Assessment of Collaborative Interactions Based on the Multimodal Dataset MULTISIMO

Shenghan Hou
Trinity College Dublin
Dublin ,Ireland
housh@tcd.ie

Devang Kankaria
Trinity College Dublin
Dublin ,Ireland
kankarid@tcd.ie

Naveen Prakash Ramesh
Trinity College Dublin
Dublin ,Ireland
rameshn@tcd.ie

Alok Singh
Trinity College Dublin
Dublin ,Ireland
alsingh@tcd.ie

Zitong Liang
Trinity College Dublin
Dublin ,Ireland
ziliang@tcd.ie

Mingyang Xia
Trinity College Dublin
Dublin ,Ireland
mxia@tcd.ie

***Abstract*—This work aims to explore the nuances of cooperative interactions included in the multimodal dataset of the MULTISIMO corpus. a careful examination of the relationship between the personalities of participants and their cooperative behavior in a group setting. Our research team plans to examine the connections between these characteristics and how they affect teamwork using data analysis on a range of subjects including visual recordings aural signals and body language. These results provide new insights into the dynamic relationships among interactions which can polish group work effectiveness and collaborative and instructional approach design.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Collaboration refers to the process wherein two or more people work together to achieve a shared task (Davin Donato, 2013). Studies have been conducted in various fields, such as schools and workplaces, to analyse how individuals' personalities relate to their collaborative behaviours in a group setting (Matthias et al., 2019; Hotmaulina et al., 2023). These studies can shed light on education planning and the design of collaboration tools that help people better understand group dynamics and facilitate collaboration. In this study, we aim to further the current research on the topic by delineating the participants' data proposed by the MULTISIMO corpus, to better understand the subtle interactions between distinct psychological profiles and how group interactions demonstrate temperament in light of their profiles. We will examine the impact of personality traits and previous events/historical records of the individuals on collaboration. The study aims to unfold the framework of productive coordination entangled with interpersonal relationships, all while emphasising personal disparities and the respective roles assigned in group work.

II. Dataset Review

The MULTISIMO corpus we have chosen represents an innovative approach to studying human social behaviour through multimodal interactions in a preset situation that requires participants' collaboration (Koutsombogera Vogel, 2017). This corpus is intended to enhance the understanding of interaction flows, communicative intent, and the affective states of speakers within a structured setting. In this section, I will provide a brief introduction to the data collection and processing procedures of the dataset we selected for the study. The data of the MULTISIMO project was collected through an experimental scenario aimed at fostering collaboration among participants, consisting of two players and one facilitator, who engaged in a game-like task. In total, 49 participants across 23 sessions participated in the study. The participants were asked to complete the Big Five Inventory and post-session experience surveys to assess the relationship between personality and collaborative dynamics. The Big Five Inventory measures personality of the participants from five dimensions: extraversion vs. introversion, agreeableness vs. antagonism, conscientiousness vs. lack of direction, neuroticism vs. emotional stability, and openness vs. closedness to experience (John et al., 1991). Participants were assigned tasks involving determining and ranking the three most popular responses to a set of questions, based on a survey of 100 individuals. The setup aimed to trigger collaborative behaviour among the group of three people, 2 participants and 1 facilitator. The process was recorded by HD and 360 cameras, head-mounted and omnidirectional microphones, and a Kinect sensor for skeletal tracking. The final dataset included visual recordings from multiple angles, individual audio recordings, and motion tracking. The multimodal dataset serves as the foundation for analysing physical and verbal interaction dynamics that can help the researchers to study their behaviours. The participants, ranging from 19 to 44 years old, were randomly grouped and largely unfamiliar with each other, representing a balanced gender distribution (F=25, M=24) and diverse nationalities (eighteen nationalities), with one-third of the participants being native English speakers. This diversity allows for an analysis of the correlation between demographic data — gender, language, and familiarity — and collaboration (Koutsombogera Vogel, 2018). The corpus is further processed at various levels, including both manual annotation and automatic measurements. The manual annotation includes segmentation of speaker turns, labelling of gaze focus, head movement, and communicative functions that these entail. Annotators also assess the perceived dominance and personality traits of the session participants. The automatic measurements are audio and video features that are automatically extracted by software such as Praat (Boersma Weenink, n.d.) and iMotions (2016). Some of the data collected includes voice pitch, loudness, facial action units, and visual emotion valence. These processed, structured data allow us to better uncover the correlations between multimodal behaviours and collaboration effectiveness (Koutsombogera Vogel, 2018).

