HMMOCS 2022
International Workshop "Hybrid methods of modeling and optimization in complex systems"

CREDIT CARD ATTRITION CLASSIFICATION THROUGH NEURONETS

Spyridon D. Mourtas (a)*, Vasilios N. Katsikis (b), Romanos Sahas (c)
*Corresponding author

(a) Laboratory "Hybrid Methods of Modelling and Optimization in Complex Systems", Siberian Federal University, Prosp. Svobodny 79, 660041 Krasnoyarsk, Russia, Department of Economics, Division of Mathematics-Informatics and Statistics-Econometrics, National and Kapodistrian University of Athens, Sofokleous 1 Street, Athens, 10559, Greece, spirosmourtas@gmail.com
(b) Department of Economics, Division of Mathematics-Informatics and Statistics-Econometrics, National and Kapodistrian University of Athens, Sofokleous 1 Street, Athens, 10559, Greece, vaskatsikis@econ.uoa.gr
(c) Department of Economics, Division of Mathematics-Informatics and Statistics-Econometrics, National and Kapodistrian University of Athens, Sofokleous 1 Street, Athens, 10559, Greece, romanossahas@econ.uoa.gr

Abstract

Credit card customers comprise a volatile subset of a banks' client base. As such, banks would like to predict in advance which of those clients are likely to attrite, so as to approach them with proactive marketing campaigns. Neuronets have found great application in many classification problems. Credit card attrition is a poorly investigated subtopic that poses many challenges, such as highly imbalanced datasets. The goal of this research is to construct a feed-forward neuronet that can overcome such obstacles and thus accurately classify credit card attrition. To this end, we employ a weights and structure determination (WASD) algorithm that facilitates the development of a competitive and all around robust classifier whilst accounting for the shortcomings of traditional back propagation neuronets. This is supported by the fact that when compared with some of the best performing classification models that MATLAB's classification learner app offers, the power softplus activated WASD neuronet demonstrated either superior or highly competitive performance across all metrics.

Keywords: Neuronets, weights and structure determination, classification, credit card attrition

1. Introduction

The growth of a company depends significantly on the efforts that the entity makes towards: maintaining and growing its existing customer base, acquiring/keeping up with new technology, focusing on specific market segments, and improving its productivity and efficiency. This is especially true in highly competitive and mature business sectors, such as the banking sector (Hu, 2005). It is asserted that the first of these criteria is also the most obvious. In particular, more and more businesses are realizing that their current client base is their most valuable asset (García et al., 2017). It is not surprising that service providers in the financial sector go to considerable lengths to entice customers away from their rivals while minimizing their own losses (He et al., 2014).

Nevertheless, clients themselves have also grown aware of the value of their customer base, not only the banks. Another factor enhancing the already competitive atmosphere is the latter party's growing knowledge of the quality of the services offered. That is, a customer will frequently abruptly stop doing business with a bank and go to a rival company because of considerations like accessibility or even a more enticing interest rate (Farquad et al., 2014). Because of the impact that even a slight increase in customer retention can have on the banks' income statement along with the well-established facts that maintaining is significantly less expensive than re-acquiring lost customers, financial institutions have been motivated to gradually change their focus from attracting new customers to retaining as many of their existing ones (Tang et al., 2014).

The importance of knowing ahead of time who of their clients, starting with the high return on investment clients, are likely to quit has increased for banks (He et al., 2014). When a financial institution has this information, it may run targeted marketing programs that have been proven to be quite successful at keeping customers (Hu, 2005). Due to the importance and high demand from financial institutions, this research introduces a neuronet (also known as neural network) for credit card attrition classification. It is worth mentioning that from a business standpoint, attrition is the gradual but intentional loss of customers or staff that happens as they leave a company and are not replaced.

Artificial neuronets have been successfully applied to a wide spectrum of fields, including but not limited to medicine, such as in the prediction of breast cancer (Ed-daoudy & Maalmi, 2020) and economics and finance, such as in the classification of firm fraud (Simos et al., 2022), in the analysis of time series (Zhang et al., 2019), in portfolio optimization (Mourtas & Katsikis, 2022) as well as in the prediction of various macroeconomic measures (Zeng et al., 2020). Furthermore, there is an abundance of applications in problems stemming from the various engineering disciplines. For example, artificial neuronet models have been applied to performance analysis of solar systems (Esen et al., 2009), remote sensing multi-sensor classification (Bigdeli et al., 2021), predict flow behavior of alloy (Huang et al., 2018), performance analysis of heat pump systems (Esen et al., 2017). This research introduces and investigates a novel feed-forward neuronet (FFN), which uses a weights and structure determination (WASD) training algorithm, for classifying credit card attrition.

