Attempted Prediction of Emotional Valence from EEG Using Multidimensional Directed Information

Clayton A. Baker

***Abstract*—Quantitative measurement of a person's emotional state can aid performance in a number of areas, such as human-machine interactions, and psychological research. Electroencephalogram (EEG) data has shown potential as a predictor of emotional valence based on asymmetric activation patterns between the left and right hemispheres of the prefrontal cortex. Multidimensional directed information (MDI) is a computational tool that allows the measurement of information content transferred between different signals in a connected system, and has previously seen applications in EEG-based affective measurement in order to detect the presence of an emotional response [1]. This study aimed to use MDI with EEG data from published datasets in order to derive a directional bias metric as a predictor for emotional valence based on frontal hemisphere asymmetry. Two methods of MDI computation were attempted; significant differences were observed in results between the two, suggesting possible errors in implementation. Neither method yielded output correlating with valence.**

I. Introduction

A number of applications stand to benefit from the ability to quantitatively derive a person's affective state. For example, automated evaluation of user affect could improve systems involving human-machine interaction [2], [3], and quantitative measures of psychophysiological state have recently been applied to study the interaction of partners in abusive relationships [4]. Multidimensional directed information (MDI) analysis is a potentially powerful tool for deriving such measures, as it enables analysis of causal temporal relationships between related signals, such as the channels of an electroencephalogram (EEG). Previously, MDI and EEG have been used with the frontal brain asymmetry principle to derive a metric for gauging the presence of an emotional response [1].

Based on the frontal brain asymmetry principle, a strong positive or negative emotion correlates respectively with a left- or right-hemisphere bias in frontal cortex activity. [1] demonstrated that a differential measurement of leftward vs rightward directed information in frontal EEG channels could aid in assessing the presence or absence of a strong emotion. This study attempts to adapt the approach taken by [1] in order to approximate a measure of emotional valence from EEG using frontal asymmetry and directed information, with testing and evaluation performed on two existing multimodal emotion recognition datasets: DEAP [5] and DREAMER [6].

II. Multidimensional Directed Information

A. Mutual Information

Directed information is a concept from information theory, and can be considered a temporally asymmetric variant of mutual information, which is a measure of the shared information content between two signals or random variables. Mutual information $I(X;Y)$ can be described as a measure of "non-independence" between signals X and Y, defined as the divergence between their joint distribution and product of their marginal distributions as shown below [7].

$$I(X;Y) = D_{KL}\big(p(x,y) \parallel p(x)p(y)\big) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (1)$$

Here, D_{KL} is the Kullback-Leibler divergence; $p(x)$, $p(y)$, and $p(x,y)$ are the marginal and joint probability distributions of X and Y. Mutual information can also be written in terms of joint and marginal entropy.

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \quad (2)$$

Furthermore, mutual information can be conditioned on one or more additional variables, allowing us to determine the amount of information shared by X and Y but not shared by some third signal Z, as defined in eq. 3 [7].

$$I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z) \quad (3)$$

B. Directed Information

As conditional mutual information allows us to measure information content shared *exclusively* between two signals in the presence of others, we are able to measure information temporally propagated from one signal to another in the following manner. Let us suppose each of our signals is a time-series of sample measurements within some epoch, with x_i being the i'th sample of signal X. Let us also define k as some time-index of interest, with P prior samples and M following samples within the epoch. Let X^P represent the pre-k section of signal X, and let X^M represent the post-k segment (extend this notation to Y, Z, etc.). We can then define the directed information from X to Y at time some time-index k as the information content shared by x_k and Y^M, but not by any other signal region within the epoch (eq. 4) [8].

$$I(x_k \to Y^M) = I(x_k; Y^M | X^P, Y^P, y_k) = \sum_{m=1}^{M} I(x_k; y_{k+m} | X^P, Y^P, y_k) \quad (4)$$

Much like eq. 3, this formula can be conditioned on some additional signal(s) Z; this allows us to better isolate the causal influence of X on Y by excluding information that appears to be shared in an $X \to Y$ manner, but is actually shared in a $Z \to (X, Y)$ manner [8]. This conditioned definition is known as multidimensional directed information (MDI).

$$I(x_k \to Y^M | Z) = \sum_{m=1}^{M} I(x_k; y_{k+m} | X^P, Y^P, y_k, Z^P, z_k) \quad (5)$$

Combining equations 3 and 5, we can express this in terms of the joint entropy values among X, Y, and Z segments.

