Machine Learning Project 2 - Using Machine Learning Algorithms to Analyse Interview Content of Out-of-Body Experiences

Guilherme Costa Ferreira, Mehdi Amor, Rodrigo Anjos
Machine Learning Fall 2024, EPFL, Switzerland

***Abstract*—Machine Learning models can also be leveraged as analytical tools, providing insights into how some set-up works. By applying these algorithms to study the valency (VADER), emotion (BERT) and topic (Top2Vec) content in interviews for Out-of-Body research done by the LNCO Lab at EPFL, we were able to validate the usefulness of our pipeline. We delved into group analysis to find differences between the experiments realised. We discovered that OBE1 and OBE2 interventions had a higher presence of surprise as the dominant emotion and joy appeared less frequently than for the control condition. When applying the intervention to OBE1 and OBE2 participants, meditation and feeling topics are more common, while for Compassion experiments, meditation and feeling are more frequent in control than in intervention. By applying an individual analysis, we verified the presence of interviewer bias and found that it affects the participant's response. Surprisingly, it is also influenced by the participant's interactions, proving a bidirectional trade of sway.**

I. Introduction

Machine learning is a predictive tool which has seen its use grow in the last few years. Its advantages are well known, with models trained to predict all sorts of outcomes. However, artificial intelligence can be more than just predicting results; it can also be a tool to understand and study different contexts. Machine learning is not just a predictive device but also an analysis device. Leveraging this analytic property, one could achieve an "objective analysis". This characteristic is not always present in a human analysis setting.

The mentioned characteristics can be fundamental in studies where researchers gather information in more subjective settings, such as interviews. When we have these conversations, the interviewer and participant will produce varied information in multiple forms.

This information can have an emotional tone. Both participant and interviewer could experience different emotions during these interviews and influence each other. Multiple topics could be present, and identifying them could give us a good insight into a study's particularities. It's valuable that we can extract that valuable knowledge without adding an extra layer of subjectivity and bias.

With all of this in mind, in this project, we propose a machine learning pipeline to extract all sorts of information from transcribed interviews. We hope anyone could use this tool to peek into what is happening under the hood during their discussions.

We didn't train any model as we didn't have much data nor had labelled data. This issue led us to go over pre-trained models. Even though we could only find models trained on internet texts, this solution was the best picked.

For this pipeline, we used three different pre-trained models to analyse our data: Vader Sentiment Analysis (for valency analysis) [1], BERT (for emotional analysis) and Top2Vec (for topic analysis) [2]. We started our research with VADER, the model with the least complex information, just positivity VS negativity. This feature allowed us to understand more glaring trends. This intel enabled us to approach emotional and topic analysis in more detail. The pipeline architecture reflects precisely that. More functions were designed for VADER as we didn't know what to look for. For the rest, the analysis was more targeted.

Because most of the time when working with this type of research, we don't have much data; there was only much information we could extract from group analysis. We also focused on individual analysis, as it captures finner details.

II. Dataset Analysis

This project, conducted with the LNCO lab at EPFL, which specialises in cognitive neuroscience research, explores immersive meditation experiences. Participants engaged in an out-of-body experience (OBE), viewing their body from an external perspective or a self-compassion scenario involving a virtual clone through a VR experience. The study produced quantitative data, such as questionnaire scores and EEG measurements, analysed statistically, and qualitative data from post-experiment interview transcripts, which are the focus of this analysis.

The objective is to explore self-consciousness by examining how individuals react to being disconnected from their bodies through an external perspective or by interaction with a virtual representation.

The dataset has the individual entries for each interview and the general overview. This setting also provides information about which condition took part (control or intervention) and the order of the conditions. We will leverage this in our favour to achieve a complete analysis.

III. Models and Methods

A. Vader Model

VADER (Valence Aware Dictionary for sEntiment Reasoning) [1] is not a machine learning model. However, although we designed this pipeline for the Machine Learning Course, we included VADER in our research because it helped us understand our dataset and guided our approach with the following machine learning models.

Briefly, this model works on a rule basis. It has some rules to classify internet texts' positivity vs negativity components. They used an already existing lexicon. A lexicon is a grouping of texts where each word has its valency already labelled. The conventions used include: (1) the use of contrastive conjunction (like "but") as a words' valency changer; (2) the use of exclamation points as an intensity amplifier; and so on.

