STOCK MARKET PREDICTION USING GENETIC ALGORITHM

[1]Mohana Sowmya Bezawada, [2]Nagamanikanta Amudalapalli, [3]Lalitha Priyanka Bonthu, [4]Dr. Narayana Satyala
[1]Under Graduate Student,[2]Under Graduate Student, [3]Under Graduate Student,[4]Professor and Mentor (AS &A)
[1]Computer Science and Engineering,
[1]Gudlavalleru Engineering College, Gudlavalleru, India

Abstract: The stock market is where shares of publicly traded companies are exchanged. The Exchange enables dealers to trade shares in companies and other securities. Forecasting the price of shares is one of the most extensive studies and the most difficult problems, attracting researchers from many fields, including economics, history, finance, mathematics, and computing. The volatile nature of the stock market makes the application of simple time series or regression techniques challenging. Our project attempts to forecast the fluctuation of the stock market index using deep learning techniques. Moreover, we optimize the topology of the CNN network to improve the performance of the models. This study proposes a method to systematically optimize the parameters of the CNN model using the genetic algorithm (GA). The experimental results show that GA-CNN surpasses comparative models and demonstrates the effectiveness of the GA-CNN hybrid approach.

Introduction:

A stock exchange is a collection of buyers and distributors of stocks that represent the ownership of a company; those might also additionally encompass securities indexed on a public inventory trade in addition to the ones best traded privately. Exchanges facilitate the brokering of shares and other securities. Publicly traded ordinary shares and other types of securities, for example, corporate bonds and convertible bonds. Market forecasts are intended to determine the future value of a company or other publicly traded financial instrument. The prediction of stock prices is one of the most in-depth studies and complex problems, attracting researchers from many fields, including economics, history, finance, mathematics, and computing. Successfully predicting the future price of a security will maximize the profits of investors.

Due to the volatility and non-linear nature of these returns, it is challenging to accurately predict market returns. An efficient forecasting system is required for the successful analysis of future stock prices for each enterprise. It is more complex for researchers to analyze future prices for major measurements in order to obtain greater precision. But in recent times stock market predictions are gaining more attention, maybe due to the fact that the market trends are predicted successfully and investors are guided better. Investment and stock exchange earnings are highly dependent on predictability. If a system is in place to consistently predict the direction of the dynamic equity market, system users will be able to make informed decisions. Furthermore, the expected market trends will assist regulators in taking corrective measures.

Recently, Convolutional Neural Networks (CNN) have been applied to various time-series problems of a diverse nature, such as voice recognition and natural language processing, and many studies have demonstrated its efficiency in time-series data. As CNN has the advantage of extracting the local features of the data, by capturing the time attributes of a given data set, an effective method for predicting time series problems. Moreover, in this study, we aim to optimize the topology of the CNN network in order to improve the performance of the model. While numerous studies have demonstrated the effectiveness of CNN models in various classification problems, there are several gaps in the construction and use of the model.

We propose a method to systematically optimize the parameters of the CNN model using the genetic algorithm (GA). We employ GA in the search for the optimal CNN architecture that leverages the ability of neural networks to model complex non-linear functions while automatically solving the optimization issue. Numerous studies have applied GA in conjunction with other AI and machine learning techniques such as ANN, vector support machine (SVM), and case-based reasoning (CBR). However, few studies have attempted to integrate GA and CNN, despite the potential for successful applications in this area.

Related work:

The following provides a brief description of each ANN-related study’s unique research focus and findings.

- Jasic and Wood (2004) developed an artificial neural network for predicting daily returns of stock indices using data from several global stock markets. Emphasis is placed on supporting the cost-effective trade. One method is introduced based on the univariate neural networks using unconverted data to provide short-term predictions of the yield of the stock index.

- Enke and Thawornwong (2005) use machine learning and information acquisition techniques to assess predictive relationships among numerous financial and economic variables. The calculation of the information gain for each variable in the model provides a classification of the variables. In addition, a cross-validation technique is used to improve the generalizability of several models.
- Convolutional Neural Network is another deep learning algorithm applied in stock market prediction after LSTM and MLP while its ability to extract efficient features has been proven in many other domains as well. In (DI Persio & Honchar, 2016), CNN, LSTM, and MLP were applied to the historical data of 175 closed prices of the S&P 500 Index. The findings showed that CNN outperformed LSTM and MLP.
- The authors of (Zhong and Enke, 2017) applied the CPA and two variants to extract better characteristics. A collection of different variables was used as 135 input data while an ANN was utilized for prediction of S&P 500. The results showed an improvement in the forecast by using the characteristics generated by PCA over the other two variations of that.

