathematics and a minor in computer science from Fordham University. His research interests include data science, problem solving, and performance optimization.

Kilby Sanchez received a B.A. and an M.S., both in computer science, from Fordham University. His research interests include the application of combinatorial techniques in econometrics, economic forecasting, decision-making, and financial modeling. He is also interested in the design of intelligent motion and signal detection systems using machine learning and data mining.

D. Frank Hsu Ph.D., is the Clavius Professor of Science and a Professor of Computer and Information Science at Fordham University in New York, NY. He received an M.S. from the University of Texas at El Paso, Texas, and a Ph.D. from the University of Michigan. His research areas include combinatorics and algorithms; network interconnections and communications; and informatics and analytics. He has co-developed "combinatorial fusion algorithm" and showed its applications in a variety of domain areas including cognitive science, bioinformatics, virtual screening, target tracking, and multimodal feature fusion. He has served on several editorial boards including Brain Informatics, Health Information Science (book monograph series), IEEE Transactions on Computer, IEEE Transactions on Reliability, International Journal of Foundation of Computer Science, Journal of Advanced Mathematics and Applications, Journal of Interconnection Networks (JOIN), and Networks.