

Discernibility in the Analysis of Binary Card Sort Data

Daryl H. Hepting* and Emad H. Almestadi

Department of Computer Science, University of Regina
3737 Wascana Parkway, Regina, SK, S4S 0A2 Canada
`{hepting|almestae}@cs.uregina.ca`

Abstract. In an open card sorting study of 356 facial photographs, each of 25 participants created an unconstrained number of piles. We consider all 63,190 possible pairs of photos: if both photos are in the same pile for a participant, we consider them as rated similar; otherwise we consider them as rated dissimilar. Each pair of photos is an attribute in an information system where the participants are the objects. We consider whether the attribute values permit accurate classification of the objects according to binary decision classes, without loss of generality. We propose a discernibility coefficient to measure the support of an attribute for classification according to a given decision class pair. We hypothesize that decision class pairs with the support of many attributes are more representative of the data than those with the support of few attributes. We present some computational experiments and discuss opportunities for future work.

1 Introduction

Card sorting [7] is an accessible technique to elicit data about participant impressions of various stimuli. We consider the analysis of data from a card sorting study of 356 facial photographs (178 Caucasian and 178 First Nations). The photographs were laminated on 5 by 4 inch cards. Participants were asked to view photos one at a time and place each photo on a pile with photos which they judged to be similar, without disturbing existing piles. The number of piles was not constrained. Within the 25 participants, the number of piles made ranged between 4 and 38. For each participant, photos in the same pile were considered to be rated as similar (distance of 0) and photos in different piles were considered to be rated as dissimilar (distance of 1). In this way, we attached a rating to each of the 63,190 pairs that can be made from 356 photos. Participants rated the similarity of each photo in relation to other photos. The smallest unit of this similarity judgement is the photo pair, so therefore the photo pairs are the attributes in this information system. Only a small fraction of these comparisons were made directly, specifically amongst the photo being placed and whichever photos were visible at the tops of existing piles. The study and a preliminary

* This paper benefitted from discussions with Dominik Ślęzak

analysis has been described elsewhere [4]. From that preliminary analysis, it was hypothesized that different strategies for sorting the photos may be used amongst the participants studied.

We continue to work at identifying and understanding the different strategies that may be at work. The earlier work looked for meaningful ways to distinguish between 2 groups. In particular, we looked at various qualities inherent in or identified about the photos as the basis for constructing decision class pairs. In this paper, we continue the search for identifiable strategies from a quantitative perspective. Although we still consider binary decision classes, each of these decision classes may be later further subdivided as required.

Gathering the ratings for each pair of photos (attribute) from each participant led to a binary vector of length 25 that became associated with the attribute. Some photos were not recorded during data entry, so the distance for pairs formed with these photos was -1. Our approach reported here replaced each -1 within these binary vectors with 0 and 1 in turn to generate all possible alternative patterns in new binary vectors. In cases when an attribute had an incomplete original binary vector (containing -1 values), the attribute became associated with all newly generated binary vectors. Any duplicate binary vectors were removed with the associated attributes moved to the single remaining instance of the vector. The result of this process was a list 28,379 unique binary vectors. Following Table 1, each of the vectors was assigned an ID. None of the unique vectors was the inverse of another vector in the list.

Table 1. Sample binary vectors. Each bit position represents a participant (object). The table shows the IDs associated with binary vectors: interpreted as integers, vectors are valued from 0 to $2^{n-1} - 1$ on the left and from $2^n - 1$ to 2^{n-1} on the right. Interpreted as a decision class specification, all objects with the same value are assigned to the same decision class. Therefore, a vector and its inverse have the same ID. The first row does not have an ID because the vector and its inverse do not contain both 0 and 1.

ID	Binary Vector	Inverse Vector
-	000	111
1	001	110
2	010	101
3	011	100

Each bit position in the binary vector represents a participant (object). These binary vectors have 2 possible interpretations. On the one hand, each vector represents the values for a particular attribute. A zero (0) indicates that the particular participant judged the photo pair to be similar (distance = 0). A one (1) indicates that the particular participant judged the photo pair to be dissimilar (distance = 1). On the other hand, each vector represents a possible way to assign the objects into 2 decision classes. Participants with the same value are assigned to the same decision class. (See Table 1 for more detail.)

The unique vectors distilled from participant data represent only a very small fraction of the total possible ways to divide 25 participants into 2 groups, yet

they are an appealing starting point because they record real participant behaviour. If we look amongst them for evidence of differing strategies employed by participants in the judgement of facial similarity, we may be encouraged to find attributes that allow a highly accurate classifier to be built for a particular binary decision class specification. We suggest that this is a necessary but not a sufficient criterion for identification of “good” decision class pairs. We suggest that a better indication of “good”-ness for a decision class pair is the number of attributes from which a highly accurate classifier can be built.

