Time series prediction by using Convolutional Neural Networks

Ronny Velastegui[1][0000−0001−8628−9930],
Luis Zhinin-Vera[1,2][0000−0002−6505−614X], Gissela E. Pilliza[1][0000−0001−6386−9254]
and Oscar Chang[1][0000−0002−4336−7545]

[1] School of Mathematical and Computational Sciences, Yachay Tech University, 100650, Urcuqui, Ecuador.
[2] Model Intelligent Networks Development - MIND Research Group
mind-researchgroup.com.
ronny.1908@hotmail.com

Abstract. All companies need an effective method to predict future sales, and several classic statistical methods exist and are heavily used in the industry. This work proposes a novel sales prediction method based on Convolutional Neural Networks. This type of neural network is generally used for image processing tasks. But in this work, we explore new applications and develop models that produce good results in sales prediction for real pharmaceutical product data. Also, we implemented several classical and statistical prediction methods, and we compared them with our proposed model. For this, we used three comparison metrics: prediction accuracy, number of weights, and number of iterations. Finally, we proceeded to determine which prediction method is better both in accuracy and efficiency terms.

Keywords: Sales Forecast · Artificial Neural Networks · Convolutional Neural Networks · Deep Learning.

1 Introduction

Sales forecasts are a fundamental part of any company, since, these companies can plan and make important decisions that determine the future success or failure of the company [8]. This is because a good prediction will allow us to correctly decide important actions such as purchase of raw material, increase of personnel, preparation of advertising, acquisition of equipment, among others. While there are several methods to predict sales, used by companies today, it is not always possible to obtain a completely accurate result, however, a good approximation can make a difference in the future of the company. The methods of sales prediction are classified mainly into two large groups, qualitative methods, and quantitative methods [12]. The difference between both methods lies in the presence of past data. On the one hand, qualitative methods are used to predict sales when there is no previous data. On the other hand, quantitative methods are used when a certain amount of past data is available and we assume that there is a possible trend in them. Depending on the situation, each method may work better than the other.

Quantitative methods are subdivided into classical statistical methods and methods based on artificial intelligence. Some examples of classical statistical methods are the Naive approach, the Average approach, and the Moving average approach. The Naive approach is the simplest prediction method of all, but not necessarily the most accurate. The other two mentioned methods use averages to do the predictions, but, like the naive method, both accuracies obtained are very low. The advantage of these methods is that they are very easy to understand and implement, that is why they are still used in some companies [3].

While it is true that classical statistical methods such as the Naive approach, Average approach, Moving average approach, among others, have been widely used by companies for years. Currently, companies are opting to use more modern methods based on artificial intelligence [2]. One of these methods is artificial neural networks. Thus, the Feedforward Neural Networks and Recurrent Neural Networks have demonstrated good precision when facing time series prediction problems. But, it is important to mention that, when trying to make very complex predictions, where the data does not show a very clear trend, these methods are not very effective. This is because to predict complex time series, you usually need to use a large number of hidden layers and neurons, this confronts the well-known Vanishing Gradient problem, which affects deep networks [6].

It is for this reason that, in this work, it is proposed to develop a novel sales prediction method based on Convolutional Neural Networks. While this type of neural network is generally used for image processing tasks [7], in this work we develop new applications in sales prediction. For this, we will first choose a sales database corresponding to an Ecuadorian Pharmacy Franchise. This database contains weekly sales of products for 4 years. Once the data set is defined, several classical prediction methods and also prediction methods based on artificial intelligence were implemented. Then, our novel method based on convolutional neural networks was proposed and implemented. After this, all these prediction methods were compared using three metrics: prediction accuracy, number of weights and number of iterations. Thus, we proceeded to determine which prediction method is better both in terms of accuracy and efficiency, also, the advantages and disadvantages of each prediction method were determined.

2 Related Works

In this section, we will summarize the most important aspects of some related works previously developed. Table 1 shows a brief summary of related works.

