Proceedings of the 4th International Conference on Statistics: Theory and Applications (ICSTA'22)
Prague, Czech Republic – July 28- 30, 2022
Paper No. 110
DOI: 10.11159/icsta22.110

Mutual Information in the Analysis of Trust Gains from Subsets of Information

Ronit Bustin and Claudia V. Goldman
General Motors, R&D Technical Center
Herzliya, Israel
ronit.bustin@gm.com; claudia.goldman@gm.com

Abstract **-** Information can increase trust of humans in automated machines. However, assessing the impact of all combinations of information pieces on the trust level of humans might not be practical. This paper assumes that data can be collected from human participants having interacted with some automated machine. We assume a two-stage study in which the participants initially submit their ranking (trust level) when no information is provided, and then provide additional independent rankings for each piece of additional information. The goal is to determine the best combination of information pieces over all combinations without directly asking the participants to rank the possible combinations. The impact of the combinations on the trust ranking is evaluated using the mutual information quantity. We further consider the question of statistical significance in this unique setting, and suggest an optimization objective that examines the trade-off between the impact of the subset on the trust measure, on the one hand, while considering the complexity of the subset, measured by the size of the subset (number of additional pieces of information), on the other hand. We provide a numerical example that shows all aspects discussed in this work.

Keywords**:** G-test, statistical significance, mutual information, information theory

1. Introduction

Consider a user study where the goal is to understand the effect of different pieces of information, that are not necessarily independent, on some trust measure. The trust measure could be the participants understanding of a given scenario (*e.g.*, some contextual experience resulting from the participant interaction with the machine in an actual setting or a simulated setting) and the pieces of information could be explanations of that scenario. In addition, let us now assume that we are also interested in understanding the effect of the possible combinations of these pieces of information on the trust measure. Such a study can be designed to examine the user-interface of an automated machine. The machine may provide explanations or instructions to the user in different forms and in different scenarios. These explanations/instructions can be stand-alone, or they could be combined. The goal of the study could be to measure, for example, the users' understanding of the scenario given different combinations of explanations/instructions [1].

One straightforward approach would be to ask the participants to rank the trust measure for each combination, however, this approach becomes infeasible when the set of possible pieces of information is larger than 4, since the number of possible combination is two to the power of the number of pieces of information (thus, in the case of 4 pieces of information we have 16 possible combinations, but if we have 5 pieces of information we already have 32 combinations! We cannot ask the participants to provide ranking for each possible subset). In this paper we formalise this problem, suggest an information theoretic approach to resolve it, provide claims and tools to determine statistical significance measures, and finally suggest an optimization objective that takes into account the trade-off between the additional information gained by the combination of pieces of information and the complexity (measured by the number of pieces of information combined).

We begin with a user-study with the following structure. Participants are asked to rank a scenario. The ranking is over some finite set, denoted as $\mathcal{D}$. This initial ranking will be denoted $rank_0$. After doing so, they are presented with K options for additional information (for example, a sentence describing a behavior occurring in the scenario, or some highlighting of certain objects in the scenario) denoted as e_i. Such additional information might affect the trust level of the participants as it may clarify confusion. The set of these additional pieces of information is denoted as $\mathcal{S}$. For each e_i the participant is then requested to rank again the scenario, now assuming the additional information is provided. Their ranking for each of the additional pieces of information, e_i, is denoted as $rank_i$. The above process has two important aspects. First, this is a two-stage process. The participant is presented with the same question without e_i and then with e_i. Second, the participant is presented with multiple options of additional information from the set $\mathcal{S}$ and is requested to rank each one separately (independently).

The fact that the study has two stages opens the door to examine the change in the participants response (in their trust measure). The change itself then becomes the target of the analysis. In other words, after measuring the change, one can evaluate the probability for a participant (with specific attributes, such as gender or age) to have a specific change in their ranking of the specific scenario. This approach allows to distinguish between low impact additional pieces of information, that provided no change or a small change, as opposed to high impact explanations that caused a participant to change their response considerably. Still this approach views each of the additional explanations as a separate explanation and does not examine the possible dependencies. For more on this we refer the reader to [2].

