NEURAL NETWORKS AND THE AUTOMATIC SELECTION OF FINANCIAL RATIOS

Duarte Trigueiros
ISCTE,
Av. Forças Armadas,
1600 Lisbon, Portugal
Phone: +351 (1) 793 50 00 Fax: +351 (1) 796 47 10,
E-Mail: dmt@sara.inesc.pt

Robert H. Berry
University of East Anglia,
School of Information Systems,
Norwich NR4 7TJ,
United Kingdom
Phone +44 (603) 56 161 Fax: +44 (603) 50 77 20

April 20, 1991

Abstract

In this paper we show that Neural Networks can automatically build optimal structures similar to financial ratios, thus avoiding the empirically-based search of the best ratio for a given task and providing ready-to-use models based on past experience. We use a well-known example to illustrate this mechanism.

1 Introduction

Accounting reports are an important source of information for managers, investors and financial analysts. Ratios are the usual instruments for extracting this information. They are supposed to control for the effect of size and to highlight some noteworthy features of the firm. However, as N accounting variables can generate up to $N^2 - N$ ratios and many of them carry similar information, it is often difficult to select the appropriate ratio for a given task.

Statistical techniques have been used to discover appropriate ratios. An early attempt is that of Beaver (1966) [1] who discovered by trial which ratios would best predict financial distress. However, later on, accounting research avoided the problem of selecting appropriate ratios by using a large number of them as input variables in statistical models. It is easy to find in the literature models with forty and more predictors. Chen and Shimerda (1981) [2] review this problem.

The Multi-Layer Perceptron [6] (MLP) is a supervised learning Neural Network. Topologically it is a layered feed-forward configuration: Nodes are arranged in layers and each node's output is connected to next layer's inputs. The MLP is different from the algorithms intended to learn a relation input-outcome from a set of examples, in that it approaches relations by steps. In each step the MLP creates new sets of variables corresponding to different stages of the process of modelling the desired

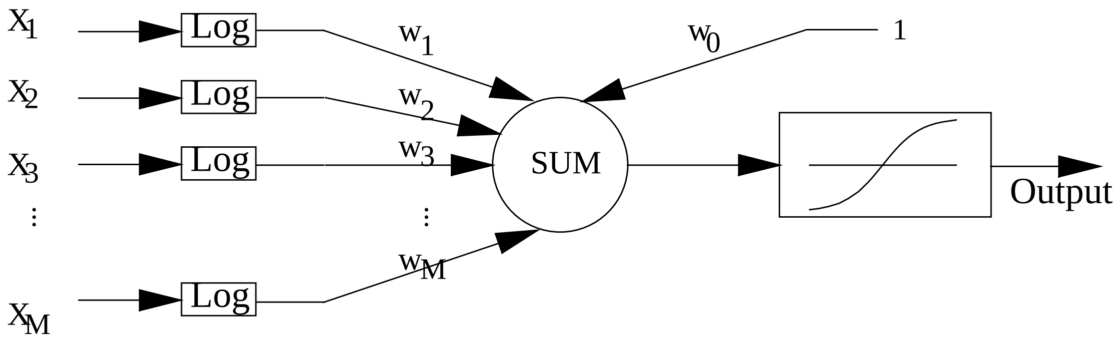

Figure 1: A node able to form a ratio in the first hidden layer of a MLP.

relation. A particular step uses the variables from the previous one as input. Then, it makes an improvement towards the final modelling of the relation and it outputs a new set of variables to be used as input for the next stage. These intermediate variables are often referred to as internal representations.

In this paper we show that the ability of the MLP to build internal representations make it possible to automatically build optimal ratios directly from raw data as found in accounting reports, thus avoiding the need for selecting ratios. Section 2 describes how reports can be used as direct inputs for an MLP. Section 3 introduces a well-known problem of modelling with accounting data. Using such problem as an example, section 4 explains the departures from ordinary techniques we introduced in the training of the MLP. Finally we discuss the results.

Acknowledgement: This study has been carried out in the University of East Anglia, in the U.K. We acknowledge the guidance and continued support of Professor Luis B. Almeida at INESC, Portugal.

2 Ratios as Internal Representations

The problem of learning from examples using ratios can be formalized in this way: Let x and y be two items from the same report, j. They form the ratio $y/x = r$. For learning, we have a sample containing $1, \cdots, j, \cdots, N$ examples of these two observations, plus t, the vector of related outcomes. If we assume the existence of a map $\mathcal{W}$ such that $\mathcal{W} : r \mapsto t$, then we learn it by finding a $\mathcal{W}$ which is optimal in some sense.