Table 1. Summary table of related works

Name of work	Author	Year	Proposed Prediction Architecture	Approach
"Sales forecasting using neural networks"	Thesing and Vornberger	1997	Feed-forward multilayer perceptron (Shallow-MLP)	Regression
"Forecasting with artificial neural networks: The state of the art"	Zhang, Patuwo and Hu	1998	Feed-forward multilayer perceptron (Shallow-MLP)	Classification, Regression
"An artificial neural network (p, d, q) model for timeseries forecasting"	Khashei and Bijari	2010	ARIMA + ANN	Regression
"Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations"	Jie Wang and Jun Wang	2016	Elman Recurrent Neural Network (ERNN)	Regression
"A deep learning algorithm to forecast sales of pharmaceutical products"	Chang, Naranjo and Guerron	2017	Deep MLP + Autoencoder	Regression
"Forecasting stock prices from the limit order book using convolutional neural networks"	Tsantekidis, Passalis, Tefas, Kanniainen, Gabbouj and Iosifidis	2017	Convolutional Neural Network (CNN)	Classification
"Sales demand forecast in e-commerce using a long short-term memory neural network methodology"	Bandara, Shi, Bergmeir, Hewamalage, Tran and Seaman	2019	Long-Short Term Memory (LSTM)	Regression

3 Technical Background

3.1 Time Series Prediction

The understanding of the past is the key to predict future [11]. So, time series prediction or time series forecasting is the process of analyzing data of a time series, and propose a model to predict the future of that time series [4].

3.2 Time Series Prediction Methods

Naïve Approach. The Naïve Approach is the simplest method to forecast time series. This method consists on predict the next future value of the time series using only the last value observed [5]. So, this method uses the following equation:

$$\hat{y}_{t+1} = y_t \tag{1}$$

y_t: Time series value at time t
$\hat{y}_{x+1}$: Predicted time series value at time $t+1$

As we can see in the equation, this method estimates that the value to be predicted is equal to the last value observed. As we know, time series rarely follow that simple pattern [3], for this reason, this method is not so accurate when making predictions.

Simple Average Approach. This method uses all past data of the time series to do the prediction. So, this method uses the following equation:

$$\hat{y}_{x+1} = \frac{1}{x}\sum_{i=1}^{x} y_i \tag{2}$$

y_i: Time series value at time i
x: Number of total time steps
$\hat{y}_{x+1}$: Predicted time series value at time $x+1$

As we can see in the equation, this method estimates the future value using the total average of all past data of time series [3].
Due that this method uses more information to make the prediction, it may seem that this method is always better than the previous one, but this is not always true. This method works best when the time series has a constant average, otherwise, its accuracy may be worse than the Naive Method.

Moving Average Approach. This method is an improvement to the Simple Average Approach. So this method does not necessarily use all the observations of the time series, instead, it only uses a subset of it [8]. This method uses the following equation:

$$\hat{y}_t = \frac{1}{p}(y_{t-1} + y_{t-2} + y_{t-3}... + y_{t-p}) \tag{3}$$

$\hat{y}_t$: Predicted time series value at time t
y_{t-1}: Time series value at time $t-1$
p: Window size of time steps

As we can see in the equation, this method estimates the future value using the average of the last p observations of time series. Note that p is a fixed parameter that we must adjust.

Simple Exponential Smoothing. As we saw in the previous methods, when we make time series predictions, the use of the latest time series data leads to better prediction accuracy. Instead, the use of very old time series data does not contribute significantly to the improvement of accuracy [2].
Based on this idea, the Simple Exponential Smoothing method uses all observations of the time series to make a forecast, but the contribution of each observation decreases exponentially as we move into the past.

$$\hat{y}_{t+1} = \alpha * y_t + (1 - \alpha) * \hat{y}_t \tag{4}$$

y_t: Time series value at time t
$\hat{y}_{x+1}$: Predicted time series value at time $t + 1$
α: Smoothing factor, $0 < \alpha < 1$

4 Methodology

In this section, all the procedures followed to reach the final results will be shown. First, the database, data normalization and data input format used will be detailed. Second, we will explain architectures configurations of the four prediction methods to use: Shallow MLP, Deep MLP, CNN. This section also includes a detailed explanation of all the setup of the network's parameters. Third, metrics and formulas used to perform the comparison of the prediction methods implemented are described in detail.