The following are the main points of this research: a novel power softplus WASD (PS-WASD) neuronet for addressing classification problems is presented; the PS-WASD neuronet's performance is compared to some of the best performing MATLAB's classifiers, as well as a Bernoulli WASD (B-WASD) neuronet (Zhang et al., 2019), on a publicly accessible credit card attrition dataset.

2. Problem Statement

The most popular training algorithm for FFNs is the back-propagation error algorithm. However, traditional back-propagation neuronets come with a few shortcomings, namely: they are characterized by high computational complexity, especially on the grounds of obtaining an optimal structure; the selection of suitable hyper-parameters, a rather difficult task, can often make or break the neuronet's performance; there is high risk of converging to a sub-optimal solution. To address these shortcomings, the alternative training algorithm WASD is proposed (Zhang et al., 2019).

Newly implemented WASD training algorithms offer a feature that their predecessors lack. Namely, the weights direct determination (WDD) process, inherent in any WASD algorithm, facilitates the direct computation of the optimal set of weights, hence allowing one to avoid getting stuck in local minima and all in all contributing in the achievement of lower computational complexity (Simos et al., 2022; Zhang et al., 2019). In this paper, a PS-WASD neuronet will be applied to a publicly available credit card attrition dataset and its' performance will be compared to other popular classifiers that MATLAB's classification learner app offers, namely linear support vector machines (SVM), fine k-nearest neighbors (KNN), kernel naive bayes (KNB), as well as the B-WASD neuronet.

3. Research Questions

This research proposes a PS-WASD neuronet to deal with the problem of classifying credit card attrition. Does the PS-WASD training algorithm, however, allow for the construction of a neuronet that matches (or potentially surpasses) the performance of benchmark classifiers, such as models coming from MATLAB's classification learner app? The findings of this paper demonstrate that this is in fact feasible. We construct a 3-layer PS-WASD based FFN that, as far as the dataset in question is concerned, outperforms a number of well-established high performing. Additionally, we study the question if the PS-WASD algorithm perform even better than the already established WASD, such as those in (Zhang et al., 2019). The results of the experiments show that the PS-WASD is more accurate and precise.

4. Purpose of the Study

In this study, a FFN is used to classify credit card customers that are likely to attrite. The purpose is to develop a WASD classifier that can rise to the task. As a result, financial institutions that can acquire that knowledge for their client base obtain a competitive advantage in the form of operating cost reduction.

5. Research Methods

This section describes the 3-layer PS-WASD based FFN model which has m input and n hidden layer neurons. Particularly, Layer 1 is the input layer that receives and allocates the input values $X_1, X_2, \ldots, X_m$ to the corresponding neuron in Layer 2, which has a maximum number of activated neurons of n, with equal weight 1, whereas Layer 3 is the output layer and has one activated neuron. The weights $W_i, i = 1, 2, \ldots, n-1$ in neurons link between Layer 2 and Layer 3 are obtained via the WDD process. By using the PS-WASD algorithm, the neuronet model can achieve minimal hidden layer utilization.

5.1. The PS-WASD Based Neuronet

Let $X = [X_1, X_2, \ldots, X_m] \in \mathbb{R}^{s \times m}$ be the normalized input matrix in the interval $[-0.5, -0.25]$ with s denoting the number of the samples and $Y \in \mathbb{R}^s$ be a binary target vector. Notice that in order to avoid over-fitting, the inputs X must be normalized in the interval $[-0.5, -0.25]$, which can be done by applying the following transformation (Zhang et al., 2019): $X_{nor} = \frac{X - \min(X)}{4(\max(X) - \min(X))} - \frac{1}{2}$. It should also be noted that all operations on X are element-wise in this section.

Assume that there exists an underlying relationship f such that $Y = f(X_1, X_2, \ldots, X_m)$. The neuronet aims at approximating f through a linear combination of n activation functions F, where $F_d(X) = \ln(1 + e^{X^d})$ is the power softplus activation as in (Simos et al., 2022), with $d = 0,1, \ldots, n-1$. For each activation F_d, let q_d denote the image of X under F_d. Thus, $q_d \in \mathbb{R}^{s \times m}$. Assuming a weights vector $W = [W_0, W_1, \ldots, W_{n-1}] \in \mathbb{R}^{mn \times n}$, a linear combination of all n images may be written as $\sum_{i=0}^{n-1} W_i q_i = \hat{Y}$, where $Q = [q_0, q_1, \ldots, q_{n-1}] \in \mathbb{R}^{s \times mn}$. Note that $\hat{Y}$, the neuronet's output, must be converted to binary. To this end, an element-wise function B is employed so that the final output of the neuronet is $B(\hat{Y})$ with

$$B_i(\hat{Y}) = \begin{cases} 1, & \hat{Y}_i \geq -0.375 \\ 0, & \hat{Y}_i < -0.375 \end{cases} \quad \text{for} \quad i = 1,2, \ldots, s,$$

where 1 implies a customer that is likely to attrite and vice versa. On the convergence of the neuronet, the following Theorem 1 and Proposition 1 should be noted.

