Review Comments of

Design and Implementation of Big Data Analysis Method for Short-term Stock Trend Prediction

Author: 李澍培

Reviewer: 任泽华

3121154002 5.17.2022

CONTENTS

1. Work Overview

Design and Implementation of Big Data Analysis Method for Short-term Stock Trend Prediction

This paper proposes a stock price prediction algorithm based on LSTM (Long Short Term Memory) classifier. First, the continuous data is discretized, then 14 groups of discrete data types are reduced by decision tree algorithm to eliminate useless data types, and then the reduced attributes are put into the LSTM classifier for training and prediction. In addition, non-technical indicators are added to the model to improve the accuracy of the model.

2. Writing Evaluation

Strengths

1. The research problem is meaningful,
   and the proposed solution is technically correct.
2. This paper is clear and easy to understand.
3. The article combines the results of the experimental analysis to make the conclusion more convincing.

Weakness

1. Formatting problems, 2. Express problem, 3. Formula problem,
4. Grammatical errors, 5. Typography problems.

2. Writing Evaluation

A. Formatting problems

① The first letter of the word after the semicolon should be lowercase.

using decision tree algorithm to stock data of condition attributes in rough set attribute reduction; Then the LSTM algorithm is used to predict the rise and fall of the stock. In addition, non-technical indexes are added to improve the accuracy of the model. Taking

② The font size of these two paragraphs is not uniform.

study is to predict the rise and fall of the stock, so the rise and fall rate of the forecast day is selected as the decision-making attribute of this study, and the other attributes as conditional attributes. Label the remaining 14 conditional attributes in sequence.

Since the initial data set obtained in this experiment is continuous data, in order to facilitate computer processing, the data needs to be discretized first. Data discretization refers to the sampling and quantization of each attribute value without changing the relative size of data. The

2. Writing Evaluation

B. Express problem

The paper doesn't list what the other properties are, which will confuse readers.

> the actual research purpose, that is, the rise and fall of the stock. The purpose of this study is to predict the rise and fall of the stock, so the rise and fall rate of the forecast day is selected as the decision-making attribute of this study, and the other attributes as conditional attributes. Label the remaining 14 conditional attributes in sequence.

C. Formula problem

The symbol *a* is used in the left of equal sign but it does not appear in the right side, while there are several D^v in the right. Consider change the superscript into a or explain in the text.

$$\mathrm{Gain}(D,a)=\mathrm{Ent}(D)-\sum_{v=1}^{V}\frac{|D^v|}{|D|}\mathrm{Ent}(D^v) \quad (2)$$

2. Writing Evaluation

D. Grammatical errors

① There are some "A" in this paragraphs which is strange. Consider remove some of them and change others into lowercase.

often used to process sequential data. Cyclic neural networks have A loop structure that allows information to persist. A group of neural networks A receives some input and outputs A value.

② The last highlighted word might be non-technical according to context.

3. Results and Discussion

Table 1 shows the classification accuracy without attribute reduction and with different number of conditional attributes, as well as the classification accuracy after adding non-technical indicators. The classification data set adopts Microsoft dataset. "A" means "non-technical indicators are not added" and "B" means "technical indicators are added".

2. Writing Evaluation

E. Typography problems

Please use the IEEE two-column template and translate references into English.

Bibliography

[1] 王天娥, 叶德谦, 季春兰. 粗糙集 attribute 约简方法在股票预测中的应用研究[J]. 计算机工程与应用, 2009, 045(030):227-229,236.

[2] 张泽亚, 黄丽明, 陈翀,等. 基于新闻特征抽取和循环神经网络的股票预测方法[J]. 文献与数据学报, 2020(1).

[3] Weng B, Ahmed M A, Megahed F M. Stock market one-day ahead movement prediction using disparate data sources[J]. Expert Systems with Applications, 2017, 79(AUG.):153-163.

[4] 郑乾裕. 基于技术指标的股票专家系统研究[D]. 浙江大学.

[5] Hartigan J A, Wong M A. A k-means clustering algorithm[J]. Applied Statistics, 1979, 28(1).

3. Technical Evaluation

Questions

①There are lack of results in 14 attribute situation when non-technical indicators are added into the model. **Is it a technical reason?** Hope to explain in the text.

Tab 1 Classification result

Discretization algorithm	Attribute numbers	A Accuracy	B Accuracy
Equidistant discretization	6	50.98%	52.94%
Equidistant discretization	14	49.02%	\
One-dimensional clustering	6	56.86%	60.78%
One-dimensional clustering	14	47.06%	\

3. Technical Evaluation

②**How does the non-technical indicator part affect the forecast results?** How do these unquantifiable factors act quantitatively on the predictive model?

The *Equidistance bulk chemical method* is not explicitly stated in the text. Please explain it in revised manuscript.

In this paper, the variation of the stock equity structure as technical indicators, using Microsoft's stock data as data sets, from flush website related indicators, and USES the equidistance bulk chemical method for discretization and join the technical index data set.

3. Technical Evaluation

Some doubts

This method tries to fit the stock trend curve, but there are many random factors in the stock market, and this method does not seem to take the effect of randomness into account. **Combining stochastic process analysis to give confidence intervals seems to be a better approach.** Otherwise, the model may be overfitting, and some features of random interference may be learned and cannot make more accurate predictions.

Thank you!

5.17.2022