4.1 Dataset

The data set used in this project for training and testing of the ANN predictions models was obtained from an Ecuadorian Pharmacy Industry. So, for this project, we use sales databases of the pharmacy chain Farmaenlace. So, this data set is composed of weekly sales of 100 different products over a period of 5 years. Each time series of each product information are stored in one *.txt* file and includes the name of the product, code of product, and a total of 200 weekly sales.

So, based on the work [1], the distribution of the data set that we use is the following: 75% for training and 25% for test. That means that we will predict last year of sales based on the previous four years approximately.

4.2 ANN Prediction Models

In this section, the three prediction time series architecture used in this project will be defined. First, we will explain 2 common neural network architectures used actually in sales forecasting, these are Shallow MLP and Deep MLP. After that, we will see in detail the proposed CNN architecture for sale forecasting. So, for each of these three architectures, we propose some different models.

Also, is important to mention that all the prediction models used in this work are Multi-step ahead prediction methods. That means that, during the test phase, each model predicts future sales using the previously predicted values.

Shallow Multi-Layer Perceptron Architecture. The first forecasting architecture that we will use is the Shallow Multi-Layer Perceptron (Shallow - MLP). We started with this architecture because it is one of the simplest and easiest to understand ANNs. Due that MLP is a universal function approximator, we can consider the sales time series like functions where time correspond to X-Axis and sales correspond to Y-Axis. In this way, we can address this problem as a regression problem. So, the objective is to forecast the weekly future sales, based on n lasts weeks' sales.

The principal structure of this shallow architecture is based on [9]. From this previous investigation, we extracted the most important parameters of the neural network such as the number of neurons, learning coefficient, loss function, etc. While a few other parameters were calculated experimentally for our purposes. Table 2 shows a brief summary of this model.

Table 2. Shallow MLP Model 1 parameters

Input layer size (n)	16
Hidden layer size	10
Output layer size	1
Time series size	200
Percentage of training data	75%
Percentage of test data	25%
Initial learning rate	0.005
Learning rate drop factor	0.1
Learning rate drop period	20
Transfer function	RELU
Optimization algorithm	SGDM, Backpropagation
Mini-batch size	32
Momentum	0.9
Loss function	Half-mean-squared-error
Max. Epochs number	125
Iterations per epoch	4
Max. Iteration	500
Regularization technique	Dropout 0.5

Deep Multi-Layer Perceptron Architecture The second approach of forecasting architecture that we will use is the Deep Multi-Layer Perceptron (Deep MLP). This architecture works very similar to the previous method; the only difference is the number of hidden layers. Shallow MLP only has one hidden layer, but if the number of hidden layers is more than 1, then the architecture is considered a Deep MLP. So, when we add hidden layers, the ANN acquires more computational power and abstraction ability. Though, the adding of many hidden layers can be counterproductive and reduce the accuracy of the prediction model. This is because when we add many hidden layers in a MLP architecture, it is possible to occur vanishing gradient problem.

The principal structure of this deep architecture is based on [1]. From this previous investigation, we extracted the most important parameters of the neural network such as the number of layers, number of neurons of each layer, learning coefficient, loss function, etc. While a few other parameters were calculated experimentally for our purposes. Table 3 shows a brief summary of this model.

Table 3. Deep MLP Model 1 parameters

Input layer size (n)	16
First hidden layer size	10
Second hidden layer size	10
Output layer size	1
Time series size	200
Percentage of training data	75%
Percentage of test data	25%
Initial learning rate	0.005
Learning rate drop factor	0.1
Learning rate drop period	20
Transfer function	RELU
Optimization algorithm	SGDM, Backpropagation
Mini-batch size	32
Momentum	0.9
Loss function	Half-mean-squared-error
Max. Epochs number	125
Iterations per epoch	4
Max. Iteration	500
Regularization technique	Dropout 0.5

Convolutional Neural Network Architecture Finally, this section details the sales prediction architecture based on CNN. This architecture will work similar to the Deep-MLP; the only difference is the addition of convolutional layers. These convolutional layers replace the fully connected layers, and therefore reducing the number of weights in the architecture [10]. We will propose different configurations of this architecture, in each configuration we will change: number of convolutional layers, number of filters, number of neurons and number of pooling layers. Thus, after performing the training and the test of each configuration, we will choose the best one in terms of precision and efficiency. Finally, these CNN architecture models will then be compared against the other models based on Shallow MLP and Deep MLP.

