Sequential Models for Automatic Personality Recognition from Multimodal Information in Social Interactions

Jeanfed Ramírez Lima, Hugo Jair Escalante Balderas, Luis Villaseñor Pineda
Instituto Nacional de Astrofísica Óptica y Electrónica

Abstract

We study the problem of recognizing personality from videos depicting users' social interaction. Multimodal information is represented by using pretrained models and multi stream sequential models are considered for prediction. Experimental results of the proposed method in the recently released UDIVA dataset are reported and compared to related work. We show that the proposed methodology is competitive with the state-of-the-art while using less complex models.

Introduction

The **automatic personality recognition** task has become one of the topics of higher interest in the affective computing field.

The principal purpose of the task is to find patterns of behavior leveraged from human interactions, e.g., from **social environments**.

Personality can be measured using **The Big Five** personality traits theory.

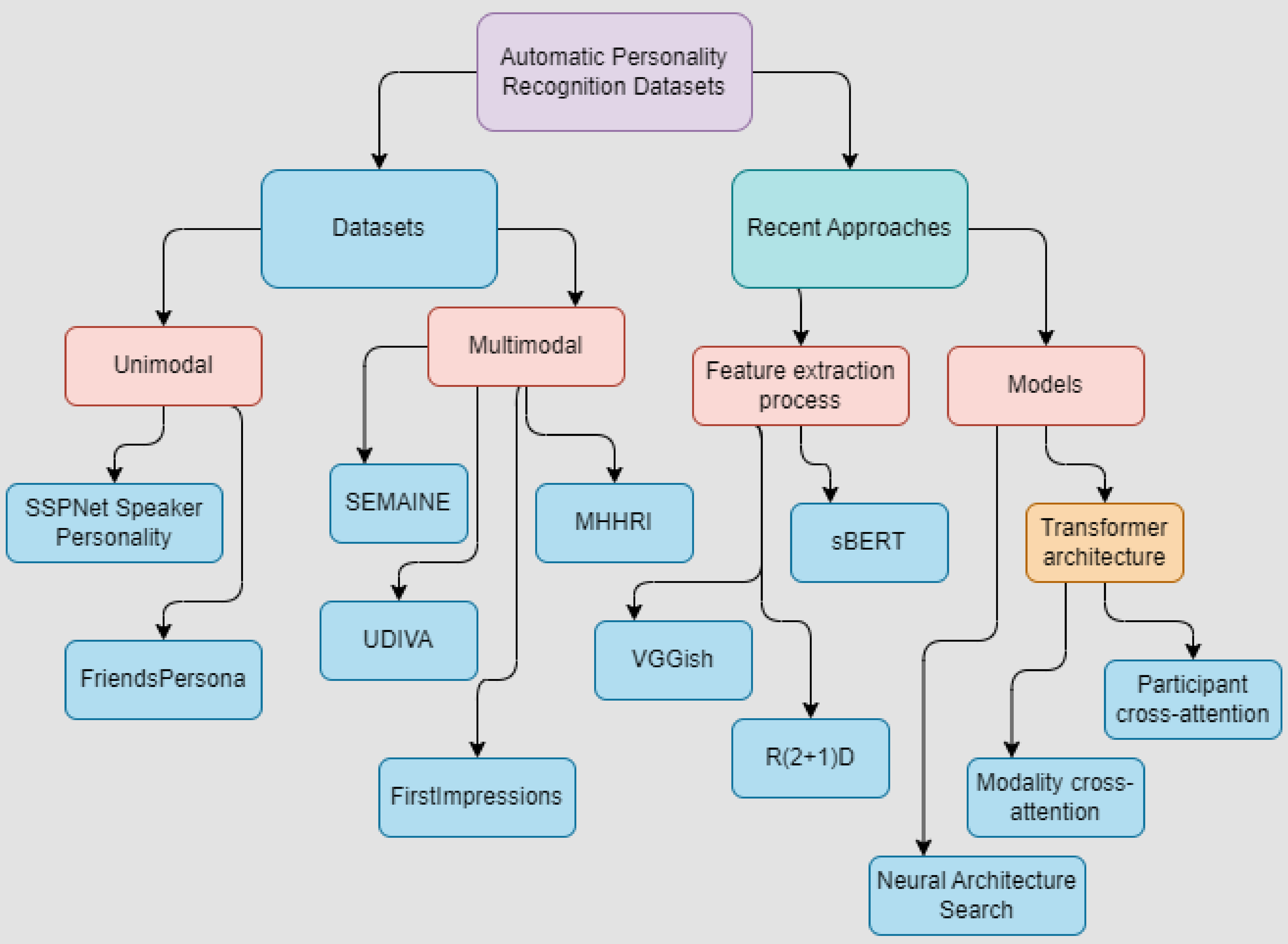
- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

One of its most interesting applications is to provide social mechanisms to computers, in such way that the human-computer interaction becomes more natural.

