www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 12(2), May-Jun 2021, 7-15. ISSN: 2007-1558.

Comparative analysis of the prediction and classification accuracy of artificial neural networks with respect to traditional statistical methods

Carlos E. Belman-López [a], José A. Jiménez-García [b], José A. Vázquez-López [b]*

[a] PhD Department in Engineering Sciences, Tecnológico Nacional de México in Celaya.
[b] Department of Industrial Engineering; Tecnológico Nacional de México in Celaya.
carlosbelman@gmail.com; alfredo.jimenez@itcelaya.edu.mx; antonio.vazquez@itcelaya.edu.mx

Abstract. ANNs are flexible tools that have shown utility as function estimators and prediction methods in Data Mining, Machine Learning, among others. The aim of this research is to carry out comparative analysis of the prediction and classification accuracy between artificial neural networks and traditional statistical methods. To carry out this comparative analysis, different cases of prediction and supervised classification were selected. The samples of the different cases were divided into samples for training and validation. Finally, a comparative analysis of the prediction and classification accuracy between artificial neural networks and traditional statistical methods was obtained, using the sum of squared errors (SSE), the coefficient of determination (R^2), among others.

Keywords: Artificial Neural Networks, statistic, multilayer perceptron, regression, discriminant analysis

Article Info
Received July 20, 2020
Accepted April 14, 2021

1 Introduction

Today, many situations are needed to solve problems involving large sets of variables and data samples with some relation [25]. This can lead to the development of a prediction model, that is, a procedure to estimate the response given a set of independent variables. There are many examples of this requirement, in areas such as customer churn prediction [23], manufacturing [26,28,29], sales [16], marketing, financial sector [3], among others. This is also a fundamental case in areas of recent creation such as statistical learning [5], machine learning [29], Big Data [18,22,30] and data mining [13,27].

The use of artificial neural networks (ANNs) applied to statistical inference has grown widely in recent years. In statistical terms, ANNs are non-parametric estimators. For example, the traditional least squares regression method is a parametric estimator, because a starting model (the straight line) is imposed on the problem and the parameters are going to be adjusted according to these rules [2].

Unlike parametric statistical methods, multilayer perceptron (MLP) and other ANNs do not impose any functional starting form. The MLP has been related to traditional statistical models, such as regression or discriminant analysis, since it performs a type of multidimensional and non-linear regression, in which the initial assumption of a certain functional form is not necessary. It also can provide better results in non-linear problems and high dimension spaces [9].

The comparison between the MLP and different statistical models is of great importance in understanding how neural models operate. In fact, the relationship between ANNs and statistics is today one of the most important areas of work, both in terms of the number of publications and their impact. ANNs are currently considered one more set of methods to be added for data processing, which, depending on the conditions, can provide better results than conventional statistical methods [8]. For these reasons, in the present work ANNs were used in prediction and classification scenarios, to carry out a comparative analysis that measured the efficiency between traditional statistical methods and ANNs.

The document is organized as follows: the remainder of this section provides an explanation about artificial neural networks and the statistical methods used in this research. Section 2 explains the method and materials. Section 3 presents the results found. Section 4 presents the discussion. Section 5 presents the conclusions.

1.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) are structures with a mathematical and statistical behavior with the ability to learn, this is, the acquisition of knowledge that in most cases is based on examples. This learning is produced by a computer style called in parallel, that tries to simulate some of the capabilities in our brain, for this reason they are defined as artificial neural networks to distinguish them from biological models [20].

ANNs are a mimic of the human brain, an interconnected network of neurons, in which there are weighted communication channels between neurons [8]. Fig 1. shows a basic artificial neural network. In ANNs, neurons can react to multiple stimuli from neighbouring neurons and the whole network can change its state according to different inputs from the environment [2]. As a result, ANNs can generate outputs as a response to environmental stimuli, just as the human brain reacts to different environmental changes. ANNs are typically layered structures of various configurations. The most straightforward approach is to implement ANNs on a general-purpose central processing unit (CPU) in a multithread or multicore configuration from a hardware perspective. Furthermore, graphical processing units (GPUs), which are good at convolutional computations, have been found to be advantageous over CPUs for large-scale ANNs. CPU and GPU co-processing has turned out to be more efficient than CPU alone [15].