III. Corpus Feature Extraction (Literature Review Part 2)

Three studies were centred around the MULTISIMO corpus. (Koutsombogera, Costello, Vogel) The team extracted from the corpus the number of speeches, words per minute, total number of rounds, average duration, personality traits, and gaze characteristics. How dominance is quantified in multi-person conversations was investigated, with a particular focus on the relationship between linguistic (verbal and non-verbal) features and personality traits and their impact on perceived dominance ratings. The association between dominance ratings and personality traits and linguistic characteristics was explored through ordinal regression modelling. It was found that higher dominance ratings were associated with uttering more words per minute and higher extraversion and openness traits, while lower dominance ratings were associated with higher pleasantness. Also based on the relationship between gaze information and dialogue dominance, they found that participants' dialogue roles were important in interpreting gaze focus. (Koutsombogera, Sarthy, Vogel) The team used audio and verbal characterisation of participants' speech and transcripts using self and observer personality reports. Results highlight that verbal content best predicts traits The multimodal nature of the personality calculations suggests that content is less important than acoustics: with the exception of two cases based on acoustic traits alone or in combination, models verbal traits, outperformed models based on verbal traits alone; results also suggest that there is no single optimal choice of model or set of traits to use for predicting cross-personality traits Reported that different models are best suited for different traits. (Vogel, Koutsombogera, Reverdy) The team extracted features such as turn order, time, and word count from a corpus. As well as personality features, the team investigated the correlation between dialogue content features and interaction features related to collaborative assessment. It was found that there was a correlation between whether the dialogue content was unbalanced and the first speaker's assessment of collaboration, and that when the first speaker took up more time it affected the other members' participation in collaboration, that the duration before the second speaker's first speech and the degree of imbalance in the number of words produced by the speaker were negatively correlated with the assessment of collaboration, and that focusing on monotonous increase in duration in successive dialogue segments was positively correlation.

IV. Research Questions

How do non-verbal cues (i.e., gaze, hand gestures, and head pose) correlate with observed levels of collaboration and dominance in group interactions? This question aims to explore the relationship between non-verbal behaviors and the external assessments of collaboration and dominance among participants during the quiz-solving tasks. Non verbal cues like gaze direction, gestures and head movements act as unseen yet pivotal communicators, in the domain of human interactions. These cues mold the dynamics of the collaboration and assertion in a group interaction.Analyzing those cues and gestural communications, using the MULTISIMO corpus, help us to decrypt the complex ballet of gestures that support the group interactions. This study tries to dissect how these non-verbal cues impacts the impression of leadership and teamwork during collaborative tasks, providing a thorough understanding of the implicit rules in the social settings. The key goal of this study is to decipher the complex relationship between non verbal gestures in collaborative conversations, and the level of participation of each participant in the same. By thoroughly evaluating the nuances of body language and other gestures identified in the MULTISIMO dataset, the goal of this study is to find out the existing patterns and correlations that hints at effective cooperation and communication between the participants. Such analysis might shed light on the subtle ways in which people negotiate their roles and responsibilities in a group, improving our understanding of social conversations and collaborations. The outcome of this research is intended to have a deep impact on comprehending the cohesive collaboration, collaborative environment and understanding of non verbal communications. By exploring the enigma of body language, gestures, and non verbal cues in collaborative exchange.

Upon advancing in the evaluation of the subject at hand, there's an urge to unbox another question, which falls into a multilateral spectrum of examination:

- How, in group discourses, do reserved beings (introverts) and sociable beings (extroverts) have disparate measures of contributions?
- Are dominant attributes of a role and greater score in conscientiousness in direct correlation in decision-making processes?
- What is the reliability of past positive endeavours, amidst group situations, in the anticipation of a relatively more collaborative behaviour?

The facets of MULTISIMO corpus, viz:

- Audio-visual recordings
- Non-verbal behaviour interpretation
- Comprehensive data exploratory analysis.

give a stage for an overall analysis of this intricate assessment.

