Exploring Psychological Data by Integrating Explanatory and Predictive Approaches through Artificial Neural Networks: A Brief Overview of Current Applications

Monica Casella [1], Pasquale Dolce [2], Raffaella Esposito [1], Maria Luongo [1], Davide Marocco [1], Nicola Milano [1], Michela Ponticorvo [1] and Roberta Simeoli [3]

[1] *NAC "Orazio Miglino" Lab, Department of Humanistic Studies, University of Naples Federico II, Naples, Italy*
[2] *Department of Public Health, University of Naples Federico II, Naples, Italy*
[3] *Neapolisanit S.R.L. Rehabilitation Center, Ottaviano, Italy*

Abstract

Traditional psychometric data analysis techniques belong to an explanatory approach and often require several assumptions on data. In contrast, Machine Learning (ML) techniques belong to a predictive approach and necessitate mild assumptions on input data. These predictive techniques aim to identify data patterns and generate accurate predictions of output values based on input values of new observations. In recent years, several works proposed the integration of explanatory and predictive approaches. This paper provides an overview of the works carried out at the Natural and Artificial Cognition "Orazio Miglino" Lab, discussing various applications of Artificial Neural Networks in psychology. The discussed studies highlight the promising outcomes of integrating machine learning techniques into traditional psychometric data analysis. Specifically, due to their flexible assumptions on input data, their ability to handle different types of input data, and their ability to model complex and nonlinear relationships between variables, the integration of ML techniques could complement and enhance psychological data analysis.

Keywords

Artificial Neural Networks, Autoencoders, Psychometrics, Dimensionality Reduction

1. Introduction

Psychometric data analysis has traditionally relied on explanatory modelling techniques, which aim to test hypothesized relationships between variables and study intercorrelations using methods such as correlation or latent variable analysis [1]. However, these techniques assume that the relationships among variables are linear and can be limited when this assumption is violated. Several studies have indicated that ignoring nonlinear relationships among variables can lead to biased results and misleading interpretations [2, 3].

In contrast, Machine Learning (ML) techniques belong to a predictive approach, where researchers aim to identify patterns in data and generate accurate predictions of output values based on input values of new observations [4]. Unlike traditional methods, ML techniques require milder assumptions on input data and can handle nonlinear relationships among variables.

Although explanatory modelling remains the dominant approach, several studies have

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29-31, 2023, Pisa, Italy
EMAIL: monica.casella@unina.it (A. 1); pasquale.dolce@unina.it (A. 2); raffaella.esposito22@studenti.unina.it (A. 3); maria.luongo@unina.it (A. 4); davide.marocco@unina.it (A. 5); nicola.milano@unina.it (A. 6); michela.ponticorvo@unina.it (A. 7); roberta.simeoli@unina.it (A. 8)
ORCID: 0000-0002-6017-602X (A. 1); 0000-0002-7588-6067 (A. 2); 0009-0002-6582-0002 (A. 3); 0009-0001-8780-6327 (A. 4); 0000-0001-5185-1313 (A. 5); 0000-0002-1604-5161 (A. 6); 0000-0003-2451-9539 (A. 7); 0000-0001-7564-6816 (A. 8)

proposed an integration of Machine Learning techniques, such as Artificial Neural Networks, to complement traditional psychometric methods and improve research efficiency, facilitate model performance evaluation, and increase interpretability [5, 6]. In particular, Artificial Neural Networks have been applied in various areas of psychology, including the validation and development of psychometric tests, the development of short-form of existent psychometric tests, and the identification of disorders such as autism. The following section provides an overview of these applications.

2. Applications

2.1. Artificial Neural Networks for Validating Psychometric Tests

The study of Dolce et al. [7] presents a novel procedure that integrates explanatory and predictive modelling to develop new psychometric questionnaires based on psychological and neuroscientific theories. The procedure involves a series of steps that select the most predictive items while preserving the factorial structure of the scale. This approach combines both the explanatory power of the theory and the predictive power of modern computational techniques, such as exploratory data analysis and artificial neural networks (ANNs).