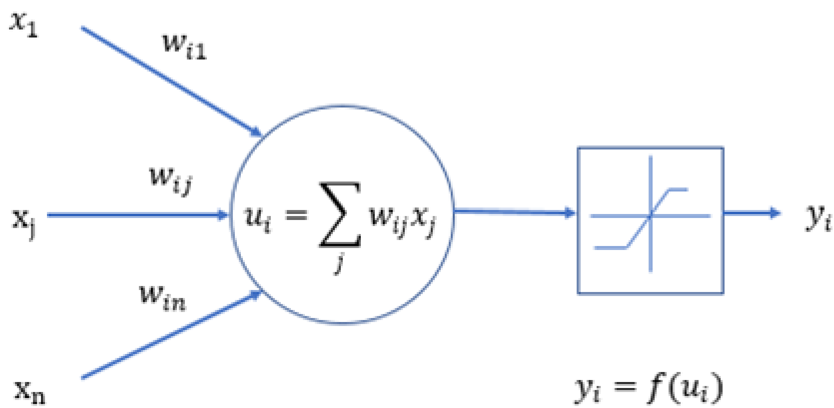

Fig 1. Basic model for an artificial neural network

MLP is a kind of artificial neural network conformed by multiple layers and it has the capability to act as a universal function approximator through the backpropagation algorithm. The backpropagation algorithm has the capability to provide the network with generalization so that the model obtains a correct output for an input data set that had not been used before [17].

1.2 Linear regression

This is the classic statistical method for predicting an output variable when there are one or more regressors. It is also a parametric method because assumes about the functional form of the relationship between the response with the regressors. This means, linear regression assumes the relationship between the response with the regressors is linear [7].

Because the relationship had been assumed as linear, estimating the coefficients is greatly simplified. The most common approach to fitting the model is known as least squares method. It is evident that the adjusted regression is an estimate of the true regression function [25].

1.3 Discriminant analysis

Discriminating functions are linear combinations of variables that most closely separate groups. The term group is used to represent either a population or a sample of the population. There are two main objectives in group separation:

1. Description of group separation is the creation of linear functions of *n* variables (discriminant functions) that are used to describe the differences between 2 or more groups. The discriminant functions should identify the relative contribution of the n variables to the separation into groups and find the optimal plane in which the points can be projected to illustrate the groups' configuration best.
2. Prediction or assignment of new observations in groups, in which functions of n linear variables are used to assign an individual sample unit in one of the groups [14].

The purpose of linear discriminant analysis is to find the linear combination of variables that provides the best possible separation between the groups in the data set. The maximum number of useful discriminant functions that can separate a data set is the minimum between the number of groups minus one and the number of variables used for classification [4].

1.4 Logistic regression

In real life, there are many situations in which it is evident that the response is not normal. For example, there are many applications where the response is binary (0 or 1). In the social sciences, a problem could be developing a model that can predict whether an individual is risky for a loan (0 or 1), or in manufacturing, whether an item is defective or not. The most popular approach to modelling binary responses is logistic regression [25]. This same approach can be applied to the problem of classifying two populations. One way to approach the problem is to define a classification variable, which takes the value zero when the element belongs to the first population, and the value of one when it belongs to the second population [11]. The logistic regression model can be generalized to the case where it is attempted to explain more than two discrete options.

This type of regression, where the dependent variable has more than two categories is known as multinomial logistic regression.

2 Methods and materials

To carry out this study, several case studies were selected to be analyzed, these cases are listed below:

- Case 1. Simple linear regression vs MLP + BP
- Case 2. Multiple linear regression vs MLP + BP
- Case 3. Binary logistic regression and linear discriminant analysis vs MLP + BP
- Case 4. Multinomial logistic regression and discriminant analysis vs MLP + BP

The method used for research is a staged method, the stages are described below.

Stage 1. Software selection.

R [12] was used as a tool to build traditional statistical methods and the training and testing the ANNs.