The principal structure of this CNN architecture is based on [10]. From this previous investigation, we extracted the most important parameters of the neural network such as the filter size, filter number, number of layers, number of neurons of each layer, learning coefficient, loss function, etc. While a few other parameters were calculated experimentally for our purposes. Table 4 shows the parameters of this architecture and Figure 1 shows a representation model.

Table 4. CNN Model 1 parameters

Input layer size (n)	16
1D Convolutional Layer 1	Number of filters: 4, Stride = 1, Filter size: [7 x 1], Padding: Same
1D Max Pooling Layer 1	Pooling filter size: [2 x 1], Stride = 2
1D Convolutional Layer 2	Number of filters: 8, Stride = 1, Filter size: [5 x 1], Padding: Same
1D Max Pooling Layer 2	Pooling filter size: [2 x 1], Stride = 2
1D Convolutional Layer 3	Number of filters: 16, Stride = 1, Filter size: [3 x 1], Padding: Same
1D Max Pooling Layer 3	Pooling filter size: [2 x 1], Stride = 2
1D Convolutional Layer 4	Number of filters: 32, Stride = 1, Filter size: [1 x 1], Padding: Same
1D Max Pooling Layer 4	Pooling filter size: [2 x 1], Stride = 2
Dropout	Dropout factor = 0.2
Fully Connected Layer	10
Output layer size	1
Time series size	200
Percentage of training data	75%
Percentage of test data	25%
Initial learning rate	0.005
Learning rate drop factor	0.1
Learning rate drop period	20
Transfer function	RELU

5 Results

The objective of this section is to describe the advantages and disadvantages of each prediction model. For this, we will analyze and compare the models based on the metrics: accuracy, number of weights and iterations. In this way, we can choose the model that achieves the highest prediction accuracy using the lowest number of weights in its structure.

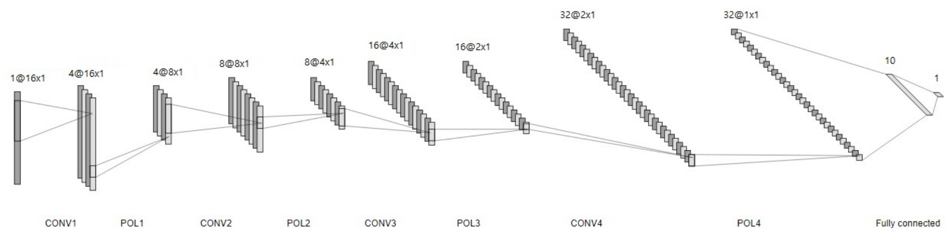

Fig. 1. CNN - Model 1

5.1 Comparison of the models

Once we have chosen the best models of the three main architectures, then we will proceed to compare them. For this, we will use three important metrics: prediction accuracy, number of weights and iterations. Also, we will compare these prediction methods based on neural networks, with the classical quantitative prediction methods: Naïve, Average, and Moving Average.

Table 5. Comparison of prediction models using Time Series 1

Time Series 1				
	Number of weights	**Train RMSE**	**Test RMSE**	**Iterations**
Naïve method	-	-	45.52	-
Average method	-	-	47.58	-
Moving average method	-	-	33.39	-
Shallow-MLP Model 3	17000	14.48	11.42	200
Deep-MLP Model 4	570	19.37	15.41	250
CNN Model 1	478	16.58	14.11	200
CNN Model 2	313	18.23	15.36	200

