

A Multimodal Causal-Inspired Approach towards Affect Modeling

Kevin Farrugia*

Department of Artificial Intelligence,
Faculty of Information and
Communication Technology
University of Malta
Msida, Malta
kevin.j.farrugia.15@um.edu.mt

Athanasios Papathanasiou*

Department of Artificial Intelligence,
Faculty of Information and
Communication Technology
University of Malta
Msida, Malta
athanasios.papathanasiou@um.edu.mt

Konstantinos Makantasis*

Department of Artificial Intelligence,
Faculty of Information and
Communication Technology
University of Malta
Msida, Malta
konstantinos.makantasis@um.edu.mt

Abstract

This study introduces a novel causal-inspired approach to affect modeling that leverages distribution shifts and emotional subjectivity as advantages rather than obstacles. Our methodology employs causal graph discovery algorithms to identify relationships between affect measurements and emotional states across different environments, defined by user-task-annotator combinations. We implement constraint-based (PC) and score-based (GES) algorithms to discover environment-specific causal graphs for constructing Bayesian networks for affect prediction. The approach is validated on the RECOLA dataset, demonstrating competitive performance with neural networks and random forests. A key finding reveals a consistent directed edge between arousal and valence nodes, extending Russell's circumplex theory by identifying conditional dependencies when affect measurements are explicitly observed. This work contributes by integrating causal inference in affect modeling, implementing multiple graph discovery approaches, validating on a benchmark dataset, and establishing groundwork for developing causal models of affect.

CCS Concepts

• Human-centered computing → Empirical studies in HCI.

Keywords

affect modelling, causal graphs, bayesian networks, graph discovery

ACM Reference Format:

Kevin Farrugia, Athanasios Papathanasiou, and Konstantinos Makantasis. 2025. A Multimodal Causal-Inspired Approach towards Affect Modeling. In *The Pervasive Technologies Related to Assistive Environments (PETRA '25)*, June 25–27, 2025, Corfu Island, Greece. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3733155.3737909>

*All authors contributed equally to this research.



This work is licensed under a Creative Commons Attribution International 4.0 License.

PETRA '25, Corfu Island, Greece

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1402-3/25/06

<https://doi.org/10.1145/3733155.3737909>

1 Introduction

Affect modeling refers to the process of mapping measurable manifestations of affect, derived from various user input modalities—such as facial expressions, speech, and physiological data—to affective labels [29]. These labels typically correspond to continuous values of arousal and valence. The affect modeling task is commonly approached as a supervised learning problem, which necessitates the use of labeled data. There are two primary methods for collecting labeled affective corpora: the first-person annotation protocol and the third-person annotation protocol (see Fig.1).

In the first-person annotation protocol, a participant is exposed to a stimulus—such as the context of an interactive task—that elicits affective reactions. During this process, the context is recorded alongside other user-related modalities, such as physiological signals and facial expressions. After the interaction, the participant self-reports their affective state in terms of arousal and valence, which are used as the target variables. An example of a recently published dataset collected using this protocol is the AGAIN dataset [23]. The third-person annotation protocol, as the name implies, differs primarily in how the labels are created. In this approach, a third-party annotator (e.g., an external observer) evaluates the participant's reactions—such as facial expressions—and infers their affective state, which is then labeled accordingly. Notable datasets collected using this protocol include SEWA [16] and RECOLA [33].

Any affect modeling task has two primary objectives. The first is to establish accurate mappings between measurable affective manifestations and corresponding affect labels. To achieve this, machine learning techniques are utilized to develop models that ideally generalize across diverse participants and interaction contexts. The second objective is to create interpretable mappings, providing insights into the mechanisms underlying human affective manifestations [43].

When it comes to interpretability, white-box models—such as linear models or decision trees—are often used due to their transparency. However, these models typically lack the discrimination power needed to capture the complex patterns present in affective data. In contrast, black-box models, such as advanced neural network architectures, can identify subtle patterns in affect data but are inherently difficult to interpret. Moreover, interpretation techniques (e.g., counterfactual explanations, LIME, SHAP) aim to uncover cause-effect relationships between inputs and outputs within the “world” of the model [26]. However, these relationships do not necessarily correspond to real-world cause-effect dynamics between the observed variables.

In terms of accuracy, current affect modeling approaches typically follow the statistical learning paradigm [19, 22, 30, 41], which

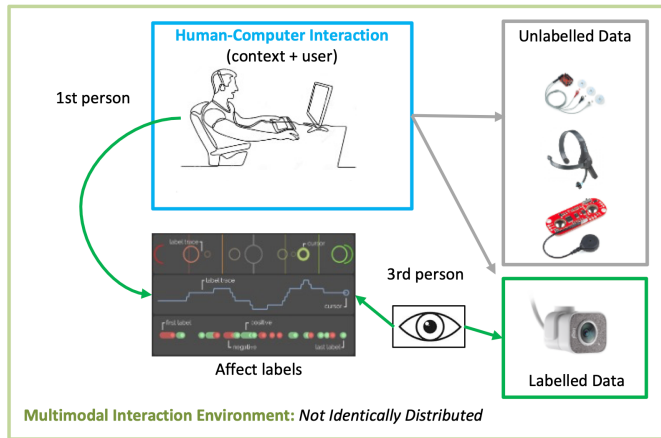


Figure 1: The two dominant approaches for collecting affect corpora: the first-person annotation protocol and the third-person annotation protocol.

relies on the fundamental assumption that data is identically and independently distributed (IID). However, as evident from the discussion on data collection protocols, affective data rarely, if ever, satisfies this assumption. Each unique combination of user, task, and annotator defines its own data generation distribution, making the data non-identically distributed. As a result, current approaches are susceptible to modeling spurious correlations in the data, which undermines both interpretability and generalizability. To address this challenge, incorporating inductive biases into models is essential to mitigate the impact of spurious correlations.