Theorem 1: Let K be a nonnegative integer and assume a target function $f(\cdot)$ that has the $(K + 1)$-order continuous derivative on interval $[a, b]$, then for $x \in [a, b]$,

$$f(x) = P_K(x) + R_K(x), \tag{1}$$

where $P_K(x)$ is a polynomial used to approximate $f(x)$ and $R_K(x)$ is the error term.

Given a constant value $h \in [a, b]$, $f(x)$ can be approximated as:

$$f(x) \approx P_K(x) = \sum_{i=0}^{K} \frac{f^{(i)}(h)}{i!}(x - h)^i, \tag{2}$$

where $f^{(i)}(h)$ is the value of the i-order derivative of $f(x)$ at h, $i!$ implies the factorial of i, and $P_K(x)$ signifies the K-order Taylor polynomial of function $f(x)$.

Proposition 1: Theorem 1 can be used to approximate multivariable functions (Zhang et al., 2019). For a target function $f(x_1, x_2, \ldots, x_g)$ with $(K + 1)$-order continuous partial derivatives in an origin's neighborhood $(0, \ldots, 0)$ and g variables, the K-order Taylor polynomial $P_K(x_1, x_2, \ldots, x_g)$ to the origin is:

$$P_K(x_1, x_2, \ldots, x_g) = \sum_{i=0}^{K} \sum_{i_1 + \cdots + i_g = i} \frac{x_1 \cdots x_g}{i_1 \cdots i_g} \left(\frac{\partial^{i_1 + \cdots + i_g} f(0, \cdots, 0)}{\partial x_1^{i_1} \cdots \partial x_g^{i_g}} \right), \tag{3}$$

where $i_1, i_2, \ldots, i_g$ are nonnegative integers.

Based on the aforementioned analysis, the optimal weights W can be obtained through the WDD process (Zhang et al., 2019) as follows:

$$W = Q^\dagger Y, \tag{4}$$

where $Q^\dagger$ denotes the pseudo-inverse of Q.

5.2. The PS-WASD Algorithm

With the computation of the optimal weights being handled by (4), the next concern is the determination of the optimal size for the neuronet. This is done through the PS-WASD algorithm that is employed in conjunction with cross validation so as to ensure that the model generalizes beyond the set of training. The PS-WASD algorithm can be described in the following three steps.

Step 1: Considering the normalized input matrix X, the training and testing sets are determined by a random $80\% - 20\%$ split. The training set is once again divided into two sets of samples, one for model fitting and the other for validation. The parameter $p \in (0,1] \subseteq \mathbb{R}$, which is set by the user, represents the percentage of samples that is to be allocated for model fitting purposes.

Step 2: Assuming that the training set consists of $J = \frac{80}{100} s$ samples, we use the first $J_1 = pJ$ samples for fitting the model, and the rest $J_2 = J - J_1$ samples for validation. Starting from the power $d = 0$ and an empty matrix Q, the PS-WASD algorithm grows Q by adding the respective columns q_{d+1}, whilst incrementing d. In each iteration, the optimal weights are computed with respect to the training set as in Eq. 4 and the current performance of the neuronet is assessed on the validation set through a metric of choice, namely the mean-absolute-error (MAE $= \frac{1}{J_2} \sum_{j=1}^{J_2} |\hat{Y}_j - Y_j|$). Added columns are only kept if their addition has resulted in a performance boost; that is, a lower MAE.

Step 3: This process is repeated until d reaches a predefined threshold, which we set to be $d_{max} = 30$. In reaching d_{max}, the training and validation sets are brought together and the final set of weights is computed relative to all available training and validation samples.

To conclude, the PS-WASD training algorithm not only guarantees that the weights are optimal, but also facilitates future computation by growing the neuronet's structure to the smallest size possible. In combination with cross validation as described above, the PS-WASD training algorithm is indeed a powerful tool.

