Representations of Personality Features using Modified Autoencoders

Brandon Cheng
Queen's University
brandonkhcheng123@gmail.com

Chloe Atherton
Queen's University

Darwin Chen
Queen's University

Molly Shillabeer
Queen's University

***Abstract*—Complex systems can often be represented in lower dimensions, and doing so can often grant important insights to the system by making it easier to analyze. However, many data compression and reduction techniques face a trade-off between generalizability and accuracy. In this study, we develop a model that aims to tackle generalizability without sacrificing accuracy, and test this model by training it to learn a representation of personality features using self report survey data. Overall, results were inconclusive, but the model shows promise and testing suggests that the theory behind the model is sound. As such, this project will be worth furthering through testing and refinement.**

I. Introduction

Complex systems are unavoidable in the realm of data science. Be it the stock market, the weather, or the physical condition of a human body, raw data taken from these systems can often be large, messy, and difficult to comprehend. However, some of these systems, like the human body, can be represented using a much simpler set of factors, such as the state of critical organ systems within the body.

A. Motivation

One such system that particularly interests us is human behaviour. Various efforts have been made to simplify and categorize human behavior into intuitive groups, but with varying degrees of success. One well-known example is the MBTI test [1], which has been criticized for its lack of scientific backing and questionable reliability. Despite this, some attempts to factor personality have proven to be more successful, such as the Big 5 personality test that aims to categorize personality into five distinct traits and has shown promising levels of validity. However, a common criticism for tests of this nature is their lack of reliability, especially given that the criteria and factors are often based on the creator's subjective intuition.

The objective of this research is to develop a customized model(outlined in the Methodology section) and assess its effectiveness in generating meaningful data representations. The performance of the model will be assessed based on its ability to learn a generalizable personality feature representation using an existing self-report survey dataset. Additionally, this study endeavors to create a personality factor representation that is entirely isolated from human bias.

B. Related Works

Autoencoders [2] are used to learn alternative or compressed representations of data. It does this using two neural networks, one that converts the data to a latent representation, and another that reconstructs the input data from the latent representation.

A common use for autoencoders is in data denoising. By training an autoencoder to represent the data in a limited latent space, noise that does not conform to the learned representation is discarded from the input data.

Another common use for autoencoders is in decomposing and representing complex physical systems in lower dimensions [3]. This utilizes an autoencoder to extract modes of nonlinear systems like fluid flow, allowing a linear transformation to be applied to the latent representation to reflect a similar transformation in the original system.

An example of this method uses an autoencoder to learn the sinusoidal properties of a turbulent flow system from a video recording of the system. A transformer can then be applied to the latent representation and reconstructed into a video frame representing the system after a fixed time interval.

C. Problem Definition

Autoencoders are generally bounded by a bias-variance trade-off. If an autoencoder's capacity to fit to the data is too high, the learned representation may not be meaningful and can instead overfit to the peculiarities of the dataset. This can make it difficult to analyze or transform the representation effectively. On the other hand, if the capacity of the autoencoder is too limited, the representation may not be accurately reconstructed to the original data.

In this study, a customized autoencoder-like model will be constructed in an attempt to overcome the flaws and trade-offs of traditional autoencoders. The aim of this model is to create generalized, meaningful representations without sacrificing the ability to accurately reconstruct data.

II. Methodology

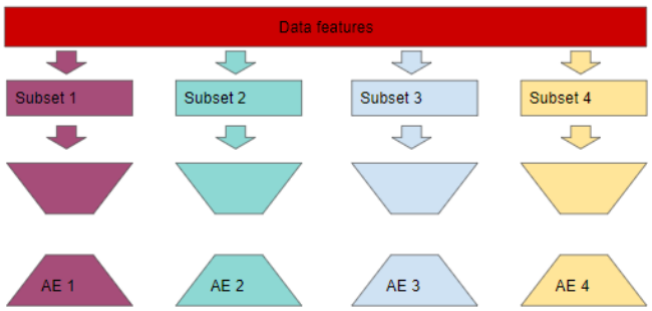

Fig. 1. Visual representation of how data is split and assigned to different encoders and decoders.

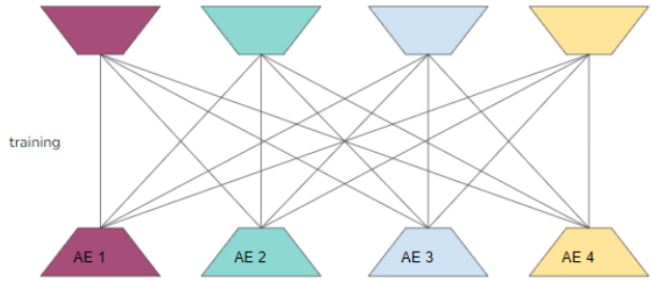

Fig. 2. Visual representation of encoders and decoders being cross trained.