In this study, we introduce inductive bias into the affect modeling process by leveraging graphs that represent the relationships between the variables involved. Drawing inspiration from causal discovery methods, we utilize both constraint-based [13] and score-based [12] causal graph discovery algorithms. To enhance the robustness of graph discovery, we partition the affect data into subsets, where each subset corresponds to a unique user-task-annotation triplet and is assumed to be generated from the same distribution (i.e., environment). This inherent characteristic of affect corpora enables us to discover one graph per environment. Subsequently, we construct Bayesian networks that reflect the structure of these graphs and aggregate them into an ensemble model for making predictions.

The structure of the graphs produced by causal graph discovery methods is designed to capture cause-effect relationships between the variables, effectively suppressing or even eliminating the influence of spurious correlations. Additionally, the resulting Bayesian Networks, which replicate the structure of these discovered graphs, can be seamlessly translated into Structural Causal Models (SCMs) [4], enabling the analysis¹ of interventions and even counterfactual scenarios [37]. By combining these two aspects, our approach becomes both fully transparent and interpretable, offering valuable insights into the mechanisms underlying human affective manifestations while remaining robust against spurious correlations.

¹The analysis of interventions and counterfactual scenarios is beyond the scope of this paper.

This study advances the field of affect modeling in four key ways. First, to the best of our knowledge, it is the first that integrated concepts and tools from causal inference and causal graph discovery using observational data in affect modeling. Second, to ensure methodological rigor, we implement two distinct algorithmic approaches for graph discovery: a constraint-based method and a score-based method. Third, we validate our methodology using the RECOLA dataset [33], a widely recognized benchmark in international affect modeling research. Last but not least, we establish a foundation for developing SCMs of affect, which enable researchers to analyze both interventional and counterfactual scenarios—a significant step forward in understanding emotional processes.

2 Related Work

This section examines state-of-the-art methodologies in affect modeling and their inherent limitations. The discussion then explores initial efforts to incorporate causal theory frameworks into affect modeling, representing an emerging direction in the field.

2.1 Current Affect Modeling Approaches

Affect modeling encompasses various human-machine interactions, from robot interactions to movie viewing, chatbot conversations, and advertisement consumption. The generalization capabilities of affect models face three critical challenges [34, 36, 44]. First, human-machine interactions are dynamic processes where both context and user emotions continuously evolve. Second, emotions are inherently subjective, lacking universal definitions. Third, affect datasets exhibit distribution shifts, as each unique combination of task, user, and annotator generates its own distinct data distribution. These interrelated challenges—dynamic interactions, emotional subjectivity, and distribution variations—fundamentally limit the generalization and performance of affect models.

Current affect modeling research employs four strategies to address the aforementioned challenges. The first strategy reframes affect modeling as a ranking or preference learning problem to minimize subjectivity bias in annotations [2, 18, 27, 28]. This approach has proven particularly effective in customer service interactions, where ordinal learning helps capture emotional nuances [10], and through specialized ranking networks that leverage privileged information to enhance model performance [17, 19].

The second strategy focuses on the collection and annotation of large-scale data assuming extensive datasets can better represent statistical patterns in emotional expressions [15, 25, 31]. EmotioNet [5] provides automatic annotation capabilities for facial expressions in natural settings, processing over a million expressions in unconstrained environments. Similarly, AffectNet [25] offers a comprehensive database for facial expression analysis containing more than 1M facial images, while AFEW-VA [15] specifically addresses valence and arousal estimation in real-world scenarios providing frame-by-frame annotations for 600 video clips.

The third strategy focuses on multimodal integration, where diverse sensor types are combined to capture the multifaceted nature of emotional states and enhance model generalization [29, 42]. Recent developments include GameVibe [1], a multimodal affective game corpus that enables sophisticated emotion analysis in gaming

contexts. Research has demonstrated that combining audiovisual data significantly improves emotion detection accuracy [19], while advanced transfer learning methods have successfully bridged the gap between laboratory-controlled environments and real-world applications [20, 22]. The effectiveness of multimodal approaches is evident in recent achievements in long-term player engagement prediction, where researchers have successfully combined visual data with gamepad interaction patterns [32].

The fourth strategy emphasizes user-specific modeling, developing individualized approaches that account for personal variations in emotional expression. This approach recognizes that emotional responses are inherently individual, leading to adaptive recognition models that can accommodate user-specific patterns and variations in emotional expression [45].