This research question demands the area of analysis dealing with how the different amplitudes of personalities, namely:

- Introversion
- Extroversion
- Conscientiousness

mould the magnitude of candidate participation and the arc of leadership in a group environment. The factors are analysed thoroughly by MULTISIMO corpus, which holds the multi-modal data, facilitating a greater understanding. Observation of the contrast in the behaviour and inheritance of roles inside groups are the goals to focus on, which will then render with the analytic data that'll speak about how experiences and personality affect collaborative consequences.

V. Methodology Flow Chart for Studying Collaboration

A. Data Collection

- We will be utilizing MULTISIMO corpus from multi-modal data which incorporates audio and visual recordings of various group interactions.
- In the MULTISIMO corpus Non verbal behaviour compromises the significant portion of data collected which includes gestures head poses and gaze during a group dialogue.
- The MULTISIMO Corpus aims to capture the characteristics of interactions among individuals with different psychological profiles.
- The dataset pays a particular attention to different personality attributes which influences the group dynamics like introversion and extroversion.

B. Feature Engineering

- We will Extract features that are quantifiable from the multimodal data that reflects personality traits, conscientiousness level and past records of collaborative efforts of an individual.
- Features which corresponds to non-verbal communication like frequency and gesture types gaze direction and movement of head will be extracted from data.
- Features that we can correlate to leadership skills and decision making process in a group will be identified and extracted.
- We will develop a framework with which we will analyze how past behaviour will help to predict collaborative behaviour of people.

C. Data Pre-processing

- Normalization and standardization of raw data will be undertaken to ensure analysis is consistent and accurate.
- We will be segmnenting the raw data to isolate the relevant behaviour cues and interactions to examine behaviour in detail.
- For computational analysis and recognising patterns in data we will be encoding non-verbal behaviour into a structured format .
- For Exploratory data analysis dataset will be prepared such that it is free from redundancies and properly annotated for processing it further.

D. Neural network design

The diagram illustrates a neural network composed of an input layer, hidden layer(s), and an output layer. Here is a description of each level:

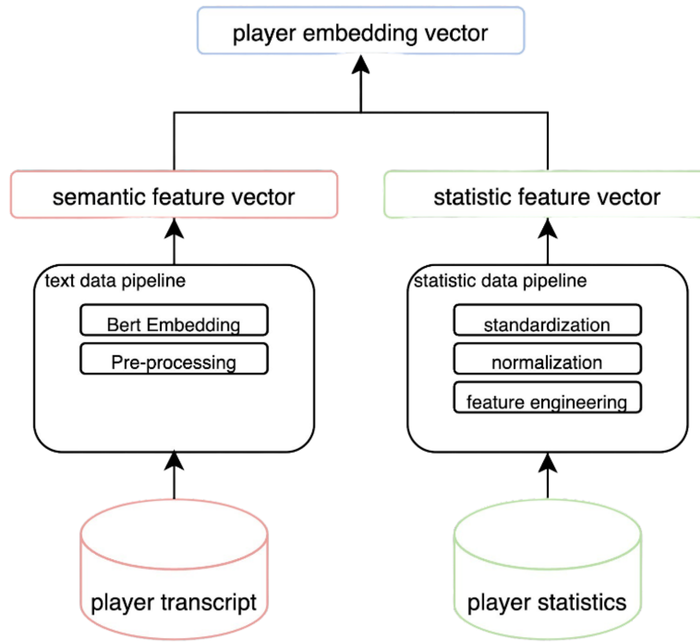


Fig. 1. Data Pre-processing Pipeline

1) Input Layer: It contains multiple input units, representing the features of the neural network. Each input unit (e_p1, e_p2, ..., e_m) corresponds to a feature. The e_p refers to player's feature and the e_m refers to the host of the game.
2) Hidden Layer: In this example, as a draft, only a basic structure of the neural network is presented, which includes the following components:
   - Fully Connected Layer (FC): This is a typical neural network layer where each neuron is connected to all neurons in the previous layer.
   - Activation Function (ReLU): This is a non-linear activation function used to introduce non-linearity into the network, allowing it to solve non-linear problems.
   - Dropout: This is a regularization technique used to prevent overfitting in neural networks. It randomly drops (temporarily removes) connections to certain neurons during training.
3) Output Layer: Softmax is commonly used for multi-class classification problems as it can convert output values into a probability distribution. In this case, we utilize the same five categories ranging from 0 to 4 as used in the dataset's collaboration assessment for prediction.