6. Findings

With the neuronet having obtained it's final structure, all that is left is to apply it on the testing set. It is worth mentioning that the credit card attrition dataset used for evaluating the neuronets performance is available at https://www.kaggle.com/datasets/jaikishankumaraswamy/customer-credit-card-churn, and includes 33 variables and 1 target. Figures 1a and 1b depict the training error path the classification results on the testing set of the PS-WASD under $p = 0.75$, respectively, whereas Tables 1 and 2 evaluate the PS-WASD performance against other popular classifiers. It bears mentioning that the abbreviation 'CCC' in the rightmost Figure stands for credit card customer.
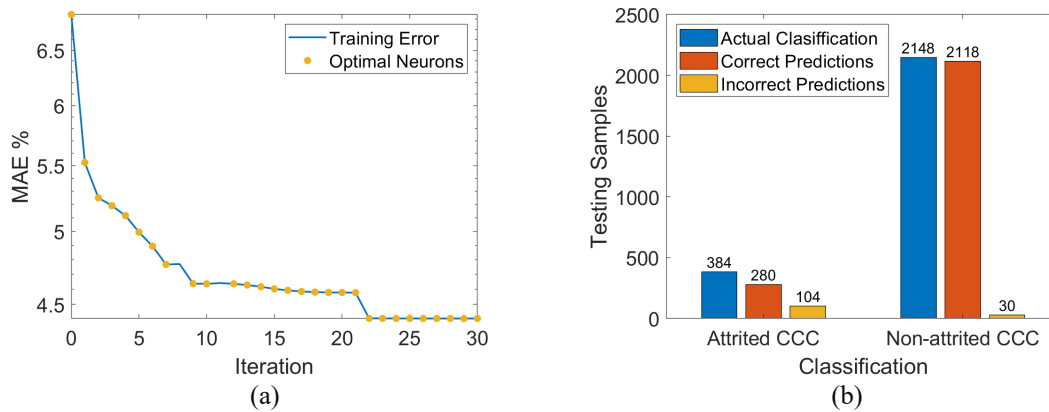
Figure 1. Training error path (a) and testing set classification (b) of the PS-WASD

To be able to extract meaningful conclusions as to the competency of the proposed PS-WASD neuronet, a comparison with other well performing classifiers is in order. To that end, we picked three popular models from MATLABs' classification learner app, namely, KNB, Linear SVM and Fine KNN, as well as the B-WASD. All models are evaluated on the basis of nine metrics, namely, MAE, TP, FP, TN, FN, Precision, Recall, Accuracy and F-measure, where TP, FP, TN, FN stand for true positive, false positive, true negative and false negative rate, respectively. Except for MAE, the rest of the metrics are calculated by means of those last four quantities.

Table 1. Performance comparison between all classifiers

Model	PS-WASD	KNB	Linear SVM	Fine KNN	B-WASD
MAE	0.052923	0.14652	0.092022	0.15008	0.19155
TP	0.98603	1	0.96648	0.92551	0.79562
FP	0.013966	0	0.03352	0.074488	0.20438
TN	0.72917	0.033854	0.58073	0.42708	0.88021
FN	0.27083	0.96615	0.41927	0.57292	0.11979
Precision	0.98603	1	0.96648	0.92551	0.79562
Recall	0.78452	0.50861	0.69744	0.61766	0.86914
Accuracy	0.94708	0.85348	0.90798	0.84992	0.80845
F_measure	0.87381	0.67428	0.81021	0.74088	0.83076

Table 2. Results of McNemar test

PS-WASD	McNemar test	
vs.	Null Hypothesis	p-value
KNB	Rejected	6.1951e-49
Linear SVM	Rejected	1.3301e-13
Fine KNN	Rejected	1.801e-45
B-WASD	Rejected	4.0024e-61

7. Conclusion

The results of the collective testing are shown in Table 1, where the bolded numbers correspond to the best outcomes for each metric. The PS-WASD definitely stands out as it scores the lowest MAE, the

highest F-measure and Accuracy. In the departments of Recall and Precision it takes second place to the B-WASD and KNB, respectively. However, high performance at certain metrics does not seem to come at a cost of unjustifiably low scores on the rest of metrics, as is the case with the latter two models. Thus, the PS-WASD proves itself a reliable classifier that does not simply excel at one metric but rather performs exceptionally well all around.