The aforementioned approaches exhibit fundamental limitations in two key areas. They largely treat human-machine interaction dynamics, emotional subjectivity and distribution shifts as obstacles to be eliminated rather than essential features to be incorporated. This simplification enables the development and application of sophisticated deep learning algorithms but overlooks crucial aspects of emotional expression. On top of that, while current successes in affect modeling primarily stem from large-scale pattern recognition within the statistical learning framework, this approach is inherently flawed. The core assumption of independent and identically distributed data is violated in practical affect modeling scenarios, making purely statistical approaches insufficient for addressing the fundamental challenges of emotion recognition (see Fig.1).

2.2 Causality Tools in Affect Modeling

Initial attempts towards employing causation tools in affect modeling rely mainly on Granger causality [6]. Indicatively, Jinliang et al. [14] introduced a novel approach combining Granger causality with capsule neural networks for EEG-based emotion recognition. Gao et al. [7] developed a single-trial EEG emotion recognition system that uniquely combines Granger Causality with Transfer Entropy analysis. Chen et al. [3] proposed a sparse Granger causality model that leverages sensor correlations in EEG data, specifically addressing the challenge of high-dimensional data in emotion recognition tasks. However, instead of discovering cause-effect relations between emotion and epiphenomena (e.g. physiological reactions caused by the felt emotions), Granger causality tests whether temporal information about the epiphenomena can forecast the felt emotion. Nevertheless, these initial attempts highlight our study's significance.

Beyond Granger causality, the integration of other causal analysis tools in affect modeling remains limited. Notable exceptions include Affect2MM [24], a framework that incorporates emotion causality analysis for multimedia content understanding, demonstrating how causal relationships can enhance affective analysis in multimodal scenarios. A different causal perspective was introduced by Makantasis et al. [21] who exploited distributional variations in affect data to develop robust cleaning mechanisms for affect corpora, thereby enhancing model reliability.

This study extends current approaches by investigating how causal graph discovery algorithms can enhance affect modeling

and exploring the potential of graph structural knowledge as an inductive bias in emotional state prediction.

3 Proposed Methodology

This section establishes the mathematical framework for affect modeling and presents our methodological approach. We begin with a formal problem formulation, develop the foundations of our affect models, and then detail their practical implementation for emotional state prediction.

3.1 Causal-Inspired Affect Modeling

Let $\mathbf{X} = (X_1, \dots, X_d)$ represent a set of d affect measurements, such as facial expressions, audio features, and physiological signals. Let $\mathbf{Y} = (Y_1, \dots, Y_l)$ denote the corresponding continuous affect labels (e.g., arousal, valence). As mentioned above, in affect corpora, data is generated under different environments e , each defined by a unique combination of user, task, and annotator. For environment e , we denote the corresponding subset of the affect corpus as $\{(X^{e,n}, Y^{e,n})\}_{n=1}^{N_e}$, where $e = 1, \dots, E$ and N_e represents the number of examples generated in that environment. For notational convenience, we use $\mathbf{Z} = [\mathbf{X} : \mathbf{Y}]$ to represent the complete set of predictor and target variables.

The objective of affect modeling is to learn a mapping function $f : \mathbf{X} \rightarrow \mathbf{Y}$ from affect measurements to affect labels. In machine learning frameworks, this function represents the conditional expectation $\mathbb{E}(\mathbf{Y}|\mathbf{X})$, which is based on the conditional probability:

$$\mathbb{P}(\mathbf{Y}|\mathbf{X}) = \mathbb{P}(\mathbf{Y}|X_1, X_2, \dots, X_d). \quad (1)$$

Using all predictors in this conditional probability can lead to models that are vulnerable to spurious correlations and collinearities in the data limiting their generalization capacity. Our approach addresses this limitation by focusing only on predictors that have direct *causal* influence on the target variables.

Towards this endeavor, we need to derive a *disentangled factorization* [37] of the joint probability $\mathbb{P}(\mathbf{X}, \mathbf{Y})$. Let us denote $\mathbf{PA}_i$ the variables that have a direct influence on variable Z_i . We will refer to $\mathbf{PA}_i$ as the *parents* of variable Z_i . Knowing the parents of every variable Z_i , the disentangled factorization of the joint probability $\mathbb{P}(\mathbf{X}, \mathbf{Y})$ is given by

$$\mathbb{P}(\mathbf{X}, \mathbf{Y}) = \mathbb{P}(\mathbf{Z}) = \prod_{i=1}^{d+l} \mathbb{P}(Z_i|\mathbf{PA}_i). \quad (2)$$

This factorization uniquely decomposes the joint probability into conditional probabilities that account for all statistical dependencies among observed variables. These conditionals represent causal mechanisms [38] that describe the underlying phenomenon. In contrast to (1) which uses all predictors indiscriminately, the disentangled factorization in (2) represents the joint probability distribution as a product of causal mechanisms.

In this study, we derive one disentangled factorization for each environment. To identify the parents $\mathbf{PA}_i^e$ of the variable Z_i^e in the environment e , we utilize causal graph discovery algorithms. These algorithms aim to identify and map out cause-effect relationships between variables in observational data, typically representing these relationships as a Directed Acyclic Graph (DAG) [8]. Nodes

correspond to variables, while directed edges to cause-effect relationships uncovering, this way, the underlying causal structure that governs how different variables influence each other. However, it is important to note that multiple causal structures may be consistent with the same observational data. In this study we use two different algorithms for graph discovery. Specifically, we use the Peter-Clark (PC) algorithm [40] which is a constraint-based method, and the Greedy Equivalence Search (GES) algorithm [12], which is a score-based method.

