Research Article

Design of Accounting Earnings Forecasting Model Based on Artificial Intelligence

Jia Liu,1 Shi Wu,1 and V. T. Pham1,2

1Business School, Fuzhou Institute of Technology, Fuzhou 350506, China
2Saigon University, Ho Chi Minh, Vietnam

Correspondence should be addressed to V. T. Pham; vuongt.pham@sg.edu.vn

Received 23 April 2022; Revised 18 May 2022; Accepted 23 May 2022; Published 7 June 2022

Copyright © 2022 Jia Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The increasing complexity of the international situation intensifies the changes of the economic environment. People’s demand for information represented by accounting earnings, such as judging the profitability and risk coefficient of the company, is becoming more and more urgent. This study puts forward the theory of predicting accounting earnings through accounting earnings factors in a nonlinear way and designs an accounting earnings forecasting model based on artificial intelligence. Integrating LSTM, seq2seq, and reinforcement learning and combining with self-attention like mechanism, a complex multifactor time series forecasting model is established, and reinforcement learning is used to stabilize the model to prevent overfitting, which puts forward a new solution to the multifactor time series forecasting problem of complex relationship. The experimental results and comparative analysis show the effectiveness of the enhanced recurrent neural network accounting earnings prediction model designed in this study.

1. Introduction

As a discipline that provides economic information reflecting the financial status and operating results of enterprises, accounting reflects the performance of the entrusted responsibilities of the enterprise management by providing the users of financial accounting reports with accounting information. No accounting method can get the real income of an enterprise, which is helpful for the users of financial accounting reports to make economic decisions [1–5].

Accounting surplus is the most important concept and index in accounting information. Its decision usefulness is the foundation of financial accounting and the main means to judge the value of the company [6].

In the information view of accounting, it is considered that the market is incomplete and full of uncertain factors [7]. No accounting method can get the real income of an enterprise, but accounting earnings information is a signal to investors that is helpful to judge and estimate economic income, which can improve the accuracy of investors’ prediction of the future situation of the company. The valuation view of accounting further complements the role of accounting earnings information. It is believed that investors will take the corresponding accounting data (with profit as the core) as the model change when valuing the company, so that the accounting earnings information and stock price affect each other. In other words, through the value of accounting earnings, we can infer a lot of information related to accounting earnings and use the earnings information to analyze the company’s operation, risk, future profitability, stock price change trend, etc. Therefore, the prediction of accounting earnings and the analysis of earnings information have always played a very important role in corporate management, investment, and other economic behaviors.

The research of earnings information system can be traced back to the relationship between the intensity of expected income change and stock price adjustment. From the beginning of being concerned to today’s research, many scholars have been exploring the correlation between accounting earnings and earnings information based on the company’s stock price. In these studies [1, 8–11], a large
number of theoretical and empirical studies show that the relationship between accounting earnings and stock returns (usually expressed by earnings response coefficient) changes alternately. Many studies also found that the stock price will fluctuate in the window period of accounting earnings information announcement such as annual report. In other words, earnings information will have an impact on the expected stock price in the future. At the same time, the stock price will also affect the future surplus.

In the current research, considering the weakness of the linear correlation between accounting earnings and stock return and the limitations of relevant assumptions, scholars have been seeking to establish a nonlinear correlation to uncover the secrets between accounting earnings, stock price, earnings announcement, annual report, and other factors, so as to judge the value of the company and help investors analyze and make decisions. The characteristics of this nonlinear system coincide with the nonlinear properties of neural networks [12, 13]. Moreover, accounting earnings and related data are time series data. As a neural network that can display dynamic time series and use its internal memory to process the input sequence of any time series, recurrent neural network (RNN) [14–20] can predict earnings fluctuation and reflect the stock price behavior at this stage in combination with the influence of different time series data. Reinforcement learning can intelligently solve complex problems, get rid of the constraints of the current theoretical analysis of accounting earnings value, and bring more possibilities for the research and development of earnings information.