On the other hand, asking a ranking for each individual piece of information, under the independence assumption, allows us to estimate the ranking of each subset. We focus on the following question: *What is the "best" subset of explanations for the specific scenario, with respect to the given rank/trust measure?* Since the goal of this method is to obtain a subset of explanations, rather than a single explanation, the dependencies among the explanations become relevant. A central measure used in our determination of the "best" subset is the mutual information measure. Mutual information measures the mutual dependency between any two sets of random variables, or, in other words, the amount of information the sets provide on one another [3]. Assume a pair of discrete random variables (X, Y) (taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively), the mutual information is defined by their join probability $\Pr(\mathrm{X}, \mathrm{Y})$ as follows:

$$I(X,Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \Pr(X = x, Y = y) \log \frac{\Pr(X = x, Y = y)}{\Pr(X = x)\Pr(Y = y)}. \tag{1}$$

In order to make such a measure relevant for our purpose we need to carefully define the random variables it will consider. The main issue is that mutual information measures dependency, whereas we are after the subset of additional pieces of information that maximize the rank (or expected rank).

Another question that must be considered is that of statistical significance. Recall that we are comparing the above-mentioned mutual information quantity but defined over different subsets of additional pieces of information. Thus, our support is not the same. Still we are interested in determining whether the values we obtain for the mutual information are indeed statistically distinguishable. Finally, we suggest a specific optimization objective that takes into consideration the mutual information quantity but not only. The optimization objective explores the trade-off between the amount of information the subset provides with respect to the rank/trust measure, and the complexity of the subset, which is measured by the cardinality of the subsets, *i.e.*, the number of additional pieces of information contained in the subset.

2. Assumptions

In this section we provide the basic assumptions required from the design of the user study. We explain the basis for these assumptions, the need for these assumptions and whether they can be validated from the data collected.

A basic required assumption is the following:

Assumption 1: Assume the initial ranking, for a specific participant, for a given scenario, is $rank_0$ and their ranking in the second stage, given each one of the additional pieces of information in $\mathcal{S} = \{e_1, e_2, \dots, e_K\}$ is denoted as $rank_i$ for $i \in \{1.2. \dots, K\}$, respectively. Then, if

$$\mathrm{rank}_0 \le \mathrm{rank}_\mathrm{i}, \quad \forall \mathrm{i} \in \{1,2, \dots \mathrm{K}\} \tag{2}$$

the study is monotonic.

The monotonicity assumption can easily be verified and shown to hold with high probability (we will show this later on in our example user study). Thus, this could be a required property rather than an assumption. The next two assumptions differ in this sense, meaning one cannot easily verify them. The second assumption is that of independence.

Assumption 2: The participants rank each additional piece of information independently from their ranking of each of the other pieces of information.

The basis for this assumption is an explicit request from the participant to rank the additional pieces of information independently. Note that the first two assumptions are characteristics that either exist or do not exist in the performed study. The last assumption differs in this sense, as it extends beyond the scope of the performed study. Our third and main assumption is the following:

Assumption 3: The rank provided by a participant when presented with a combination of several pieces of additional information can only increase compared to their ranking when presented with the individual pieces of information.

The above assumption requires foremost that the explanation set $\mathcal{S}$ will not contain contrasting/contrudicting pieces of information. However, it goes beyond this, since as the explanation becomes more complex, containing more and more pieces of information it will begin to confuse and reduce the overall trust/understanding. We take this effect into account in the sequel.

As mentioned, asking participants to rank every combination of the additional pieces of information is infeasible. The above assumption allows us to estimate the value a participant would provide for a combination of additional pieces of information using his/her ranking of each individual piece of information. For example, the ranking a participant would provide if presented with the following set of additional pieces of information $\{e_i, e_j, e_k\}$ would be:

$$rank_{e_i,e_j,e_k} = \max\{rank_i, rank_j, rank_k\}. \tag{3}$$

3. Definitions and Basic Quantities

In this section we provide the basic definitions that allow us to extract from the user study the quantities of interest. Recall that our goal is to extract mutual information (1) quantities that reflect the impact of combinations of explanations on some trust measure (given in the value of $rank$). Once we carefully define the relevant random variables, we require the calculation of their joint probability so as to calculate the mutual information. We show how these quantities are to be calculated.

We begin with several basic definitions:

Definition 1: The probability of receiving any one of the single pieces of information is denoted as P_e.