The functional relationship yielding ratios can be generalized so as to assume, in logarithmic space, the form

$$\log y_j - b \times \log x_j = w_0 + \varepsilon_j \quad \text{or similar. In ordinary space,} \quad \frac{y_j}{x_j^b} = \exp(w_0) \times f_j \quad (1)$$

(see [11]). This relationship is different from the one we outlined above, linking accounting features and outcomes. However, these two relations are not independent. Outcomes are dictated by internal features of the firm which, we believe, are captured by appropriate ratios.

In the accounting statistical models used so far, the former relation is embedded in the choice of the input data — ratios. In the framework presented here we let the MLP form both such relations. Appropriate ratios are discovered and used to approach the outcomes, as part of a unique optimization process. Firstly, we let the raw data from financial reports of firms be the input to an MLP. Then, the first hidden layer forms ratios that best approach the outcomes. Other steps follow. Finally, the outputs model the desired relation.

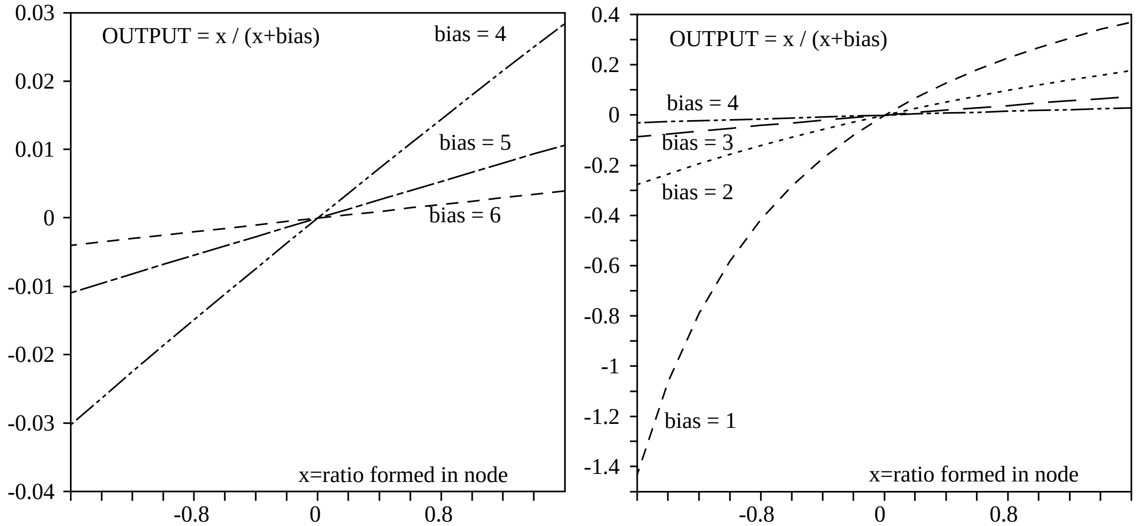

Figure 2: The output of each node in the log MLP will be a concave function approaching linearity for increasing values of the bias. On the left, a magnified view.

A multivariate generalization of (1), able to account for both common and particular components of the variability of M accounting items is

$$\log r = \sum_{i=1}^{M} w_i \times \log x_i . \tag{2}$$

The residuals are omitted. Notice that this expression, an inner product, is the same as a Neural Network node's output. Our approach consists of letting w_i be the adjustable connections or weights linking the inputs of an MLP with the nodes in the first hidden layer. The inputs are the logs of the accounting items, x_i. Thus, we create in each node of the first hidden layer an internal representation with the form of a ratio. Next layers use such ratios to approach the outcomes.

By using an appropriate training scheme these ratios can be set to assume a simple and interpretable form. If the overall model discovered by the MLP is optimal, the discovered ratios represent an optimal choice of combinations of variables as well.

Figure 1 is a representation of a node intended to form ratios. The logistic function

$$f(x) = \frac{1}{1 + \exp(-x - \theta)} \tag{3}$$

which is standard in Multilayer Perceptrons as a transfer function, will bring back the extended ratios from logarithmic space and will also provide a controlled amount of non-linearity for the lower values of r.

$$f(r) = \frac{1}{1 + \exp(-\log r - \theta)} = \frac{r}{r + \exp(-\theta)}$$

θ is the bias. Large negative values of θ yield a linear relation between r and the output of the node. Smaller values introduce a concavity affecting small r.

Figure 2 shows the way θ controls the output of its node. For increasing bias the node's response is linear. In general, the training of the bias is directed by the optimization algorithm so that the output is linear. Therefore, the first hidden layer is not apportioning non-linearity to the model. Non-linearity can be introduced in next layers.