Our approach, reported here, has been to develop a measure of discernibility that can be easily computed and used to quantitatively assess how well a particular attribute can be used to discern objects according to a given decision class pair. These results were calibrated in a small test with the Rough Set Exploration System [2].

The rest of the paper is organized in the following way. Section 2 discusses discernibility and develops new measures related to discernibility. Section 3 details some computational experiments, including the use of the Rough Set Exploration System [2]. Section 4 presents some conclusions based on the obtained results and discusses some opportunities for future work.

2 Discernibility

Discernibility is a key idea in rough set theory [6, 8], and it can be applied here to understand participant judgements in 2 ways:

- by examining all judgements made by pairs of participants (objects): It is possible for a pair of participants to disagree about every attribute, in which case the participants would be readily discernible. It is also possible for a pair of participants to agree about every attribute, in which case the participants would be indiscernible.
- by examining all judgements made about each attribute: It is possible for all participants to agree with each other about an attribute, in which case the attribute would not contribute to the discernibility of the participants. It is not possible for all participants to disagree with each other about an attribute, because each participant rates an attribute as either “Similar” (0) or “Dissimilar” (1). For a given vector, the product of the number of 0’s and the number of 1’s indicates the amount of “disagreement” (discernibility). Equation 1 defines the maximum discernibility possible within a binary vector of length n .

We focus our attention here on those vectors with maximum discernibility (which contain either 12 zeroes and 13 ones or 13 zeroes and 12 ones). In this way, we hope to focus on the most informative attributes [1]. By doing so, we are left with 1705 vectors out of the total 28,379 with which we began.

In these vectors, 156 out of the 300 possible pairs of participants are different (either 01 or 10) and only $300 - 156 = 144$ of the possible pairs of participants are the same (either 00 or 11). Beginning with a vector that specifies the binary

decision classes, we wish to compare it with attribute vectors to see how well the decision class pair represents the observed data.

Choi *et al.* [3] present 75 different ways to assess the similarity between 2 binary vectors. The task of assessing the discernibility of a binary vector with respect to another is somewhat different. As outlined in Table 1, we consider that a vector and its inverse represent the same assignment of objects to decision classes. This interpretation is different than Janusz and Ślęzak [5], for example, who regarded inverse vectors as complementary rather than similar. In our case, we are concerned with values on diagonals of the contingency table (see Table 2).

Table 2. Contingency table consistent with Choi *et al.* [3]. Rows labelled as x_0 and x_1 indicate respectively 0's and 1's in vector x . Columns labelled as y_0 and y_1 indicate respectively 0's and 1's in vector y . $D_{\text{coeff}}(x, y) = 1$ if $a + d = n$ or $b + c = n$.

	y_0	y_1	sum
x_0	a	c	$a + c$
x_1	b	d	$b + d$
sum	$a + b$	$c + d$	$a + b + c + d = n$

$$D_{\max} = \begin{cases} \left(\frac{n}{2}\right)^2 & \text{when } n \text{ is even} \\ \left(\frac{n}{2}\right) \times \left(\left(\frac{n}{2}\right) + 1\right) & \text{when } n \text{ is odd} \end{cases} \quad (1)$$

$$D_{\text{coeff}}(x, y) = \frac{ad + bc}{D_{\max}} \quad (2)$$

$$D_{\text{dist}}(x, y) = 1 - D_{\text{coeff}}(x, y) \quad (3)$$

Given n objects, there will be $\binom{n}{2}$ pairs of objects. Consider that each of these objects is assigned to 1 of 2 decision classes. Pairs of objects from different decision classes will be discernible with respect to an attribute if the values for that attribute are different for these pairs of objects. Equation 1 defines the maximum number of object pairs with objects from different decision classes. $D_{\text{coeff}}(x, y)$, defined in Equation 2, compares 2 binary vectors (one a decision class specification and the other containing attribute values) and computes the number object pairs from different decision classes that have different attribute values over the maximum number of such pairs. The range for the $D_{\text{coeff}}(x, y)$ is $[0, 1]$. Notice that $D_{\text{coeff}}(x, y) = D_{\text{coeff}}(y, x)$ for any pair of vectors, x and y . The coefficient is meant to answer the question “Does the attribute with values given by x help to discern objects in decision classes specified by y ?” If $D_{\text{coeff}} = 1$ (or close to it), the answer is “Yes”. Either ad or $bc = D_{\max}$, which means that the attribute values match the decision class specification (or its inverse) exactly (what was earlier called a “splitting pair” [4]). If $D_{\text{coeff}} = 0$ (or close to it), the answer is “No”. Either ac or $bd = D_{\max}$, and the attribute contributes nothing to the discernibility of the decision classes. This value will only occur if all of the attribute values are the same. Equation 3 defines a distance in terms of D_{coeff} .