This structure allows the model to run fast and produce precise and accurate results. It was evaluated on a Twitter dataset and obtained a 0.96 F1 score.

We ran it so each speech segment in the transcriptions would be assigned a compound sentiment score, ranging from -1 (most negative) to +1 (most positive). For the group analysis, the average sentiment scores were calculated. After identifying key features, we ran it for condition-specific settings, as we wanted to see how it changed from control to intervention. These steps are performed by `vader_participant_analysis` `vader_interviewer_analysis`. Once we knew this, we advanced to subject analysis and checked how valency evolved during the full interview. Not only that, but we also extracted, for both speakers, the most extreme entries to have a better insight into what was shaping these polarities. `vader_subject_analysis` does this analysis.

B. Bert Emotion Model

BERT (Bidirectional Encoder Representations from Transformers) [3] [4] is a pre-trained model that captures word meanings by analysing their full context, simultaneously considering both preceding and following words while using a transformer architecture. It is trained on large text corpora using two tasks: Masked Language Modeling (MLM), where it masks random words in a sentence and predicts using surrounding context, and Next Sentence Prediction (NSP), which determines if two sentences are logically connected. BERT uses WordPiece tokenisation to break words into smaller units, enabling the handling of uncommon words. It includes specialised tokens like [CLS] for sentence-level tasks and [SEP] to mark sentence boundaries.

Like in VADER, the training set was not necessarily composed of interviews, but we still decided to go with it since it achieved a record F1 score of 93.2% on SQuAD v1.1 and an average GLUE benchmark score of 82.1%.

The same analysis structure used for VADER was also applied to BERT, as both share similarities in the type of information they provide, with BERT offering additional complex details. Functions like `bert_participant_analysis`, `bert_interviewer_analysis` and `bert_subject_analysis`.

C. Top2Vec Model

Top2Vec is an unsupervised learning algorithm that identifies latent topics in text documents. In the algorithm's first stage, it creates representations of documents and words as vectors in a high-dimensional semantic space, using pretrained embeddings such as Doc2Vec [5] and Word2Vec [6].

Afterwards, the high dimensional document embeddings are reduced, using Uniform Manifold Approximation and Projection (UMAP) to capture the essential local and global semantic structures. Since proximity is a measure of similarity in the semantic space, Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [7] is then applied to identify dense regions (clusters) of semantically similar documents, identifying each cluster as a topic.

It computes the topic vector of each cluster as its centroid and identifies words closest to the centroid as the most representative of the topic. This step enables the automatic identification of key topic words for each document.

This model distinguishes itself from the rest by its ability to automatically find the number of topics, producing hierarchical relations, and not requiring extensive preprocessing such as stemming, lemmatisation and stop-word removal to be able to run. Its main limitation, however, is assigning only one topic to each document. This architecture excels at understanding text by capturing context and meaning, achieving a BERTScoreF1 of 0.409 on topic modelling tasks.

IV. Group Analysis and Discussion

In this section, let's dive into what we found in the dataset using our pipeline. This analysis focused on a broader inquiry of the data. That is, the analysis is more group-based rather than subject-based. To read about subject-based analysis, go to section V.

Here, we will see the contrasts between OBE1, OBE2 and Compassion while differentiating between participant and interviewer. We also see how control and intervention affect the results and how the order might skew the analysis.

A. Vader Analysis

The VADER analysis evaluates sentiment across three experiments: OBE1, OBE2, and Compassion. In the control condition (C), the average compound scores tended to cluster around slightly positive values, reflecting a baseline emotional tone (0.35, 0.35 and 0.37, respectively). In contrast, the intervention condition (I) showed a noticeable shift, with participants exhibiting more positive sentiment on average

(0.49, 0.49, 0.41). This shift suggests that the intervention elicits stronger emotional reactions, potentially making the experience more impactful or engaging. These distribution plots can be found in appendix figure 5.

However, after careful consideration, this effect seems less predominant. When looking at the order of the conditioning, we could observe a reduction of this difference when comparing deltas. When the intervention was before the control, the average subject delta (difference between intervention and control) was -0.01, while 0.1 for the opposite. These values elucidate the novelty effect: people would be more open to a new intervention after experiencing the baseline. To inspect these deltas please refer to appendix figure 6.