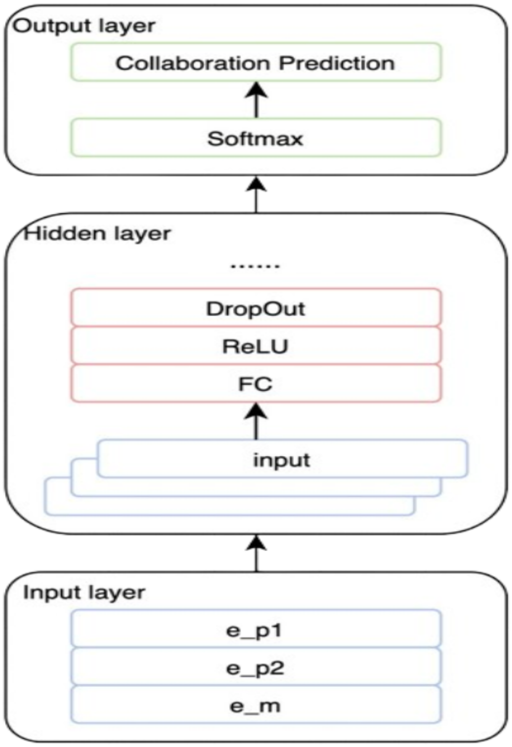


Fig. 2. Neural Network Layers

VI. Evaluation methods

Evaluation Metrics Overview: In evaluating the neural network, we will begin with a concise overview of key metrics. This includes accuracy for a snapshot of overall performance, precision and recall for insights on class-specific outcomes, and the F1 score to balance the two. Cross-Validation with K-Fold: Due to the limited size of the dataset, K-fold cross-validation will be pivotal. This method enhances the reliability of the evaluation by using different partitions of the data as both training and validation sets. This approach maximizes the usage of available data for a more comprehensive performance assessment. Ablation Studies: Finally, ablation studies will be conducted to determine the influence of individual features on the target variable, the collaboration assessment. By systematically removing features and observing the resulting impact on model performance, we can identify the most influential factors contributing to the assessment of collaboration levels. We can even observe changes in model accuracy by deleting or replacing the feature vectors representing host roles, thereby analyzing the impact of the game host on the success level of collaboration.

References

[1] Boersma, P. Weenink, D. (n.d.). Praat: doing phonetics by computer. https://www.fon.hum.uva.nl/praat/

[2] Davin, K., Donato, R. (2013). Student collaboration and teacher-directed class- room dynamic assessment. Foreign Language Annals, 46, 5–22. https://doi.org/ 10.1111/flan.12012

[3] Hotmaulina, Sihotang., Leony, Sanga, Lamsari, Purba., Renatha, Ernawati. (2023). School Commitment with Collaboration from a Cultural and Personality Perspective. Tadbir, doi: 10.29240/jsmp.v7i1.6623

[4] iMotions A/S. (2016). iMotions Biometric Research Platform 6.0. Copenhagen, Denmark

[5] John, O. P., Donahue, E. M., Kentle, R. L. (1991). Big Five Inventory (BFI) [Database record]. APA PsycTests. https://doi.org/10.1037/t07550-000

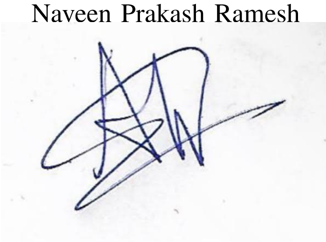
[6] Koutsombogera, M. Vogel, C. (2017). The MULTISIMO multimodal corpus of collaborative interactions. In 9th ACM International Conference on Multimodal Interaction (ICMI2017), pp. 502–503. ACM, ACM.

[7] Koutsombogera, M. Vogel, C. (2018). Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus. In Calzolari, N. C., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., Tokunaga, T. (Eds.), Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pp. 2945–2951 Paris, France. European Language Resources Association (ELRA) Matthias, Stadler., Katharina, Herborn., Maida, Mustafic., Samuel, Greiff. (2019). Computer-Based Collaborative Problem Solving in PISA 2015 and the Role of Personality. Journal of Intelligence, doi: 10.3390/JINTELLIGENCE7030015

Declaration of Contribution

All the member of group have contributed equally to this work .Everyone was actively involved in research, discussion and the meetings which led us to coming up with this paper .

Naveen Prakash Ramesh



Alok Singh

Mingyang Xia

Devang Kankaria

Shenghan Hou

Zitong Liang