



Artificial Intelligence with Knowledge Based Systems

Review 1

Topic: Stock Prediction of Tesla and Ferrari

Team:

Shashank Singh 21MIA1110

Dazzle A J 21MIA1119

Introduction:

The dataset is specifically focused on capturing the share prices of Ferrari and Tesla, which are two prominent companies in the automotive industry. It is designed to provide researchers, analysts, and data scientists with a comprehensive set of data points to explore and analyze the stock performance of these companies.

Typically, stock price datasets include various attributes such as the date of the recorded price, the opening and closing prices for the trading day, the highest and lowest prices reached during the day, and the trading volume. These attributes enable the analysis of historical price trends, volatility, and trading activity.

By working with the "Ferrari and Tesla Share Prices 2015-2023" dataset, we can conduct a range of analyses and studies. We can examine the price movements of both companies over time, identify patterns and trends, and explore potential relationships between the stock prices of Ferrari and Tesla.

Objectives:

The objectives of this dataset are:

1. **Stock Price Analysis:** The dataset provides an opportunity to analyze and understand the historical stock price movements of Ferrari and Tesla. Users can explore trends, patterns, and fluctuations in the stock prices over the specified time period.
2. **Comparative Analysis:** Since the dataset includes stock prices for both Ferrari and Tesla, users can compare the performance of these two companies. They can identify similarities, differences, and potential correlations between the stock prices of these automotive industry giants.
3. **Financial Modelling and Forecasting:** The dataset can be used to develop financial models and predictive algorithms to forecast future stock prices of Ferrari and Tesla. By analyzing historical data, users can build and test models that aim to predict the future performance of these stocks.
4. **Investment Strategies:** Investors and financial analysts can utilize the dataset to evaluate investment strategies for Ferrari and Tesla stocks. They can back test trading strategies, assess risk-reward ratios, and explore potential investment opportunities based on historical stock price data.
5. **Machine Learning and Data Analysis:** The dataset serves as a valuable resource for data scientists and machine learning practitioners interested in developing predictive models or applying data analysis techniques to stock market data. It can be used to explore various algorithms, feature engineering techniques, and evaluation methods for stock price prediction.

Literature Review

- **Metaheuristic Algorithms:**

1. **Genetic Algorithm:**

- a. **Parameters Identification of Continuous System Based on Hybrid Genetic Algorithm**

Abstract: A new hybrid genetic algorithm is provided by adding up the advantages of the genetic algorithm and gradient algorithm, as uses the results of gradient algorithm improving the populations of genetic algorithm, and selects the best point as the start point of gradient algorithm next time by comparing the best point of genetic algorithm with the last results of gradient algorithm. Applying the method to estimating the parameters of continuous system, the simulation results show it is more quickly than genetic algorithm and owes better anti-noise ability, and improves the defects of genetic algorithm with slower searching ability near a point, and it provides a new method for the parameters estimation of continuous system.

- b. **Designing embedded parallel systems with parallel genetic algorithms**

Abstract: Generic parallel genetic algorithms are developed with reference to the example of the real-time path planning problem for mobile robots. Most robot motion planners are used off-line: the planner is invoked with a model of the environment; it produces a path which is passed to the robot controller which in turn executes it. In general, the time necessary to achieve this loop is not short

enough to allow the robot to move in a dynamic environment (moving obstacles). The goal is to try to reduce this time in order to be able to deal with real time path planning in dynamic environments. The authors use a method, called 'Ariadne's CLEW algorithm', to build a global path planner based on the combination of two parallel genetic algorithms: an EXPLORE algorithm and a SEARCH algorithm. The purpose of the EXPLORE algorithm is to collect information about the environment with an increasingly fine resolution by placing landmarks in the searched space. The goal of the SEARCH algorithm is to opportunistically check if the target can be reached from any given placed landmark.

c. A Task Scheduling Algorithm Based on Genetic Algorithm and Ant Colony Optimization Algorithm with Multi-QoS Constraints in Cloud Computing