In Table 5, the prediction results of the time series 1 are shown. First, we can see that the three classical quantitative prediction methods reach the lowest prediction accuracy. So, the Naïve prediction method reaches a prediction accuracy Test RMSE of 45.52. The Average method reaches an accuracy of 47.58. Finally, the Moving average method reaches an accuracy of 33.39.
Regarding prediction models based on ANN, we can see that the model with the best prediction accuracy is the Shallow-MLP Model 3. This model reaches a Test RMSE of 11.42. But, the problem with this model is that to reach that high accuracy, the model needs to adjust 17000 weights. That means a high computational cost during training compared to the other models.
Now, the prediction model based on ANN with the lowest prediction precision is the Deep-MLP Model 4. This model reaches a Test RMSE of 15.41. But, the

main difference of this model is that it uses less weights compared to Shallow-MLP Model 3. The CNN Model 1 is more accurate than Deep-MLP Model 4, but less accurate than Shallow-MLP Model 3. However, the biggest advantage of this model is that it uses a smaller amount of weights than the other two models, so it needs to adjust 478 weights during training. Finally, CNN Model 2 is the model with the smallest amount of weights of all other models. While this model is not as accurate as the CNN Model 1, this model only uses 313 weights. This means a great reduction of computational cost at the moment of adjusting weights during training.

Table 6. Comparison of prediction models using Time Series 2

Time Series 2				
	Number of weights	**Train RMSE**	**Test RMSE**	**Iterations**
Naïve method	-	-	5.801	-
Average method	-	-	5.780	-
Moving average method	-	-	5.703	-
Shallow-MLP Model 3	17000	0.546	4.295	200
Deep-MLP Model 4	570	1.461	4.709	250
CNN Model 1	478	1.250	3.640	200
CNN Model 2	313	1.280	3.840	200

In Table 6, the prediction results of the time series 2 are shown. First, we can see that the three classical quantitative prediction methods reach the lowest prediction accuracy. So, the Naïve prediction method reaches a prediction accuracy Test RMSE of 5.801. The Average method reaches an accuracy of 5.780. Finally, the Moving average method reaches an accuracy of 5.703.

Unlike the results obtained in time series 1, in this case, Shallow-MLP Model 4 is not the most accurate model. So, despite that it has a huge number of weights and neurons to process, this model only reaches a Test RMSE of 4.209. Also, the prediction model based on ANN with the lowest prediction precision is the Deep-MLP Model 4. This model reaches a Test RMSE of 4.709. But, the main difference of this model is that it uses fewer weights compared to Shallow-MLP Model 3. The CNN Model 1 is the most accurate prediction model for time series 2. This model reaches a Test RMSE of 3.640. Also, the biggest advantage of this model is that it uses a smaller amount of weights than the other two models, so it needs to adjust 478 weights during training. Finally, CNN Model 2 is the model with the smallest amount of weights of all other models. While this model is not as accurate as the CNN Model 1, this model only uses 313 weights. This means a great reduction of computational cost at the moment of adjusting weights during training.

In table 7, the prediction results of the time series 3 are shown. As we already know, due to the irregular and random behavior of this time series, the prediction models used in this project reach a very low precision. Regardless of this, we will

Table 7. Comparison of prediction models using Time Series 3

Time Series 3				
	Number of weights	**Train RMSE**	**Test RMSE**	**Iterations**
Naïve method	-	-	25.21	-
Average method	-	-	24.18	-
Moving average method	-	-	24.58	-
Shallow-MLP Model 3	17000	2.29	23.48	200
Deep-MLP Model 4	570	5.94	21.18	250
CNN Model 1	478	4.36	19.97	200
CNN Model 2	313	4.77	20.22	250

analyze the performance of these models anyway using the comparison metrics already mentioned.

First, we can see that the three classical quantitative prediction methods reach the lowest prediction accuracy. So, the Naïve prediction method reaches a prediction accuracy Test RMSE of 25.21. The Average method reaches an accuracy of 24.18. Finally, the Moving average method reaches an accuracy of 24.58.

