INFORMATION DECOMPOSITION BASED ON COMMON INFORMATION

Pradeep Kr. Banerjee
Max Planck Institute for Mathematics in the Sciences,
Leipzig, Germany,
e-mail: pradeep@mis.mpg.de

The total mutual information (MI) that a pair of predictor random variables (RVs) (X_1,X_2) convey about a target RV Y can have aspects of *synergistic information* (conveyed only by the joint RV (X_1X_2), denoted $SI(\{X_1X_2\};Y)$), of *redundant information* (conveyed identically by both X_1 and X_2, denoted $I_\cap(\{X_1,X_2\};Y)$), and of *unique information* (conveyed exclusively by either X_1 or X_2, denoted resp. $UI(\{X_1\};Y)$ and $UI(\{X_2\};Y)$). We have [1]

$$\begin{aligned} I(X_1X_2;Y) &= I_\cap(\{X_1,X_2\};Y) + SI(\{X_1X_2\};Y) + UI(\{X_1\};Y) + UI(\{X_2\};Y) \\ I(X_i;Y) &= I_\cap(\{X_1,X_2\};Y) + UI(\{X_i\};Y), \;\; i = 1,2 \end{aligned} \tag{1}$$

In this note, we show that a recently proposed measure of $I_\cap$ inspired by the information-theoretic notion of *common information* (due to Gács and Körner) cannot induce a nonnegative decomposition of $I(X_1X_2;Y)$.

Consider the AND mechanism, $Y = \text{AND}(X_1,X_2)$, where $X_i = \text{Bernoulli}(\frac{1}{2})$, $i = 1,2$ and joint pmf $p_{X_1X_2Y}$ is such that $p_{(000)} = p_{(010)} = p_{(100)} = p_{(111)} = \frac{1}{4}$. The decomposition evinces both synergistic and redundant contributions to the total MI. First note that $X_1 \perp X_2$, but $X_1 \not\perp X_2|Y$ since $I(X_1;X_2|Y) = +.189 \neq 0$. Fixing Y induces correlations between X_1 and X_2 when there was none to start with. The induced correlations are the source of positive synergy. The redundancy can be explained by noting that if either $X_1 = 0$ or $X_2 = 0$, then both X_1 and X_2 can exclude the possibility of $Y = 1$ with probability of agreement one. Hence the latter is nontrivial information shared between X_1 and X_2. For independent X_1 and X_2, when one can attribute any nonzero redundancy entirely to the mechanism, there is some consensus that $I_\cap(\{X_1,X_2\};Y) = \frac{3}{4}\log\frac{4}{3} = +.311$ and $SI(\{X_1X_2\};Y) = +.5$ [2].

Given two RVs (X,Y), Gács and Körner (GK) defined the notion of a *common RV* to capture the dependence between X and Y and showed that in general, common information does not account for all the mutual information between X and Y. A measure of $I_\cap$ was defined in [2] to measure how well the redundancy that X_1 and X_2 share about Y can be captured by a RV.

$$I_\cap(\{X_1,X_2\};Y) := \max_{\substack{Q:\, Q-X_1-Y \\ Q-X_2-Y}} I(Q;Y), \tag{2}$$

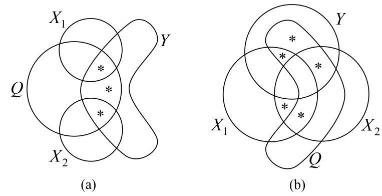

Figure 1: I-diagrams for proof of Lemma 1. Denoting the I-Measure of RVs (Q,X_1,X_2,Y) by μ^*, the atoms on which μ^* vanishes are marked by an asterisk.

where we consider only X_1, X_2 and Y with finite alphabets $\mathcal{X}_1$,$\mathcal{X}_2$ and $\mathcal{Y}$ resp. and it suffices to restrict ourselves to $p_{Q|X_1X_2Y}$ with alphabet $\mathcal{Q}$ such that $|\mathcal{Q}| \leq |\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}| + 3$. Intuitively, if Q specifies the optimal redundant RV, then conditioning on any predictor X_i should remove all the redundant information about Y, i.e., $I(Q;Y|X_i) = 0$, $i = 1,2$. We show that for independent X_1 and X_2, if Y is a function of X_1 and X_2 when any positive redundancy can be attributed solely to the function (e.g., as in the AND mechanism above), $I_\cap$ defined as per (2) fails to capture a nonnegative decomposition.

For finite RVs, there is a one-to-one correspondence between Shannon's information measures and a signed measure μ^* over sets, called the I-measure. We denote the I-Measure of RVs (Q,X_1,X_2,Y) by μ^*. We use X to also label the corresponding set in the Information or I-diagram. The I-diagrams in Fig. 1 are valid diagrams since the sets Q,X_1,X_2,Y intersect each other generically and the region representing the set Q splits each atom into two smaller ones.

Lemma 1. *(a) If* $X_1 \perp X_2$*, then* $I_\cap(\{X_1,X_2\};Y) = 0$*. (b) If* $X_1 - Y - X_2$*, then* $SI(\{X_1X_2\};Y) \leq 0$*.*

Proof. (a)The atoms on which μ^* vanishes when the Markov chains $Q - X_1 - Y$ and $Q - X_2 - Y$ hold and $X_1 \perp X_2$ are shown in the generic I-diagram in Fig. 1(a); $\mu^*(Q \cap Y) = 0$ which gives (a).
(b)The atoms on which μ^* vanishes when the Markov chains $Q - X_1 - Y$, $Q - X_2 - Y$ and $X_1 - Y - X_2$ hold are shown in the I-diagram in Fig. 1(b). In general, for the atom $X_1 \cap X_2 \cap Y$, μ^* can be negative. However, since $X_1 - Y - X_2$ is a Markov chain by assumption, we have $\mu^*(X_1 \cap X_2 \cap Y) = \mu^*(X_1 \cap X_2) \geq 0$. Then $\mu^*(Q \cap Y) \leq \mu^*(X_1 \cap X_2)$, which gives $I_\cap(\{X_1,X_2\};Y) \leq I(X_1;X_2)$. From (1), if $X_1 - Y - X_2$, then the derived synergy measure is $SI(\{X_1X_2\};Y) = I_\cap(\{X_1,X_2\};Y) - I(X_1;X_2) \leq 0$ which gives the desired claim. □

References

[1] P. L. Williams and R. D. Beer (2010), Nonnegative decomposition of multivariate information, *arXiv:1004.2515*.

[2] V. Griffith, T. Ho (2015), Quantifying redundant information in predicting a target random variable, *Entropy*, vol. 17, no. 7, pp. 4644-4653.