1) Data:
The main dataset we will be using is the 16PF survey result dataset [4].
The data will first be cleaned by discarding unwanted features like country, name, gender, etc. entries that contain missing data or only contain one response value will be discarded.
Next, the data will be split into 3-5 subsets using randomized stratified sampling to ensure each subset contains a somewhat balanced set of questions based on the subgroups the survey questions were labeled with.
2) Proposed solution:
The custom model will consist of 3-5 encoder decoder pairs, each matched with a data subset, and all having the same latent space size. (see Figure 1-2)
When training, the model will iterate through all possible encoder decoder pairs for each train step, with the loss evaluating the mse between the decoder output and the data it is supposed to reconstruct. This is done to force all the encoders and decoders to learn the same latent space, as each encoder must represent the data in a way that can be decoded by all the decoders, and each decoder must learn to decode the latent representation of all the encoders. This also ensures that the encoders represent the data in a way that is generalizable and isn't easily affected by peculiarities that are specific to any one subset of data.
3) Evaluation:
The model will be evaluated on two metrics, which are applied to all encoder decoder combinations except the combination containing the encoder and decoder from the same set. This is done to focus the metrics on measuring how well the model generalizes the latent representation to non-identical data. The first metric is accuracy. This is measured by rounding and clipping the model's output to a valid survey response. It is then compared to an actual user's response to determine the accuracy of the model. The second metric is the model's binary loss, where the model will only be evaluated on how accurately it predicts the net sentiment(1-2 vs 4-5) of the survey response whenever the response isn't null(3).
4) Details/Analysis:
The optimal size of the latent space was determined by testing each latent space within a reasonable range(3-19) and recording the binary accuracy of the model trained with that latent space. It was determined that a latent space of 8 would be optimal as further increasing the latent space would result in a negligible gain in accuracy. (see Figure 3)
Using a fixed latent space, different model shapes and layer depths were tested. With a latent space of 8, it was found that the optimal shape was to have no hidden layers. However, hidden layers were used when using a latent space of 3.
The training progress with a latent space of 3 was also plotted to analyze how well each encoder's latent space was converging. The results of this analysis confirmed the theory that the different encoders would learn the same latent representation through the customized training function. (see Figure 4)

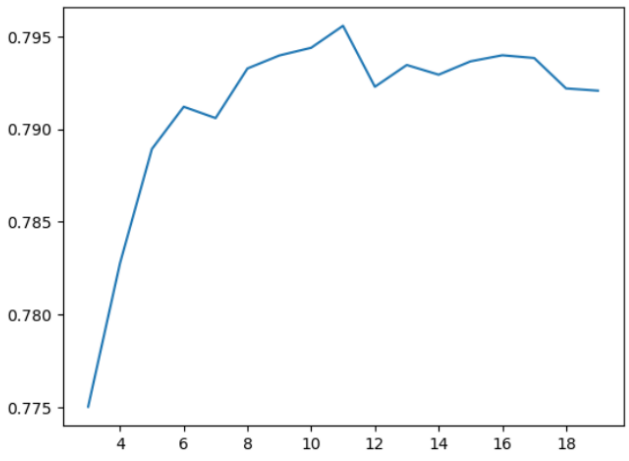

Fig. 3. binary accuracy of models trained with different latent space sizes.

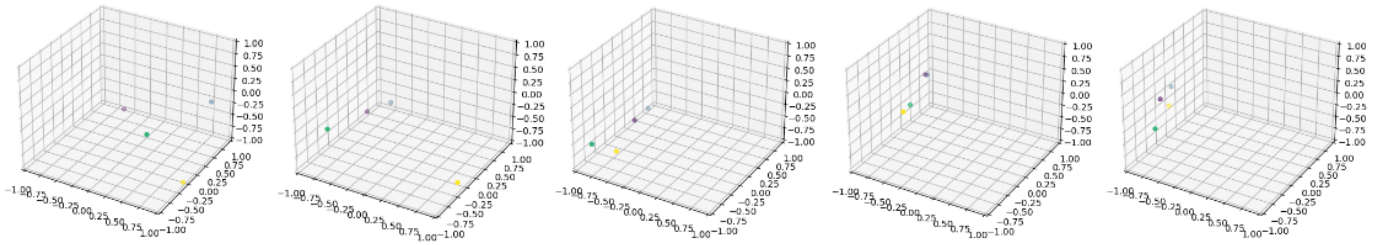

Fig. 4. A representations of four subsets in a data point converging as the model trains.

III. Results

TABLE I
MODEL RESULTS.

Latent space	Hidden layers	Accuracy	Binary Accuracy
8	0	40%	79%

As shown in table 1, the model appears to have performed poorly in terms of the accuracy metric, but decently well in terms of the binary accuracy. This shows us that the model could not predict the exact score a subject would respond to a survey question with, but could predict, generally, whether the subject agreed or disagreed with the survey question. While the model did not perform as well as expected, its binary accuracy shows promise, as it implies that the latent representation generated could represent the presence of personality features decently well, if not the exact degree to which they are present.

IV. Conclusion

Overall, this study has shown that the customized model is a viable method for creating alternate representations of data. It also shows that the theory behind the model is sound, as Figure 4 demonstrates that the latent spaces of the encoders and decoders converge to learn a shared representation of the data.

However, it has yet to be definitively shown that the customized model can create representations of data that surpass traditional autoencoders in terms of generalizability and meaningfulness. It has also yet to be shown how well the personality representation generated in this study compares to traditional personality tests like the Big 5.

In conclusion, this study has served as a good starting point for the development of the customized model. We have shown its viability, demonstrated the soundness of the theory behind it, and utilized it to create a representation of personality from a self report survey dataset to some degree of success. More testing and refinement will have to be done in order to polish and prove the customized model, and additional data processing and comparison will be required to fully develop an adequate personality representation.

References

[1] Stein, R, Swan, AB. Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology. Soc Personal Psychol Compass. 2019; 13:e12434. https://doi.org/10.1111/spc3.12434

[2] Bank, D., Koenigstein, N., & Giryes, R. (2020). Autoencoders. doi:10.48550/ARXIV.2003.05991

[3] Lusch, B., Kutz, J.N. & Brunton, S.L. Deep learning for universal linear embeddings of nonlinear dynamics. Nat Commun 9, 4950 (2018). https://doi.org/10.1038/s41467-018-07210-0

[4] Open psychology data: Raw Data from online personality tests. (n.d.). Retrieved March 13, 2023, from http://openpsychometrics.org/_rawdata/