In order to properly assess whether one can conclude to a model that is better suited for predicting credit card attrition in the specific dataset, we add to the previous analysis a statistical element, in the form of a mid-p-value McNemar test (Simos et al., 2022). The purpose of this test is to check whether two classifiers demonstrate equal accuracy (null hypothesis) or unequal accuracy (alternative hypothesis) in predicting the true class. We took all possible combinations of pairs consisting of the PS-WASD and one of the other four models and conducted the mid-p-value McNemar test. The results are shown in Table 2. At the 5% significance level, there is strong evidence to reject the null hypothesis for all pairs. The superior accuracy of the PS-WASD is thus put into firm ground.

To conclude, this paper introduced a PS-WASD neuronet in view of classifying credit card attrition. When applied to a publicly available credit card attrition dataset, the PS-WASD demonstrated either superior or highly competitive performance capabilities, compared to a number of well-established and well-performing classifiers. The credibility of those results is further supported by the conduction of a mid-p-value McNemar test through which we have been able to assert that the difference between the predictive accuracy of the PS-WASD and the rest of the models is in fact statistically significant. We may thus conclude that the PS-WASD classifier is a reliable classifier that can successfully take up the task of classifying credit card attrition.

Acknowledgments

References

Bigdeli, B., Pahlavani, P., & Amirkolaee, H. A. (2021). An ensemble deep learning method as data fusion system for remote sensing multisensor classification. *Applied Soft Computing, 110,* 107563. https://doi.org/10.1016/j.asoc.2021.107563

Ed-daoudy, A., & Maalmi, K. (2020). Breast cancer classification with reduced feature set using association rules and support vector machine. *Network Modeling Analysis in Health Informatics and Bioinformatics, 9*(1), 34. https://doi.org/10.1007/s13721-020-00237-8

Esen, H., Esen, M., & Ozsolak, O. (2017). Modelling and experimental performance analysis of solar-assisted ground source heat pump system. *Journal of Experimental & Theoretical Artificial Intelligence, 29*(1), 1-17. https://doi.org/10.1080/0952813x.2015.1056242

Esen, H., Ozgen, F., Esen, M., & Sengur, A. (2009). Artificial neural network and wavelet neural network approaches for modelling of a solar air heater. *Expert Systems with Applications, 36*(8), 11240-11248. https://doi.org/10.1016/j.eswa.2009.02.073

Farquad, M. A. H., Ravi, V., & Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing, 19,* 31-40. https://doi.org/10.1016/j.asoc.2014.01.031

García, D. L., Nebot, À., & Vellido, A. (2017). Intelligent data analysis approaches to churn as a business problem: a survey. *Knowledge and Information Systems, 51*(3), 719-774. https://doi.org/10.1007/s10115-016-0995-z

He, B., Shi, Y., Wan, Q., & Zhao, X. (2014). Prediction of Customer Attrition of Commercial Banks based on SVM Model. *Procedia Computer Science, 31,* 423-430. https://doi.org/10.1016/j.procs.2014.05.286

Hu, X. (2005). A Data Mining Approach for Retailing Bank Customer Attrition Analysis. *Applied Intelligence, 22*(1), 47-60. https://doi.org/10.1023/b:apin.0000047383.53680.b6

Huang, C., Jia, X., & Zhang, Z. (2018). A Modified Back Propagation Artificial Neural Network Model Based on Genetic Algorithm to Predict the Flow Behavior of 5754 Aluminum Alloy. *Materials, 11*(5), 855. https://doi.org/10.3390/ma11050855

Mourtas, S. D., & Katsikis, V. N. (2022). Exploiting the Black-Litterman framework through error-correction neural networks. *Neurocomputing, 498,* 43-58. https://doi.org/10.1016/j.neucom.2022.05.036

Simos, T. E., Katsikis, V. N., & Mourtas, S. D. (2022). A multi-input with multi-function activated weights and structure determination neuronet for classification problems and applications in firm fraud and loan approval. *Applied Soft Computing, 127,* 109351. https://doi.org/10.1016/j.asoc.2022.109351

Tang, L., Thomas, L., Fletcher, M., Pan, J., & Marshall, A. (2014). Assessing the impact of derived behavior information on customer attrition in the financial service industry. *European Journal of Operational Research, 236*(2), 624-633. https://doi.org/10.1016/j.ejor.2014.01.004

Zeng, T., Zhang, Y., Li, Z., Qiu, B., & Ye, C. (2020). Predictions of USA Presidential Parties From 2021 to 2037 Using Historical Data Through Square Wave-Activated WASD Neural Network. *IEEE Access, 8,* 56630-56640. https://doi.org/10.1109/access.2020.2982192

Zhang, Y., Chen, D., & Ye, C. (2019). *Deep neural networks: wasd neuronet models, algorithms, and applications.* CRC Press.