Feature	Ratio	Tr.	Feature	Ratio	Tr.
Operating	NW	Log	Fixed Capital	FA/TA	Sqrt
Scale	S	Log	Intensity	$S/$ Av. FA	Log
Labour-Capital	W/TA	Sqrt	Short Term	D/CA	None
Intensity	$VA/$ Av. TCE	Sqrt	Asset Intensity	D/I	Log
Profitability	OPP/S	Sqrt	Asset Turnover	DD	None
	$EBIT/S$	Log		$S/$ Av. TA	Log
	$OPP/$ Av. TCE	Sqrt		S/I	Sqrt
	$EBIT/$ Av. TCE	Sqrt			
Net Trade	D/C	Sqrt	Financial	$DEBT/NW$	Sqrt
Credit			Leverage	$DEBT/TCE$	None

Table 1: Ratios used in the original study and their transformations.

3 A Real-World Example

In this section we apply our framework to a known accounting modelling problem, the test of the separability of the components of an industrial grouping. This test has been carried out by Sudarsanam and Taffler (1985) [10].

All companies quoted on the London Stock Exchange are classified into different industry groups according to the Stock Exchange Industrial Classification (SEIC). We selected 14 manufacturing groups according to the SEIC criteria. After discarding some firms (see below) we got accounting information on 500 cases belonging to reports from the year of 1984.

The processing usual in finance research consisted of "forming 18 financial ratios chosen as to reflect a broad range of important characteristics relating to the economic, financial and trade structure of industries (...) [10]" and extracting from them the eight principal components. These new variables were then used as inputs for a Fisher's Multiple Discriminant Analysis (MDA). A description of the modelling procedure can be found in [10]. Table 1 reproduces the original 18 ratios along with the transformations applied. DD is the ratio Debtors Days.

In our approach, 8 accounting items were used directly, not in the form of ratios. The selected items were Fixed Assets, Inventory, Debtors, Creditors, Long Term Debt, Net Worth, Wages and Operating Expenses less Wages. All these variables were present in the original 18 ratios, along with others like Earnings, Value Added, Total Capital Employed and Total Assets which we didn't use. The criteria for selecting the new variables was twofold. Firstly, they should have been present in the original set in order to allow the comparing of results. No new information was to be introduced in the problem. Secondly, the input dimension should be eight or less. The number of common factors extracted from ratios in the original study was eight. Eight items or less wouldn't allow a larger flow of information.

The choice of EX and Wages instead of Sales and Operating Profit stems from the same reasoning. The discarding of Earnings stems from not being appropriate for the log transformation. The information contained in $EBIT$ could be introduced by Sales and $COGS$ but for this particular model the residual $EBIT$ didn't seem important.

4 Improving Generalisation and Interpretability

The characteristics which make our MLP different from the standard algorithm are:

- The use of two samples, one to learn and another one to assess the classification performance.
- The training finishes before completion.

N.	Group Code	Group Name	N. Cases	Proportion
1	14	Building Materials	31	6.2%
2	32	Metallurgy	19	3.8%
3	54	Paper and Pack	46	9.2%
4	68	Chemicals	45	9.0%
5	19	Electrical	34	6.8%
6	22	Industrial Plants	17	3.4%
7	28	Machine Tools	21	4.2%
8	35	Electronics	79	15.7%
9	41	Motor Components	23	4.6%
10	59	Clothing	42	8.4%
11	61	Wool	19	3.8%
12	62	Miscellaneous Textiles	30	6.0%
13	64	Leather	16	3.2%
14	49	Food Manufacturers	80	15.9%

Table 2: Industrial groups and the proportion of each one in our sample.

- The penalization of small weights.
- Learning rates particular to each weight as described in Silva and Almeida [7].
- Likelihood maximization instead of Least-Squares minimization.

The first characteristic relates to improvements in the ability to generalise. It is a particular implementation of a known procedure, the Cross-Validation [8] [9]. The random penalization of errors is specific to this study. It allows the interpretability of results. We now comment on these features.

Two samples: In order to obtain an estimate of the generalisation capacity of a model, the original samples were divided randomly into two sub-samples of approximately equal size. All models were constructed twice, first with one half of the sample and a check carried out with the other half, and again reversing the roles of the two half data sets. Results were considered conclusive if both models, when validated with the half-sample not used to build them, produced consistent results.