$$\begin{aligned} I(x_k \to Y^M | Z) = & H(X^P, x_k, Y^P, y_k, Z^P, z_k) \\ & + H(X^P, Y^P, y_k, y_{k+m}, Z^P, z_k) \\ & - H(X^P, x_k, Y^P, y_k, y_{k+m}, Z^P, z_k) \\ & - H(X^P, Y^P, y_k, Z^P, z_k) \end{aligned} \quad (6)$$

The total information shared in this manner, would be defined as the sum of this value over all k.

$$S^{XY|Z} = I(X \to Y|Z) = \sum_{k} I(x_k \to Y^M | Z) \quad (7)$$

III. Methods

A. Data and Processing

Subjective response ratings and EEG data were taken from the DEAP (32 subjects, 40 trials) and DREAMER (23 subjects, 18 trials) datasets, comprising a total of 1694 trials across 55 participants [5, 6]. In both sets, trials consisted of audiovisual stimulus in the form of a music video or film clip, with EEG recording at or 128Hz; while the DEAP data was originally acquired at 512Hz, this study uses the preprocessed version available, downsampled to 128Hz. Subjective responses to each trial in terms of valence, arousal, and dominance, were acquired via self-assessment on a 9-point (DEAP) or 5-point (DREAMER) scale utilizing a Self-Assessment Manikin [9].

In order to exploit frontal brain asymmetry as per [1], EEG signals were selected from prefrontal and frontal regions about the midline: channels F3, F4, Fp3, and Fp4, present in both datasets. All EEG data was filtered to the α-β range (8-30Hz) using a 12th order Butterworth bandpass filter. Outliers were detected in filtered signals based on median absolute deviation; outliers detected concurrently on multiple channels are interpreted as measurement artifacts, and the erroneous samples are zeroed on all channels. Valence scores from the DEAP dataset were scaled to coincide with the range of the DREAMER set's 5-point scale. Target values for epoch duration and step length were defined in temporal units and approximated according to the sample rate, as listed in Table 1. For the full stimulus period of each trial, the following shared information values were computed to represent to total leftward and rightward directed information between the prefrontal channels, conditioned on differential measurement of the frontal channels as in [1], with epoch parameters listed in Table I.

TABLE I
EPOCH PARAMETERS

		Desired	Samples	(Actual)
P	(epoch memory)	100ms	13	(101.6ms)
M	(epoch future)	100ms	13	(101.6ms)
k_{step}	(inter-epoch step)	50ms	7	(54.7ms)

The desired and actual parameters used for epoch definition as constrained by the 128Hz sampling rate. The total duration of each epoch can be determined as $(P + M + 1)$ samples, or 258ms.

$$S^L = S^{\text{Fp4}\to\text{Fp3}|(\text{F4}-\text{F3})}, \qquad S^R = S^{\text{Fp3}\to\text{Fp4}|(\text{F4}-\text{F3})} \quad (8)$$

From this, a normalized relative measure of directional bias B was computed, such that $B > 0$ implies greater rightward information flow, and $B < 0$ implies greater leftward flow.

$$B = \frac{S^R - S^L}{S^R + S^L} \quad (9)$$

All processing was performed using MATLAB (Mathworks, Natick, MA, USA).

B. MDI Calculation – Covariance

Two different methods were implemented for computing the values defined in eq. 9. The first method, used by [4,8], utilizes the following theorem whereby the joint entropy of a collection of Gaussian stochastic variables $s_1, \dots, s_n$ can be expressed in terms of the determinant of their covariance matrix R.

$$H(s_1, \dots, s_n) = 0.5 \log(2\pi e^n |R(s_1, \dots s_n|) \quad (10)$$

Applying this formula to the right hand side of equation 6 yields the following…

$$I(x_k \to Y^M | Z) = \frac{1}{2} \log\left(\frac{|R(S_1)||R(S_2)|}{|R(S_3)||R(S_4)|}\right) \quad (11)$$

where S_1 through S_4 represent the four collections of terms within the entropy functions in equation 6. As these inputs are a mix of both scalar (y_k, y_{k+m}, etc.) and vector (X^P, Y^P, etc.) terms, the methodology for computing the covariance matrices is detailed below. Beginning from the definition of the R as a matrix of pairwise covariances between a collection of signals

$$R(s_1, \dots, s_n) = \begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \cdots & \sigma_{n,n} \end{bmatrix} \quad (12)$$

with $\sigma_{i,j}$ being the covariance of signals s_i and s_j

$$\sigma_{i,j} = \sum_n (s_i[n] - \mu_i)(s_j[n] - \mu_j) \quad (13)$$

In order to accommodate scalar inputs to equation 13, the mean μ for each signal (X, Y, or Z) was computed over the entire the epoch. All inputs derived from a given signal used its corresponding epoch-mean in equation 13. For example, the value σ for the pairwise matchup of X^P and y_{k+m} would be computed as $\sum_{n=1}^{P}(X^P[n] - \mu_X)(y_{k+m} - \mu_Y)$.