The PC algorithm discovers a graph in three main steps. Starting with a fully connected undirected graph, the algorithm removes edges between variables that are conditionally independent. Then, it identifies and orients collider structures where two variables point to a common variable. Finally, additional edges are oriented using orientation rules based on the identified colliders and the assumption that cycles are not allowed. The score-based GES algorithm also operates in three sequential phases. Starting from an empty graph, edges are iteratively added between variables to improve the Bayesian Information Criterion (BIC) score until no single edge addition can increase the score further. Then, edges are systematically removed from the graph when their removal improves the BIC score. Finally, edge orientations are flipped to improve the BIC score while maintaining the equivalence class structure.

The causal discovery algorithms output a DAG, which identifies the parent variables PA_i^e for each variable Z_i^e in environment e . After PA_i^e identification, we can estimate the disentangled factorization in (2) for environment e by building a Bayesian network that mirror the discovered graph's structure. Bayesian networks, despite their limited learning capacity, provide three crucial advantages: interpretable decision-making processes, resilience to noise and missing data, and natural compatibility with SCMs. While their conversion to SCMs enables intervention and counterfactual analysis, such applications lie beyond the scope of this study.

3.2 Affective State Prediction

Let $\{(X^{e,n}, Y^{e,n})\}_{n=1}^{N_e}$, where $e = 1, \dots, E$, represent training data used to develop E distinct Bayesian networks. For an unseen environment u , we aim to predict target variables $\{Y^{u,n}\}_{n=1}^{N_u}$ given new observations $\{X^{u,n}\}_{n=1}^{N_u}$. We propose two distinct approaches for the prediction problem. The first one creates an ensemble of environment-specific Bayesian networks, while the second develops a single Bayesian network based on the averaged causal graph across all environments.

The ensemble approach can be formalized as follows. Let $f^e : X \rightarrow Y$ represent the mapping function from affect measurements to affect labels learned by the Bayesian network for environment e . The ensemble prediction function is then defined as:

$$f(X) = \frac{1}{E} \sum_{e=1}^E f^e(X). \quad (3)$$

The ensemble averaging in (3) reduces prediction variance while maintaining bias (see Chapter 10 in [11]), leading to more robust and generalizable affect models.

The second approach operates directly on the discovered DAGs from each environment. Let $G^e = (Z, D^e)$ denote the DAG for

environment e , where Z represents the nodes (predictors and target variables) and D^e represents the edges. For each potential edge $D_{i \rightarrow j}$ connecting nodes i and j , we count its occurrences $N_{i \rightarrow j}$ across all environments and compute its probability to be included in the average graph as:

$$P(D_{i \rightarrow j}) = \frac{1}{E} N_{i \rightarrow j}. \quad (4)$$

We establish a threshold α computed as:

$$\alpha = \frac{1}{E} P(D_{i \rightarrow j}). \quad (5)$$

Edges with probabilities below α are removed, resulting in an average graph $G = (Z, D)$ containing only edges where $P(D_{i \rightarrow j}) \geq \alpha$. If the resulting graph G contains cycles, we iteratively remove edges with the lowest probabilities until achieving acyclicity. This approach assumes that edges appearing frequently across environments are highly likely to appear in the unseen environment as well carrying important predictive information. The final step involves constructing a Bayesian network that mirrors the averaged graph structure and training it on data from all environments. This network then serves as the prediction model for unseen environments.

4 Experimental Protocol

This section describes our experimental protocol, beginning with the description of the data set and the pre-processing steps. We then detail our evaluation protocol and conclude with an analysis of the experimental results.

4.1 The RECOLA Dataset

The RECOLA Database [33], used for evaluating our methodology, contains multimodal recordings of online dyadic interactions between 34 participants. The dataset comprises 130 audio features derived from voicing-related descriptors (based on the COMPARE challenge [39]), 40 video features capturing action units, head pose estimation, texture, and optical flow, and 116 physiological features extracted from electrodermal activity (EDA) and electrocardiography (ECG). The EDA features include signal variability, zero-crossing, spectral entropy, and coefficients, while ECG features encompass spectral entropy and mean frequency.

Six annotators provided continuous arousal and valence ratings in the range at 25Hz. RECOLA has been used for audiovisual emotion recognition challenges in which 11 participants serve as an evaluation set, and thus their data is not publicly available. From the original 34 participants, we utilize 18 recordings after excluding non-public data and sessions with missing information. For pre-processing, we segment each recording into overlapping time windows of 3 seconds with 400ms sliding steps. Features are averaged within each window to create single feature vectors, while annotation values are, first, aggregated using the median to address inter-annotator variability, and then averaged within each window. For each modality (visual, audio, and physiological), we construct separate Bayesian networks after dimensionality reduction. Principal Component Analysis (PCA) reduces the feature space while preserving 95% of the explained variance, resulting in 32 components for visual, 62 for audio, and 46 for physiological data. This reduction accelerates the graph discovery process without significant information loss.

