Automated assessment of symptoms of anxiety in children from parent-child interaction videos

Aristophanes Albertus Alvin,
Supervisors: Mang Ning, MSc, and Dr. Itir Önal Ertuğrul

Utrecht University, The Netherlands
a.aristophanesalbertusalvin@students.uu.nl
{m.ning, i.onalertugrul}@uu.nl

Abstract. Anxiety symptoms are expressed more saliently through certain modes of communications and contexts. In this thesis, we applied machine learning for the detection of anxiety symptoms in 9-year-old children. Our models are trained on features extracted from video recordings of interactions between 9-year-olds and their parents. Results suggest that symptoms are most noticeable in conflictual interactions and are conveyed through the hand movements, facial expressions, and word choice. To our knowledge, this is the first study that applies machine learning for detecting anxiety symptoms in children, utilizes multimodal features, and considers the interactional context.

Keywords: anxiety · children · automatic · symptoms · assessment · detection · machine learning

1 Introduction

Current anxiety diagnostic methods rely mostly on the reported level of subjective distress and suffering to distinguish normal anxiety from pathological anxiety [1, 2]. This approach can introduce bias into diagnoses and makes early detection less accessible to some segments of patients. This work aims to demonstrate the feasibility of applying machine learning for the detection of anxiety symptoms in 9-year-old children using multimodal behavioral features.

2 Methods

Features were extracted from four modalities: acoustic, linguistic, facial and bodily expressions. From the child's voice, OpenSMILE extracted a set of 88 acoustic parameters of the *extended Geneva Minimalistic Acoustic Parameter Set* (eGeMAPS; [3]). Linguistic Inquiry and Word Count (LIWC, or the latest software iteration, LIWC-22) text analysis toolkit [4] extracted 85 linguistic parameters. Py-Feat [5] detected the probability of activation of 20 Facial Action Units (AUs) on the face of the subject. Also, OpenPose [6] detected the positions of 9 Key Points on the subject's body, from which 15 affective body pose features were derived.

Performance of three models (Gradient Boosting Classifier, Light Gradient Boosting Machine, and Support Vector Machine) were compared across types of feature modalities (unimodal and multimodal), fusion types (early and late fusion), and interaction scenarios (conflictual and cooperative). Multimodal models used features from different modalities, which were merged using early and late fusion techniques. In early fusion, feature sets from all modalities were concatenated horizontally, i.e., extending the feature vector of each data point. A model trained on this combined feature set therefore considers information from various modalities when making predictions. In contrast, features from different modalities remained separated and were used to train individual models in the late fusion approach (see Fig. 1).

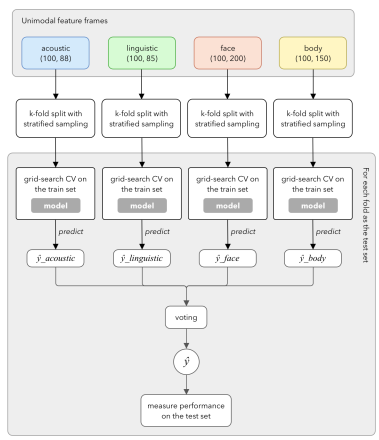

Fig. 1. Late fusion model pipeline.

3 Results

Our GBC models were able to automatically assess anxiety symptoms from multimodal behavioral features of children. The best performance was produced using multimodal features merged using early fusion technique (accuracy = 0.66%, MCC = 0.29), and features from conflictual interaction videos performed significantly better than cooperative interaction (p = 0.001). As an explainable methodology, the most significant indicators of anxiety symptoms were identified using Shapley values that are derived from game theory to determine the contribution of individual players. These key behavioral features are hand movements, body posture, facial expression (more specifically, the mouth area), and choice of words.

References

1. Knepley, M. J., Kendall, P. C., Carper, M. M.: An Analysis of the Child Behavior Checklist Anxiety Problems Scale's Predictive Capabilities. Journal of Psychopathology and Behavioral Assessment **41**(2), 249–256 (2019)
2. Read, K. L., Settipani, C. A., Peterman, J., Kendall, P. C., Compton, S., Piacentini, J., McCracken, J., Bergman, L., Walkup, J., Sakolsky, D., Birmaher, B., Albano, A. M., Rynn, M., Ginsburg, G., Keeton, C., Gosch, E., Suveg, C., Sherrill, J., March, J.: Predicting Anxiety Diagnoses and Severity with the CBCL-A: Improvement Relative to Other CBCL Scales? Journal of Psychopathology and Behavioral Assessment **37**(1), 100–111 (2015)
3. Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., Truong, K. P.: The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. IEEE Transactions on Affective Computing **7**(2), 190–202 (2015). https://doi.org/10.1109/TAFFC.2015.2457417
4. Boyd, R. L., Ashokkumar, A., Seraj, S., Pennebaker, J. W.: The development and psychometric properties of LIWC-22. University of Texas at Austin, Austin, TX (2022)
5. Cheong, J. H., Jolly, E., Tiankang, X., Byrne, S., Kenney, M., Chang, L. J.: Py-Feat: Python Facial Expression Analysis Toolbox. Affective Science **4**, 781–796 (2023). https://doi.org/10.1007/s42761-023-00191-4
6. Cao, Z., Hidalgo, G., Simon, T., Wei, S., Sheikh, Y.: OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. IEEE Transactions on Pattern Analysis & Machine Intelligence **43**(1), 172–186 (2021). https://doi.org/10.1109/TPAMI.2019.2929257