The reported results concern the test set, not the training set. That is, they were obtained by measuring the rate of correct classification in the half-set not used for learning. The classification performance on the set used for learning depends solely on the number of free parameters and can be increased simply by introducing more nodes on the net.

The procedure adopted, will, with a large enough data set, produce unbiased estimates (see [3] [8]). When combined with incomplete training, it improves the generalisation of the MLP.

Incomplete training: Since the MLP seeks an optimum iteratively, we can stop its training when an optimum is obtained in the test set rather than in the training set. In doing so we prevent this powerful algorithm from over-fitting the data.

Back-Propagation seeks the modelling of progressively smaller or less important features of the relation during the learning process. Firstly, broad features are accounted for: The mean, a linear trend. Then, more detailed ones are modelled. Hence, the effective degrees of freedom the MLP engages can be viewed as increasing during learning [12].

Assuming that the topology of the net contains plenty of free parameters, the MLP will be able to model, not only the desired features but also the undesirable random uniqueness of a particular sample. We prevent it from doing this by stopping the process before finishing. The appropriate moment for stopping is when the results, as measured by the test set, are optimal. For a good topology, the fact that the learning stops before a minimum is reached in the learning set clearly enhances the generalisation.

Penalization of small weights A major goal of this study was to evaluate the power of Neural Networks in knowledge acquisition. Multi-Layer Perceptrons are often considered as not ideal in applications where self-explanatory power is required. However, in the case of accounting variables it seems possible to interpret the way the relation has been modelled by looking into the weights connecting input variables with the first hidden layer's nodes. These weights are the slopes of ratios.

In order to enhance interpretability we introduced during training a random penalization of weights with small absolute values. A weight is inhibitory when its absolute value is smaller than the unit. If the input variables were very differently scaled, inhibition values in the input weights could just mean that the MLP was trying to scale down a particular variable. Since the log items used as input to the MLP are mean-adjusted and have very similar spread the only reason for any such weights to remain smaller than the unit throughout the learning is to try to diminish the importance of one variable in the output of the node it belongs to.

In a Neural Network each node acts as a modelling unit with a certain amount of free parameters. The same output can be obtained with very different combinations of weights. Inhibition weights connecting inputs with the first hidden layer appear when the node tries to weaken the contribution of a variable. If we randomly introduce a small penalization of such weights during the training, as the correction of weights is proportional to the input variables, the weights smaller than the unit tend to remain small. In the same way, the large weights tend to have their values strengthen.

The final result is a contrasted set of weights: The first layer now contains only very large or very small weights. The information concerning the modelled relation is concentrated in a few weights instead of distributed by all of them. If the relation to be modelled is consistent with such a contrast, then there is no reason to expect that the described manipulation will damage the performance of the model.

The procedure to achieve interpretability involves these steps:

- Let one node in the first hidden layer model the size effect [11] and introduce it in subsequent layers. Input variables not convenient for the modelling of size (Debt is an example) have weights connecting to this node set to zero. The others have fixed and equal weights.

- During training, and whenever a new presentation of the entire training set is to begin, one of the remaining nodes of the first layer is randomly selected. Their weights are examined and those with inhibitory weights are penalized by a small factor, typically 0.98.

- Before the end of training, all the weights connecting inputs to the first layer and exhibiting very small values are set to zero and fixed.

This procedure is applied only after discovering the topology yielding the best results. Just by dedicating one node of the first hidden layer to the modelling of size we noticed an improvement in speed of convergence and in the final generalisation. Adding the random penalization of inhibitory weights both speed and generalisation received a further, significant, improvement. When the topology is not the best this procedure can worsen the generalisation. Other popular methods for pruning the MLP are the “Skeletonization” [5] and “Optimal Brain Damage” [4]. However, they aren't appropriate for this task: The first one is intended to reduce the number of nodes, not weights and the second one is too general.

Variable Node Number	2	3	4	5	6
Long Term Debt			-6		
Net Worth	8				
Wages	1			-6	
Inventory	8				
Debtors	2				-2
Creditors				3	
Fixed Assets	-9	-4		6	-4
Operating Expenses less Wages	-10	4	8	-2	3

Table 3: Approximate values of weights connecting input variables with nodes in the first hidden layer after training with random penalization.

5 Results:

When the training finishes the number of variables influencing each node is small and characteristic. Looking at the non-zero weights it is possible to understand, in accounting terms, what the ratios formed in each node are modelling. Table 3 shows the extended ratios formed in a net with 8 inputs, 6 nodes in one hidden layer and 14 output nodes. The emerging organization reproduces the way an expert in ratio analysis chooses variables. It is usual to build several ratios around one or two variables judged as important to capture a relation. As an example, efficiency is modelled around capital turnover, stock turnover and so on. Analysts put together several points of view around a few significant variables. Extended ratios seem to be trying the same. The item EX has been used in all hidden nodes as a contrast to the other ones.