Contrary to the expected, the interviewer was also not as close to neutral as expected. Even though it seems not to be affected by the order (with deltas of 0), in each experiment, the average polarity was far from 0, with Compassion reaching the maximum mean value of 0.43 (appendix figure 7. These results raised discussion around the possible effect of the interviewer's bias.

B. Bert Analysis

When we ran the group emotional analysis, the results weren't as promising as the ones in VADER. It's important to remember that as we moved from one model to the next, we increased the number of features while maintaining the same number of samples. Nonetheless, we could still find different maximum emotions per experiment. OBE1 saw a prevalence of disgust (mean value of 0.19), OBE2 of surprise (0.2) and Compassion of joy (0.16). The significant variability among participants highlights the highly subjective nature of these responses, making it challenging to draw universal conclusions. While overall trends provide some insights, they may overlook individual differences.

C. Topic Analysis

Overall, the most prevalent topics revolve around meditation, what they felt, and vision. In the OBE1 and OBE2 experiments, meditation and feeling topics are much more present during intervention than in control, suggesting interventions elicited more subjective experiences, provoking the expression of one's states. Interestingly, though, for the compassion experiment, it is the other way around. Meditation is at a much higher presence in control than in intervention, while feelings don't vary much. The eye and image topics are also much higher in control, revealing that participants might have become more focused on visual aspects in this setting than in the intervention.

An analysis of which emotions each topic could elicit was also conducted (figure 1). We can see that the forest topic, likely relating to nature, evokes joy. Another interesting finding is that the body topic seems more related to surprise and disgust than other emotions, excluding neutral.

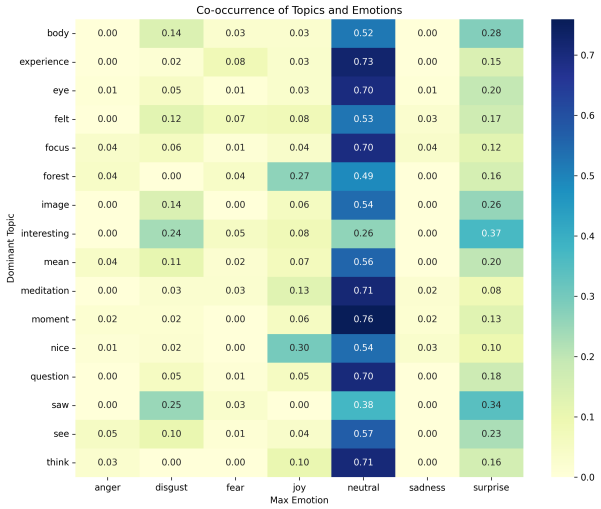

Figure 1. Co-ocurrence of topics and emotions in the whole dataset

V. Individual Analysis and Discussion

As we observed during group analysis, many aspects of qualitative data are lost if the participants don't share those features consistently. This loss of information can be due to the high subject variance or the lack of data. Nonetheless, we must understand what each subject's data looks like to be able to observe more subtle trends. The subject analysis step is crucial to check if our models produce coherent results.

To do so, we decided to check specific subjects' emotional and topic contents.

A. Bert Analysis

When we looked into each participant's emotional analysis, relevant observations could be made. Let's take the subject's 24 disgust evolution over time (figure 2).

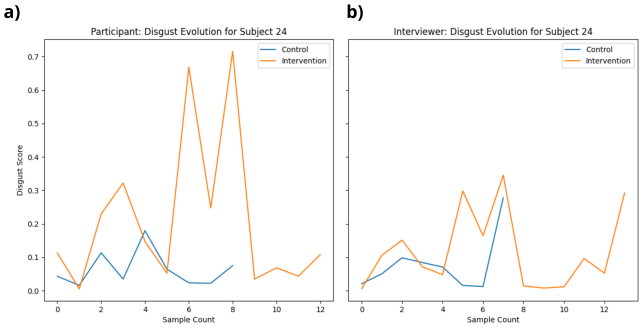

Figure 2. Disgust Time Analysis for Subject 24; Experiment: OBE1; Condition Order: Control-Intervention **(a)** Participant Analysis **(b)** Interviewer Analysis

In the left panel (figure 2a)), we can observe how disgust evolved during both control and intervention for subject 24.

