Machine Learning Based Future Priced Prediction of the Products

Ishan Pandey[1], Mudimela Sathish Kumar Reddy[2], Mukkamalla Dharani Vamsidhar Reddy[3], Karumoju Krishna Babu[4], Harshpreet Kaur[5]

[1-5]School of Computer Science and Engineering Lovely Professional University Phagwara, Punjab, India

---

Abstract: *Price prediction is a useful feature for consumers as well as businesses. A price prediction tool motivates users to engage with a brand or evaluate offers in order to spend their money wisely. Price prediction enables businesses to set pricing in a manner that builds customer engagement and loyalty. With Machine Learning (ML) technology a price prediction problem is formulated as a regression analysis which is a statistical technique used to estimate the relationship between a dependent/target variable and single or multiple independent (interdependent) variables. In regression, the target variable is numeric. This project will focus on ML algorithm used for price prediction.*

1. Introduction

With the rapid development of the Internet and the rise of e- commerce, more and more data are generated on e-commerce websites. In the face of such huge data, reasonable utilization will produce great value. For example, the development trend of Tmall platform can be estimated through the full amount of commodities calculated on Tmall platform. For the commodities of a certain industry and company, by calculating its GMV (Gross Merchandise Volume) at a specific time, and then doing year-on-year and sequential analysis, the development trend of the industry and company can be analyzed, which can be used as important references for Chinese concept stocks.

In recent years, time series prediction based on neural networks has been popular. Unlike traditional models, deep neural networks have several distinct advantages as non-parametric, self-learning, non- assumption and noise tolerant, which are unavailable in traditional models (Haykin & Network, 2004; Hochreiter & Schmidhuber, 1997; Sharda & Patil, 1992). Therefore, deep neural networks might be more effective in forecasting agricultural commodity futures price in comparison to traditional models (Kaastra & Boyd, 1995; Zhang & Hu, 1998).

1.1 Existing System

In general, many researchers at home and abroad have proposed many prediction algorithms with good analysis effect and broad application scenarios in the field of price prediction in recent years, but there are still some problems, such as how to take the most effective way to generate the final effective data set for data processing [13]. Due to the different methods and models used, the accuracy of the price measurement is not the same

The prediction of commodity prices has a long history. People make predictions based on experience at the earliest, which has certain subjectivity and limitations. Subsequently, domestic and foreign experts gradually began to simply sort the price data according to time, so as to establish a model to analyze the change trend of the price in the future [8]. With the rapid development of machine learning and big data in recent years, researchers have proposed many effective commodity price prediction algorithms. YE Lu and LI Yuping[10] used BP neural network combination algorithm based on particle

1.2 Proposed System

This paper introduces the calculation of correlation coefficient on the basis of RNN algorithm, so that the algorithm can consider the weight of each data dimension and eliminate the influence of R value to a certain extent by statistical method. On the basis of the improved RNN algorithm to classify the range of commodity prices, the prediction value of commodity prices is given by using decision tree regression algorithm. The accuracy of this result is better than the direct application of decision tree regression

2. Module Description

2.1 Data Collection

Through the construction of scrappy framework, the second- hand notebook sales data on did e-commerce platform are collected, including description information, price, store, evaluation number, flash purchase, free mail, coupons, gifts, freight insurance and other information. There were 6,000 pieces of data collected.

2.2 Improvement of Parameter k value Selection (IPKS)

The traditional selection of k value can only be based on experience. In order to eliminate this effect, this paper will delete the large value with abnormal distance. In statistics, values whose mean value is more than 3 standard deviations away from the mean value are defined as outliers. This method is adopted in this paper to remove values with abnormally large distances.

2.3 Classification Module

The idea of KNN algorithm is: if most of the k most similar samples of a sample in the feature space belong to a certain category, then the sample also belongs to this category. In KNN algorithm, the selected neighbors are all objects that have been correctly classified. K neighborhood algorithm has the characteristics of stability, simplicity and efficiency, but in the analysis of a large number of data sets, the algorithm has the defects of K value and size dimension. In order to overcome these two defects, the following improvements are made on the basis of this algorithm:

2.4 Regression Module/Prediction

According to the above process, after the improvement of KNN algorithm, the pre-processing data set will produce three categories of higher, appropriate and lower classification results. If the price of a second-hand laptop is marked as appropriate, the decision tree regression modeling will only be conducted on the part of the data set marked as appropriate to obtain the final predicted price of the product.

2.5 Euclidean distance

We introduce the correlation coefficient p of the feature and target variable, and takes the product of the distance difference of each dimension to measure the contribution of each dimension to the target variable.

The class diagram is a static diagram. It represents the static view or the overview of an application. Class diagram is not only used for visualizing, describing and documenting different aspects of a system but also for constructing and designing the executable code of the software application on which we are working on.

The class diagram describes the attributes or properties and operations of a class and also the different constraints imposed on the system. The class diagrams are widely used in the modelling of object oriented systems, so called the OOPS system because they are the only UML diagrams which can be mapped directly with object oriented languages.