Definition 2: We define binary random variables b_i. $b_i = 1$ with probability P_e (see Definition 1) and $b_i = 0$ with probability $1 - P_e$. When $b_i = 1$ we have that e_i is presented to the participant, whereas if $b_i = 0$ we have that e_i is not presented to the participant.

Definition 3: The collection of all b_is is denoted as $\boldsymbol{B} = \{b_1, b_2, \dots, b_k\}$.

Definition 4: The power set of $\boldsymbol{B}$, meaning the set of all subsets over $\boldsymbol{B}$ is denoted by $\mathcal{B}$. The power set includes 2^K elements. These elements are denoted by B.

Before proceeding to the definition of the random variable $rank$ we require the definition of $rank_{\boldsymbol{b}}$. Consider a specific realization of $\boldsymbol{B}$, meaning some value from $\{0,1\}^K$, which we denote as $\boldsymbol{b}$, $rank_{\boldsymbol{b}}$ is evaluated using Assumption 3 by taking the maximization over $rank_{e_i}$for each $b_i = 1$ and $rank_0$ for each $b_i = 0$. For example, if $\boldsymbol{b} = \{b_1 = 0, b_2 = 1, b_3 = 1, b_4 = 0\}$ our ranking will be $rank_{\boldsymbol{b}} = max\{rank_0, rank_{e_2}, rank_{e_3}, rank_0\}$. Note that due to Assumption 1 we don't need to include $rank_0$ for each i for which $b_i = 0$, but in the case where all $b_i = 0$ this transition guarantees that $rank_{\boldsymbol{b}}$ is well defined.

Given the above definition we proceed to the definition of the random variable $rank$. This random variable takes values from the set $\mathcal{D}$. Its probability is extracted from the data collected in the study and the assumptions provided in Section 2.

We assume N participants in the study. For a specific realization of $\boldsymbol{B}$, meaning some value from $\{0,1\}^K$ (denoted as $\boldsymbol{b}$) and $\mathcal{D} = \{1,2,3,4\}$ the conditional probability Pr(rank|**B**) is defined as follows (recall that $rank$ takes values over $\mathcal{D}$):

$$rank = \begin{cases} 1, w.p. \dfrac{\#pct. with\, rank_{\mathbf{b}} = \mathbf{1}}{N} \\ 2, w.p. \dfrac{\#pct. with\, rank_{\mathbf{b}} = \mathbf{2}}{N} \\ 3, w.p. \dfrac{\#pct. with\, rank_{\mathbf{b}} = \mathbf{3}}{N} \\ 4, w.p. \dfrac{\#pct. with\, rank_{\mathbf{b}} = \mathbf{4}}{N} \end{cases} \tag{4}$$

This can be extended to any finite set $\mathcal{D}$.

Equation (4) defines the conditional probability $\Pr(rank|\mathbf{B} = \mathbf{b})$. To complete this and obtain the joint probability we use Definitions 1 and 2 together with Assumption 2 of independence, allowing us to obtain the probability $\Pr(\mathbf{B} = \mathbf{b})$ for any $\boldsymbol{b} \in \{0,1\}^K$, as follows:

$$\Pr(\mathbf{B} = \mathbf{b}) = \prod_{i \in [1,K]} P_e^{b_i}(1 - P_e)^{1-b_i} \tag{5}$$

where b_i is the i^{th}element in $\boldsymbol{b}$. There are two options here. The first picking a uniform probability over the subsets of explanations. This would be the case if on top of Assumption 2 we would also assume that each explanation is given or not given with equal probability, meaning $P_e = 0.5$. The second option is to assume a different probability for giving or not giving any one of the explanations, $P_e \neq 0.5$. In our study we used the later as will be explained in Section 6.

So far, we have obtained the joint probability $\Pr(\text{rank}, \mathbf{B})$. We wish to extend this to any $B \in \mathcal{B}$ using standard marginalization. Assume some subset $B \in \mathcal{B}$, and denote its complementary subset $\overline{\mathrm{B}} = \mathbf{B} \setminus \mathrm{B}$. The calculation is as follows:

$$\Pr(rank, B) = \sum_{\bar{B}} \Pr(rank, \mathbf{B}) = \sum_{\bar{B}} \Pr(rank|\mathbf{B}) \Pr(\mathbf{B}) \tag{6}$$