Recent methods that achieved state-of-the-art results utilize custom Transformer architectures and Neural Architecture Search. Some of these methods ignore the sequential nature of the data.

Our approach uses a simpler architecture and considers data sequentially, obtaining competitive results.

Related work



Methodology

1 Dataset

We used the **UDIVA v0.5** dataset since it contains **multimodal information** derived from **social interactions.**

2 Information representation

Sequences of utterances for each of the modalities.

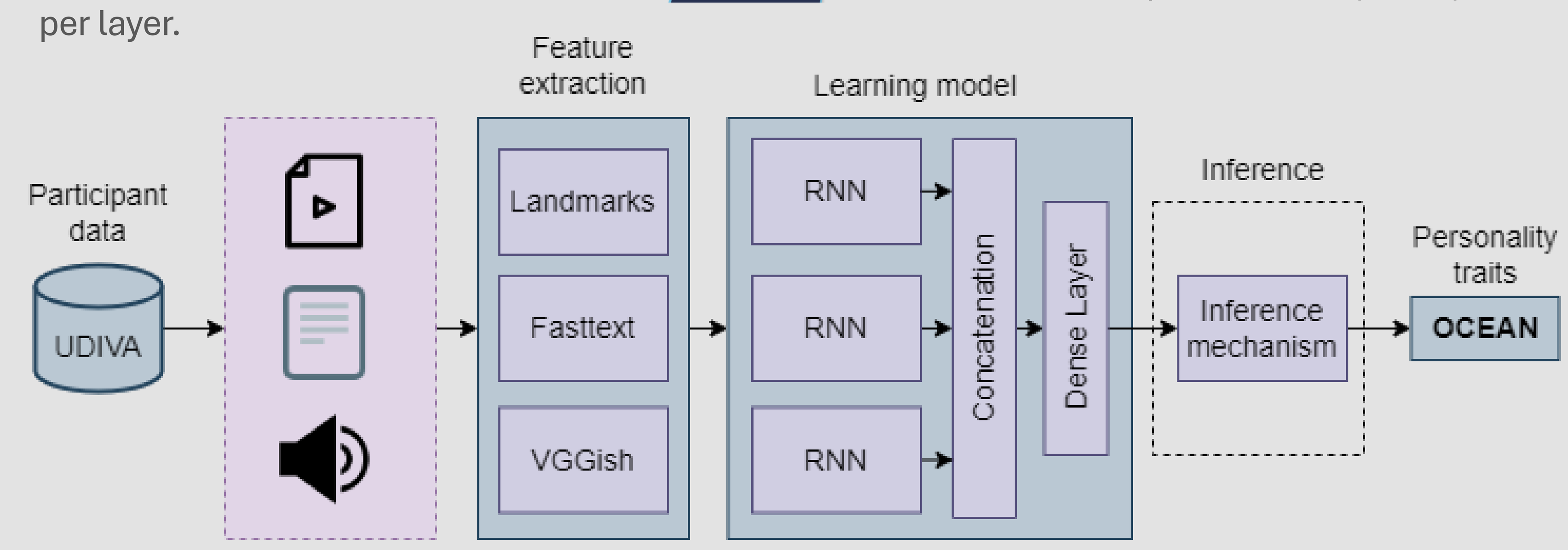
3 Feature extraction

We opted to use pretrained models to extract features from each of the modalities.

- Fasttext (text)
- VGGish (audio)
- 3DDFAv2, MeTRAbs, FrankMokap, ETH-Gaze (visual)

4 Architecture design

Defined as a basis a RNN architecture we used **AutoKeras to find the best model hyperparameters**, such as the number of layers and number of units per layer.