Stage 2. Sample selection

In this stage, different data samples were selected for the different scenarios to carry out the comparative analysis between the traditional statistical method and the ANNs. In each case, 2/3 of the data set was used for training, and the remaining 1/3 was used for testing.

a) Case 1. Simple linear regression vs MLP + BP

The data sample for this scenario was selected from [25]. This consists of data from 30 chipboards with densities ranging from approximately 8 to 26 pounds per cubic foot and their rigidity was measured in pounds per square inch. It is important to estimate the relationship between the density of a wood product and its rigidity. Fig. 2 shows the scatterplot for the values of ‘x’ and ‘y’. The correlation between ‘x’ and ‘y’ is 0.893, which implies a high degree of linear association between ‘x’ and ‘y’.

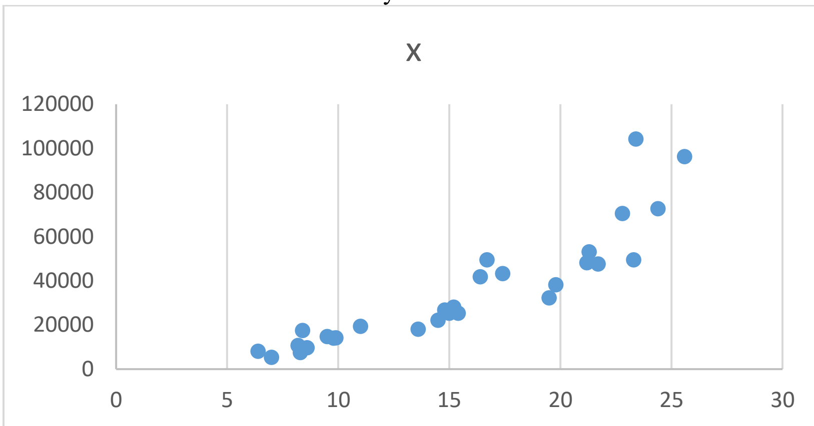

Fig. 2. Scatter diagram for the values ‘x’ and ‘y’ with correlation of 0.893.

b) Case 2. Multiple linear regression vs MLP + BP

This data set consists of a group of light trucks with engines that use diesel as fuel to find out if humidity, air temperature and barometric pressure influence the amount of nitrous oxide they emit (in ppm). Emissions were measured at different times and under various experimental conditions. The data set was taken from [25].

c) Case 3. Binary logistic regression and linear discriminant analysis vs MLP + BP

For this case, a public domain data set was used [10]. This data set is known as “Bank Marketing Dataset”, the data is related to the direct marketing campaigns (telephone calls) of a Portuguese banking institution. The objective of the classification was to predict whether the client will subscribe (“yes” or “no”) a bank deposit (variable y). The data set is made up of 45,211 observations and 16 input and 1 output variables.

d) Caso 4. Regresión logística multinomial y análisis discriminante vs MLP + BP

For this case, a public domain data set was used [21]. This data set is known as the “Seeds Dataset”, which consists of the measurements of 7 geometric properties of the seeds belonging to three different varieties of wheat, in addition to the response variable (the class to which it belongs).
A non-destructive X-ray technique was used to obtain the true value of seven measured attributes. 70 elements of each class were observed, obtaining 210 observations in total.

Stage 3. Development of statistical models.

a) Case 1. Simple linear regression vs MLP + BP

The concept of regression analysis refers to finding the best relationship between ‘y’ and ‘x’. The regression analysis was adjusted by the least squares method using the 'lm' function from R. This function is already included in the basic packages, so it is not necessary to install additional packages.

b) Case 2. Multiple linear regression vs MLP + BP

Multiple linear regression analysis was carried out by the least squares method using the ‘lm’ function from R. This function is included in the basic packages already included when installing R, so it is not necessary to import any additional package.

c) Case 3. Binary logistic regression and linear discriminant analysis vs MLP + BP

Discriminant Analysis was performed using the ‘lda’ function in “MASS” [24]. The logistic regression model was generated using the ‘glm’ function. This function is included in the basic packages already included with R, so it is unnecessary to import any additional package.

d) Case 4. Multinomial logistic regression and discriminant analysis vs MLP + BP

Discriminant Analysis was performed using the ‘lda’ function in “MASS” [24]. The multinomial logistic regression model was generated using the ‘vglm’ function of the “VGAM” package [19].