We have now obtained the joint probability $\Pr(rank, B)$ for all $B \in \mathcal{B}$. We can now use it to calculate the mutual information:

$$I(rank; B) = \sum_{rank \in \mathcal{D}} \sum_{B \in \{0,1\}^{|B|}} \Pr(rank, B) \log \frac{\Pr(rank, B)}{\Pr(rank)\Pr(B)}. \tag{7}$$

The first summation is over all values of $rank$. The second summation is over all possible realizations of B (which depends on the number of elements in B, denoted as $|B|$). As an example for the calculation of the join probability in (6) consider the set $\boldsymbol{B} = \{b_1, b_2, \dots, b_8\}$ and $B = \boldsymbol{B} \setminus \{b_6\}$, meaning we exclude a single element from the set $\boldsymbol{B}$. In this case $\bar{B} = \{b_6\}$ and we the summation in (6) is only over $b_6 = 0$ and $b_6 = 1$. Surely, if we choose to extract other pieces of information we will need to take that into account in the above summation and perform the summation over all combinations of the extracted b_is.

4. Statistical Significance

In the previous section we have shown how to calculate the mutual information quantity between $rank$ and any $B \in \mathcal{B}$ from the data collected during the user study. We now want to understand how to determine statistical significance for this calculation, or more specifically, to determine whether two mutual information quantities that we obtain are significantly different.

Let us assume that we are comparing two sets B_1 and B_2 both of the same cardinality $K - x$ (meaning that we extracted from $\boldsymbol{B}$ x random variables). Following the calculation of the conditional probability $Pr(rank|\boldsymbol{B})$ and the derivation of the join probability for any B in (6) we obtain two joint probabilities:

$$P(rank, B_1) \equiv \Pr(rank, B_1) \;\; and \;\; Q(rank, B_2) \equiv \Pr(rank, B_2). \tag{8}$$

We denote them as P and Q to simplify our derivation in the sequel. Each one of these joint probabilities is basically a table with $\mathcal{D}$ rows, for each value of $rank$, and 2^{K-x} columns for each possible combination of the binary random variables in B_1or in B_2. The tables are similar in size, and even in the values (the columns are combinations from $\{0,1\}^{K-x}$), but B_1 and B_2 are different sets of random variables. Consider for example a scenario where $K = 8$ and $x = 4$, and we are comparing $B_1 = \{b_1, b_2, b_3, b_4\}$ with $B_2 = \{b_5, b_6, b_7, b_8\}$. If we preserve the order in the above sets as is, that means that we have mapped as follows: $b_1 \leftrightarrow b_5$, $b_2 \leftrightarrow b_6$, $b_3 \leftrightarrow b_7$ and $b_4 \leftrightarrow b_8$. This is just one of 4! possible mappings. These different mapping correspond to a different order of the columns in the join probability table. We refer to each such mapping as a *match* between B_1 and B_2. Thus, the question of statistical significance in this setting requires some additional thought.

Our main result is the following:
Theorem 1: The G-value used for determining the statistical significance of the difference between $I_P(rank; B_1)$ and $I_Q(rank; B_2)$ can be bounded by:

$$G - value \geq 2N\left(I_P(rank; B_1) + I_Q(rank; B_2) - 2H(rank) - 2H(B_1) + 2\min_{\mathcal{E}}\left\{H_{\frac{1}{2}(P+\mathcal{E}(Q))}(rank, B_1)\right\}\right) \tag{9}$$

Where N is the number of participants, and $\mathcal{E}$ is any permutation of the elements in B_2 (a mapping), and H denotes the entropy [3].

Essentially, we lower bound G-value, by finding the match $\mathcal{E}$ between B_1 and B_2 that minimizes the entropy over the average joint probability. This average is our reference join probability. By obtaining a lower bound we ensure that we comply with the required confidence levels.

Proof Sketch: The proof requires two important observations (proofs are omitted due to the page limit):
Lemma 1: The entropy of $rank$ is independent of the set B and its cardinality, meaning $H_P(rank) = H_Q(rank) \equiv H(rank)$ for any P and Q, defined for sets B and $\tilde{B}$ of cardinalities $K - x$ and $K - \tilde{x}$, respectively.

This is as expected and validates our construction.
Lemma 2: Given Assumption 2 of independence, the entropy of B of a given cardinality $K - x$ is independent of the specific set, meaning, $H_P(B_1) = H_Q(B_2) \equiv H(B_1)$ for any P (joint probability over set B_1) and Q (joint probability of set B_2), given that the two sets are of the same cardinality.