Therefore, driven by artificial intelligence [12–20], this paper constructs an overall model of various accounting earnings value related factors, such as earnings (here refers to the specific number of earnings, which can be equivalent to profits), earnings announcement, stock price, assets and liabilities, and company cash flow, based on the neural network for time series and combined with parameter self-tuning means such as reinforcement learning. Based on earnings forecasting, an enhanced RNN earnings forecasting model is proposed, which can be nonlinear and can automatically adjust the importance of factors related to earnings value through model learning.

2. Accounting Earnings Value Forecast Model

2.1. The Basic Idea. The current two types of models (time series analysis model [21, 22] and multiple cross-sectional regression model [23, 24]) have their advantages and disadvantages. The time series analysis model can have a more stable output because it considers the dependencies that pass over time. Still, it is also difficult to obtain its data, and the model is too ideal. The multiple regression model based on cross-sectional data starts from reality, considers the correlation between earnings factors, and can achieve a more accurate output than the time series analysis model. However, the limitation of the linear model makes it impossible to include the time-dependent relationship at the same time, and the effect on multiple factors is also limited, so it is limited to instability and further development.

Due to the superiority of RNN in processing time-series data and the excellent ability of the neural network in fitting complex models, we hope to combine the advantages of the time series analysis model and multivariate cross-section regression model. The mainstream research factors of the accounting earnings system are unified into a nonlinear earnings forecasting model. The purpose is to realize an earnings forecast model that can reflect the relationship between accounting earnings factors and output more accurate accounting earnings forecast results by improving and enhancing the RNN infrastructure.

As a model that can transmit data time relationships and input multiple influencing factors simultaneously, RNN has many points to pay attention to when used as the basis of the whole model. The simplest RNN is prone to failure to converge when the factors are too complex; therefore, in the impact of multiple accounting earnings factors, next, this paper forms a preliminary model framework inspired by the multistep prediction based on seq2seq. Here, we use the seq2seq structure to improve LSTM. Because the seq2seq structure is a structure of encoder-decoder, after using seq2seq to improve LSTM, our earnings forecasting model can implement variable-length inputs. In this way, even if the value of each factor related to the value of accounting earnings that we choose is not null, it will not affect the model’s output. Another feature of this improved seq2seq structure is that it can use the joint probability of previous values to predict the next value, making the entire prediction model more stable and reflecting the relationship between earnings-related factors.

With the improvement of seq2seq [25, 26], we can incorporate many factors related to earnings value into the earnings forecast model. However, many parameters still need to be manually adjusted in the entire forecast model. From this point of view, the model will have similar problems as the multisecton regression model: too many parameters, challenging to adjust, and then affect the effect of the model. To address this shortcoming, we use reinforcement learning to improve the whole prediction model. Through reinforcement learning, we can make the whole prediction model adjust through prediction, receiving feedback and feedback. This idea realizes the self-tuning of the prediction model and avoids the problem that the prediction model is difficult to complete the training caused by too many parameter adjustments. Under the guidance of the above improvement ideas, the accounting earnings forecast model is designed as shown in Figure 1, which can represent our earnings forecast’s primary process and key steps.

2.2. Accounting Earnings Forecast Algorithm Model. In the above, we have proved the basic idea of the overall model and the main architecture and algorithms used. In this part, we will explain the structure and improvement of each module in detail and introduce the basic model used, the way of improvement, the flow of data in it, process functions etc. We start with the seq2seq structure in the model, first explain the most basic data processing framework, then
introduce the reinforcement learning algorithm based on seq2seq, and further explain the key functions in the model and the improved weighting method.

2.2.1. Improved Seq2seq Model. In the above, we summarized the types of accounting earnings factors and analyzed the relationship between different accounting earnings factors. We explained in Section 2.2 that seq2seq could use the joint probability of prior values to predict the characteristics of the following value. The commonly used seq2seq has two structures. Here, we use the second structure of seq2seq to simulate the characteristics that accounting earnings factors affect each other and act on the final forecast value. Combined with our first three-level relationship of the surplus factors analyzed in the chapter, we explained the reason for using LSTM in the seq2seq structure before, so we use Figure 2 to represent our final seq2seq structure.