Abstract: Task scheduling problem in cloud computing environment is NP-hard problem, which is difficult to obtain exact optimal solution and is suitable for using intelligent optimization algorithms to approximate the optimal solution. Meanwhile, quality of service (QoS) is an important indicator to measure the performance of task scheduling. In this paper, a novel task scheduling algorithm MQoS-GAAC with multi-QoS constraints is proposed, considering the time-consuming, expenditure, security and reliability in the scheduling process. The algorithm integrates ant colony optimization algorithm (ACO) with genetic algorithm (GA). To generate the initial pheromone efficiently for ACO, GA is invoked. With the designed fitness function, 4-dimensional QoS objectives are evaluated. Then, ACO is utilized to seek

out the optimum resource. The experiment indicates that the proposed algorithm has preferable performance both in balancing resources and guaranteeing QoS.

d. Hybrid Optimization Method Based on Genetic Algorithm and Cultural Algorithm

Abstract: Knowledge about evolutionary information is not used in genetic algorithms effectively. Cultural algorithms with dual inheritance structure converge slowly because only mutation operator is adopted in the population space. A novel hybrid optimization method is proposed using genetic algorithm in population space. Four kinds of knowledge and two phases are abstracted. Steps of the algorithm are described in detail. Simulation results on the benchmark optimization functions indicate that the method converges faster than traditional cultural algorithms. In iteratively dynamic situation, results show that experience knowledge in the knowledge space is benefit to apperceive the change of situation and has the ability in memory, which increases the speed of convergence in a certain situation.

2. Ant Colony Optimization Algorithm

a. Research on Path Planning of AGV Based on Improved Ant Colony Optimization Algorithm

Abstract: Path planning is a key problem in the motion control of mobile robot. In order to solve the problem that the traditional storage mode of automatic container terminal affects the overall operation efficiency, this paper puts forward a matrix yard storage mode, which is transformed into grid map model, and then uses ant colony optimization algorithm to plan the path of AGV. Aiming at the shortcomings of traditional ant colony optimization algorithm (ACO) in global path planning, such as slow convergence speed and weak optimization ability, an improved ant colony path planning algorithm is proposed. Firstly, the grid map is established, and the fruit fly optimization algorithm (FOA) is used for fast pre-search on the grid map to generate the original pheromone distribution required by the ant colony optimization algorithm, and then the ant colony optimization algorithm is used for global path planning. At the same time, in order to solve the problem of many path turning angles and large cumulative turning angles in the planning, the path smoothing is carried out. The simulation results show that the improved algorithm has fewer turns and smoother path, and the improved ant colony algorithm has a greater improvement in path search speed and accuracy than the traditional algorithm.

b. A Modified Pareto Strength Ant Colony Optimization Algorithm for the Multi-objective Optimization Problems

Abstract: Ant colony optimization is a meta-heuristic that has been widely used for solving combinatorial optimization problems, and most real-world applications are concerned with multi-objective optimization problems. The Pareto strength ant colony optimization (PSACO) algorithm, which uses the concepts of Pareto optimality and also the domination concept, has been shown to be very effective in optimizing any number of objectives simultaneously. This paper modifies the PSACO algorithm to solve two combinatorial optimization problems: the travelling salesman problem (TSP); and the job-shop scheduling problem (JSSP). It uses the random-weight based method as an improvement. The proposed method achieved a better performance than the original PSACO algorithm for both combinatorial optimization problems and obtained well-distributed Pareto optimal fronts.

c. Hybrid Algorithm Combining Ant Colony Optimization Algorithm with Genetic Algorithm

