Multi-Source Asynchrony Time Series Integration*

Abstract

Keywords:

1 Introduction

Wearable devices enable non-intrusive measurement of physiological biomarkers that correlate with stress levels, emotional states, and other biological responses. Those measurements often include heart rate Variability (HRV), Electrodermal Activity (ADA), Heart Rate (HR), and three-axis acceleration (ACC) [1]. Advances in machine learning have allowed us to predict emotional states from these biomarkers, reflecting a shift toward recognizing mental well-being as an integral component of human health.

Authors in [2] evaluate a set of traditional machine learning algorithms to predict people's stress based on EDA activity, including K-Nearest Neighbor, Support Vector Machine (SVM), Naive Bayes, Logistic Regression, and Random Forest. They trained models on both statistical features and raw sensor readings, finding that SVM achieved the highest accuracy, although performance varied inconsistently between feature-based and raw-data approaches.

Despite their utility, shallow models often lack expressiveness and capacity to generalize well [3]. Moreover, features often rely on statistics, forgetting sequential dependencies in the data. A closer overview dives us into a multi-modality scenario, where signals are sampled at different frequencies, introducing additional challenges for feature extraction and fusion.

Deep learning approaches address these limitations by automatically leveraging data structures as time dependencies for sequential recordings or spatial patterns for images through feature representation from multiple data entities. However, a key challenge lies in effectively combining heterogeneous data sources [4, 5].

The work developed by [6] demonstrated the power of multimodal fusion by integrating autoencoders for genetic data with 3D CNNs for imaging, outperforming shallow and single-modality baselines. In the domain of physiological sensing, [7] proposed a CNN-based feature extractor for time-series sensor data, while [8] developed an attention-based LSTM framework to fuse smartphone and wearable signals for emotion recognition. Prior studies by [9, 10] further highlight the effectiveness of LSTM architectures in modeling inter-participant variability and integrating multiple modalities.

*Under grants provided by the research project 111091991908, funded by MINCIENCIAS.

2 Mathematical Framework

2.1 Problem Definition

Consider a set of P variables, where the p-th variable contains L_p observations as $\boldsymbol{x} = \{(t_l^{(p)}, x_l^{(p)})\}_{l=1,p=1}^{L_p,P}$, being $x_l^{(p)} \in \mathbb{R}$ the corresponding observation at time $t_l^{(p)} \in \mathbb{R}$. A graph representation of that structure with $P = 3$ is plotted in Figure 1. Each $\boldsymbol{x}$ has its own target variable $\boldsymbol{y} \in \mathbb{R}^D$, leading to a collection of N input-output i.i.d. pairs denoted as $\mathcal{D} = \{\boldsymbol{x}_n, \boldsymbol{y}_n\}_{n=1}^N = \{\boldsymbol{X}, \boldsymbol{Y}\}$ called training set. The task is to generalize the map from each input $\boldsymbol{x}$ to its corresponding target output $\boldsymbol{y}$ in a stochastic fashion.

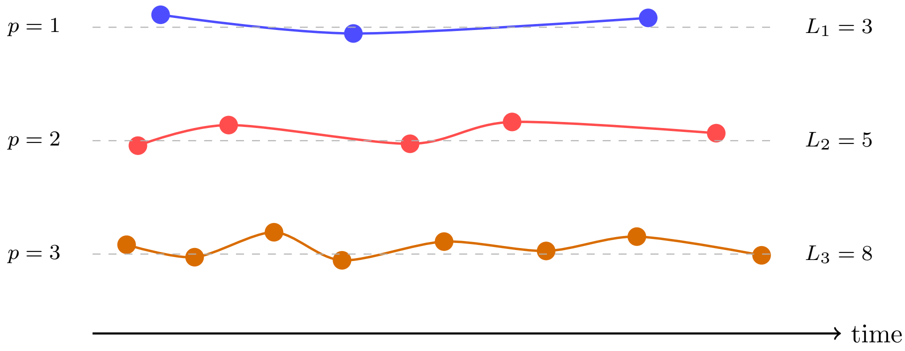

Figure 1: Structure of input samples. Each dot represent a pair $(t_l^{(p)}, x_l^{(p)})$.

2.2 Likelihood Model

For this propose, consider a likelihood functions that rule the generation of recorded targets $\boldsymbol{Y}$ from inputs $\boldsymbol{X}$ through some set of parameters $\boldsymbol{\theta} \subseteq \mathbb{R}^J$

$$p\left(\boldsymbol{Y} \mid \boldsymbol{\theta}(\boldsymbol{X})\right) = \prod_{n=1}^{N} p\left(\boldsymbol{y}_n \mid \boldsymbol{\theta}(\boldsymbol{x}_n)\right). \tag{1}$$

Each element of $\boldsymbol{\theta}(\boldsymbol{x})$, denoted as $\theta_j(\boldsymbol{x})$, could be restricted to some subset of $\mathbb{R}$. To handle that, we model $\theta_j(\boldsymbol{x}) = h_j(f_j(\boldsymbol{x}))$ as a transformation of an unrestricted latent variable $f_j(\boldsymbol{x})$ via a link function h_j. Our task boils down to finding the latent vector function $\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), \cdots, f_J(\boldsymbol{x})]^\top \in \mathbb{R}^J$.

3 Results

4 Conclusions

References

[1] G. Vos, K. Trinh, Z. Sarnyai, and M. R. Azghadi, "Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review," 5 2023.

[2] L. Zhu, P. Spachos, P. C. Ng, Y. Yu, Y. Wang, K. Plataniotis, and D. Hatzinakos, "Stress detection through wrist-based electrodermal activity monitoring and machine learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, pp. 2155–2165, 5 2023.

[3] K. Yang, C. Wang, Y. Gu, Z. Sarsenbayeva, B. Tag, T. Dingler, G. Wadley, and J. Goncalves, "Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, pp. 1082–1097, 4 2023.

[4] T. Baltrusaitis, C. Ahuja, and L. P. Morency, "Multimodal machine learning: A survey and taxonomy," 2 2019.

[5] P. P. Liang, A. Zadeh, and L. P. Morency, "Foundations & trends in multimodal machine learning: Principles, challenges, and open questions," *ACM Computing Surveys*, vol. 56, 6 2024.

[6] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of alzheimer's disease stage," *Scientific Reports*, vol. 11, 12 2021.

[7] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Networks and Applications*, vol. 25, no. 2, p. 743 – 755, 2020. Cited by: 470.

[8] K. Yang, C. Wang, Y. Gu, Z. Sarsenbayeva, B. Tag, T. Dingler, G. Wadley, and J. Goncalves, "Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 2, p. 1082 – 1097, 2023. Cited by: 45.

[9] G. Zhang and A. Etemad, "Capsule attention for multimodal eeg-eog representation learning with application to driver vigilance estimation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, p. 1138 – 1149, 2021. Cited by: 65; All Open Access, Gold Open Access, Green Open Access.

[10] Q. Li, J. Tan, J. Wang, and H. Chen, "A multimodal event-driven lstm model for stock prediction using online news," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 10, p. 3323 – 3337, 2021. Cited by: 122; All Open Access, Bronze Open Access.