The LSTM cell starts from reading the input data and the two states $c_t$ and $h_t$ of the previous cycle. $z_f$, $z_i$, and $z_o$ are all formed by the input data $x_t$ and the previous cycle $h_{t-1}$ after splicing according to different weights through the activation function $\sigma_\theta$. Calculated, these three variables are gated states, the LSTM cell starts from reading the input data $x_t$ and the two states $c_t$ and $h_t$ of the previous cycle. $z_f$, $z_i$, and $z_o$ are all formed by the input data $x_t$ and the previous cycle $h_{t-1}$ after splicing according to different weights through the activation function $\sigma_\theta$. Calculated, these three variables are gated states, but in fact, the actual input is $z$ calculated by the activation function $\phi_\theta$.

The first step of LSTM is controlling the input information by the forget gate. We discard the relatively unimportant part of the input data through this step. In this process, the LSTM cell controls the previous cell state $c_{t-1}$ by calculating $z_f$ as the gate of the forget gate removal and retention of information in $c_{t-1}$. Then, the LSTM cell processes the input, where the input gate $z_i$ will control the selective input of $x_t$. In this process, the previously calculated $z$ is performed together with $z_o$, which we express with (2)-(1). The symbol $\odot$ represents the multiplication of corresponding elements in the operation matrix.

$$c_t = z_f \odot c_{t-1} + z_i. \quad (1)$$

The output stage will determine all the outputs regarded as the current cycle. Here, $z_o$ is used as the output gate. The calculation methods of $h_t$ and $y_t$ are listed in (2) and (3). In general, $y_t$ is obtained by $h_t$ transformation.

$$h_t = z_o \odot \phi_\theta(c_t), \quad (2)$$

$$y_t = \sigma_\theta(w^v \cdot h_t). \quad (3)$$

After processing the encoding part, the encoding vector enters the decoder for actual prediction, and the value of the decoder part is calculated according to the process in Figure 2. Assuming that the encoder part obtains the final hidden layer state value is $h_{t_{\text{en}}}$, the decoder at the time of $t+1$. The state value of the part is calculated by (4), and the predicted value $\hat{y}_{t+1}$ at this moment is calculated by (5).

$$s_{t+1} = \phi_\theta(y_{t_{\text{en}}}^v, h_{t_{\text{en}}}^h, c_t). \quad (4)$$

$$\hat{y}_{t+1} \sim \pi_\theta(y_{t+1} | \hat{y}_t, s_{t+1}). \quad (5)$$

If we make the actual value to be predicted $y$, then the loss function at this time can be expressed by (6), and $X$ represents the input data sequence.
\[ L_\theta = -\sum_{t=1}^{T} \log \pi_\theta(y_t | y_{t-1}, s_t, c_{t-1}, X). \] (6)

Considering that the surplus data is negative or positive and the characteristics of LSTM itself, we choose the sigmoid function as the \( \sigma \) activation function (7), the tanh function as the \( \varphi_\theta \) function (8), and the SoftMax function as the \( \pi_\theta \) function (9).

\[
f(x) = \frac{1}{1 + e^{-x}}. \tag{7}
\]

\[
f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \tag{8}
\]

\[
f(x) = \frac{e^x}{\sum e^x}. \tag{9}
\]

If the accounting earnings factor is put into the model, we use Figure 3 to represent the input and relationship composition of the earnings factor in seq2seq.

The encoding vector is composed of the three output vectors of the encoder part. The importance of these three types of accounting factors is different, and the three types of accounting factors can be roughly determined through existing model research. Therefore, in order not to further complicate the model, we only draw on the idea of attention mechanism (10), directly weight these three vectors, and then obtain the input of the decoder part.

\[
\text{Attention score(query, source)} = \sum_{i=1}^{T} \text{similarity}^{imery}(ey_{yi}) \cdot \text{value}_i.
\] (10)

However, under the seq2seq model at this time, the optimization problem of the loss function is still not solved. Under such conditions, it is difficult to make the model have the best optimization strategy, making the instability produced by the model indistinguishable from the multiple regression cross section models. Therefore, we use reinforcement learning to improve this problem of the seq2seq model, hoping to make the prediction results more accurate.