C. MDI Calculation – Entropy Estimation

The second method employed was to directly compute entropy values from an approximation of the signals' joint probability distribution, enabling the use of equation 6 for MDI calculation. Given a collection of random variables or signals $S_1, \dots, S_n$ described by joint probability function $p(s_1, \dots, s_n)$, their joint entropy is defined as

$$H(S_1, \dots, S_n) = -\sum p(s_1, \dots, s_n) \log\big(p(s_1, \dots, s_n)\big) \quad (13)$$

over all possible values of $\langle s_1, \dots, s_n \rangle$, with the convention that $0\log(0) = 0$ [7].

For every processed trial t, a joint distribution $p_t(x, y, z)$ was approximated for the three signals of interest, based on a three-dimensional joint histogram function $N_t[i, j, k]$ derived from that trial..

$$p_t(x, y, z) \approx p_t[i_x, j_y, k_z] \quad (14)$$

$$p_t[i, j, k] = \frac{N_t[i, j, k]}{\sum_{i,j,k} N_t} \quad (15)$$

Here, the indices i_x, j_y, k_z refer to the coordinates of the bin corresponding to a value observation $\langle x, y, z \rangle$; $p_t[i, j, k]$ is the discretized probability distribution determined from the bin counts N_t throughout the trial.

The joint distribution for a given combination of epoch segments (such as X^P, Y^P, y_k, etc.) was calculated as follows. Let $\hat{X}$, $\hat{Y}$, and $\hat{Z}$ represent any arbitrary subsets of X, Y, and Z respectively. Let $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ be the sets of bin indices along each dimension of the histogram, corresponding to the values contained in $\hat{X}$, $\hat{Y}$, and $\hat{Z}$, with duplicates omitted. For example, if we defined $\hat{Y} \coloneqq \{Y^P, y_k, y_{k+m}\} \equiv \{y_{k-P}, \dots, y_k, y_{k+m}\}$, then $\hat{\jmath}$ would contain all unique indices from $\{j_{y_{k-P}}, \dots, j_{y_k}, j_{y_{k+m}}\}$. The joint distribution $\hat{p}_t$ of $\hat{X}, \hat{Y}, \hat{Z}$ would then approximated from the subset of the histogram encompassed by $\hat{\imath}, \hat{\jmath}, \hat{k}$…

$$\hat{p}_t(x, y, z) \approx \frac{N_t[i, j, k]}{\sum_{\hat{\imath},\hat{\jmath},\hat{k}} N_t}, \qquad \text{for } \begin{matrix} x \in \hat{X}, y \in \hat{Y}, z \in \hat{Z} \\ i \in \hat{\imath}, j \in \hat{\jmath}, k \in \hat{k} \end{matrix} \quad (15)$$

where $\sum_{\hat{\imath},\hat{\jmath},\hat{k}} N_t$ is the total number of observations in the $\hat{\imath}, \hat{\jmath}, \hat{k}$ subset of the histogram. Combining this with equation 14, we can define the entropy for this subset of the distribution, allowing computation of I via equation 6.

$$H(\hat{X}, \hat{Y}, \hat{Z}) = -\sum_{i \in \hat{\imath}} \sum_{j \in \hat{\jmath}} \sum_{k \in \hat{k}} \frac{N_t[i, j, k]}{\sum_{\hat{\imath},\hat{\jmath},\hat{k}} N_t} \log\left(\frac{N_t[i, j, k]}{\sum_{\hat{\imath},\hat{\jmath},\hat{k}} N_t}\right) \quad (16)$$

IV. Results

Although there was a noticeable difference in the behavior of the two computation methods, neither method yielded an observable relationship between valence score and the directional bias B, as demonstrated in Figure 1. Figure 2 shows an overall comparison of normalized S^R and S^L values for the two methods and datasets, revealing the notable discrepancy between methods. Results acquired from the covariance method tend to cluster about the origin, accounting for the wide range of B values exhibited on the left of Figure 1. Results from the entropy estimation method tend to cluster along the positive diagonal, and consequently exhibit a much narrower range in Figure 1, where $B = 0$ would correspond to $S^R = S^L$.