By integrating these techniques, the study aims to derive theoretical insights on the characteristics of the items selected and their conformity with the theoretical framework of reference. Additionally, the procedure selects those items that are most relevant in terms of prediction by considering their relationship with the actual psychopathological diagnosis. This approach helps to construct a diagnostic tool that conforms both to the theory and to the individual characteristics of the population under study.

The proposed procedure involves constructing an ANN capable of predicting the diagnosis of a group of subjects based on their responses to a questionnaire. The most predictive items are then automatically selected while preserving the factorial structure of the scale. Results of the study indicate that the machine learning procedure selected a set of items that drastically improved the prediction accuracy of the model. Specifically, the selected set of 167 items resulted in a prediction accuracy of 88.5%, compared to the original set of 260 items, which resulted in a prediction accuracy of 74.4%. Moreover, the procedure reduced the redundancy of the items and eliminated those with less consistency. Overall, the procedure presents a promising approach for developing psychometric questionnaires that both conform to theory and accurately predict psychopathological diagnoses.

2.2. Artificial Neural Networks for Dimensionality Reduction and Short-Form Development of Psychometric Tests

Determining the number of dimensions in a dataset requires researchers to make several important decisions: in particular, the choice of dimensionality extraction method and the decision of the number of dimensions to retain are considered among the most critical in the development of psychometric tests [8].

One of the most widely used statistical techniques for dimensionality reduction in test construction is principal component analysis (PCA) [9]. However, the assumptions underlying PCA, in particular the linearity of the relationships between variables, are not always verifiable in the psychological domain. Therefore, the principal component analysis may not always be the most appropriate analysis method. An interesting alternative to PCA is the autoencoder.

Autoencoders were first introduced by LeCun in 1987 [10] and developed by Baldi and Hornik in 1989 [11].

They are a particular type of artificial neural network with the same number of outputs and inputs and a smaller number of hidden units, trained to learn the identity function, such that the output is as similar as possible to the input. Perfect reconstruction of the input vectors is not possible, since the central layer is smaller than the others. Because of this, it is in this layer that the most relevant information for input reconstruction is encoded.

In its simplest form, the weights in the network are trained to minimise the squared reconstruction error and because of this learning strategy, it can be demonstrated that a linear autoencoder, with *n* features, converges to the *n-th* dimensional PCA subspace [12]. However, there are some important differences among these methods: PCA is a linear transformation and assumes the normality of observed data; autoencoders can deal with non-linear relations [13] and make no assumptions about data distribution. On the other hand, the principal components are orthogonal and sorted in order of decreasing variability. So, the autoencoder is a more flexible and powerful method because it can learn more complex relations present in data but, unlike PCA, is less interpretable.

A study by Pham et al. [14], proposes a PCA-autoencoder for image recognition. This PCA-autoencoder has independent latent space and ordered internal nodes, so it makes explicit PCA's main characteristics. When used with linear activation only and one hidden layer, the PCA-autoencoder proposed in this study is totally equivalent to PCA. [15]

In a recent study by Casella et al. [16], autoencoders have been applied to the development of a short-form of psychological tests. In fact, despite the potential advantages of using shorter tests, their development shows several limitations: firstly, it is a relatively laborious process, and secondly, researchers usually consider only a small part of the possible short-forms of a test. In this context, machine learning techniques [17, 18] and, in this case, neural networks, can help automate and optimize this process.

In particular, the results of this work show that an autoencoder trained on an existing long-form of psychometric test could be useful to select a short-form that best reconstructs the original test responses, among the many possible alternative short-forms. Furthermore, maintaining the number of neurons in the internal layer equal to the number of dimensions of the original test help in choosing short-forms that preserve the original test's dimensionality.

2.3. Artificial Neural Networks to Identify Children with Autism Spectrum Disorder

Another recent study by Simeoli et al. [19] suggests that motor abnormalities can provide a computational marker for autism, which could lead to more objective and efficient assessment and diagnosis processes. The researchers used a software tool on a smart tablet device to capture detailed information about children's motor patterns, comparing the movement trajectories of autistic children and typically developing children during a cognitive task.