Overall, emotion scores tend to follow this trend: scores close to zero most of the time, with some abrupt spikes here and there. This example is the case. However, the most interesting thing is how the participant and interviewer influence each other. In this case, as the interviewer showed disgust, the participant followed. The opposite is also the case in other examples. The interviewer also tended to follow the participant's emotional lead.

This emotional bi-directional connection seems to corroborate the empathetic nature of the human self. Where one went in the emotional spectrum, the other followed. The interviewer-participant direction is highly concerning since that could mean the research could, without noticing, bias the participant into feeling one way or another. The other way seems to be more positive. By reinforcing the participant's emotions, one could feel more confident and produce better content for the study.

Another trend we could get from this analysis was the comparison between control and intervention. The most present emotion per experiment (like disgust in OBE1) showed a considerable increase between control and intervention, as it's possible to verify in figure 3a).

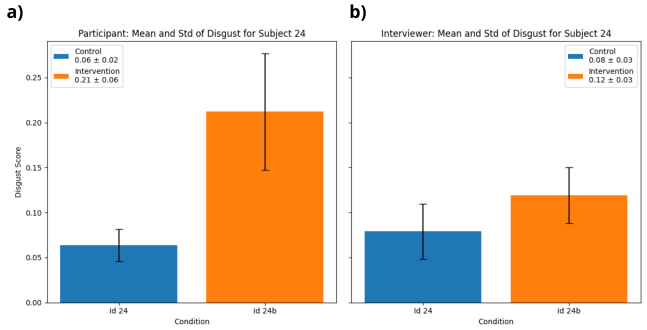

Figure 3. Disgust Mean Analysis for Subject 24 in both Conditions; Experiment: OBE1; Condition Order: Control-Intervention **(a)** Participant Analysis **(b)** Interviewer Analysis

B. Topic Analysis

Through analysis of an individual subject, in this case, S302 (figure 4), we can see that the interviewer seems to be able to introduce topics, sometimes repetitively asking about the same thing, likely to direct the conversation toward topics they think are the most valuable. In S302man, continuous mention of the "saw" topic leads the participant to change his discourse to "see". We can confirm that the interviewer guides the participant, who then expands on each theme, maintaining extended mentions, whereas the interviewer's topics are least consistent.

We also observed that the control condition showed a narrower range of topics, whereas the intervention shows Participants reflecting more on their experience.

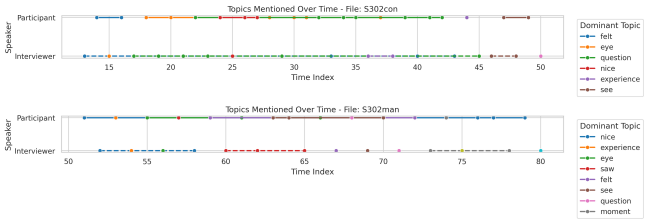

Figure 4. Topic Analysis of interviews with subject 302. S302man is Intervention, S302con is Control

VI. Future Work

As we developed this pipeline, ideas and solutions were discussed that didn't make it to this project. Some of them represent what we believe are the necessary future features for this tool.

One logical step would be to further delve into topic analysis and find concrete metrics to correlate with emotions, both at the subject and at the experiment level.

We also found the idea of using this pipeline to improve how research is made promising. In a scenario where the interviewer is conditioning the participant, as in the one we presented, it might be key to have proper solutions to avoid this error. By inserting the proposed intervention in the pipeline, one could receive back a more neutral and less influenceable approach, effectively improving the condition of the interview. We can gather this feedback by generating multiple similar interventions with a generative LLM, like ChatGPT. Then, using either VADER or BERT, the system would pick the most neutral course of action.

VII. Conclusion

With our project, we demonstrated that latent information is present in the interview text content. Even though we could not use the tone and mannerisms, as well as the facial expressions of both speakers, we could still identify clear and plausible trends, like differences between intervention and control in all experiments, but also the influence of the order on the subject's perception of the effectiveness of the approach. We could understand how topics changed over the interview and how they correlated with the identified emotions. We could also perceive how both speakers could influence each other in emotional content and spoken themes.