In this study, we define environments based on individual participants, as participant variation is the primary source of distribution shifts in our dataset, while task parameters and annotators remain constant. Consequently, our analysis encompasses 18 distinct environments, each corresponding to a different participant.

4.2 Evaluation Protocol

Since all variables—predictors and targets—involved in this dataset are continuous, we build Gaussian Bayesian networks for modeling and predicting affective states. Gaussian Bayesian networks model continuous variables by assuming that each node follows a conditional Gaussian distribution given its parents. For each variable Z_i in the network, its conditional distribution is specified by a linear regression model where the predictors are its parent variables PA_i , with the regression coefficients representing the strength of relationships [9]. The joint distribution of all variables is then factorized according to (2).

We evaluate our method using leave-one-participant-out cross-validation and measure performance using Pearson’s Correlation Coefficient (PCC) between predicted and ground-truth affect labels. PCC is widely used in affect modeling evaluation [16, 22] as it measures the linear relationship between predicted and ground truth emotional states, making it particularly suitable for continuous-valued affect predictions. Additionally, it is invariant to scaling and shifting of the predictions, which is crucial in affect modeling where absolute values are often less important than capturing the correct patterns of emotional change over time. Finally, we compare the PCC score of our method against two baseline models—random forests and neural networks—of higher learning capacity than Bayesian networks. To ensure methodological rigor, all models—including baselines and Bayesian networks—are trained and evaluated using identical datasets.

5 Results

As mentioned above, the evaluation employs a leave-one-participant-out cross-validation methodology. This protocol involves training and testing each model through 18 independent iterations, corresponding to the number of environments, yielding over 600 experimental runs. The performance assessment utilizes the PCC score, averaged across all environments for each model. Table 1 summarizes the comparative results of this analysis.

For arousal prediction, our proposed model demonstrates comparable performance to more sophisticated architectures: random forest and neural network models. While neural networks exhibit marginally superior performance, our approach consistently maintains the second-highest performance, with the average graph model slightly outperforming the ensemble variant. In the context of valence prediction, our methodology achieves superior performance across two of the three investigated modalities. It is noteworthy that previous investigations utilizing the RECOLA dataset² have established valence prediction as a more challenging task compared to arousal prediction. Nevertheless, our causal-inspired framework demonstrates robust predictive capability for valence through its focus on robust relationships among predictors and target variables.

²<https://diuf.unifr.ch/main/diva/recola/news.html>

		Visual	Audio	Physiology
Arousal	AVG _{GES}	0.18	0.44	0.36
	AVG _{PC}	0.30	0.47	0.35
	ENS _{GES}	0.26	0.44	0.36
	ENS _{PC}	0.28	0.46	0.34
	Random Forest	0.29	0.46	0.34
	Neural Network	0.32	0.49	0.38
Valence	AVG _{GES}	0.22	0.20	0.36
	AVG _{PC}	0.37	0.23	0.36
	ENS _{GES}	0.33	0.24	0.36
	ENS _{PC}	0.36	0.24	0.35
	Random Forest	0.33	0.23	0.33
	Neural Network	0.32	0.25	0.34

Table 1: Performance comparison across different modalities and methods for Arousal and Valence prediction in terms of average PCC score. AVG_{GES} and AVG_{PC} correspond to the Bayesian network that mirrors the structure of the average graphs discovered by the GES and PC algorithms, respectively. ENS_{GES} and ENS_{PC} correspond to the ensemble method where the Bayesian networks mirror the structure of the environment-specific graphs discovered by the GES and PC algorithms, respectively.

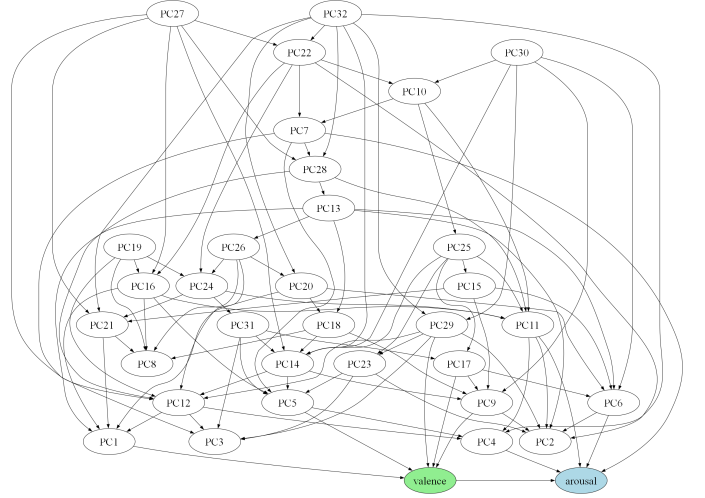


Figure 2: The graph discovered by the PC algorithm for the visual information of participant with 37.

The empirical evidence demonstrates that our methodology consistently surpasses random forest models in predictive accuracy across both affect dimensions—arousal and valence—despite the fact that the learning capacity of the developed Bayesian networks is lower. Furthermore, our approach offers a distinct advantage over both random forests and neural networks through its inherent transparency, providing interpretable insights into emotional manifestation mechanisms. These findings underscore the significance of identifying and leveraging causal relationships in affect modeling, particularly exploiting the inherent property of affect corpora being captured under different environments.