Recall that we have two sets of random variables B_1 and B_2 of the same cardinality $|B_1| = |B_2|$, thus we have two joint probability tables, denoted as P and Q. We apply the G-test, due to its popularity and its clear connection to information theoretic quantities [4]. We want to emphasize that our goal is to test the statistical significance of one joint probability compared to another. It is well known that the G-value can be written using known information quantities. Our contribution here is that in our specific setting a different manipulation is required in order to simplify the problem and extract a lower bound on the G-value.

In the G-test we compare the observed value with an expected value. The expected value when we compare two joint probabilities is set to be an average of their observations for each value in the support. We assume a given match between the two sets, denoted by $\mathcal{E}$. This match is a permutation on the set B_2. Once this permutation is performed we assume that the two sets are matched, meaning $B_1 = \mathcal{E}(B_2)$. We calculate the expected value as follows:

$$\frac{\mathrm{N}}{2}\Big(P(rank, B_1) + Q\big(rank, \mathcal{E}(B_2)\big)\Big). \tag{10}$$

We can manipulate the above to obtain mutual information quantities (1) and entropy quantities that are independent of $\mathcal{E}$, and obtain the expression in (9) (before minimization). To obtain this, we require Lemma 1 and 2. Note that in (9) the dependency on the specific match, $\mathcal{E}$, between the two join probabilities P and Q appears only in the last argument. The smaller G-value is, the less statistical significant the results are. Thus, our goal is to obtain the minimizing value of (9) over the possible matches $\mathcal{E}$. This will be attained by minimizing $H_{\frac{1}{2}(P+Q)}\left(rank, \mathcal{E}(B_2)\right)$. In other words, we want the entropy of the match to be as small as possible (furthest away from the uniform distribution), and this will provide us with a lower bound on the G-value. This concludes the proof.

■

Surely, we degrade the G-value, which is not necessary if one knows the matching. However, in our setting the correct matching is unknown, and we must take this minimization in order to guarantee that our results comply with a required confidence level.

5. Selection of Optimal Subset

The above mutual information provides us with a measure of the dependency between the random variable $rank$ and the specific set of binary random variables in B. A high value would indicate that $rank$ is highly dependent on the set of random variables in B, whereas a low value would indicate otherwise. This allows us to examine which subset of additional pieces of information has higher impact on the value of rank. From the chain rule of mutual information we have that:

$$I(rank; \mathbf{B}) = I(rank; B) + I(rank; \bar{B}|B) \tag{11}$$

thus, $I(rank; \mathbf{B}) \geq I(rank; B)$. Similarly, the above is true for any sequence of inclusions, meaning $\tilde{B} \subset B$ we have that $I(rank; B) \geq I\left(rank; \tilde{B}\right)$. Thus, there is no point in maximizing the mutual information alone, as the maximizing set would be $\boldsymbol{B}$. Given the above discussion we propose the following optimization:

$$\max_{B \in \mathcal{B}} I\,(rank; B) - \beta|B| \tag{12}$$

In this optimization there is a penalty due to the cardinality of the set B, denoted by $|B|$. In Section 6 we show the application of this optimization in a specific user study example.

6. Example User Study

In this section we describe, as an example, a user study design, and provide some numerical results that allow us to calculate the above quantities. In this study the participants were asked to answer a question considering their trust in the scenario. The possible answers translate to a ranking over $\mathcal{D} = \{1,2,3,4\}$, 1 corresponding to the lowest level of trust and 4 to the highest level of trust. When a participant's initial ranking ($rank_0$) was either 1or 2 they were presented with additional information $\mathcal{S} = \{e_1, e_2, \dots, e_8\}$, and were requested to rank each additional piece of information independently. Note that we gave the additional explanations only to those participants that $rank_0 \in \{1,2\}$, meaning those that gave an initial low ranking. This determines the probability to receive the additional information, meaning P_e (as opposed to setting it to equal 0.5). This required us to make an additional assumption (extension of the monotomicity assumption, Assumption 1). We recommend putting all participants through both stages to get the full resolution in the collected data.