Table 3. Three sample binary vectors, labelled as A, B, and C for brevity, are compared. (Following the convention outlined in Table 1, their numerical IDs are as follows: A = 350655, B = 350639, and C = 11184810.) To the right of each pair of vectors is the contingency table for the comparison. $D_{\text{coeff}}(A, B) = 144/156 = 0.923$ and $D_{\text{coeff}}(A, C) = (56 + 25)/156 = 0.519$.

ID	Values																
A	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	0	1
B	1	1	1	1	1	0	1	0	1	0	0	1	1	0	0	1	0
ID	Values																
A	0	0	0	0	0	1	0	1	0	1	1	0	0	1	1	0	1
C	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

		B_0	B_1
A_0	0	12	
A_1	12	1	

		C_0	C_1
A_0	7	5	
A_1	5	8	

3 Experimentation

Each of the 1705 vectors, in turn, was interpreted as the decision class specification, in preparing to apply the rough set attribute reduction methodology [6]. D_{coeff} was computed for all attribute vectors with respect to the given decision class specification, and the average coefficient was computed for each candidate decision class specification. We then chose the vectors with the maximum and minimum average, represented in Table 4. In addition to the interpretation of support for a decision class specification, the average coefficient can also be interpreted as a measure of the importance of the attribute(s) associated with each vector. In this case, the coefficient answers the question “Is this attribute important in discerning objects?” If $D_{\text{coeff}} = 1$ (or close to it), the answer is “Yes”. If $D_{\text{coeff}} = 0$ (or close to it), the answer is “No”.

Table 4. Max. (D_{coeff} average = 0.590) and Min. (D_{coeff} average = 0.515) vectors, compared. $D_{\text{coeff}}(\text{max}, \text{min}) = 0.5$.

ID	Values																
Max.	1	1	1	1	1	0	1	0	1	0	0	1	1	0	0	1	0
Min.	0	0	1	0	1	0	1	1	0	1	0	1	0	0	1	1	0
ID	Values																
	Max ₀	Max ₁															
Min ₀	7	6															
Min ₁	6	6															

Hepting *et al.* [4] focused on reducing the number of attributes required as input to RSES [2] in order to accurately classify participants according to a decision class pair. Instead of looking only for the existence of an accurate classification via RSES, this work is concerned with exploring the limits of an accurate classification: how many different attributes support accurate classification according to a specified decision class pair? We hypothesize that candidate decision class pairs with the support of many attributes are more representative of the data than those with the support of few attributes.

For both of the Min. and Max. vectors, we created bins for D_{coeff} between 0.5 and 1.0 in increments of 0.05. RSES input files were generated to test the

classification accuracy using up to 25 attributes from only 1 specified bin. The bins, the number of attributes in each bin, and the average coefficient value for the bin are indicated in Table 5.

After creating the various input files, we followed a standard procedure with RSES [2], as follows: Preprocessing: split the input table of 25 objects into 2 equal parts (1 for training and 1 for testing); Training: calculate up to 10 reducts from the training data using the genetic algorithm in RSES; Testing: generate rules from the reducts and test the results by classifying the testing data. Each input file was processed 10 times and the averages are reported in Table 5: average accuracy (including standard deviation), average reduct length, and average number of rules. All results had 100 percent coverage, which means that the classifier based on the reducts generated from an ensemble of reducts was able to recognize everything, which is valuable in itself.

Table 5. Summary of results from 2 sets (Max., left and Min., right) of runs of RSES. Data from each bin was run 10 times and averages are reported. Dashes indicate that the bin had no data.

Bin	Nbr. Attr.	Avg. Coeff.	Avg. Acc.	Std. Dev.	Avg. Red.	Avg. Rule	Bin	Nbr. Attr.	Avg. Coeff.	Avg. Acc.	Std. Dev.	Avg. Red.	Avg. Rule
0.95	0	-	-	-	-	-	0.95	0	-	-	-	-	-
0.90	5	0.923	1	0	1.32	12	0.90	0	-	-	-	-	-
0.85	15	0.853	0.977	0.037	1.88	33.8	0.85	0	-	-	-	-	-
0.80	25	0.845	0.992	0.024	1.61	29.4	0.80	0	-	-	-	-	-
0.75	25	0.788	0.969	0.040	1.85	35.8	0.75	0	-	-	-	-	-
0.70	25	0.731	0.915	0.067	2.46	51	0.70	0	-	-	-	-	-
0.65	25	0.692	0.862	0.087	1.92	37.6	0.65	5	0.673	0.823	0.089	2.36	19.4
0.60	25	0.636	0.846	0.103	2.9	65.7	0.60	25	0.634	0.838	0.133	2.67	57.1
0.55	25	0.596	0.808	0.075	2.6	54.8	0.55	25	0.596	0.769	0.103	2.84	62.8
0.50	25	0.545	0.746	0.115	3.53	85	0.50	25	0.545	0.662	0.121	3.64	82.2

Max.