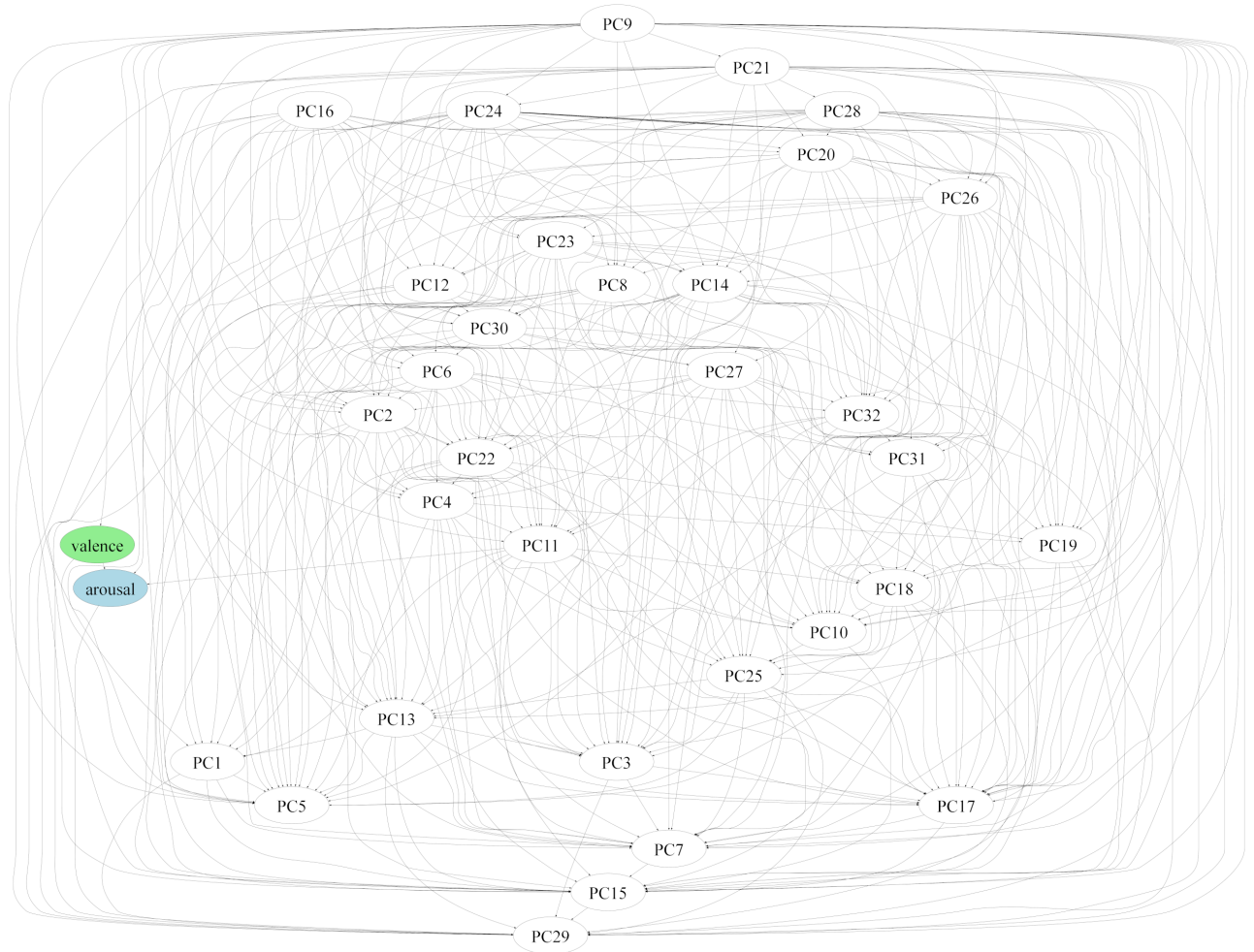


Figure 3: The graph discovered by the GES algorithm for the visual information of participant 37.

Figure 3 and Figure 2 illustrates two causal graphs generated by distinct graph discovery algorithms. The PC algorithm yields a notably sparser graph structure compared to the GES algorithm, reflecting their fundamentally different approaches to causal discovery. This divergence in graph structures is theoretically sound, as multiple causal structures can be consistent the same observational data [37].

A particularly significant finding evident in these figures is the presence of a directed edge between the nodes representing arousal and valence. This edge is consistently present across nearly all environments in our dataset. While this might appear to conflict with Russell’s affect circumplex theory [35], which posits independence between these affect dimensions, our finding actually extends the theoretical framework. Specifically, while arousal and valence may be independent when affect measurements remain unobserved, their relationship changes when those measurements are explicitly observed. This novel observation reveals a conditional dependency structure that has remained unidentified and unexploited in previous affect modeling research.