The ratios the MLP discovers are not simple. According to table 3, our interpretation of the ratios formed in the hidden nodes is:

In the 2^{th} node: $\dfrac{NW \times I}{FA \times EX}$ In the 3^{th} node: $\dfrac{EX}{FA}$

In the 4^{th} node: $\dfrac{EX}{DB}$ In the 5^{th} node: $\dfrac{FA \times C}{W \times EX}$

In the 6^{th} node: $\dfrac{EX}{\sqrt{FA \times D}}$.

We tested the performance of such ratios when used as inputs for linear classifiers in the described problem. The five ratios plus the size variable, classify the 14 industrial groups with the same accuracy as the original 18 variables.

The gain in performance by using the MLP is, of course, much more visible. While the usual method, using 18 ratios as input variables, achieved a success of 29% (groups correctly classified), the MLP, using 8 items, classified correctly 37% of the cases. Apart from its non-linear modelling capacity — which in this particular problem didn't seem to be very important — such a gain is due to its superior generalisation. Analytic tools cannot control the relative importance of parameters during training nor stop the optimization process before its end, to avoid overfitting.

6 Discussion

So far, expectations about Neural Networks are related to the modelling of difficult relations (pattern recognition) or the mimicking of brain functions. There has been little emphasis in their potential explanatory power. Here we argue that some statistical problems requiring self-explanatory power can take advantage from the existence of meaningful internal representations.

Numerical, continuous-valued observations such as those found in stock returns, or data organized in accounting reports, cannot be efficiently used by actual expert systems as a source of knowledge. Algorithms intended to automatic extraction of rules from examples cannot perform efficiently with non-symbolic, non-hierarchical data. Neural Networks can now be seen as an alternative self-explanatory tool. In our example, hidden units were able to form more appropriate ratios than those commonly used.

The developments of this study are closer to Beaver's original works than its successors. Beaver tried to discover the most appropriate ratios to model a relation. The goal was not just an efficient modelling. It was mainly the discovering of simple tools for doing the job. After him, the practice of using multi-variate techniques and a large amount of ratios as inputs made any interpretation of results impossible.

The emphasis on interpretation should not hide the other findings of our study. The MLP proved able to outperform the classification performance of a traditional discriminant analysis approach. The MLP achieved a better performance with half the number of input variables and within a much simpler framework. Several accounting variables used to form the 18 original ratios were not present in our 8 variable set.

References

[1] W. Beaver. Financial ratios as predictors of failure. *Journal of Accounting Research, Suplement. Empirical Research in Accounting: Select Studies*, pages 71–111, 1966.

[2] K. Chen and A. Shimerda. An empirical analysis of useful financial ratios. *Financial Management*, pages 11–26, Spring 1981.

[3] R. Eubank. *Spline Smoothing and Non-Parametric Regression*. Marcel Dekker, Inc., 1988.

[4] Y. LeCun. Optimal brain damage. In *Neural Information Processing Systems*, volume 1. Morgan Kaufmann, Denver 1990.

[5] M. Mozer and P. Smolensky. Skeletonization: A technique for trimming the fat from a network via relevance assessment. In *Neural Information Processing Systems*, volume 1. Morgan Kaufmann, Denver 1988.

[6] D. Rumelhart, G. Hinton, and R. Williams. Learning internal representations by error propagation. In *Parallel Distributed Processing*, volume 1. MIT Press, 1986.

[7] F. M. Silva and L. B. Almeida. Acceleration techniques for the backpropagation algorithm. In *Neural Networks*. EURASIP Workshop, Sesimbra, Portugal, 1990. L. B. Almeida and C. J. Wellekens (Eds.), Springer-Verlag.

[8] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, B-36, 1974.

[9] M. Stone. Cross-validation: A review. *Math. Operationsforsch. Statist., Ser. Statistics*, 9(1), 1978.

[10] P. Sudarsanam and R. Taffler. Industrial classification in UK capital markets: A test of economic homogeneity. *Applied Economics*, 17:291–308, 1985.

[11] D. Trigueiros. The cross-sectional characterization of acounting data. Technical report, ISCTE, Lisbon, Portugal, 1992.

[12] A. Weigend, B. Huberman, and Rumelhart D. Predicting the future. a connectionist approach. Technical Report PARC-SSL-90-20, Stanford University, PDP Research Group, 1990.