Multimodal Dialogue
Emotion expression through spoken language in Huntington disease

Ariel R. Ramos Vela
Institut Polytechnique de Paris
ariel.ramosvela@ip-paris.fr

Abstract

The development of machine learning models to assess data in the Healthcare field offers the opportunity to discover hidden patterns, often not easily recognised by clinical observation [3]. Indeed, they can be valuable in the study of neurodegenerative disorders such as Huntington's disease to better understand its impact on a person's functional abilities. In this report, a summary of the paper "Emotion expression through spoken language in Huntington disease" [1] is provided along with a critical analysis of it. Finally we extend it further by discussing future ideas that could enhance the performance of the experiment done in the paper. The report can also be found at https://github.com/arielramos97/Multimodal_Dialogue.

1 Summary

The paper is divided into 5 sections: Introduction, Methodology, Results, Discussion and Conclusions.

The Introduction section presents an overview of Huntington's disease (HD), a neurodegenerative disorder that affects the nerve cells in the brain. Motor disorders produced by this disease have been largely studied as opposed to communication disruptions. Indeed, the paper states that no previous study has assessed the ability of HD patients to express emotions through spoken language. Therefore, their objective is to investigate how the expression of emotions through spoken language (speech and linguistic content) is impaired in HD patients.

For this study, authors analysed speech and linguistic content (also referred as language) independently. Speech features include fundamental frequency, energy or speech rate whereas linguistic content refers to the choice of words and syntactic structure. Thus, they developed separate machine learning emotion recognition models (instead of a human jury) to assess emotion expression. For this experiment, neutral and emotional speech recordings were collected from healthy controls, preHD individuals, and HD individuals. To obtain these recordings, individuals were asked to narrate stories associated with fear, anger and joy. Lastly, the interviews were segmented and transcribed to then be processed and fed to the machine learning models.

Moving on to section 2, the paper describes the methodology followed to perform the experiment. Six sets of data were collected; 3 groups: healthy controls, preHD, HD and for each one of them 2 modalities: speech and language. Subsequently, each of these datasets were used to train and test one emotion classifier. In addition, the paper outlines that 115 participants (all French native speakers) were included in the experiment; 25 controls, 22 preHD and 68 HD. As previously mentioned, participants were asked to narrate non personal stories that made them feel sad, angry and happy respectively whereas to record a neutral emotion they were asked to describe their routine in the latest 24h.

In terms of data preprocessing, speech samples were transcribed, split in stretches using the software Praat, and labelled according to the task where they were collected (neutral ,sad, angry or happy).

Afterwards, relevant features from both audio stretches and linguist content were extracted. For the audio stretches, the Extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) was employed to obtain affective information and the openSMILE toolkit to retrieve features related to energy, pitch and articulation. On the other hand, to extract features from the linguistic content, the LASER (Language-Agnostic SEntence Representation) sentence embedding model was used. It was selected due to the scarcity of training data and because of its performance in multilingual classification tasks.

In addition, this section highlights that Random Forest classifiers were employed to perform the emotion recognition task at hand and that 10-fold cross validation was used to evaluate the performance of the models. Then, a Kruskal-Wallis test on the three series of accuracies (controls, preHD and HD) was used and after a significant difference among the groups was reached (within each modality), post-hoc Wilcoxon Mann Whitney tests were conducted to compare the accuracies group-wise. Lastly, with the aim to check the consistency of the models, human scorers were asked to classify both audio and language transcripts, which then were used to compute the agreement measure between human groups and the models.

In section 3, the paper demonstrates that the model's accuracies were significantly lower for HD patients. This suggests that the ability to express emotions through voice and language is impaired in HD individuals. Interestingly, the accuracies for preHD and control groups were equivalent meaning that there is no sign of impairment in early stages of the disease. Regarding the comparison of the method with human juries, results indicate that the models are better discriminant and thus, perform a better classification for both voice and text modalities.

Proceeding to section 4, authors emphasise that the experiment enriches the theories of emotional processing in HD by adding evidence of the effects of the disease in the production of spoken language. The paper claims that, by the date of the paper, it is the first to employ machine learning models to identify emotions (using spoken language) of individuals with neurological disorders. Besides, the method offers the advantages of not being sensitive to speech motor impairments given that the data of each group (control, preHD and HD) is treated separately, and not being biassed by the differences of individuals at the moment of speaking which are unrelated to the expression of emotions.

Similarly, the authors highlight that the obtained results on emotion expression using the voice modality support the hypothesis made by previous studies which suggests that the sensorimotor representation of emotions is affected by the deterioration of motor processing components, which in turn, causes the impairments in the expression and recognition of emotion. Nonetheless, the results using the language modality deviate from the joint perception-production impairment by the theory of embodied cognition. The authors point out that this might happen if the conceptual channel remains functional without simulating emotions. An alternative explanation is that the incongruence between the perception and production of emotional language does not stem from emotional processing.

Finally in section 5, the paper concludes that the use of machine learning was useful to analyse voice and language features individually. Additionally, it describes some of the limitations such as the amount of data collected from each participant and it opens up opportunities for future research mentioning that models could be extended by including the analysis of bodies or faces.

Main scientific questions

- Is linguistic expression of emotions impaired in Huntington disease (HD)?
- Is vocal expression of emotion impaired in Huntington disease (HD)?

2 Critical analysis

The paper introduces a novel method that employs machine learning algorithms to investigate whether the linguistic and vocal expression of emotions is impaired in HD patients. In this section, the pros and cons of the paper will be discussed along with its impact in the medical community.

