VISUALIZING INFORMATION CONSERVATION AND DECOMPOSITION VIA THE INFORMATION MATRIX

Dor Tsur & Haim Permuter
School of Electrical Engineering
Ben-Gurion University of the Negev
{dortz,haimp}@post.bgu.ac.il

ABSTRACT

We introduce a novel framework for visualizing information conservation, decomposition and transfer in time-series data, termed the Information Matrix ($\boldsymbol{I}^{XY}$). Our approach, grounded in information theory, focuses on mutual information (MI), directed information (DI), and transfer entropy (TE) to analyze sequential data. This framework not only offers theoretical insights into information dynamics in sequential systems but also provides a simple visualization of information flow in such systems. We demonstrate the utility of the Information Matrix to the analysis of sequential real world data.

1 INTRODUCTION

Information theoretic quantities, such as mutual information (MI) play a key role in the analysis and design of machine learning systems across most domains Haussler et al. (1994); Goldfeld & Polyanskiy (2020); Shwartz-Ziv & LeCun (2023), and was shown useful for causal analysis in machine learning tasks Peters et al. (2017). Contemporary machine learning tasks consider non-i.i.d. time-series data, e.g. video, speech and text. With MI being a measure between two random variables, it often fails to properly quantify dependence between two sequences, as it involves conditioning of current events on future events. To this end, the literature of information theory and neuroscience consider the two time-series generalizations of MI, termed directed information (DI) Massey et al. (1990) and transfer entropy (TE) Schreiber (2000), which are given by

$$I(X^m \to Y^m) := \sum_{i=1}^{m} I(X^i; Y_i | Y^{i-1}), \quad T_m^{X\to Y}(k,l) := I(X_{m-k}^{m-1}; Y_m | Y_{m-l}^{m-1}), \tag{1}$$

respectively. These measure had already shown benefit for the analysis of times series data in numerous fields Raginsky (2011); Zhou & Spanos (2016); Tiomkin & Tishby (2017); Kalajdzievski et al. (2022); Bonetti et al. (2023). We believe that the increase of sequential machine learning will escalate the utility of sequential information theoretic frameworks. Thus, the focus of this work is to study an information theoretic data-structure, yielding appealing visualizations of the underlying information transfer that provide a deeper insight into the temporal-dynamics.

The proposed framework allows for the visualization of existing information theoretic conservation laws, which are usually based on algebraic tricks applied to the underlying KL divergence representations. These visualization rely on simple matrix entry coloring arguments by transferring temporal relations into visual patterns.

2 INFORMATION MATRIX

Consider a time series $(X^n, Y^n) \sim P_{X^n,Y^n}$ and define the Information Matrix (InfoMat) as an $n\times n$ matrix of conditional mutual information terms, given by $\boldsymbol{I}_{i,j}^{X,Y} := I(X_i, Y_j | X^{i-1}, Y^{j-1})$, where we take $X_k = Y_k = \emptyset$ for $k \in \mathbb{N} \setminus [1, \dots, n]$. Following the chain rule for MI Cover & Thomas (2006), we observe the following relation

$$I(X^n; Y^n) = \sum_{i=1}^{n}\sum_{j=1}^{n} I(X_i, Y_j | X^{i-1}, Y^{j-1}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{I}_{i,j}^{X,Y}. \tag{2}$$

The first utility of $\boldsymbol{I}^{X,Y}$ stems from its ability to simplify the visualization of information conservation, due to equation 2. We can thus represent information conservation laws (see Massey & Massey (2005); Amblard & Michel (2011)) by identifying DI and TE as subsets of entries of $\boldsymbol{I}^{X,Y}$, i.e.,