V. Discussion

It should be noted that this study made use of a single epoch window throughout, with a symmetric memory and future window of ~100ms. Given the results, it is worthwhile to examine the influence of these parameters, and the possible side-effects of poor selection.

The M parameter determines the window of the target signal over which $I(x_k; y_{k+m} | \dots)$ is summed (eq. 5); if a signal is propagated from X to Y, its information will be present in the summation as long as it arrives at Y within this window. For

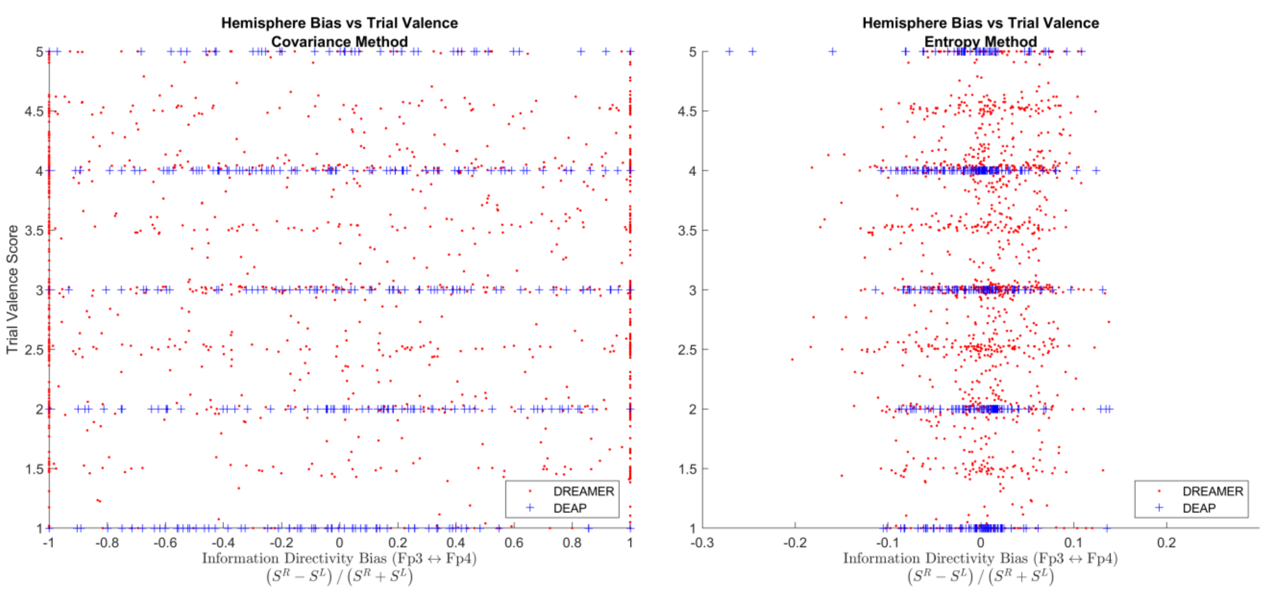

Fig. 1. Participant valence score compared to information directivity bias B, using the covariance approach (left) or entropy histogram approach (right). Note that valence (y-axis) was measured on a continuous scale in the DEAP dataset, but a discete scale in the DREAMER set.