Abstract: By use of the properties of ant colony algorithm and genetic algorithm, a hybrid algorithm is proposed to solve the traveling salesman problems. First, it adopts genetic algorithm to give information pheromone to distribute. Second, it makes use of the ant colony algorithm to get several solutions through information pheromone accumulation and renewal. Finally, by using across and mutation operation of genetic algorithm, the effective solutions are obtained. Compare with the simulated annealing

algorithm, the standard genetic algorithm, the standard ant colony algorithm, and statistics initial ant colony algorithm, all the 16 hybrid algorithms are proved effective. Especially the hybrid algorithm with across strategy B and mutation strategy B is a simple and effective better algorithm than others.

d. Research on Optimization of Human-Skilled Matching of SMEs Based on Ant Colony Optimization Algorithm

Abstract: Human resource management plays a key role in the development of enterprises since the capital theory proposed. In this context, in this paper, based on the optimization of traditional ant colony algorithm, it is applied to enterprise human resource management, and its development mode optimization is studied. First of all, the ideas and advantages and disadvantages of the ant colony optimization algorithm are analyzed. Secondly, the optimized ant colony algorithm for human resource management is optimized, the Bayesian model of talent training constructed. Thirdly, through the expert assessment of the matching of people and posts, the ability index system is obtained. At last, the raw data of the algorithm test is used to test the algorithm. The test results show that the generalized regression ant colony algorithm optimized by ant colony algorithm is used to match the actual situation of the ant colony algorithm, which has certain advantages in solving such problems, playing a certain reference for the application of enterprise human resource management and computer algorithm.

- **Machine Learning Algorithm:**

- 1. Random Forest:**

- a. Prediction Analysis using Random Forest Algorithms to Forecast the Air Pollution Level in a Particular Location**

Abstract: To forecast the degree of air pollution in a specific area of a region using techniques Innovative Random Forest against Naive Bayes. Two groups of algorithms are Random Forest and Naive Bayes. The technique was developed and tested on a 32516-record dataset. In a programming experiment, each approach was iterated $N=5$ times to identify different levels of air pollution. The threshold value is 0.05 percent, and the confidence interval is 95%. The G-power test is around 80% effective. When compared to Naive Bayes, the innovative Random Forest method (98.26%) offers higher accuracy (97.32%). Random forest has the highest accuracy in comparison to the Naive Bayes algorithm. Significance value for accuracy is $0.056(p>0.05)$, Precision $0.02(p<0.05)$ and recall $0.01(p<0.05)$ based on 2-tail analysis.

- b. Heritage Properties Price Prediction Using Random Forest Classifier**

Abstract: Recently, machine learning has been gaining popularity for its use and because it holds a great promise for real estate valuation and prediction. However, the application of machine learning in predicting the prices of heritage property has limited adoption. Therefore, the present study aims to test the best machine

learning in predicting the price of heritage properties. This paper provides a report from the data collection method, preliminary data analysis with statistical method, and the experimental implementation of the machine learning algorithms from training and validating to the testing stage. The results show that the Random Forest model produced better performance on heritage property, which was tested with another five hold-out data. The testing results show that the machine learning approach with Random Forest was able to predict the prices of heritage property within the acceptable valuation.

c. Typical and Non-Typical Diabetes Disease Prediction using Random Forest Algorithm

Abstract: A non-communicable disease Diabetes is increasing day by day at an alarming rate all over the world and it may cause some long-term issues such as affecting the eyes, heart, kidneys, brain, feet and nerves. It is really important to find an effective way of predicting diabetes before it turns into one of the major problems for the human being. If we take proper precautions on the early stage, it is possible to take control of diabetes disease. In this analysis, 340 instances have been collected with 26 features of patients who have already been affected by diabetes with various symptoms categorized by two types namely Typical symptoms and non-typical symptoms. The purpose of this study is to identify the Diabetes Mellitus type accurately using Random Forest algorithm which is an Ensemble Machine Learning

technique and we obtained 98.24% accuracy for seed 2 and 97.94 % for seed 1 and 3.

