Hierarchical Attention Network in Stock Prediction

Liming Huang1, Hongfei Yan1,3, Siping Ying1, Yansong Li1, Rui Miao1, Chong Chen2, and Qi Su1
1 Peking University, 2 Beijing Normal University, 3 National Engineering Laboratory for Big Data Analysis and Application Technology,

论文摘要

Stock price prediction refers to the behavior of predicting the direction and possibility of future stock price movements by analyzing the historical information of the stock market. This article focuses on predicting the stock price will rise or fall, which is a binary classification problem. With the circulation of stocks, stock prices will generate a series of fluctuations. In each trading day, opening price, closing price, high price, low price, trading volume, trading value, and other data are usually used to summarize the price fluctuations of the day. And the transaction situation, in which the closing price is generally used to compare with the previous trading day, as a measure of the stock price rise or fall in the day. Firstly, given the stock s and the trading day t, the rise or fall of stock each day is defined as follows:

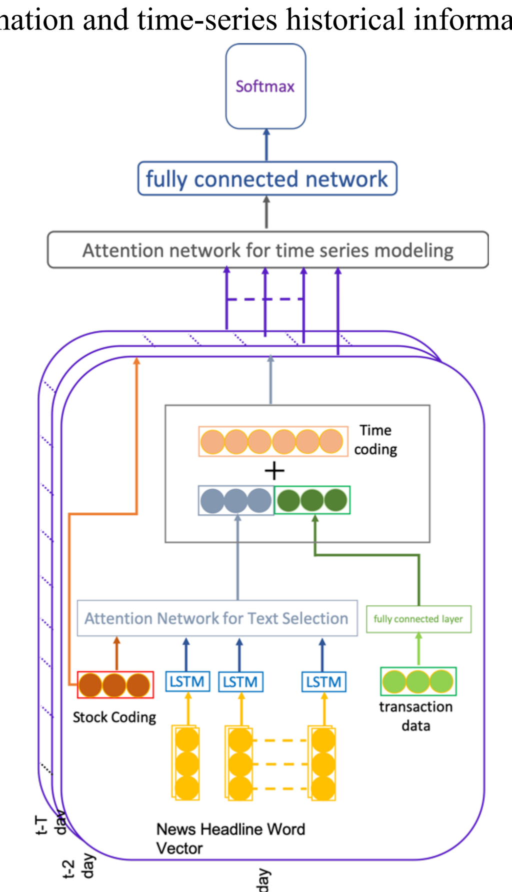
$y_s^t = \begin{cases} 0, close_t^s < close_{t-1}^s \\ 1, close_t^s > close_{t-1}^s \end{cases}$ Where close represents the closing price of stock s on trading day t. When the closing price of y on trading day t is higher than the previous trading day t–1, y = 1,indicating that the stock price of stock s on trading day t increases; Conversely, if the closing price is lower than or equal to the previous trading day, then y = 0, indicating that the stock price has fallen or remained flat.

In addition to the relevant data on the trading day, text information related to stock-related financial news, institutional research reports, etc. may be generated every day, which can be used to understand the current stock market or event information related to certain stocks. The transaction data of stock s on trading day t is P, and the related news information is D . The stock forecasting tasks studied in this paper are defined as follows:

Given a stock s and an arbitrary trading day t, given a historical time series of length T-1 [t − T, ..., t − 1]. Use the stock trading information [P t−T , ..., Pt−1] and related news texts [D t−T , ..., D t−1] to predict the rise or fall of the closing price of stocks s on trading day y: $\hat{y}_t^s = f([P_{t-T}^s, \dots, P_{t-1}^s], [D_{t-T}^s, \dots, D_{t-1}^s])$

系统模型

Inspired by multi-head attention mechanism and hierarchical attention network, this paper proposes a stock forecasting model based on hierarchical multi-head attention network, which applies two different levels of attention network to the modeling of serialized news text information and time-series historical information, respectively. Think of stock prediction as a binary classification task.



Stock forecasting model based on hierarchical multi-way attention network

As shown in Figure 1, the model proposed in this paper mainly includes the following five components:

Input layer: The length of historical time we used is T, that is, when the stock price of the t-th trading day is predicted to rise or fall, the input layer will use the information in [t − T, ..., t − 1] as input.