6.1. Monotonicity

In the above we mentioned the monotonicity assumption (Assumption 1). As mentioned, in a two-stage user study such an assumption can be examined and validated simply by comparing each participant's answer in the first stage to their answer in the second stage. In our case, approximately 10% of the answers do not comply with the monotonicity assumption. Due to this, these answers were not considered in the calculation of the mutual information.

6.1. Mutual Information Calculation

We calculated the mutual information of different sets B and the $rank$ measure. Figure 2 shows the results of calculating the mutual information from our collected data following two greedy methods, which we refer to as max-greedy and min-greedy. In both methods we begin from $I(rank; \boldsymbol{B})$. In the max-greedy method, at each step, we remove the additional information that caused the least impact on the value of the mutual information. In the min-greedy method, at each step, we remove the additional information that had the largest impact on the value of the mutual information.

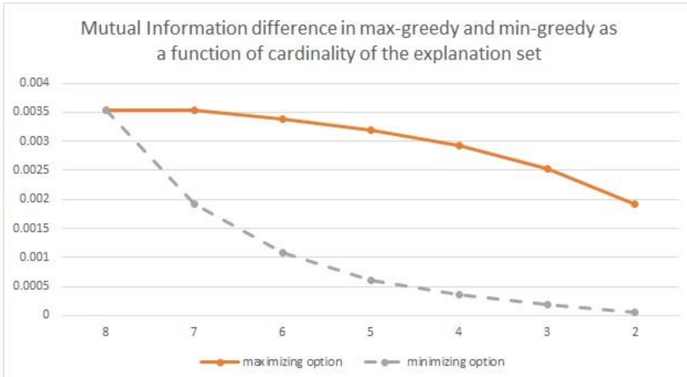

Fig. 1: The mutual information value as a function of cardinality, $|B|$. In orange (solid) we observe a series of mutual information values that were chosen according to the max-greedy method. In gray (dashed) we observe a series of mutual information values that were chosen according to the min-greedy method.

6.2. Statistical Significance

We want to examine our main result, given in Theorem 1, regarding the statistical significance of the mutual information values. We begin by setting our significance level to $\alpha = 0.1$ (meaning a confidence level of 90%). As shown in Lemma 1 the entropy of the $rank$ is a constant value throughout the sets (regardless of the joint probability or the cardinality), and in our setting its value is $H(rank) = 1.84798139428937$ bits. As shown in Lemma 2 the entropy of the sets is constant for a given cardinality. If we examine a set of cardinality 3 (and we will soon explain why we pick this cardinality) the entropy equals 2.96478666326454 bits.

We now want to examine the statistical significance of our mutual information values for the following two sets: $B_1 = \{b_3, b_4, b_7\}$ and $B_2 = \{b_1, b_5, b_6\}$. The mutual information values that we obtain for these two sets are the ones that you see in Figure 1 and are: $I(rank; B_1) = 0.00252712$ bits and $I(rank; B_2) = 0.000178968$ bits. In order to evaluate a lower bound on the G-value, according to Theorem 1, we need to examine all possible permutations, meaning compare 3! = 6 values for the entropy of the averaged joint probabilities. We provide all values in Table 2. The minimizing combination results with an entropy of 4.81181579481787 bits.

Table 1: Entropy of the averaged distribution as a function of the match $\mathcal{E}$.

Match $\mathcal{E}$	$\mathrm{H}_{\frac{1}{2}(\mathrm{P}+\mathcal{E}(\mathrm{Q}))}(\mathrm{rank}, \mathrm{B}_1)$
$\mathrm{b}_1, \mathrm{b}_5, \mathrm{b}_6$	4.8118846744431 bits
$\mathrm{b}_1, \mathrm{b}_6, \mathrm{b}_5$	4.81181579481787 bits
$\mathrm{b}_6, \mathrm{b}_1, \mathrm{b}_5$	4.81183638223602 bits
$\mathrm{b}_6, \mathrm{b}_5, \mathrm{b}_1$	4.81183638223602 bits
$\mathrm{b}_5, \mathrm{b}_6, \mathrm{b}_1$	4.8118258541467 bits
$\mathrm{b}_5, \mathrm{b}_1, \mathrm{b}_6$	4.81187573400737 bits

Placing all the above values into our lower bound we obtain that $G - value \geq N \cdot 0.00160312$. Since we are examining a set of cardinality 3 we have 2^3 columns, and 4 rows ($|\{D\}|$), and our degrees of freedom are: $(2^3 - 1) \cdot (4 - 1) = 21$. Due to the high degrees of freedom, we need N to be considerably high, $N = 19{,}000$, to ensure the statistical significance here, meaning to ensure that the resulting p-value is lower than our significance level of $\alpha = 0.1$.