Stage 4. Training and optimization for the ANNs

ANNs are not easy to train and before starting the training it is necessary to do some preparation for the data. It is recommended to normalize the data before training the ANNs. Avoiding normalization can lead to useless results or a very difficult training process. Different methods can be chosen to scale the data, for example, z-normalization or min-max scale. In this study was chosen the min-max method and scale the data in the interval [0,1]. In general, the scale in the intervals [0,1] or [-1,1] tends to give better results [1]. Therefore, the first step before training the ANNs was to normalize the data by the min-max method. For the different scenarios, the

ANNs were trained to use the 'neuralnet' function belonging to 'neuralnet' package [6]. The ANNs were trained using k-fold with k equal to 5 and the learning rate equal to 0.01. The rest of the parameters used in each scenario are the backpropagation algorithm as optimizer, hyperbolic tangent function as activation function, the sum of squared errors (SSE) as loss function for prediction and crossentropy for classification. After training, the ANNs emitted an output in a normalized way, so the values had to be returned to their original scale, in order to make a meaningful comparison in the samples for testing using the ‘compute’ function belonging to “neuralnet” package.

Selection of architectures

a) Case 1. Simple linear regression vs MLP + BP

For this case, it was used a 1:3:1 configuration. It indicates that the input layer has 1 neuron, there is one hidden layer with three neurons and one neuron in the output layer. The output was linear, it means, the activation function was not applied to the output neuron.

b) Case 2. Multiple linear regression vs MLP + BP

For this case, it was used a 3:3:1 configuration. It indicates that the input layer has three neurons, there is one hidden layer with three neurons and one neuron in the output layer. The output was linear, it means, the activation function was not applied to the output neuron. The output was linear, that is, the activation function was not applied to the output neuron.

c) Case 3. Binary logistic regression and linear discriminant analysis vs MLP + BP

For this case, it was used a 16:9:1 configuration. It indicates that the input layer has 16 neurons, there is one hidden layer with 9 neurons and one neuron in the output layer. The output was not linear, it means, the activation function was applied to the output neuron.

d) Case 4. Multinomial logistic regression and discriminant analysis vs MLP + BP

For this case, it was used a 7:10:5:1 configuration. It indicates that the input layer has seven neurons, there is two hidden layers with ten and five neurons, respectively and one neuron in the output layer. The output was linear, it means, the activation function was not applied to the output neuron.

3 Results

Using the trained models, the SSE and R^2_{pred} was calculated on the validation samples. The results are presented below.

a) Case 1. Simple linear regression vs MLP + BP

In this case, simple linear regression was addressed by the least squares method and by ANNs, specifically using a Multilayer Perceptron. The results shown in Table 1 were obtained based on the predicted values using the proposed methods.

Table 1. SSE and R^2 for the proposed models in the case 1

Correlation	**Metrics**	**Linear regression**	**MLP**	
0.893	SSE	3782270481	2508107931	
	R^2	0.797458044	0.865689911	;0816

b) Case 2. Multiple linear regression vs MLP + BP

In this case, the problem of multiple linear regression was addressed by the least squares method and by ANNs, specifically using a Multilayer Perceptron. The results shown in Table 2 were obtained based on the predicted values using the proposed methods.

Table 2. SSE and R^2 for the proposed models in the case 2

	Multiple linear regression	**MLP**
SSE	0.0564871	0.02261404
R^2	0.62432094	0.8496007

c) Case 3. Binary logistic regression and linear discriminant analysis vs MLP + BP

In this case, the multivariable data classification problem was addressed using three different supervised classification methods. In this case, the sample is made up of two classes. The methods used are binary logistic regression, discriminant analysis and artificial neural networks (MLP+BP).