In the case of news texts, the headings of all relevant news are used as input. This paper introduces an external knowledge base for expanding stock-related entity noun sets for finding news headlines related to each stock in the news data set. The daily related news headline set is represented as D, and the input set of the time period is: $\vec{D}_t^s = [D_{t-T}^s, \dots, D_{t-1}^s]$

In terms of transaction data, this article uses six indicators such as opening price, closing price, high price, low price, trading volume, and trading value. In this paper, the normalization of data is performed for each stock to eliminate the difference in numerical scale between different indicators and different stocks, the input set is: $\vec{P}_t^s = [P_{t-T}^s, \dots, T_{t-1}^s]$

论文简介

To solve the stock prediction problem, we propose a deep learning model base on a hierarchical attention network. Our model is divided into two models. The first model is the article selection attention network that transfers the news into a low dimension vector. This model could identify the important factors in the news that affect the stock price. The second model is a time series attention network which combines the output of the first model and the transaction data as input. In this model, we could figure out the potential impact between different dates and summarize the historical data to predict whether the stock price will rise or fall. The most innovative concept in this paper is stock encoding. The model learns the difference between each stock and make predictions more accurate by using the stock encoding.

算法原理

Coding layer:

Stock code: The model will receive different stock inputs simultaneously and make predictions, and each stock has its own unique nature. Therefore, the vector with dimension ds is used to learn its own feature vector Es for each stock, which helps the model to better distinguish and identify the characteristics of each stock.

Title: For each news headline entered, it is first encoded using the pretrained Word2Vec model and then processed using a Bi-LSTM neural network. Since the output of each step of the Bi-LSTM network already contains all the previous input information, only the last step of output h is used. The output of the forward sequences and reverse sequences are concatenated to obtain the low-dimensional vector features of each title.

Transaction data: We increase the linear expression ability of transaction data through a fully connected layer to obtain the feature vector PE

Transaction time: Considering that the model will follow the multi-head attention mechanism for timing modeling, Therefore, the position coding in is introduced to encode each historical event to obtain TE.

Attention Network for Text Selection:

The first level of text selection attention network calculates the correlation between each news and the current forecasted stock, removes the noise of unimportant information, and extracts the factors that really affect the price rise or fall of the stock. This layer is based on the general attention mechanism. The relationship between stock coding and news headlines are obtained through a trainable matrix W. The score of each news is: $\mathrm{score}_i = ES^s \cdot W \cdot \mathrm{titleEmbed}_i$

Finally, the weighted average of title embedding is the output TE: $TE_t^s = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \times titleEmbed_i$

Where N is the number of daily news, and a is the weight distribution of news: $\alpha_i = \frac{\exp(\mathrm{score}_i)}{\sum_{j=1}^{N}\exp(\mathrm{score}_j)}$

Attention network for time series modeling:

The input of this layer includes the low-dimensional feature vector of news, transaction information, and time information. The vector E input for each trading day is defined as follows: $E_t^s = \mathrm{concat}(DE_t^s, PE_t^s) + TE_t^s$

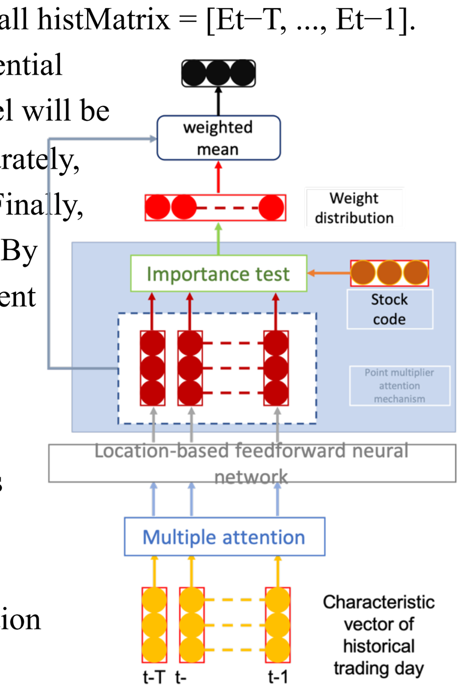
The time series modeling network is shown in Figure. It includes a multihead attention mechanism, a position-based feedforward neural network and a layer of normalized point product attention mechanism. The multi-head attention mechanism can be considered as a multi-parallel form of multiple self-attention mechanisms, which is obtained by splicing the results of multiple self-attention mechanisms. By normalizing the feature vector E of each historical trading day with all histMatrix = [Et−T, ..., Et−1].