2.2.2. Reinforcement Learning Algorithm Integrating Seq2seq.

The earnings forecast model to be established in this study is based on a long time series. During this process, the relevant data are constantly changing, and according to the Introduction section, we know that there is also a correlation between the accounting earnings data of the same year, that is, the model we need is not a mapping of inputs to outputs, but a pattern between earnings-related information. However, pure LSTM not only has independent and identically distributed input but also cannot learn this “pattern.” We have constructed the relationship dependence between surplus factors through the seq2seq structure, so we continue to use reinforcement learning to make seq2seq balance; then, after exploration and development, choose the most rewarding and most effective behavioral mode.

Because the final model is a value prediction model, we need to choose a type of reinforcement learning algorithm suitable for value analysis. We mentioned in Section 1 that the most typical reinforcement learning algorithm for value prediction is Q-learning. Therefore, we integrate the algorithm idea of Q-learning with the seq2seq model constructed in Section 2.2.1 to form the overall reinforcement learning algorithm model shown in Figure 4.

In this process, our purpose is to maximize the expectation of reward through the interaction between the agent and the environment and action. We use (11) to solve the reinforcement learning that integrates seq2seq, which is the same as Q-learning; \( Q_\pi \) represents the utility function under policy.

\[
\max E_{\tilde{y}_1, \ldots, \tilde{y}_T} \rho_\pi(\tilde{y}_1, \ldots, \tilde{y}_T) [r(\tilde{y}_1, \ldots, \tilde{y}_T)], \tag{11}
\]

\[
L_\theta = -E_{\tilde{y}_1, \ldots, y_T} \rho_\pi(\tilde{y}_1, \ldots, y_T) [r(\tilde{y}_1, \ldots, y_T)]. \tag{12}
\]

At this time, the loss function can be calculated by (12); then by (13), we can obtain the value of the corresponding maximum partial derivative, we make the output of the decoder part before SoftMax is \( O_t \), and then the partial derivative can be rewritten as (14) and solved by (15). \( R_b \) is a baseline reward value that does not depend on the seq2seq part.
∇\theta L_\theta = -\mathbb{E}_{\tilde{y}_{t:s}} \nabla_\theta \log \pi_\theta (\tilde{y}_{1:T}) r(\tilde{y}_{1:T}), \quad (13)

\nabla_\theta L_\theta = \frac{\partial L_\theta}{\partial \theta} = \sum_t \frac{\partial L_\theta}{\partial \theta_t} \frac{\partial \theta_t}{\partial \theta}, \quad (14)

\frac{\partial L_\theta}{\partial \theta_t} = (\pi_\theta (y_1; \tilde{y}_{t-1}, s_t, c_{t-1}) - y_t) (r (\tilde{y}_1, \ldots, \tilde{y}_T) - r_\theta), \quad (15)

3. Model Implementation and Experimental Analysis

Due to the large number of companies involved in the forecast results and the vast difference between the forecast results, in Figure 5, the horizontal axis is the company code, the vertical axis is the unit length, and the green polyline represents the 2018 earnings of Shanghai A shares predicted by the enhanced recurrent neural network earnings forecast model. The red polyline is the actual earnings of these companies in 2018. Figure 5 includes 1512 forecast values, and the vertical axis value of other nonexistent company codes is 0. It is impossible to judge the quality of the specific forecast situation only by the particular forecast value. Therefore, according to the calculation method in Section 2, we calculate the metric value of the forecasted earnings in 2018 by the designed model.