Machine Learning analysis of the motor patterns identified autism with 93% accuracy, indicating that the disorder can be computationally identified.

In particular, researchers use an Artificial Neural Network (ANN) to recognize autism motor signatures by processing complex and nonlinear relationships between variables. In particular, a feedforward multilayer perceptron was used to classify individuals with autism spectrum disorder and typical development. This research highlights the potential of technology in enhancing assessment and diagnosis processes for autism, as well as providing a new starting point for rehabilitation treatments.

3. Conclusion and Future Research

This paper discusses the use of machine learning techniques, specifically artificial neural networks (ANNs), in psychometric data analysis. While traditional psychometric methods rely on explanatory modelling, machine learning techniques offer a complementary approach where researchers aim to identify patterns in data and generate accurate predictions of output values based on input values of new observations.

Future research in this area will explore more powerful autoencoders, specifically Variational Autoencoders (VAE) [20], which encourage the latent space to follow a predefined distribution (e.g., normal distribution). VAE maps similar inputs closer into the latent space, allowing for greater interpretability of latent space and more direct comparison to latent variable analysis techniques such as Factor Analysis [21, 22]. Thanks to its characteristics, this neural network combines the predictive power typical of ML techniques, but also the interpretability of the explanatory approach making the integration of these approaches unified in a single method.

In particular, future research will investigate more in-depth the similarities and differences between VAE and other famous dimensionality reduction techniques such as Factor Analysis.

Furthermore, VAE will be applied to analyze movements and classify individuals with autism spectrum disorder and typical development.

One of the key advantages of Machine Learning techniques is their ability to generate accurate predictions based on new data inputs. This is particularly important in psychology, where accurate prediction of outcomes is critical for making informed decisions and guiding effective interventions. By using Machine Learning algorithms to identify patterns in data, researchers can develop more accurate models of human behaviour and cognition, which can help guide the development of new therapies and interventions.

Another advantage of Machine Learning techniques is their ability to handle large datasets and complex relationships among variables. This is particularly useful in fields like psychology, where datasets are often complex and difficult to analyze using traditional statistical techniques. By using Machine Learning algorithms to identify patterns in data, researchers can gain new insights into the underlying mechanisms of human behaviour and cognition, which can help guide the development of new theories and models.

As discussed in this contribution, Artificial Neural Networks, in particular, have shown great promise in a variety of psychological applications, from the development and validation of psychometric tests to the identification of neurological disorders such as autism. By using these powerful algorithms, researchers can gain new insights into the underlying mechanisms of these disorders, leading to more effective therapies and interventions.

In summary, the integration of Machine Learning techniques into traditional psychometric data analysis could provide new ways to analyze and interpret data, enabling the integration of explanatory and predictive modelling.

[1] L. Breiman. 2001. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science* 16, 3 (August 2001), 199–231. DOI:https://doi.org/10.1214/ss/1009213726

[2] D. J. Bauer. 2005. The Role of Nonlinear Factor-to-Indicator Relationships in Tests of Measurement Equivalence. *Psychological Methods* 10, (2005), 305–316. DOI:https://doi.org/10.1037/1082-989X.10.3.305

[3] W. C. M. Belzak and D. J. Bauer. 2019. Interaction effects may actually be nonlinear effects in disguise: A review of the problem and potential solutions. *Addictive Behaviors* 94, (July 2019), 99–108. DOI:https://doi.org/10.1016/j.addbeh.2018.09.018

[4] G. Shmueli. 2010. To Explain or to Predict? *Statistical Science* 25, 3 (August 2010), 289–310. DOI:https://doi.org/10.1214/10-STS330

[5] T. Yarkoni and J. Westfall. 2017. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspect Psychol Sci* 12, 6 (November 2017), 1100–1122. DOI:https://doi.org/10.1177/1745691617693393

[6] J. M. Hofman, D. J. Watts, S. Athey, F. Garip, T. L. Griffiths, J. Kleinberg, H. Margetts, S. Mullainathan, M. J. Salganik, S. Vazire, A. Vespignani, and T. Yarkoni. 2021. Integrating explanation and prediction in computational social science. *Nature* 595, 7866 (July 2021), 181–188. DOI:https://doi.org/10.1038/s41586-021-03659-0