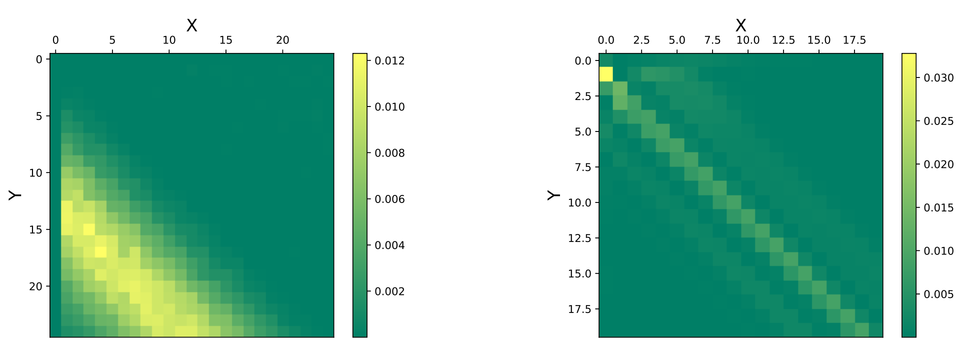
$$\begin{pmatrix} \boldsymbol{I}^{X,Y}_{1,1} & \boldsymbol{I}^{X,Y}_{1,2} & \cdots & \boldsymbol{I}^{X,Y}_{1,n} \\ \boldsymbol{I}^{X,Y}_{2,1} & \boldsymbol{I}^{X,Y}_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{I}^{X,Y}_{n-1,n} \\ \boldsymbol{I}^{X,Y}_{n,1} & \cdots & \boldsymbol{I}^{X,Y}_{n,n-1} & \boldsymbol{I}^{X,Y}_{n,n} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}^{X,Y}_{1,1} & \boldsymbol{I}^{X,Y}_{1,2} & \cdots & \boldsymbol{I}^{X,Y}_{1,n} \\ \boldsymbol{I}^{X,Y}_{2,1} & \boldsymbol{I}^{X,Y}_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{I}^{X,Y}_{n-1,n} \\ \boldsymbol{I}^{X,Y}_{n,1} & \cdots & \boldsymbol{I}^{X,Y}_{n,n-1} & \boldsymbol{I}^{X,Y}_{n,n} \end{pmatrix}, \tag{3}$$

$$\begin{aligned} I(X^n;Y^n) &= I(X^n \to Y^n) + I(\mathrm{D} \circ Y^n \to X^n) & (4)\\ &= I(\mathrm{D} \circ X^n \to Y^n) + I(\mathrm{D} \circ Y^n \to X^n) + I_{inst}(X^n, Y^n), & (5) \end{aligned}$$

where $I(\mathrm{D} \circ X^n \to Y^n) := \sum_{i=1}^n I(X^{i-1}; Y_i | Y^{i-1})$ is the time-delayed DI with $\mathrm{D} \circ X^n$ being a left concatenation of a null symbol with X^n_{n-1}, and $I_{inst}(X^n, Y^n) := \sum_{i=1}^n I(X_i; Y_i | X^{i-1}, Y^{i-1})$ is the instantaneous information exchange. To obtain a clear visualization of the relations in $\boldsymbol{I}^{X,Y}$, we color each term with its corresponds elements in $\boldsymbol{I}^{X,Y}$ (e.g. $I_{inst}(X^n, Y^n) := \sum_{i=1}^n \boldsymbol{I}^{X,Y}_{i,i}$) and consider an equality under the summation over all the matrix entries. We explore new theoretical relations and provide their visual proofs in Appendix A.

Visualization tool We demonstrate the utility of the infomat for the visualization of information flow in sequential systems. We consider two datasets. The first is a Gaussian process with memory and asymmetric relation, which is visualized in Fig. 1a. Second we consider real world data. Specifically, we visualize the relation between breath and heart rate physiological data, which we use to determine the direction and magnitude of information flow (Fig. 1b). The proposed tool allows us to, not only quantify the overall direction of effect from the entire sequence, but to observe its evolution over time. See Appendices B, C for more details and visualizations.