- The discriminant analysis correctly classified 89.6% of the observations.
- The binary logistic regression correctly classified 89.5% of the observations.
- The MLP (RNA) correctly classified 90.71% of the observations.

d) Case 4. Multinomial logistic regression and discriminant analysis vs MLP + BP

In this case, the problem of multivariate data classification was addressed using three different types of supervised classification techniques, such as artificial neural networks (MLP+BP), multinomial logistic regression (RL), and discriminant analysis (AD). In this case, the sample is made up of three classes.

- The discriminant analysis correctly classified 69 of 70 observations with a percentage of correct classifications of 98.57%.
- The multinomial logistic regression correctly classified 69 of 70 observations with a percentage of correct classifications of 98.57%.

- The MLP + BP correctly classified 67 of 70 observations with a correct classification percentage of 95.71%.

4 Discussion

a) Case 1. Simple linear regression vs MLP + BP

Based on the experimental results, it was concluded that when the data used have a moderately high correlation, the MLP + BP can significantly improve the prediction's quality according to the established measures.

b) Case 2. Multiple linear regression vs MLP + BP

Based on the experimental results, it was concluded that the MLP + BP has the ability to significantly improve the prediction accuracy for new observations, according to the established measures SSE and R^2.

c) Case 3. Binary logistic regression and linear discriminant analysis vs MLP + BP

Based on the experimental results, it is concluded that the three supervised classification techniques used in this investigation provide very good results when predicting the classification of new observations. Al realizar una prueba de comparación de las proporciones de clasificaciones correctas, como se muestra a continuación:
When performing 3-sample test for equality of proportions (the proportions of correct classifications) with p-value = 0.0002819, the hypothesis that the proportions are equal is rejected, in this case there is statistical evidence that there is a different proportion. So, it can be concluded that in this case of binary classification, ANNs are better.

d) Case 4. Multinomial logistic regression and discriminant analysis vs MLP + BP

The experimental results concluded that the three supervised classification techniques used in this investigation provide very good results when classifying new observations. When performing a comparison test of the proportions of correct classifications, it turns out that the hypothesis that the proportions are equal cannot be rejected (p-value = 0.6225). In this case there is no statistical evidence that any method is better. It should be noted that in this case, the sample size is small, so the Fisher exact test was used, in addition an exploratory study of the sample size necessary to obtain a β of the test of between .8 or .9 suggests around 600 data, a much larger amount than the 70 used in this case.

5 Conclusions

Model selection is a key step when a prediction model is developed. ANNs are structures with mathematical and statistical behavior that have the ability to learn based on data. ANNs are flexible tools that have shown utility as function estimators and now have increased the interest in using them as prediction tools in areas such as Data Mining, Machine Learning, among others. For these purposes, R language is a flexible, open source and widely used tool in data science that can be used to train statistical and artificial intelligence models. It is also easy to use, provides very good performance and does not consume a lot of computational resources.
Based on this research results, it can be concluded that MLP + BP can significantly decrease the reducible error and improve the prediction and classification accuracy according to several scenarios and different measures (SSE and R^2). It is important to highlight the prediction accuracy on new observations not used during the training. For future research, the models may be applied and compared on large volumes of data (Big Data), observing the prediction accuracy, and training and execution time. About ANNs setting, it may be considered using the MLP with other optimizers or even use other ANNs, for instance, Radial Base Function (RBF), SVM, among others. Regarding the measures used to evaluate the prediction accuracy, the PRESS statistic may be considered for assessing both statistical and ANN models.