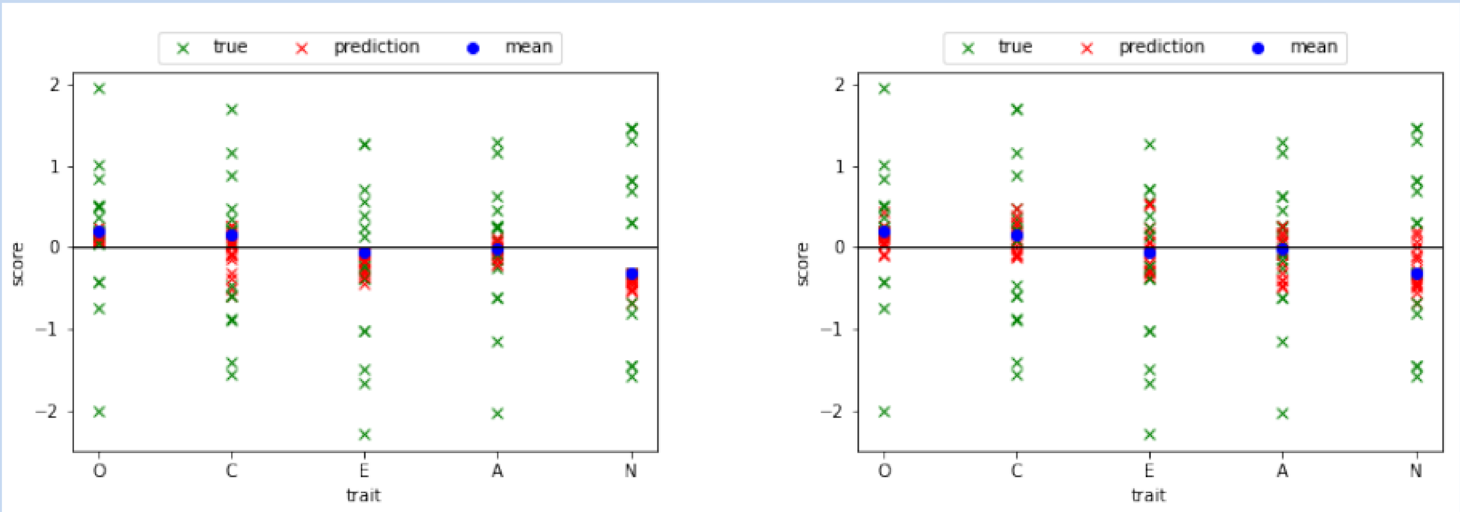


Experimental results

A comparison between our approaches (3 seconds window and utterance level window and state-of-the-art approaches.

Literature models	O	C	E	A	N	AMSE
GW bimodal NAS [4]	**0.684**	**0.588**	0.830	0.550	**0.796**	**0.690**
Dyadformer [5]	-	-	-	-	-	0.722
Transformer [3]	0.744	0.794	0.886	0.653	1.012	0.818
Proposed model ($w_L = 3$)						
LSTM_I	0.726	0.830	0.956	0.674	1.220	0.881
LSTM_T	0.740	0.814	0.942	0.677	1.226	0.888
LSTM_A	0.724	0.849	0.915	0.664	1.163	0.863
GRU_{I+T}	0.721	0.720	**0.802**	0.643	1.291	0.836
GRU_{I+A}	0.766	0.765	0.945	0.661	1.145	0.856
GRU_{T+A}	0.718	0.737	0.859	0.605	1.109	0.805
LSTM_{T+I+A}	0.737	0.869	0.955	0.662	1.156	0.888
Proposed model ($w_L = \text{utterance}$)						
GRU_I	0.724	0.848	1.004	0.663	1.146	0.877
GRU_T	0.738	0.828	0.983	0.677	1.130	0.871
GRU_A	0.730	0.790	0.876	**0.525**	0.989	0.782
GRU_{I+T}	0.737	0.816	0.883	0.632	1.171	0.848
GRU_{I+A}	0.749	0.788	0.883	0.588	1.015	0.804
GRU_{T+A}	0.777	0.714	0.855	0.526	0.905	0.755
LSTM_{T+I+A}	0.749	0.726	0.886	0.598	1.112	0.814

The dispersion of the values before and after passing through the inference mechanism, where the variance of the predictions is higher at utterance level.



(a) 3 second window (b) Utterance

Conclusion

The presented methodology for the recognition of personality from multimodal data effectively exploits sequential information. This has shown competitive results in the Extraversion and Agreeableness personality traits. As future work we plan to modify the inference mechanism for other way of regressing the last trait score.

Referencias

1. M. S. Aliakbarian, F. S. Saleh, M. Salzmann, B. Fernando, L. Petersson, and L. Andersson, "Encouraging LSTMs to Anticipate Actions Very Early," in 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, pp. 280–289. doi: 10.1109/ICCV.2017.39.
2. H. J. Escalante, M. Montes-y-Gómez, L. Villaseñor-Pineda, and M. L. Errecalde, "Early text classification: a Naive solution," Proceedings of NAACL-HLT 2016, pages 91–99. arXiv, 2015. doi: 10.48550/ARXIV.1509.06053.
3. K. Zhou, C. Shu, B. Li, and J. H. Lau, "Early Rumour Detection," in Proceedings of the 2019 Conference of the North, 2019, pp. 1614–1623. doi: 10.18653/v1/N19-1163.
4. C. Palmero et al., "Context-Aware Personality Inference in Dyadic Scenarios: Introducing the UDIVA Dataset," in 2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW), Jan. 2021, pp. 1–12. doi: 10.1109/WACVW52041.2021.00005.
5. H. Salam, V. Manoranjan, J. Jiang, and O. Celiktutan, "Learning Personalised Models for Automatic Self-Reported Personality Recognition," in Understanding Social Behavior in Dyadic and Small Group Interactions, Jun. 2022, vol. 173, pp. 53–73. [Online]. Available: https://proceedings.mlr.press/v173/salam22a.html
6. D. Curto et al., "Dyadformer: A Multi-modal Transformer for Long-Range Modeling of Dyadic Interactions," in 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Oct. 2021, pp. 2177–2188. doi: 10.1109/ICCVW54120.2021.00247