The class diagram shows a collection of classes, interfaces, associations, collaborations and constraints, properties and various types of inherited properties of an application. They are also known as structural diagrams as they describe the structure more professionally.

3.1 Purpose:

The purpose of the class diagram is to model the static view of an application to enhance the coding and designing part of an application. The class diagrams are the only diagrams which can be directly mapped with object oriented languages and thus widely used at the time of construction as already discussed.

The UML diagrams like activity diagram, sequence diagram can only give the sequence flow of the application but class diagram is a bit different. So it is the most popular UML diagram in the coder community.

So the purpose of the class diagram can be summarized as:

3.1.1 Analyze the design and overview of the static view of an application.

3.1.2 Describe the responsibilities of a system.

3.1.3 It can also act as a base for other diagrams.

3.1.4 Helps in forward and reverse Engineering.

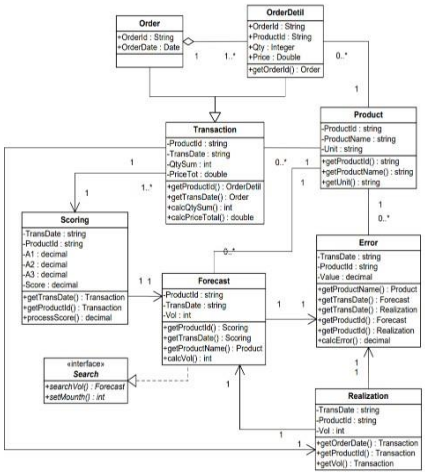

Fig 1: Class Diagram of the proposed system

4. Algorithm Used

The whole system is based on a technique called RNN where connections between the different nodes from a directed or undirected graph along a temporal sequence.

The term "recurrent neural network" or RNN is used discriminately to refer to the two different classes of networks with a similar type of structure, where one is finite and having limited impulse and the other is infinite and having unlimited impulse. Both classes of networks have temporal dynamic behavior and properties. A finite impulse recurrent network is a directed acyclic graph that can be unrolled or deployed and replaced with a strictly forward neural network, while an infinite impulse recurrent network is a directed cyclic graph that cannot be unrolled or unbacked.

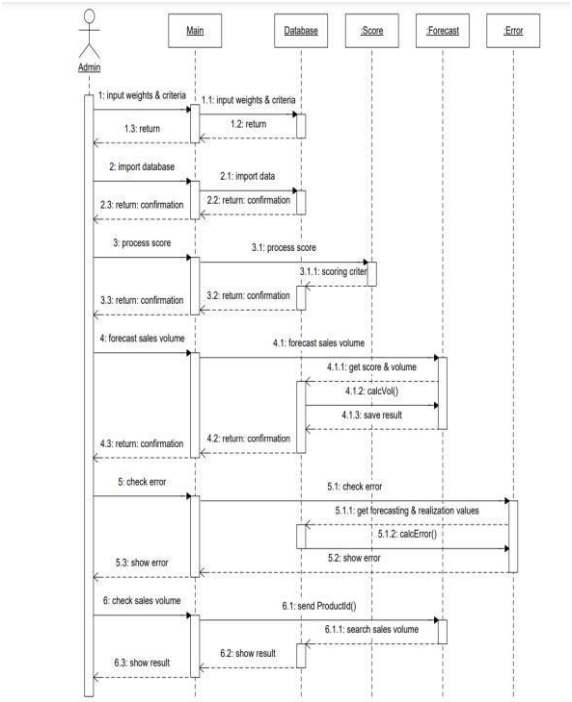

Figure 2: Use Case Diagram of Proposed System

Both of the recurrent networks can have additional stored states or registered states where they store the data, and the storage can be under direct control or indirect control by the neural network. The storage can also be replaced by another network or graph or tree, if that calling time delays or has feedback loops. Such controlled states are called or referred to as gated state or gated memory, and are part of long short- term memory networks (LSTMs) and gated recurrent units also knows as RU or recurrent units in proper. This is also called Feedback Neural Network (FNN).

Algorithm:

function FE(df)

#Apply only the meaningful methods on data df_expanded features - Max-MinScaling(df)

df expanded features - Polarizing(df)

df_expandedfeatures

CalcFluctuationPercentagerdt)

return df expanded features

end function

function RFE(df) # (Utilizing Recursive Feature Elimination function)

Train the model on all the features of the training dataset in df

Calculate performance of the model with samples from the test data

Rank the weights of different features based on testing the model

for each subset do

Retain i most weighted features

Train the model on all the features of the training dataset

Calculate performance of the model with samples from the test data

end for

Calculate the overall performance profile for each feature over samples from the test data Rank and select top ranked features

Train the model on the selected features using the training dataset in df

return df_RFE # (df RFE is the processed data frame after RFE algorithm)

end function

function PCA (df) # (Utilizing PCA to reduce dimension from i to j)

df PCA = applyPCA (n_components-j, whiten=False, copy=True, batchsize = 200)

return df_PCA #df_PCA is the optimized data frame after applying PCA algorithm) **end function**

function MAIN() # (Main function) df alldata = load data

df partition=DataPartition(df_alldata, method = resampling)

df_FE =FE(df_partition) df

RFE =RFE(df FE)

df PCA = PCA(df RFE)

return df PCA

end function

5. Implementation

Implementation is that the stage of the project where the theoretical design is clothed into a working system. Thus it is often considered to be the foremost critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, Investigation of the prevailing system and its constraints on implementation, design of methods to understand changeover and evaluation of changeover methods.