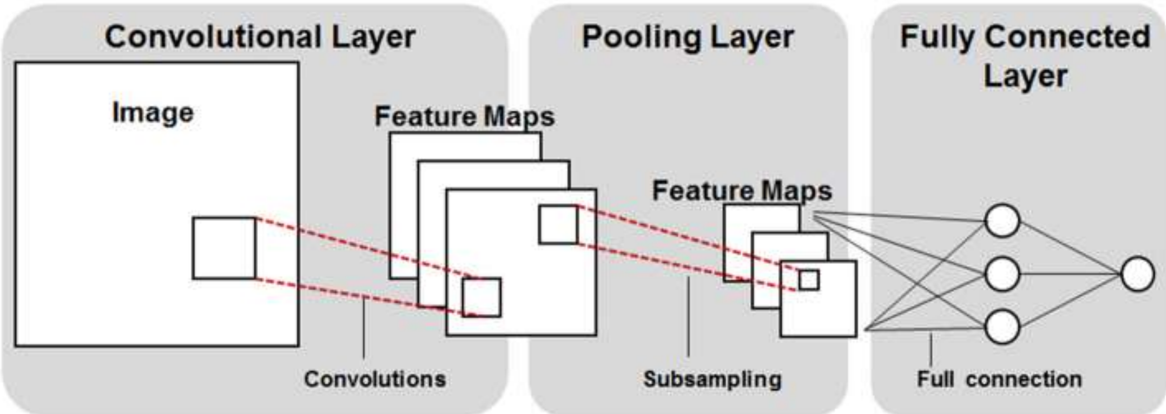
Methodology :-

Convolutional neural networks (CNN) are composed of two parts. One is the feature extraction, which is composed of convolutional and pooling layers. Another is the trainable fully connected part of the prediction or classification. As feature extraction increases the accuracy of learned models by extracting features from the input data. Meanwhile, prediction refers to the output of a trained model, representing the most likely value that will be obtained for a given input.

convolutional layers extract features that have local correlations when the positional relationships between the local features are determined. In the convolutional layer, the kernel (or filter) extracts features while moving on the input data at regular intervals.

In the pooling layer, the samples of the most representative features are extracted from the convolutional layer. The sampling methods include max-pooling and average pooling. Sampling is conducted by extracting the maximum or the average value of each interval.

The fully connected layer connect the neurons generated in the convolutional layer and the pooling layer to all the neurons of the higher layer and performs the prediction and classification operations.



Convolutional Neural Network

Genetic Algorithms (GAs) are search based algorithms based on the concepts of natural selection and genetics. GA is a subset of a much larger branch of computation known as Evolutionary Computation. In GAs, we have a pool or a population of possible solutions to the given problem. These solutions then undergo recombination and mutation (like in natural genetics), producing new children, and the process is repeated over various generations. Each individual (or candidate solution) is assigned a fitness value (based on its objective function value) and the fitter individuals are given a higher chance to mate and yield more "fitter" individuals. This is in line with the Darwinian Theory of "Survival of the Fittest". In this way we keep "evolving" better individuals or solutions over generations, till we reach a stopping criterion.

The inputs to this algorithm include the fitness function for ranking candidate hypotheses, a threshold defining an acceptable level of fitness for terminating the algorithm, the size of the population to be maintained, and parameters that determine how successor populations are to be generated:the fraction of the population to be replaced at each generation and the mutation rate.

First, a certain number of hypotheses from the current population areselected for inclusion in the next generation. These are selected probabilistically, where the probability of selectinghypothesis hi is given by

$$\Pr(h_i) = \frac{Fitness(h_i)}{\sum_{j=1}^{p} Fitness(h_j)}$$

Once these members of the current generation have been selected for inclusion in the next generation population, additional members are generated using a crossover operation.

Algorithm:

GA (Fitness, Fitness threshold, p, r, m)

Fitness: A function that assigns an evaluation score, given a hypothesis.

Fitness threshold: A threshold specifying the termination criterion.

p: The number of hypotheses to be included in the population.

r: The fraction of the population to be replaced by Crossover at each step.

m: The mutation rate.

- Initialize population: P + Generate p hypotheses at random
- Evaluate: For each h in P, compute Fitness(h)
- While (max Fitness(h)] < Fitness threshold do

    Create a new generation, P_s:

    1. Select: Probabilistically select (1-r) p members of P to add to Ps. The probability $\Pr(h_i)$ of selecting hypothesis h_i from P is given by

    $$\Pr(h_i) = Fitness(h_i)/\sum^{p}_{j=1} Fitness(h_j)$$

    2. Crossover: Probabilistically select (r-p)/2 pairs of hypotheses from P, according to $\Pr(h_i)$ given above. For each pair, (h_1, h_2), produce two offspring by applying the Crossover operator. Add all offspring to P_s.
    3. Mutate: Choose m percent of the members of P_s with uniform probability. For each, invert one randomly selected bit in its representation.
    4. Update: P -> PS

    5. Evaluate: for each h in P, compute Fitness(h)
- Return the hypothesis from P that has the highest fitness.

Data Set collection:

In this study, we consider previous stock market value for forecasting the stock value. The data set is collected from Kaggle. The data used in this model include stock indexes. The stock prices change with in time. Therefore the nature of data is noisy and nonlinear.

Data Preparation:

The raw data includes date, low price, high price, opening price and closing price. We have considered stock market 448 days in total for the prediction. 80% of data is used as training set, 20% of training data is chosen validation set. Remaining 20% is used as holdout set. We exclude date column in further steps for creating the model.