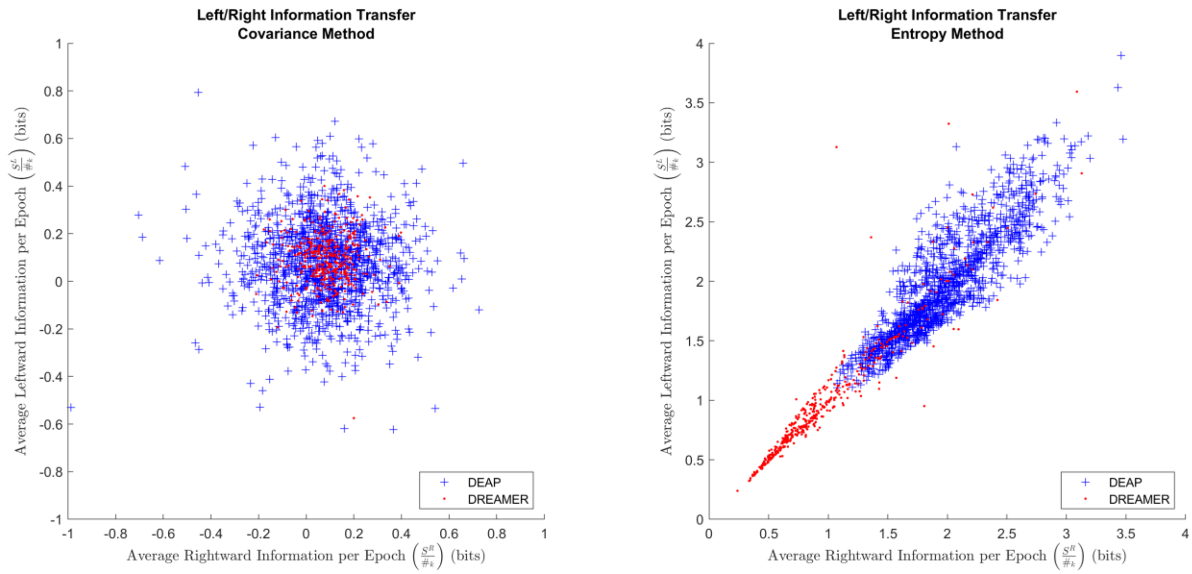

Fig. 2. Comparison between methods and datasets of the calculated per-epoch mean information transfer. Each data point represents a single trial from either the DEAP or DREAMER dataset. The x and y data are defined as the average leftward (y-axis) or rightward (x-axis) information transfer within each epoch of the trial, equivalent to S^R or S^L divided by the number of epochs in the trial. Note the method-dependent clustering.

this study, selection of 100ms for M was based on the assumption that signal propagation between the left and right hemispheres of the prefrontal cortex (Fp3 ↔ Fp4) would fall reliably within that timeframe. A longer M parameter may have improved results at the cost of computation time. Furthermore, an optimal value for M could potentially be obtained by observing the value of $I(x_k; y_{k+m}| \dots)$ as a function of m.

The P parameter determines the conditioning of $I(x_k; y_{k+m}| \dots)$ on the signals' histories; this serves to ensure that the computed value is determined only by information content that is newly observed at time k. Given an insufficient P window, lower-frequency information content originating from before the start of the epoch might be reflected in the value of $I(x_k; y_{k+m}| \dots)$, which would tend to reduce reliability and generally inflate the value of $I(X \rightarrow Y)$. These effects would be expected to apply across calculations in a fairly symmetric manner, so their influence on any relative measure between S^L and S^R would be minimal. Regardless, a P window of ~125ms or greater may be ideal, given the passband of 8-30Hz used in this study, a P window ~125ms or greater may be ideal.

Another important note is the lack of controlled validation of the MDI calculations. Given the discrepancy in results from the two computation methods, it can be confidently stated that at least one of them is providing invalid results. A method of verifying the accuracy of the code, such as the 4-channel signal propagation model used in [8], would be vital for informing any further development.

References

[1] P. C. Petrantonakis and L. J. Hadjileontiadis, "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 5, pp. 737–746, Sep. 2011, doi: 10.1109/TITB.2011.2157933.

[2] P. Rani, N. Sarkar, C. A. Smith, and J. A. Adams, "Affective communication for implicit human-machine interaction," *Proc. IEEE Int. Conf. Syst. Man Cybern.*, vol. 5, pp. 4896–4903, 2003, doi: 10.1109/icsmc.2003.1245758.

[3] P. Rani, N. Sarkar, and J. Adams, "Anxiety-based affective communication for implicit human–machine interaction," *Adv. Eng. Informatics*, vol. 21, no. 3, pp. 323–334, Jul. 2007, doi: 10.1016/j.aei.2006.11.009.

[4] H. W. Weon, Y. E. Byun, and H. J. Lim, "Quantitative EEG (QEEG) analysis of emotional interaction between abusers and victims in intimate partner violence: A pilot study," *Brain Sci.*, vol. 11, no. 5, 2021, doi: 10.3390/brainsci11050570.

[5] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012, doi: 10.1109/T-AFFC.2011.15.

[6] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 1, pp. 98–107, Jan. 2018, doi: 10.1109/JBHI.2017.2688239.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2005.

[8] O. Sakata, T. Shiina, and Y. Saito, "Multidimensional Directed Information and Its Application," *Electron. Commun. Japan (Part III Fundam. Electron. Sci.*, vol. 85, no. 4, pp. 45–55, Apr. 2002, doi: 10.1002/ecjc.1086.

[9] J. D. Morris, "OBSERVATIONS: SAM: The Self-Assessment Manikin - An Efficient Cross-Cultural Measurement of Emotional Response," *J. Advert. Res.*, vol. 35, no. 6, pp. 63–68, 1995.