Regarding prediction models based on ANN, the model with the best prediction accuracy is the CNN Model 1. This model reaches a Test RMSE of 19.97, and to reach that accuracy, the model needs to adjust 478 weights. The model with the second best prediction accuracy is CNN Model 2. This model is not so accurate like the previous model but uses less weights because the model needs to adjust 313 weights during training. Finally, the models with the worst prediction precisions are Deep-MLP and Shallow MLP based models, these models reach a Test RMSE of 21.18 and 23.48 respectively.
As we can see, in this time series example, the models that reach high accuracy are CNN Based Models. So, despite Shallow-MLP and Deep MLP based models have many weights, they do not exceed the prediction accuracy of CNN Based models.

6 Conclusion

In this paper, a CNN architecture, usually dedicated to image processing, is used to do a novel task: time series prediction. Our specific objective was to predict weekly sales of an Ecuadorian pharmacy franchise by using CNN. To analyze the effectiveness of this CNN Architecture, two different prediction architectures were also implemented: Shallow-MLP and Deep-MLP. The Shallow-MLP prediction method is formed by an artificial neural network of only one hidden layer. Deep-MLP prediction method is formed by an artificial neural network of more than one hidden layer. Also, three classical quantitative prediction methods are implemented: Naïve, Average, and Moving Average, to compare them with the proposed model. For experimental purposes, three time series examples were used to do predictions. The first time series example represents a case in which the prediction methods reach high accuracy, the second represents a case in which the prediction methods reach average accuracy and third represents a case in which the prediction methods reach low accuracy.

The next step was generating the prediction of these time series, using each one of these prediction models. After that, we select the most accurate model of each one of the three architectures. Finally, we compared the prediction results between previously selected models.

We showed that three classical quantitative prediction methods reach low accuracy compared with the ANN-based prediction methods. So, we showed that Shallow MLP and Deep MLP prediction architectures reach high prediction accuracy.

Then, we proved that the CNN Based Prediction Models reach a high accuracy using less weights amount in their structure. This means that the training phase of CNN Prediction methods has a very low computational cost, compared to the other two prediction methods. This fact is important when quick predictions for thousands of products are required, as in a typical pharmacy franchise.

In conclusion, in this work, a novel and effective approach for time series prediction was presented, producing successfully prediction sales for real pharmaceutical products.

The future scope is extend this approach to different data sets, and not just limit to pharmaceutical data sales. In this way, this CNN prediction method can be used in more fields. So, this method can be used to predict: daily temperature of a city, the daily number of dead in a specific country, etc. So, we could study the behavior of this prediction method in other different time series.

References

1. Chang, O., Naranjo, I., Guerron, C.: A deep learning algorithm to forecast sales of pharmaceutical products (10 2017)
2. Chatfield, C.: The analysis of time series: an introduction. Chapman and Hall/CRC (2003)
3. Cryer, J.D., Kellet, N.: Time series analysis. Springer (1991)
4. Farmer, J.D., Sidorowich, J.J.: Predicting chaotic time series. Physical review letters **59**(8), 845 (1987)
5. Hamilton, J.D.: Time series analysis, vol. 2. Princeton university press Princeton, NJ (1994)
6. Hassoun, M.H., et al.: Fundamentals of artificial neural networks. MIT press (1995)
7. Jain, A.K., Mao, J., Mohiuddin, K.M.: Artificial neural networks: A tutorial. Computer **29**(3), 31–44 (1996)
8. McCleary, R., Hay, R.A., Meidinger, E.E., McDowall, D.: Applied time series analysis for the social sciences. Sage Publications Beverly Hills, CA (1980)
9. Thiesing, F.M., Vornberger, O.: Sales forecasting using neural networks. In: Proceedings of International Conference on Neural Networks (ICNN'97). vol. 4, pp. 2125–2128. IEEE (1997)
10. Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., Iosifidis, A.: Forecasting stock prices from the limit order book using convolutional neural networks. In: 2017 IEEE 19th Conference on Business Informatics (CBI). vol. 1, pp. 7–12. IEEE (2017)
11. Tsay, R.S.: Financial time series. Wiley StatsRef: Statistics Reference Online pp. 1–23 (2014)
12. Wei, W.W.: Time series analysis. In: The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2 (2006)