6 References

[1] Alice, M. (2015, Septiembre 23). *R-Bloggers*. Retrieved from R-Bloggers: https://www.r-bloggers.com/fitting-a-neural-network-in-r-neuralnet-package/
[2] Beysolow II, T. (2017). *Introduction to Deep Learning Using R.* San Francisco, California, USA: Apress.
[3] Brummelhuis, R., & Luo, Z. (2017). Cds rate construction methods by Machine Learning Techniques. *Data Science Central*, 1-51. Retrieved from https://www.datasciencecentral.com/profiles/blogs/choice-of-k-in-k-fold-cross-validation-for-classification-in
[4] Coghlan, A. (2017). *A Little Book of R for Multivariate Analysis.* Cambridge: Creative Commons.
[5] Du, K.-L., & Swamy, M. (2014). *Neural Networks and Statistical Learning.* London: Springer.
[6] Fritsch, S., & Frauke, G. (2016). neuralnet: Training of Neural Networks. Retrieved from https://CRAN.R-project.org/package=neuralnet
[7] Hines, W., & Montgomery, D. (1996). *Probabilidad y estadística para ingeniería y administración.* Ciudad de México: Continental S.A de C.V.
[8] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R.* New York: Springer.
[9] Martin del Brío, B., & Sanz Molina, A. (2002). *Redes neuronales y sistemas difusos.* (2a. ed.). Madrid, España: Alfaomega & RA-MA.
[10] Moro, S., Cortez, P., & Rita, P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems. *Elsevier*(62), 22-31. Retrieved from https://archive.ics.uci.edu/ml/datasets/bank+marketing#
[11] Peña, D. (2002). *Análisis de datos multivariantes.* Madrid: McGraw-Hill / Interamericana de España, SA.
[12] R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from The R Project for Statistical Computing: https://www.R-project.org/
[13] R, S., & R, S. (2017). Data Mining with Big Data. *Intelligent Systems and Control (ISCO)* (pp. 246-250). Coimbatore, India: IEEE. doi:10.1109/ISCO.2017.7855990
[14] Rencher, A. (2002). *Methods of multivariate analysis* (2da ed.). Canadá: John Wiley & Sons, Inc.
[15] Rong, G., Mendez, A., Assi, E., Zhao, B., & Sawan, M. (2020). Artificial Intelligence in Healthcare: Review and Prediction Case Studies. *Engineering*, 291–301. doi:https://doi.org/10.1016/j.eng.2019.08.015
[16] Ruelas Santoyo, E., & Laguna González, J. (2014). Predictive comparison based in neural network versus statistical methods to forecast sales. *Ingeniería Industrial. Actualidad y Nuevas Tendencias*, 91-105.
[17] San Miguel Salas, J. (2016). *Desarrollo con MATLAB de una red neuronal para estimar la demanda de energía eléctrica (Tesis de Maestría).* Valladolid, España: Universidad de Valladolid.
[18] Spangenberg, N., Roth, M., & Franczyk, B. (2015). Evaluating New Approaches of Big Data Analytics Frameworks. *International Conference on Business Information Systems* (pp. 28 - 37). Suiza: Springer. doi:10.1007/978-3-319-19027-3_3
[19] Thomas W., Y. (2015). *Vector Generalized Linear and Additive Models: With an Implementation in R.* New York, USA: Springer.
[20] Torras P., S., & Monte, E. (2013). *Modelos neuronales aplicados en economía.* Barcelona, España: Addlink.
[21] UC Irvine. (2012, 09 29). *UCI Machine Learning Repository.* Retrieved from UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/seeds#
[22] Ularu, E., Puican, F. C., Apostu, A., & Velicanu, M. (2012). Perspectives on Big Data and Big Data Analytics. *Database Systems Journal*, 1-12.
[23] Vafeiadis, T., Diamantaras, K., Sarigiannidis, G., & Chatzisavvas, K. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 1-9.
[24] Venables, W., & Ripley, B. (2002). *Modern Applied Statistics with S.* (Cuarta ed.). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4
[25] Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias.* Ciudad de México: PEARSON.
[26] Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 144–156. doi:https://doi.org/10.1016/j.jmsy.2018.01.003
[27] Williams, G. (2011). *Data Mining with Rattle and R.* New York: Springer.

[28] Wu, D., Jennings, C., Terpenny, J., Gao, R., & Kumara, S. (2017). A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *Journal of Manufacturing Science and Engineering*, 1-10.

[29] Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, 23-45. doi:10.1080/21693277.2016.1192517

[30] Xu, L. D., & Duan, L. (2018). Big data for cyber physical systems in industry 4.0: a survey. *Enterprise Information Systems*, 1-23. doi:10.1080/17517575.2018.1442934