d. Research on used car price prediction based on random forest and LightGBM

Abstract: In recent years, while the scale of the used car market has been expanding, the price evaluation system of my country's second-hand car market has exposed the problem that it does not meet the market demand. Accurate used car price prediction can help people make correct decisions and avoid the wanton price tag of used cars in the market as much as possible. This paper uses the random forest and LightGBM algorithms to predict the price of used cars and compares and analyzes the prediction results. The experiments found that the relevant evaluation indicators of the random forest and LightGBM models are as follows: MSE is 0.0373 and 0.0385 respectively; MAE is 0.125 and 0.117 respectively; The R square of prediction is 0.936 and 0.933 respectively. Among the two prediction models, the prediction error of the LightGBM model is smaller, and it can be considered to be applied to other fields in future research.

e. Research on Random Forest and Decision Tree

Abstract: In this paper, they have compared the classification results of two models i.e., Random Forest and the Decision Tree for classifying twenty versatile datasets. They took 20 data sets available from UCI repository containing instances varying from 148 to 20000. We compared the classification results obtained from methods i.e., Random Forest and Decision Tree (J48). The classification

parameters consist of correctly classified instances, incorrectly classified instances, F-Measure, Precision, Accuracy and Recall. We discussed the pros and cons of using these models for large and small data sets. The classification results show that Random Forest gives better results for the same number of attributes and large data sets i.e., with greater number of instances, while J48 is handy with small data sets (a smaller number of instances). The results from breast cancer data set depicts that when the number of instances increased from 286 to 699, the percentage of correctly classified instances increased from 69.23% to 96.13% for Random Forest i.e., for dataset with same number of attributes but having more instances, the Random Forest accuracy increased.

2. Support Vector Regression (SVR):

a. Ensemble SVR for prediction of time series

Abstract: Recently, support vector machine (SVM) as a new kernel learning algorithm has successfully been used in nonlinear time series prediction. To improve the prediction performance of SVM, we concentrate on ensemble method. Bagging and boosting, two famous ensemble methods, are examined in this paper. Experiments on two data sets (sunspots and Mackey-Glass) have shown that bagging SVR and boosting SVR could all improve the performance when compared with single SVR. For boosting, weighted median is a better choice for combining the regressors than the weighted.

b. Multi-parameter Adaptive Prediction of Chaotic Time series based on LS-SVR

Abstract: Nowadays, fault detect and prediction is quite important for the purpose of ensuring the correct functioning of complex system; nevertheless, it is usually difficult to establish an exact mathematical model in analytical form for complex system, therefore, fault prediction of complex system always relays on the analysis of the observed chaotic time series. In order to enhance the validity and accuracy of the prediction process, all relevant multi-parameter chaotic time series information is taken into consideration in this work. Then, multi-parameter phase space reconstruction process is performed to generate training samples; and a multi-parameter adaptive

prediction model using least squares support vector regression approach is established in the end. The proposed method is based on the support vector machine prediction theory. In this manuscript, the simulation experiment of chaotic time series with three parameters of certain equipment is investigated and presented for an illustration.

c. An EMD-SVR method for non-stationary time series prediction

Abstract: In the area of prognostics and health management, data-driven methods increasingly show the superiority against model-based method due to the complex relationships and learn trends available in the data captured without the need for specific failure models. This paper uses Empirical Mode Decomposition (EMD) and Support Vector Machine (SVM) to build a model for non-stationary time series prediction. And it proves that the EMD-SVR method can solve the problem of few training samples in modelling the path of performance degradation. Then when the threshold is given, we can forecast the lifetime of engineering systems based on the performance degradation.

d. SVR Modelling and Parameter Optimization for Financial Time Series Forecasting

Abstract: Compared with developed countries, my country's financial market is not mature enough. Market forecasting technology is still in the market cultivation stage and has not been widely used in actual economic life. Therefore, SVR modelling and

parameter optimization of financial time series (FTS) forecasting are studied. The subject is of great significance. Based on the relevant theories of FTS forecasting, this paper models SVR and conducts parameter optimization research. The principal component analysis method can be used to extract financial time series. The error between the predicted value and the actual value is small.