With these numbers we get that the G-value is 30.45928 and the p-value is $0.083145 < \alpha = 0.1$ (for this we use the calculator [5]).

6.3. Selection of Optimal Subset

Following the suggested optimization criteria in Section 5 given in (12). We calculate (12) with $\beta = 0.0005$ only for the mutual information values calculated during the max-greedy algorithm (the orange sequence in Figure 1). The obtained values as a function of the cardinality of the set are given in Figure 2. We can clearly see that the maximizing option, among the seven options examined, is a set of cardinality 3. A more rigorous exploration can be performed over a larger set of options.

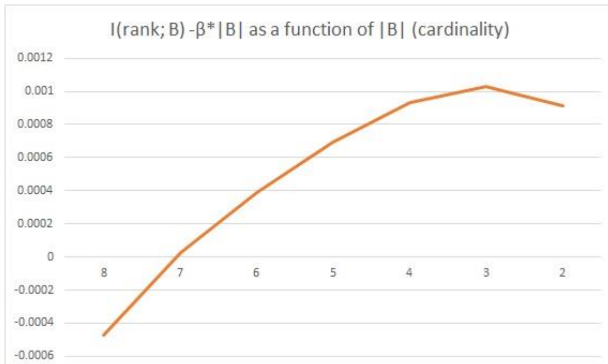

Fig. 1: The value of $I(rank; B) - \beta|B|$ for $\beta = 0.0005$ as a function of the cardinality $|B|$. We observe that the optimal cardinality in this case would be 3. This is based on the sequence of mutual information values calculated according to the max-greedy method.

6. Discussion

Mutual information is a known measure of dependency. It depends solely on the joint probability, and disregards the actual values/support of the random variables. Our goal is to find the subset of additional pieces of information that would maximize the ranking, meaning the actual value or averaged value is highly relevant. A quantity that measures the averaged value is, of course, the expectation. In our setting it would be the conditional expectation, conditioned on different subsets. Thus, an alternative (or addition) to our analysis would be to consider $\mathrm{E}\{\mathrm{rank}|\mathrm{B} = \mathbf{1}\}$ where **1** means that all binary variables in B are one. This approach, however, is limited, since conditional expectation does not measure dependency. This manifests in two important aspects. First, it can be easily shifted. Consider a scenario in which with low probability the ranking is highly influenced and becomes very high, whereas with high probability the ranking is not influenced by the specific subset and remains low. This might result with a high average ranking while the dependency on the specific subset is not very high. In other words, we examine only the average value and not the dependency. The second issue is with the possible generalization of the measure. If we want to consider the impact of the subset of additional pieces of information on several rankings, or in several scenarios the conditional expectation is no longer comparable, as different rankings may have different support.

Using the mutual information, in a specifically designed user study that is monotonic (Assumption 1), and a specific construction of the subset's ranking, allows us to obtain the best of both worlds (dependency and maximum value). We identify dependency on the one-hand, and in this design (monotonicity and the specific definition of the subset's rank) the dependency means that the shift in value can only be upwards. The generalizability of the mutual information as a measure in the determination of the subset of additional pieces of information in more complex studies, with multiple trust measures and in multiple scenarios is a matter of further research.

References

[1] S. Kujala, “User studies: a practical approach to user involvement for gathering user needs and requirements,” Thesis. Helsinki University of Technology Department of Computer Science and Engineering, June 2002.
[2] L. Zhao and S. Lipsitz, “Designs and analysis of two-stage studies,” Statistical Medicine, vol. 11, no. 6, pp. 769–782, April 1992.
[3] T. Cover and J. Thomas, Elements of information theory. John Wiley & Sons, Inc., 1991.

[4] J. H. McDonald, G-test of goodness-of-fit: Handbook of Biological Statistics, Third edition. Baltimore, Maryland: Sparky House Publishing, 2014.
[5] socscistatistics.com. (2018) Social science statistics. [Online]. Available: https://www.socscistatistics.com/pvalues/chidistribution.aspx