This mechanism allows global information to be retained for each output and captures potential interactions for better information processing. Due to the nature of its parallel computing, the model will be more efficient. The forward neural network processes the output of each historical trading day separately, and is mainly used to increase the nonlinear fitting ability of the multi-head attention mechanism. Finally, The normalized dot-product attention mechanism is similar to the text-selection attention network. By comparing the stock coding with the information of each historical trading day, we can get the current impact of each historical trading day, and then obtain the optimal feature information.

Stock Price Prediction Network:

The task of this paper is to forecast the rise or fall of stock, so the final output is y. Each element is between [0, 1], indicating the probability of price rise or fall on that day. This paper uses the Softmax function to calculate the probability of the final binary classification.

The probability that the prediction of stock s classified as J when (Pt)j is the trading day t. In this paper, the maximum probability category is used as the final prediction result, so the prediction result of the model is as follows: $\hat{y}_t^s = \max((p_t^s)_0, (p_t^s)_1)$



Attention Network for Time Series Modeling

实验仿真

Date Set:

SSE A-share and NASDAQ. Both of which include stock transaction data and textual information that may affect stock price trends.

Market	Duration	Shares	News	Fall/Rise Ratio
SSE	2013/01/01 2015/09/17	772	78,381	49.9/50.1
NASDAQ	2006/10/10 2013/11/23	161	1,745	51.6/48.4

Baseline Algorithms:

– CNN-Static: Text feature extraction and classification tasks using CNN.

– Price-SFM: As one of the variants of the RNN, the Fourier transform is used to extract the transaction information of different frequencies in the historical trading of stocks to predict the stock price.

– Event-Embedding: Structured events are extracted from news headlines, and low-dimensional vector representations of structured events are learned using tensors. Finally, CNN is used to process and make predictions.

– Price+News-PMI: The PMI algorithm is used to extract the positive and negative characteristics of related news, and then it is stitched with the transaction data, and the RNN is used for time series modeling.

– Hier-Attention: Our method.

Experiment Result:

This article studies the task of forecasting the rise or fall of stock prices, which is essentially a binary classification problem. Therefore, the index of the experimental evaluation adopts the accuracy rate of the classification method.

The performance of each research method on the two datasets is shown in Table. Since the interface used by Event-Embedding only supports English, the experimental method is not tested on the SSE A share dataset.

Model	SSE	NASDAQ
CNN-Static	0.5342	0.5304
Price+News-PMI	0.5531	-
Price-SFM	0.5671	0.5266
Event-Embedding	-	0.5264
Hier-Attention	**0.5746**	**0.5638**

Comparing the performance of different methods on the two data sets, it can be seen that the accuracy of other research methods on the US stock data set is only about 52.5, which is significantly lower than the results in the SSE. Our method in two datasets has similar performance. It is proved that our method has robustness to different datasets and achieve good performance regardless of the quality and sparseness of stock-related news text. Therefore, the experimental results fully verify the superiority of the method in this paper over other research methods in the task of stock prediction.

论文结论

This paper focuses on the application of deep learning and natural language processing technology in stock price prediction, and proposes a stock prediction method based on hierarchical attention network. Through two different levels of attention networks, the news text is analyzed and the historical information is time-series modeled, and the input information is better understood to make the best prediction results. And the introduction of stock vector coding in this article enables the model to learn the unique properties of each stock, so as to make more targeted prediction results by combining the characteristics of different stocks

This article conducts comparative experiments on the SSE A-share market and the US stock market. First, in order to verify the effectiveness of the method proposed in this paper, multiple benchmark methods are set up for analysis and comparison, and the effectiveness of the method proposed in this paper is proved. Subsequently, this article is compared with other recent stock prediction research work. The test results show that the performance of this method is significantly better than other research methods, which proves that our method makes full and effective use of input information in text feature extraction and time series modeling, and the superiority of this method in the task of stock price prediction.

西安电子科技大学
XIDIAN UNIVERSITY