Table 1 also shows the same index values calculated from the prediction results of the basic LSTM model on the same data set (because no accounting analysis is involved here, the R2, ERC, and ICC indicators of the designed model are not compared with LSTM). Due to the large amount of data involved in the experiment and carried out on a personal host, the running time of LSTM and the designed model is more than one day, but considering that the accounting earnings forecast itself is in units of years, the time required for the experiment has little impact, far less than the accuracy and other evaluation criteria, so the evaluation of the results by the time of the experiment is not considered here.

From Table 1, only through the comparison of AE, can we find that the designed model significantly improves prediction accuracy compared to the LSTM base model. At the same time, it is not difficult to find that the results predicted by the designed model are still optimistic in general (i.e., the overall forecast of earnings will be larger than the actual value). Compared with the LSTM model, the difference between AE and AAE is more significant. To a certain extent, it reflects that the designed model has a higher sensitivity when the company generates negative earnings. But at the same time, the stability (DS) of the model is not as good as that of the LSTM model, but the DS value of the designed model is still within an acceptable range.
At the same time, we compare the DE values of the prediction results of the designed model and LSTM, and it is easy to find that the designed model fits the curve of the actual value more than LSTM. The comparison between DE and ADE further shows that the overall prediction trend of the LSTM model is far more optimistic than the designed model. It also makes the accuracy of the designed model higher than that of LSTM in the case of a downward trend in earnings.

In addition to comparing the prediction accuracy with the LSTM model, this experiment further compared with the HVZ model [27] and further analyzed the prediction results of the designed model in the accounting sense. The HVZ model needs long-term data to be computationally meaningful. In the data collected in this experiment, the amount of data that meets the requirements of the HVZ model is insufficient. Therefore, we use the empirical data of Li and Mohanram [28] to compare with the designed model (Table 2), and we mainly compare the accuracy of the forecast results and whether they have accounting significance.

From Table 2, we can find that the prediction result of the designed model is more accurate than that of HVZ. When using data with a period of up to 40 years, the D$S$ value of the prediction result of the HVZ model can even be reduced to less than 10%. But overall, the designed model is more accurate than HVZ in more general cases. The fit of the HVZ model is slightly better than the designed model, but considering the limited regression parameters used by the HVZ model, this does not mean that the designed model has a worse fit. In addition, the ERC value of the designed model is slightly higher than that of HVZ. The designed model is relatively more representative of market expectations, which may be because the designed model uses the basic structure of LSTM to transmit the time relationship in the accounting earnings correlation system or because the input data includes the impact of the announcement on the market. However, the ICC value of HVZ is higher than that of the designed model, indicating that the HVZ model has a higher correlation with actual returns. It also shows that there is room for further optimization of the input data of the designed model, which needs to be further adjusted according to accounting factors.

### 4. Conclusions

An accounting earnings forecasting model is designed based on artificial intelligence. From the perspective of financial accounting, more accurate judgment of the designed model on the company’s profit trend itself is helpful for investment analysis and company decision-making. At the same time, it is difficult for traditional models to judge the sudden negative turn of company profits under volatile market conditions. The designed model can find out whether the company has profitability faster than the general model. It is easy to find out through the two common indicators of ICC and ERC. The designed model has a more sensitive response to the market and is more consistent with the company’s actual income. The forecast results are more conducive to analyzing the company’s problems in the face of adversity and can even further infer industry risks.

In general, the designed model has higher accuracy in a broader sense when forecasting earnings. It relaxes the data requirements for forecasting and the high standards for the company’s accounting years. On this premise, it can further improve the accuracy of earnings forecasts and maintain high stability. This also shows that different companies have similar accounting earnings correlation systems, and this accounting earnings correlation system can be used to build earnings forecast models for various companies. At the same time, the earnings predicted by the designed model have accounting significance, and the forecast results can reflect the market’s expectations and reflect the correlation between the forecast results and actual earnings. The experimental results also show that the selection of accounting earnings-related factors has a particular impact on the results of earnings forecasting.

### Data Availability

The data can be made available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work was supported by the 2019 Fujian Young and Middle-Aged Teacher Education Research Project, under project name: Research on Dual Line Collaborative Evolution and Innovation under the Background of “New Retail” (project no. JAT191027).

### References