5.1 Input Design

The input design is that the link between the knowledge system and thus the user. It comprises the developing specification and procedures for data preparation and other people steps are necessary to put transaction data in to a usable form for processing are often achieved by inspecting the pc to read data from a written or printed document or it can occur by having people keying the info directly into the system. The planning of input focuses on controlling the number of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the tactic simple. The input is supposed in such how so as that it provides security and straightforward use with retaining the privacy. Input Design considered the subsequent things:

5.1.1 What data should tend as input?
5.1.2 How the info should be arranged or coded?
5.1.3 The dialog to guide the operating personnel in providing input.
5.1.4 Methods for preparing input validations and steps to follow when error occur.

5.2 OUTPUT DESIGN

A quality output is one, which meets the requirements of the highest user and presents the knowledge clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it's determined how the knowledge is to be displaced for immediate need and also the text output. it's the foremost important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to assist user decision-making.

5.2.1 Designing computer output should proceed in an organized, well thought out manner; The proper output must be developed while ensuring that every output element is meant in order that people will find the system can use easily and effectively. When analysis design computer output, they need to Identify the precise output that's needed to satisfy the requirements.

5.2.2 Select methods for presenting information.

5.2.3. Create document, report, or other formats that contain information produced by the system.

6. Conclusion and Future Enhancement

In this paper, the commodity data set collected by crawlers is especially used. After data preprocessing, the improved KNN classification algorithm is firstly wont to classify commodity prices. This classification model will provide merchants with the worth trend of secondhand notebooks within the current period, which is of great significance for merchants to optimize commodity prices reasonably. Then, supported the improved KNN classification, the choice tree regression is employed to work out and predict the worth of commodities. This price model is more accurate than that given directly by the choice tree regression. However, the info set collected during this paper is little, the gathering of commodity categories is incomplete, and therefore the model used still has room for improvement from parameters. within the next step, we will consider expanding the info set and adopting better parameters.

7. REFERENCES

[1] Lihua Wu,Tian Deng. Web Data Mining and Its Implication in E- commerce[P]. Proceedings of the 2016 International Conference on Education, Management, Computer and Society,2016.

[2] Dahai Wang. Analysis of Web Mining Method Based on Intelligent E-Commerce Data[P]. Proceedings of the 8th International Conference on Social Network, Communication and Education (SNCE 2018),2018.

[3] Wang Yi. Research on Key Technologies of Data Mining System Based on Web Service[P]. Proceedings of the 2016 International Conference on Mechatronics Engineering and Information Technology,2016.

[4] Liang Zhang,Lingling Zhang,Weili Teng,Yibing Chen. Based on Information Fusion Technique with Data Mining in the Application of Finance Early-Warning[J]. Procedia Computer Science,2013,17.

[5] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "Arima models to predict next- day electricity prices," Power Systems, IEEE Transactions on, vol. 18, no. 3, pp. 1014–1020, 2003.

[6] R. C. Garcia, J. Contreras, M. Van Akkeren, and J. B. C. Garcia, "A garch forecasting model to predict day-ahead electricity prices," Power Systems, IEEE Transactions on, vol. 20, no. 2, pp. 867–874, 2005.

[7] M. ZENG, Y.-l. ZHAO, and J. ZHANG, "Electricity price forecasting methods combining arima and garch with confifidence intervals," East China Electric Power, vol. 36, no. 12, pp. 1–5, 2008.

[8] Yoon H, Hyun Y, Ha K, et al. A method to improve the stability and accuracy of ANN- and SVM-based time series models for long-term groundwater level predictions[J]. Computers & Geosciences, 2016, 90(PA):144-155

[9] Syedamahmood T. Role of Big Data and Machine Learning in Diagnostic Decision Support in Radiology.[J]. Journal of the American College of Radiology, 2018, 15(3):569-576

[10] Lu YE,Yuping LI,Qin X, et al. Vegetables Price Combination Forecasting Based on PSO-BP and RBF Neural Network[J]. Northern Horticulture, 2015

[11] Leshem G. Traffic Flow Prediction using Adaboost Algorithm with Random Forests as a Weak Learner[J]. Enformatika, 2011:193

[12] Gumus M, Kiran M S. Crude oil price forecasting using XGBoost[C]. International Conference on Computer Science and Engineering. 2017:1100-1103

[13] Mar í a Ortiz,Olatz Ukar,Filipe Azevedo,Arantza M ú gica. Price forecasting and validation in the Spanish electricity market using forecasts as input data[J]. International Journal of Electrical Power and Energy Systems,2016,77.