Is also relevant to notice that, even though we could extract reasonable trends, the variances were high. This effect is due to the high variability of a person's speech. Adding to the lack of data and the models' limitations, the meaningful information that can be extracted, especially on a group level, is much reduced. This challenge requires more robust models, maybe better tailored to the context, or more data points.

VIII. Ethical Challenges

Analysing this project's features, we faced two ethical issues, the first in the welfare and autonomy domain and the other in the privacy setting.

While the primary goal of our tool in this study is to understand the emotional and cognitive effects of out-of-body experiences, these same techniques could also be misused to compromise the welfare of the main stakeholders—the interviewees. Studying what emotional responses elicit particular topics could enable the development of strategies for manipulating emotions in various contexts, such as advertising, which can undermine individuals' autonomy, exploiting emotional vulnerabilities. Regarding this matter, we analysed the influence of topics on emotions. Still, a similar tool could even examine the impact of how individuals convey ideas or the subject's broader context, all to gain information on how that affects their emotional state.

Another ethical concern concerns the anonymisation of interview data. Today's current transformer models can infer demographic attributes such as gender, age, or cultural background based on linguistic patterns, possibly leading to the identification of participants. The erosion of anonymity violates participants' expectations of privacy and can even introduce biases in the research's results.

A possible solution to mitigate both risks would be to restrict this pipeline's use exclusively to approved research. This way, we could ensure a tighter ethical usage of this tool.

IX. Acknowledgements

This project was made in collaboration with LNCO Lab in the ML4Science project with direct input from:

- Dr. Bruno Herbelin - bruno.herbelin@epfl.ch
- David Friou - david.friou@epfl.ch

We want to thank both for the project and the help provided.

References

[1] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, May 2014. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[2] D. Angelov, "Top2vec: Distributed representations of topics," *arXiv preprint arXiv:2008.09470*, 2020.

[3] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-based sentiment analysis using BERT," in *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, M. Hartmann and B. Plank, Eds. Turku, Finland: Linköping University Electronic Press, Sep.–Oct. 2019, pp. 187–196. [Online]. Available: https://aclanthology.org/W19-6120

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 10 2018.

[5] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*. PMLR, 2014, pp. 1188–1196.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[7] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013, pp. 160–172.

Appendix

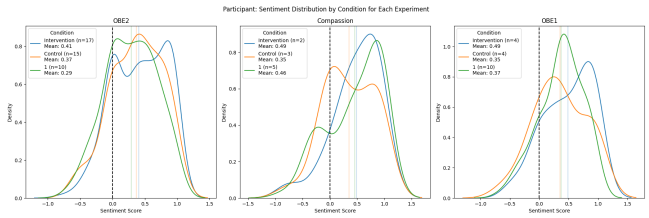

Figure 5. Participant sentiment distribution (valency) for all experiments and conditions

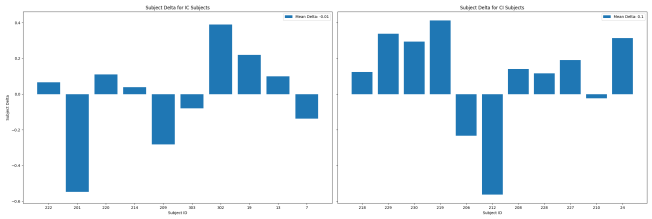

Figure 6. Comparing the difference between Intervention and Control on the different participants.

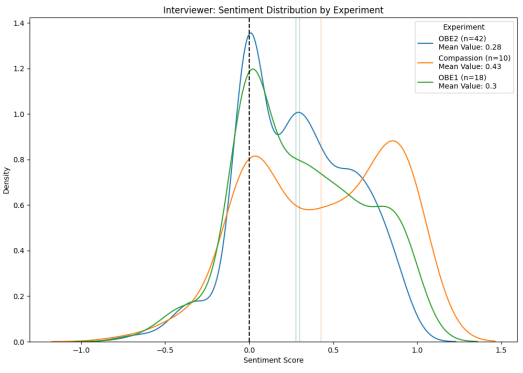

Figure 7. Overview of how the valency varies over each experiment for the interviewer