Seven technical indicators are selected as the input variables. Fund managers and traders capture important market signals through technical indicators, and many studies related to stock market prediction have proven the empirical effectiveness of these technical indicators. The seven technical indicators and their formulae:

Technical indicator	Formula
Momentum	$C_t - C_{t-n}$
Stochastic K %	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$
Relative strength index (RSI)	$100 - \frac{100}{1+\left(\sum_{i=0}^{n-1} Up_{t-i}/n\right)/\left(\sum_{i=0}^{n-1} Dw_{t-i}/n\right)}$
Moving average convergence divergence (MACD)	$EMA_{12} - EMA_{26}$
LW %R	$\frac{H_n - C_t}{H_n - L_n} \times 100$
A/D oscillator	$\frac{H_t - C_{t-1}}{H_t - L_t} \times 100$
Commodity channel index (CCI)	$\frac{M-m}{d \times 0.015} \times 100$

Training the CNN model:

For training the model, we employ a multi-channel CNN model consisting of seven channels of input variables, two filter layers, one pooling layer, and two fully connected layers. The inputs of the proposed model are multiple 1D subsequences which are separated from the multivariate timeseries. The feature learning is conducted on individual univariate time-series and we concatenate fully connected layers at the end of feature learning to perform the classification. This data scaling process increases the efficiency of training procedure of neural network model. The linearly scaled value of x^1 can be expressed as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where min(x) and max(x) are, respectively, the minimum and the maximum values of the sample x.

In contrast to image classification, we exploit the 1D kernels to extract the local temporal information and the size and the number of the kernels are searched by the GA. In the fully connected layers, we use 200 and 40 hidden nodes, respectively. We implement a rectified linear unit (ReLU) as the activation function of all the layers except the output layer. ReLU is a nonlinear function that outputs x if the input value x is positive and 0 otherwise. The ReLU function was proposed to solve the gradient vanishing problem because traditional sigmoidal functions make the back-propagated errors converge to zero with the increase in the number of layers of the neural network. The ReLU function with input x can be expressed as follows:

$$f(x) = \max(0, x)$$

To perform the binary classification, the output layer adopts the logistic sigmoid as the activation function. The sigmoid function can be defined as follows:

$$f(x) = \frac{1}{1 + e^{-x}}$$

To perform the genetic search, we organize a population of chromosomes with the encoded bits of the candidate subset of the parameters of CNN. The encoded parameters and the corresponding chromosome values used in this study are described below.

- Number of kernels in each convolutional layer: 1–63
- Kernel size of individual channels in the first convolutional layer: 1–31
- Kernel size of individual channels in the second convolutional layer: 1–31
- Window size of the pooling layer: 1–31

GA has many controlling parameters to adjust in addition to determining the fitness functions and the problem representation. There has been much debate regarding the optimal controlling parameters that should be specified for the experiment. For this experiment, we set the controlling parameters by referring to previous studies that applied GA to financial problems. In this study, the crossover rate is set to 0.7, and the mutation rate is set to 0.25.

Result:

When it comes to stock market prediction, each technical indicator has a different effective time window for the prediction, an aspect considered by only a few studies. However, the proposed model could, first, reflect the effective temporal properties of each input variable for stock market prediction through the individual GA optimization of each channel, and second, enable the finding of the best time window for each technical indicator.

Here, we can observe the actual and predicted values that are generated.

Conclusion :-

In the stock market shares are being exchanged by the companies frequently. The Exchange enables dealers to trade shares in companies and other securities. Stock market prediction is actual demand for beneficial business. Predictions always helpful to decrease risk factor in any business environment. Risk factor can be analyzed on the basis of historical data and previous business, but sometimes predicting stock prices is difficult. So we developed a model for predicting the future share value by using CNN and GA. Our model works based on the shares data that are provided by the company. By predicting the future share value one can decides whether buying that shares or not. It will help us to know about the future shares values by considering the some particular company stock values.

Future Scope :-

One of the other areas of further research is to evolve simple trading strategies after the model predicts the stock trend. It states that if the model predicts as up or down, which actions to take and how much mock to buy or sell or when to buy or sell. This is actually complimentary for the prediction model as after predictions it gives the afterward instructions to keep the investors in the best financial position Application of News Related to Automobile Industry The types of news we have used for making the prediction model are those exactly related to the political, financial, production, and other activities and policies of Iran Khodro Company. We consider that other types of news related to the automobile industry as a whole and the news related to the other automobile competitors might have an effect on Iran-Khodro stock price movement.

Hence we recommend making the prediction model applying all the types of news related to the auto industry in general and the ones related to competitors and compare the results with the current prediction model.

References :-

[1] https://doi.org/10.1016/s0305-0483(01)00026-3
[2] https://doi.org/10.1016/j.asoc.2006.03.004
[3]https://doi.org/10.1016/j.eswa.2008.08.019
[4]https://doi.org/10.1016/j.apenergy.2019.01.055
[5]https://doi.org/10.1007/s00521-011-0741-0
[6]https://doi.org/10.1016/j.jmsy.2014.11.007
[7]https://doi.org/10.1007/s00521-011-0603-