Min.

To explore some of the data in Table 5 in more detail, each of Figure 1 (for the Max. vector) and Figure 2 (for the Min. vector) illustrate the attribute(s) associated with the vector, a reduct generated from the whole (unsplit) data table taken from the respective top bin , and the rules associated with that reduct.

4 Conclusions and Future Work

The computation of the D_{coeff} is an appealing approach to understanding the structure of results of card sorting exercises because it can be done very quickly. The limited experiment presented here has provided encouraging support for our hypothesis, but more work needs to be done. For example, for the same average coefficient value, is it better to have fewer attributes with higher coefficient or more attributes with lower coefficient values?

Rule	Dec. Class	Nbr. Classified
EF = Similar & AG = Similar	0	11
EF = Similar & AG = Dissimilar	0	1
EF = Dissimilar & AG = Similar	0	1
EF = Dissimilar & AG = Dissimilar	1	12

Fig. 1. Photo pairs AB and CD are associated with vector “Max”. Pairs EF and AG are the attributes in one of the reducts, followed by corresponding rules for classification.

Rule	Dec. Class	Nbr. Classified
JI = Similar & KL = Similar & MN = Dissimilar	0	2
JI = Similar & KL = Dissimilar & MN = Dissimilar	0	8
JI = Dissimilar & KL = Dissimilar & MN = Dissimilar	0	3
JI = Similar & KL = Dissimilar & MN = Similar	1	2
JI = Dissimilar & KL = Similar & MN = Similar	1	5
JI = Dissimilar & KL = Similar & MN = Dissimilar	1	4
JI = Dissimilar & KL = Dissimilar & MN = Similar	1	1

Fig. 2. Photo pair HI is associated with vector “Min”. Pairs JI, KL, and MN are the attributes in one of the reducts, followed by corresponding rules for classification.

Many of the vectors with highest average coefficients are close to each other. This leads to opportunities to analyze the structure of the decision classes (representing strategies) that are best-supported by the data. By the same token, the similarity of the photo pair attributes associated with the Max. and Min. vectors respectively are noticeably different - something that provides more support for this approach.

In hindsight, it is becoming clear that too many (356) photos were used in the original sorting study. The process outlined here has the potential to sharply reduce the number of photos considered. If this process can successfully determine important attributes (such as photo pairs AB and CD in Figure 1), it may be possible to effectively run card sorting studies with a large number of stimuli that could be reduced based on this kind of quantitative analysis.

It is not possible to assess how well 2 decision classes are formed without testing all potential decision classes. There are 16,777,215 ($2^{24} - 1$) ways to create 2 decision classes for 25 participants, and the inexpensive computation of D_{coeff} can facilitate their review.

Acknowledgements This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Emad Almestadi acknowledges the Ministry of Higher Education in Saudi Arabia and the Saudi Arabian Cultural Bureau in Canada for their support. The comments of the anonymous reviewers were very helpful in improving the final version of this paper.

References

1. Bazan, J.G., Nguyen, H.S., Nguyen, S.H., Synak, P., Wróblewski, J.: Rough set algorithms in classification problem. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (eds.) *Rough Set Methods and Applications, Studies in Fuzziness and Soft Computing*, vol. 56, pp. 49–88. Physica-Verlag HD (2000)
2. Bazan, J.G., Szczuka, M.: The rough set exploration system. *Transactions on Rough Sets III LNCS* 3400, 37–56 (2005)
3. Choi, S.S., Cha, S.H., Tappert, C.C.: A survey of binary similarity and distance measures. *Journal on Systemics, Cybernetics and Informatics* 8(1), 43–48 (2010)
4. Hepting, D., Spring, R., Ślęzak, D.: A rough set exploration of facial similarity judgements. *Transactions on rough sets XIV* pp. 81–99 (2011)
5. Janusz, A., Ślęzak, D.: Utilization of attribute clustering methods for scalable computation of reducts from high-dimensional data. In: 2012 Federated Conference on Computer Science and Information Systems (FedCSIS). pp. 295–302 (2012)
6. Pawlak, Z.: Rough set approach to knowledge-based decision support. *European Journal of Operational Research* 99(1), 48–57 (May 1997)
7. Rugg, G., McGeorge, P.: The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems* 22(3), 94–107 (July 2005)
8. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Slowinski, R. (ed.) *Intelligent Decision Support: Handbook of Applications and Advances in Rough Set Theory*, vol. 11, pp. 259–300. Kluwer Academic Publishers (1992)