Frontline Learning Research
An official journal of EARLI

Frontline **L**earning **R**esearch 2 (2013) 99-101
ISSN 2295-3159

Modelling for Prediction vs. Modelling for Understanding: Commentary on Musso et al. (2013)

Peter Edelsbrunner[a], Michael Schneider[b]

[a]ETH Zurich, Switzerland
[b]University of Trier, Germany

Article received 11 September 2013 / accepted 12 December 2013 / available online 20 December 2013

Abstract

Musso et al. (2013) predict students' academic achievement with high accuracy one year in advance from cognitive and demographic variables, using artificial neural networks (ANNs). They conclude that ANNs have high potential for theoretical and practical improvements in learning sciences. ANNs are powerful statistical modelling tools but they can mainly be used for exploratory modelling. Moreover, the output generated from ANNs cannot be fully translated into a meaningful set of rules because they store information about input-output relations in a complex, distributed, and implicit way. These problems hamper systematic theory-building as well as communication and justification of model predictions in practical contexts. Modern-day regression techniques, including (Bayesian) structural equation models, have advantages similar to those of ANNs but without the drawbacks. They are able to handle numerous variables, non-linear effects, multi-way interactions, and incomplete data. Thus, researchers in the learning sciences should prefer more theory-driven and parsimonious modelling techniques over ANNs whenever possible.

Keywords: Artificial neural networks; Black box; Student achievement; Statistical modelling

Musso, Kyndt, Cascallar, and Dochy (2013) conducted a study in which the statistical modelling technique of artificial neural networks (ANNs) was used to predict the academic achievement of university students a year in advance. The measures used were attention, working memory, learning strategies, and demographic variables. The results were precise estimations of each student's achievement tercile after their first year at university. This is an impressive success, demonstrating the usefulness of ANNs as a statistical modelling tool.

Corresponding author: Michael Schneider, University of Trier, www.educational-psychology.uni-trier.de, m.schneider@uni-trier.de, and Peter Edelsbrunner, ETH Zurich, www.ifvll.ethz.ch, peter.edelsburnner@ifv.gess.ethz.ch
http://dx.doi.org/10.14786/flr.v1i2.74

The study has raised an important question of the preferred statistical methods used by researchers in learning sciences. Should ANNs replace conventional statistical methods such as multiple regression, discriminant analysis, and structural equation modelling? – The potential of ANNs cannot be denied especially as a tool to examine predictive patterns in complex systems. However, Musso and colleagues overestimate the ability of ANNs in their application to the learning sciences. They do not mention shortcomings of ANNs, while overemphasizing shortcomings of competing conventional methods.

ANNs are limited in at least two important ways. First, the construction of ANN models such as those used by Musso et al. is highly explorative apart from choosing relevant input and output variables (Günther, Pigeot, &Bammann, 2012; Scarborough & Somers, 2006). The connection weights, which determine how an ANN transforms input into output patterns, are not specified by the researchers or based on theory. They are set to random values and changed gradually by an optimisation algorithm. This process usually involves thousands of iterations until each input pattern leads to the desired output pattern in the training data set. ANNs, thus, cannot be entirely compared to conventional methods since the latter are aimed at confirming or disconfirming pre-specified relations and interactions. In other words, the research question should determine whether the exploratory nature of ANNs is adequate, or if a conventional, confirmatory model should be the method of choice.

Second, connection weights cannot be codified into a coherent set of rules that delineate the process by which ANNs transform input patterns into output patterns. ANNs typically have a high number of connections between neurons (e.g., 300 in ANN1 by Musso et al.). The transformation process of input into output patterns is determined by non-linear, multi-way interactions of these connection weights. Recent research has attempted to increase the interpretability of ANNs, for example with the help of visualizations for complex interactions (e.g., Cortez & Embrechts, 2013; Intrator & Intrator, 2001). However, the basic problem of how non-linear interactions between hundreds of variables can be understood and communicated in meaningful terms has not yet been solved, causing ANNs to be frequently characterised as “black boxes” (cf. Benitez, Castro, & Requena, 1997). While one can assess how well an ANN works, it is difficult to comprehensively explain *why* it performs well or not (Scarborough & Somers, 2006). To interpret their results, Musso and colleagues list an importance parameter for each predictor but these parameters do not explain interaction effects or non-linear relations among the variables. In addition, it is difficult to integrate the results of ANNs across studies and also generalise from samples to underlying populations due to the lack of output parameters such as standard errors and error probabilities.