Discussion and future work As demonstrated, the proposed visualization simplifies the analysis of information exchange and visualization thereof in sequential systems. The proposed framework can be made practical when combined with neural estimation techniques Belghazi et al. (2018); Tsur et al. (2023), which we plan to employ for the derivation of an efficient estimator of $\boldsymbol{I}^{X,Y}$ for general datasets. To address the neural estimation performance dependence on dimension, we can further consider approximate low dimensional methods Hotelling (1992); Tsur et al. (2024). Directions for future work are numerous. Applications encompass causal discovery, healthcare sensor monitoring and analysis of non-stationary sequences. Additionally, since TE and DI can be expressed with entropies, we aim to develop a corresponding framework for entropy matrices.



(a) Infomat Visualization on ARMA Gaussian process data under linear relatio with increasing weights.

(b) Infomat visualization on Physiological data. Larger effect in direction 'breath'→'heart' (below diagonal).

Figure 1: Visualization of information transfer via $\boldsymbol{I}^{X,Y}$ in several settings.

URM Statement

All the authors meet to URM criteria of the ICLR 2024 Tiny Paper Track.

References

Pierre-Olivier Amblard and Olivier JJ Michel. On directed information theory and granger causality graphs. *Journal of computational neuroscience*, 30(1):7–16, 2011.

Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In *International conference on machine learning*, pp. 531–540. PMLR, 2018.

Paolo Bonetti, Alberto Maria Metelli, and Marcello Restelli. Causal feature selection via transfer entropy. *arXiv preprint arXiv:2310.11059*, 2023.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New-York, 2nd edition, 2006.

Ziv Goldfeld and Yury Polyanskiy. The information bottleneck problem and its applications in machine learning. *IEEE Journal on Selected Areas in Information Theory*, 1(1):19–38, 2020.

David Haussler, Michael Kearns, and Robert E Schapire. Bounds on the sample complexity of bayesian learning using information theory and the vc dimension. *Machine learning*, 14:83–113, 1994.

Harold Hotelling. Relations between two sets of variates. In *Breakthroughs in statistics: methodology and distribution*, pp. 162–190. Springer, 1992.

Damjan Kalajdzievski, Ximeng Mao, Pascal Fortier-Poisson, Guillaume Lajoie, and Blake Richards. Transfer entropy bottleneck: Learning sequence to sequence information transfer. *arXiv preprint arXiv:2211.16607*, 2022.

James Massey et al. Causality, feedback and directed information. In *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, pp. 303–305, 1990.

James L Massey and Peter C Massey. Conservation of mutual and directed information. In *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, pp. 157–158. IEEE, 2005.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

Maxim Raginsky. Directed information and pearl's causal calculus. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 958–965. IEEE, 2011.

DR Rigney, AL Goldberger, WC Ocasio, Y Ichimaru, GB Moody, and RG Mark. Multi-channel physiological data: description and analysis. In AS Weigend and NA Gershenfeld (eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past*, pp. 105–129. Addison-Wesley, Reading, MA, 1993.

Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

Ravid Shwartz-Ziv and Yann LeCun. To compress or not to compress–self-supervised learning and information theory: A review. *arXiv preprint arXiv:2304.09355*, 2023.

Stas Tiomkin and Naftali Tishby. A unified bellman equation for causal information and value in markov decision processes. *arXiv preprint arXiv:1703.01585*, 2017.

Dor Tsur, Ziv Aharoni, Ziv Goldfeld, and Haim Permuter. Neural estimation and optimization of directed information over continuous spaces. *IEEE Transactions on Information Theory*, 2023.

Dor Tsur, Ziv Goldfeld, and Kristjan Greenewald. Max-sliced mutual information. *Advances in Neural Information Processing Systems*, 36, 2024.

Yuxun Zhou and Costas J Spanos. Causal meets submodular: Subset selection with directed information. *Advances in Neural Information Processing Systems*, 29, 2016.