Strengths

One noteworthy aspect of the experiment is that it was designed to avoid ending the interviews with a negative emotion for individuals. Indeed, what is prioritised is that patients feel comfortable and with the freedom to interrupt the recording at any time. In addition, the neuropsychologists conducting the interviews were properly advised not to make individuals recall personal stories which indicates that the well-being of individuals was key to obtaining quality data.

Regarding the technical pros of the paper, the fact of creating a classifier for each group of patients (control, preHD and HD) allowed the extraction of useful information for each group. If a single classifier for the 3 groups were implemented, the performance would decrease given that the model would most likely be influenced by the information of the majority class (68 HD patients). In addition, to deal with unbalanced datasets, some stretches were randomly removed to ensure that each dataset (6 in total) has the same number of stretches and the same number of emotions for both modalities (voice and text).

Another important point to mention is that authors also utilise the kappa metric to measure the prediction agreement of the decision of the 2 human groups and also the agreement of the average of these 2 groups with the machine learning model. This was important to validate the methodology with human judgement. Likewise, authors indicate that 10-fold cross validation was utilised to obtain stable results. This in turn prevents overfitting and helps the model to better generalise so it can properly adapt to unseen data.

Lastly, the use of Random Forests offers a better explainability as opposed to other approaches such as Neural Networks. This means that predictions can be better understood and explained. This is important in the medical sector to inform why a particular decision was taken and what feature information corroborated it.

Weaknesses

The paper establishes that for each dataset (6 in total; 3 groups and 2 modalities) an independent emotion recognition classifier is created. Nonetheless, it does mention the hyperparameters used in each Random Forest model such as the number of trees (n_estimators), the number of features to consider when splitting the tree (max_features) or the maximum depth of the trees (max_depth). Besides, given that the classifiers are independent, these hyperparameters will depend on the dataset being learned. Thus, they should not remain fixed if an optimal performance is sought.

Another drawback is the scarcity of data. This does not let the models achieve a higher performance and thus, more reliable results. This can be observed in Fig 2B in [1] where accuracies are relatively low for both voice and language modalities. Furthermore, the lack of data implies an impediment to explore other approaches such as Deep Learning which requires larger amounts of data to obtain a good performance.

Finally, even though the code of the experiment is open-source, it was not possible to further investigate alternative approaches given that the data is not publicly accessible. This is because of the ethical procedures that need to be followed when working with sensitive data. An attempt to contact the corresponding author by email was done but unfortunately, no response was received.

Impact and subsequent studies

The paper raises awareness of the importance of the identifiability of emotions in HD patients. Indeed, this is crucial to improve the interactions that HD patients have with their family and also with their caregivers. Besides, the paper demonstrates evidence that spoken language is also impaired by this disease which contradicts previous views. Thus, the paper opens the room for further study to better understand how communication, and particularly, the ability to express emotions, is disrupted by this disease.

Currently, the disease's follow up involves both financial and human cost. Therefore, subsequent studies and experiments (involving a larger number of participants) should be conducted to improve the reliability and performance of the machine learning models so they can be deployed in smart devices. This with the aim to monitor HD patients on a daily basis and thus, act accordingly to their needs.

3 Improvements

In order to improve the performance of the Random Forests models, a Grid Search approach [5] could be followed to find the best set of hyperparameters for each model. Similarly, to obtain more reliable results Data Augmentation techniques can be employed to generate additional synthetic data while avoiding overfitting [4]. However, this is more challenging due to the semantic and syntactic structure of the language data used in the experiment.

If the scarcity of data were not a problem, a pretrained transformer model such as RoBERTa [2] (Robustly Optimised BERT Pretraining Approach) could be used to process the linguistic content of the interviews. It has managed to achieve state-of-the-art scores in a variety of NLP tasks and can be easily modified to perform the classification task at hand. Besides, given that the data is not in English, the HuggingFace library [6] also provides an alternate version denominated 'French-Roberta' that is trained on a small French News corpus.

4 Conclusions

In this report, a summary and a critical analysis of the paper [1] was presented. Besides, future ideas were discussed which could enhance the results obtained by the experiment. Overall, the impact of the paper in the coming years is promising in terms of the awareness it raises and the benefits it can bring to HD patients and their entourage (family and caregivers). Indeed, the better interaction with patients, the better they can be assisted. That is why further research should be conducted to better understand how spoken language is impaired by the Huntington disease.

References

[1] Charlotte Gallezot, Rachid Riad, Hadrien Titeux, Laurie Lemoine, Justine Montillot, Agnes Sliwinski, Jennifer Hamet Bagnou, Xuan Nga Cao, Katia Youssov, Emmanuel Dupoux, et al. Emotion expression through spoken language in huntington disease. *Cortex*, 155:150–161, 2022.

[2] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[3] Amrita Mohan, Zhaonan Sun, Soumya Ghosh, Ying Li, Swati Sathe, Jianying Hu, and Cristina Sampaio. A machine-learning derived huntington's disease progression model: insights for clinical trial design. *Movement Disorders*, 37(3):553–562, 2022.

[4] Connor Shorten, Taghi M Khoshgoftaar, and Borko Furht. Text data augmentation for deep learning. *Journal of big Data*, 8:1–34, 2021.

[5] B Sumathi et al. Grid search tuning of hyperparameters in random forest classifier for customer feedback sentiment prediction. *International Journal of Advanced Computer Science and Applications*, 11(9), 2020.

[6] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October 2020. Association for Computational Linguistics.