The explorative and opaque nature of ANNs impedes theory-developing and limits their practical application. Each relation in a statistical model should ideally correspond to a matching relation in an educational or psychological theory that justifies and explains the assumed statistical relation. Researchers can compare competing theories and advance assumptions that are not in line with the empirical data by fitting a series of statistical models that differ in theoretically relevant aspects (Kaplan, 1990). This is not possible with ANNs because the input-output relations are implicitly coded and distributed over all connection weights, preventing researchers from being able to map elements of an ANN and elements of a theory onto each other (Luger, 2009, p. 680).

The results obtained from ANN models are also of limited use for solving real-life problems. This limitation can be illustrated in a situation where diagnosticians would have to tell certain high school students that despite achieving satisfactory levels in their current academic performances, they cannot be admitted to college because an ANN predicts low academic performance in the future. In justifying the results, the diagnosticians would have to admit that they cannot explain how the different predictors statistically combine, nor describe the causal processes that will contribute to the anticipated decrease in the students’ achievement. These limitations are unsatisfactory from diagnostic, educational, and public policymaking perspectives.

Conventional methods represent more parsimonious and theory-driven alternatives to ANNs because they use smaller numbers of parameters, which enhances the interpretability of results. Like ANNs, modern regression techniques can account for non-linear relations (Bates & Watts, 2007) and complex interactions between variables (Aiken & West, 1991). Structural equation models are built on regression techniques and allow a simultaneous analysis of numerous variables. These models can be estimated by methods that are robust to missing data and non-normal distributions, account for hierarchical data structures, and identify heterogeneous sub-populations in mixture-models (Hoyle, 2012). Especially Bayesian structural equation models represent a strong advancement in modelling non-linear relations, assessing unspecified relations and handling highly non-normal and hierarchical data (Song & Lee, 2012). In contrast to ANNs, these modelling techniques require explicit theoretical assumptions about the relationship of the variables and they allow for explicit tests of these assumptions. This might limit their predictive power compared to ANNs, but it aids theory-building, hypothesis testing, and the communication of model results in practical applications.

Keypoints

- Artificial neural networks are powerful statistical tools for pattern recognition and prediction.
- Artificial neural networks transform input patterns into output patterns by non-linear multi-way interactions between simulated neurons that are governed by information that is stored in connection weights in an implicit and distributed way.
- This "black box" nature of artificial neural networks hampers the systematic testing of theories and the communication of results in practical settings.
- More conventional regression-type models can also handle non-linear relations, interaction effects, and a high number of variables, correlated errors, missing values, and non-normal distributions.
- Artificial neural network analysis cannot replace conventional statistical methods in the learning sciences but may be applicable in specific cases.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Bates, D. M., & Watts, D. G. (2007). *Nonlinear regression analysis and its applications* (2nd ed.). Hoboken, NJ: Wiley.

Benitez, J. M., Castro, J. L., & Requena, I. (1997). Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks, 8*, 1156-1164. doi:10.1109/72.623216

Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences, 225,* 1-17. doi:http://dx.doi.org/10.1016/j.ins.2012.10.039

Günther, F., Pigeot, I., & Bammann, K. (2012). Artificial neural networks modeling gene-environment interaction. *BMC Genetics*, *13*(1), 37. doi:10.1186/1471-2156-13-37

Hoyle, R. H. (Ed.). (2012). *Handbook of structural equation modeling.* New York: Guilford Press.

Intrator, O., & Intrator, N. (2001). Interpreting neural-network results: A simulation study. *Computational Statistics & Data Analysis, 37,* 373-393. doi:10.1016/S0167-9473(01)00016-0

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research, 25,* 137-155. doi:10.1207/s15327906mbr2502_1

Luger, G. F. (2009). *Artificial intelligence: Structures and strategies for complex problem solving* (6th ed.). Boston, MA: Pearson Education.

Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research, 1,* 42-71. Retrieved from http://journals.sfu.ca/flr/index.php/journal/article/view/13

Scarborough, D., & Somers, M. J. (2006). *Neural networks in organizational research: Applying pattern recognition to the analysis of organizational behavior* (pp. 137-144). Washington, DC: American Psychological Association.

Song, X. Y., & Lee, S. Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences.* Chichester, UK: John Wiley & Sons.