6 Conclusions

This study introduces a novel causal-inspired methodology for affect modeling that fundamentally reframes how we approach distribution shifts and emotional subjectivity in affective computing. Rather than treating these characteristics as obstacles, our approach leverages them as inherent features of affect data to enhance model robustness and interpretability. Our methodology demonstrates competitive performance with more complex architectures while maintaining full transparency in its decision-making process. A significant finding of this work is the discovery of a directed edge between arousal and valence nodes, which appears consistently across different environments. This observation extends Russell’s circumplex theory by revealing conditional dependencies between these affect dimensions when measurements are explicitly observed. This insight provides new perspectives on the relationship between fundamental affect dimensions and their manifestation in observable data. Finally, this work establishes a foundation for developing SCMs in affective computing research, opening new avenues for

investigating interventional and counterfactual scenarios in affect modeling.

Acknowledgments

This study was supported by Project ERICA (GA: REP-2023-36) financed by the Malta Council for Science & Technology (MCST), for and on behalf of the Foundation for Science and Technology, through the FUSION: R&I Research Excellence Programme. AI tools have been used to detect typos, syntax, and grammar errors, as well as to suggest changes that improve the clarity of the text.

References

- [1] Matthew Barthet, Maria Kaselimi, Kosmas Pinitas, Konstantinos Makantasis, Antonios Liapis, and Georgios N Yannakakis. 2024. GameVibe: a multimodal affective game corpus. *Scientific Data* 11, 1 (2024), 1306.
- [2] Matthew Barthet, Chintan Trivedi, Kosmas Pinitas, Emmanouil Xylakis, Konstantinos Makantasis, Antonios Liapis, and Georgios N Yannakakis. 2023. Knowing your annotator: Rapidly testing the reliability of affect annotation. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 1–8.
- [3] Dongwei Chen, Rui Miao, Zhaoyong Deng, Na Han, and Chunjian Deng. 2021. Sparse granger causality analysis model based on sensors correlation for emotion recognition classification in electroencephalography. *Frontiers in computational neuroscience* 15 (2021), 684373.
- [4] Carlos Cinelli, Daniel Kumor, Bryant Chen, Judea Pearl, and Elias Bareinboim. 2019. Sensitivity analysis of linear structural causal models. In *International conference on machine learning*. PMLR, 1252–1261.
- [5] C Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez. 2016. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5562–5570.
- [6] Karl J Friston, André M Bastos, Ashwini Oswal, Bernadette Van Wijk, Craig Richter, and Vladimir Litvak. 2014. Granger causality revisited. *Neuroimage* 101 (2014), 796–808.
- [7] Yunyuan Gao, Xiangkun Wang, Thomas Potter, Jianhai Zhang, and Yingchun Zhang. 2020. Single-trial EEG emotion recognition using Granger Causality/Transfer Entropy analysis. *Journal of Neuroscience Methods* 346 (2020), 108904.
- [8] Clark Glymour, Kun Zhang, and Peter Spirtes. 2019. Review of causal discovery methods based on graphical models. *Frontiers in genetics* 10 (2019), 524.
- [9] Marco Grzegorzczak. 2023. Being Bayesian about learning Gaussian Bayesian networks from incomplete data. *International Journal of Approximate Reasoning* 160 (2023), 108954.
- [10] Wenjing Han, Tao Jiang, Yan Li, Björn Schuller, and Huabin Ruan. 2020. Ordinal learning for emotion recognition in customer service calls. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 6494–6498.
- [11] Trevor Hastie. 2009. The elements of statistical learning: data mining, inference, and prediction.
- [12] Biwei Huang, Kun Zhang, Yizhu Lin, Bernhard Schölkopf, and Clark Glymour. 2018. Generalized score functions for causal discovery. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1551–1560.
- [13] Biwei Huang, Kun Zhang, Jiji Zhang, Joseph Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. 2020. Causal discovery from heterogeneous/nonstationary data. *Journal of Machine Learning Research* 21, 89 (2020), 1–53.
- [14] GUO Jinliang, Fang Fang, Wei Wang, and Fuji Ren. 2018. EEG emotion recognition based on granger causality and capsnet neural network. In *2018 5th IEEE international conference on cloud computing and intelligence systems (CCIS)*. IEEE, 47–52.
- [15] Jean Kossaifi, Georgios Tzimiropoulos, Sinisa Todorovic, and Maja Pantic. 2017. AFEW-VA database for valence and arousal estimation in-the-wild. *Image and Vision Computing* 65 (2017), 23–36.
- [16] Jean Kossaifi, Robert Walecki, Yannis Panagakis, Jie Shen, Maximilian Schmitt, Fabien Ringeval, Jing Han, Vedhas Pandit, Antoine Toisoul, Björn Schuller, et al. 2019. Sewa db: A rich database for audio-visual emotion and sentiment research in the wild. *IEEE transactions on pattern analysis and machine intelligence* 43, 3 (2019), 1022–1040.
- [17] Konstantinos Makantasis. 2021. Affranknet+: ranking affect using privileged information. In *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 1–8.
- [18] Konstantinos Makantasis, Antonios Liapis, and Georgios N Yannakakis. 2019. From pixels to affect: A study on games and player experience. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–7.