[7] P. Dolce, D. Marocco, M. Nelson Maldonato, and R. Sperandeo. 2020. Toward a Machine Learning Predictive-Oriented Approach to Complement Explanatory Modeling. An Application for Evaluating Psychopathological Traits Based on Affective Neurosciences and Phenomenology. *Frontiers in Psychology* 11, (2020). Retrieved March 27, 2023 from https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00446

[8] M. Casella, P. Dolce, M. Ponticorvo, and D. Marocco. 2021. Autoencoders as an alternative approach to principal component analysis for dimensionality reduction. An application on simulated data from psychometric models. In *PSYCHOBIT*.

[9] H. Hotelling. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24, (1933), 417–441. DOI:https://doi.org/10.1037/h0071325

[10] Y. LeCun. 1987. Connexionist learning models. PhD thesis, Universite Pierre et Marie Curie (Paris)

[11] P. Baldi and K. Hornik. 1989. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks* 2, 1 (1989), 53–58. DOI:https://doi.org/10.1016/0893-6080(89)90014-2

[12] C. M. Bishop and N. M. Nasrabadi. 2006. *Pattern recognition and machine learning*. Springer.

[13] M. A. Kramer. 1991. Nonlinear principal component analysis using autoassociative

neural networks. *AIChE Journal* 37, 2 (1991),233–243. DOI:https://doi.org/10.1002/aic.690370209

[14] C. Pham, S. Ladjal, and A. Newson. 2022. PCA-AE: Principal Component Analysis Autoencoder for Organising the Latent Space of Generative Networks. *J Math Imaging Vis* 64, 5 (June 2022),569–585. DOI:https://doi.org/10.1007/s10851-022-01077-z

[15] M. Casella, P. Dolce, M. Ponticorvo, and D. Marocco. 2022. From Principal Component Analysis to Autoencoders: a comparison on simulated data from psychometric models. In *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 377–381. DOI:https://doi.org/10.1109/MetroXRAINE54828.2022.9967686

[16] M. Casella, P. Dolce, M. Ponticorvo, N. Milano, and Davide Marocco. In press. Artificial Neural Networks for Short-form Development of Psychometric Tests: A Study on Synthetic Populations Using Autoencoders. *Journal of Educational and Psychological Measurement*

[17] O. Gonzalez. 2021. Psychometric and Machine Learning Approaches to Reduce the Length of Scales. *Multivariate Behavioral Research* 56, 6 (November 2021), 903–919. DOI:https://doi.org/10.1080/00273171.2020.1781585

[18] T. Yarkoni. 2010. The abbreviation of personality, or how to measure 200 personality scales with 200 items. *Journal of Research in Personality* 44, 2 (April 2010), 180–198. DOI:https://doi.org/10.1016/j.jrp.2010.01.002

[19] R. Simeoli, N. Milano, A. Rega, and D. Marocco. 2021. Using Technology to Identify Children With Autism Through Motor Abnormalities. *Frontiers in Psychology* 12, (2021). Retrieved March 27, 2023 from https://www.frontiersin.org/articles/10.3389/fpsyg.2021.635696

[20] D. P. Kingma and Max Welling. 2019. An Introduction to Variational Autoencoders. *MAL* 12, 4 (November 2019), 307–392. DOI:https://doi.org/10.1561/2200000056

[21] Y. Huang and J. Zhang. 2022. Exploring Factor Structures Using Variational Autoencoder in Personality Research. *Frontiers in Psychology* 13, (2022). Retrieved March 27, 2023 from https://www.frontiersin.org/articles/10.3389/fpsyg.2022.863926

[22] C. J. Urban and D. J. Bauer. 2021. A Deep Learning Algorithm for High-Dimensional Exploratory Item Factor Analysis. *Psychometrika* 86, 1 (March 2021), 1–29. DOI:https://doi.org/10.1007/s11336-021-09748-3