- [19] Konstantinos Makantasis, Antonios Liapis, and Georgios N Yannakakis. 2021. The pixels and sounds of emotion: General-purpose representations of arousal in games. *IEEE Transactions on Affective Computing* 14, 1 (2021), 680–693.
- [20] Konstantinos Makantasis, David Melhart, Antonios Liapis, and Georgios N Yannakakis. 2021. Privileged information for modeling affect in the wild. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.
- [21] Konstantinos Makantasis, Kosmas Pinitas, Antonios Liapis, and Georgios N Yannakakis. 2022. The invariant ground truth of affect. In *2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 1–8.
- [22] Konstantinos Makantasis, Kosmas Pinitas, Antonios Liapis, and Georgios N Yannakakis. 2023. From the lab to the wild: Affect modeling via privileged information. *IEEE Transactions on Affective Computing* (2023).
- [23] David Melhart, Antonios Liapis, and Georgios N Yannakakis. 2022. The arousal video game annotation (AGAIN) dataset. *IEEE Transactions on Affective Computing* 13, 4 (2022), 2171–2184.
- [24] Trisha Mittal, Puneet Mathur, Aniket Bera, and Dinesh Manocha. 2021. Affect2mm: Affective analysis of multimedia content using emotion causality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5661–5671.
- [25] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. 2017. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing* 10, 1 (2017), 18–31.
- [26] Christoph Molnar. 2020. *Interpretable machine learning*. Lulu. com.
- [27] Srinivas Parthasarathy and Carlos Busso. 2018. Preference-learning with qualitative agreement for sentence level emotional annotations. *Interspeech 2018* (2018).
- [28] Kosmas Pinitas, Konstantinos Makantasis, Antonios Liapis, and Georgios N Yannakakis. 2022. RankNEAT: outperforming stochastic gradient search in preference learning tasks. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 1084–1092.
- [29] Kosmas Pinitas, Konstantinos Makantasis, Antonios Liapis, and Georgios N Yannakakis. 2022. Supervised contrastive learning for affect modelling. In *Proceedings of the 2022 International Conference on Multimodal Interaction*. 531–539.
- [30] Kosmas Pinitas, Konstantinos Makantasis, and Georgios N Yannakakis. 2024. Across-game engagement modelling via few-shot learning. In *2024 18th European Conference on Computer Vision (ECCV) - CV2 Workshop*.
- [31] Kosmas Pinitas, Nemanja Rasajski, Matthew Barthet, Maria Kaselimi, Konstantinos Makantasis, Antonios Liapis, and Georgios N Yannakakis. 2024. Varying the context to advance affect modelling: A study on game engagement prediction. In *Proceedings of the IEEE International Conference on Affective Computing and Intelligent Interaction (ACII)*.
- [32] Kosmas Pinitas, David Renaudie, Mike Thomsen, Matthew Barthet, Konstantinos Makantasis, Antonios Liapis, and Georgios N Yannakakis. 2023. Predicting Player Engagement in Tom Clancy's The Division 2: A Multimodal Approach via Pixels and Gamepad Actions. In *Proceedings of the 25th International Conference on Multimodal Interaction*. 488–497.
- [33] Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. 2013. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*. IEEE, 1–8.
- [34] Georgios Rizos and Björn W Schuller. 2020. Average jane, where art thou?—recent avenues in efficient machine learning under subjectivity uncertainty. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems: 18th International Conference, IPMU 2020, Lisbon, Portugal, June 15–19, 2020, Proceedings, Part I* 18. Springer, 42–55.
- [35] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.
- [36] Enrique Sanchez, Mani Kumar Tellamekala, Michel Valstar, and Georgios Tzimiropoulos. 2021. Affective processes: stochastic modelling of temporal context for emotion and facial expression recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9074–9084.
- [37] Bernhard Schölkopf. 2022. Causality for machine learning. In *Probabilistic and causal inference: The works of Judea Pearl*. 765–804.
- [38] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. 2021. Toward causal representation learning. *Proc. IEEE* 109, 5 (2021), 612–634.
- [39] Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, et al. 2013. The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France*.
- [40] Peter Spirtes, Clark Glymour, and Richard Scheines. 2001. *Causation, prediction, and search*. MIT press.

- [41] Antoine Toisoul, Jean Kossaifi, Adrian Bulat, Georgios Tzimiropoulos, and Maja Pantic. 2021. Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nature Machine Intelligence* 3, 1 (2021), 42–50.
- [42] Panagiotis Tzirakis, George Trigeorgis, Mihalis A Nicolaou, Björn W Schuller, and Stefanos Zafeiriou. 2017. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of selected topics in signal processing* 11, 8 (2017), 1301–1309.
- [43] Yan Wang, Wei Song, Wei Tao, Antonio Liotta, Dawei Yang, Xinlei Li, Shuyong Gao, Yixuan Sun, Weifeng Ge, Wei Zhang, et al. 2022. A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion* 83 (2022), 19–52.
- [44] Georgios N Yannakakis, Roddy Cowie, and Carlos Busso. 2018. The ordinal nature of emotions: An emerging approach. *IEEE Transactions on Affective Computing* 12, 1 (2018), 16–35.
- [45] Rahadian Yusuf, Dipak Gaire Sharma, Ivan Tanev, and Katsunori Shimohara. 2017. Individuality and user-specific approach in adaptive emotion recognition model. In *2017 International Conference On Biometrics and Kansei Engineering (ICBAKE)*. IEEE, 1–6.