Aggregated Financial Forecasting Calculation in Human-Computer Distributed Computing

Petar Tomov, Gergana Mateeva, and Dimitar Parvanov

Institute of Information and Communication Technologies
at the Bulgarian Academy of Sciences
Acad. Georgi Bonchev Str., block 2, office 514, 1113 Sofia, Bulgaria
petar.tomov@iict.bas.bg gergana.mateeva@iict.bas.bg dimitar.parvanov@iict.bas.bg

Abstract: Three of the most famous financial forecasting donated distributed computing projects for the last decade were MoneyBee, GStock and MQL5 Cloud Network. The idea of a desktop computer screen-saver used for artificial neural networks training was extended in VitoshaTrade project where Android Active Wallpaper is used for the same purpose. The project implementation goes on in IICT-BAS and the forecasting capabilities are evolved in the direction of human-computer based distributed computing. This research proposes financial forecasting calculation based on users' classification and votes given for the particular financial future events.

Keywords: financial forecasting, distributed computing, voting platform

1. Introduction

In recent years, many financial management problems rely on mathematical modeling. This is a complex task due to the uncertainty of probability nature of the processes, large dimensionality and complexity of the systems themselves including also the influence of the human factor. A detailed survey of different evolutionary techniques used to solve multiple objective portfolio optimization problems is described in [1]. Financial forecasting makes possible to estimate or to predict the business performance in the future and this motives authors to develop different tools. Three of the most famous financial forecasting donated distributed computing projects for the last decade were MoneyBee [2], GStock [3] and MQL5 Cloud Network [4]. MoneyBee project was organized as donated

distributed computing screen-saver. Artificial neural networks were trained to forecast financial time series during desktop computers idle usage. GStock project used donated distributed computing for trading strategies testing. Many different investment strategies were checked over historical data and the best suited were applied for buy/sell signals generation. MQL5 Cloud Network is paid distributed computing platform in which trading strategies can be tested.

The idea of a desktop computer screen-saver used for artificial neural networks training was extended in VitoshaTrade [5] project where Android Active Wallpaper is used for the same purpose. The project implementation goes on in IICT-BAS and the forecasting capabilities are evolved in the direction of human-computer based distributed computing (crowdsensing [6]). The users are asked to vote [7] for the future change in the price of Forex currency pairs. Android operating system provides the ideal user interface capabilities for such voting by its widget components. By HTTP protocol the vote of the users is collected on a remote PHP-MySQL based server. As proposed in NSFDE&A'20, users' votes are classified in four categories: 1) High voting frequency and high guess rate users; 2) Low voting frequency and high guess rate users; 3) High voting frequency and low guess rate users; 4) Low voting frequency and low guess rate users.

This research proposes financial forecasting calculation based on users classification and votes given for particular financial future event. The rest of this paper is organized as follows: Section 1 describes the problems related to financial forecasting and states the ideas for computing in a distributed environment; Section 2 proposes a practical solution for a mobile distributed voting solution; Section 3 reveals some practical experiments and related results; and Section 4 concludes with some suggestions for further work.

2. Financial Time Series

Time series are values measured in strict time order [8]. In most of the cases, measurements are done in equal time intervals, but exceptions also do exist [9]. The result of such measurements is discrete time-ordered data. There are many natural processes that can be described with time series as average daily temperature, count of sunspots number of flu sick people, values of the prices and many others [10]. Time series are useful in almost every domain where measurements can be done in temporal order. Line charts are the most used tool for visualization of time series. A time series splitting to training examples are shown in Fig. 1.

Technical analysis of financial time series uses a group of methods for temporal data analyzing in such a way that meaningful statistics and other characteristics to be extracted. Financial time series forecasting is based on a model that predicts future values according to previously observed values.
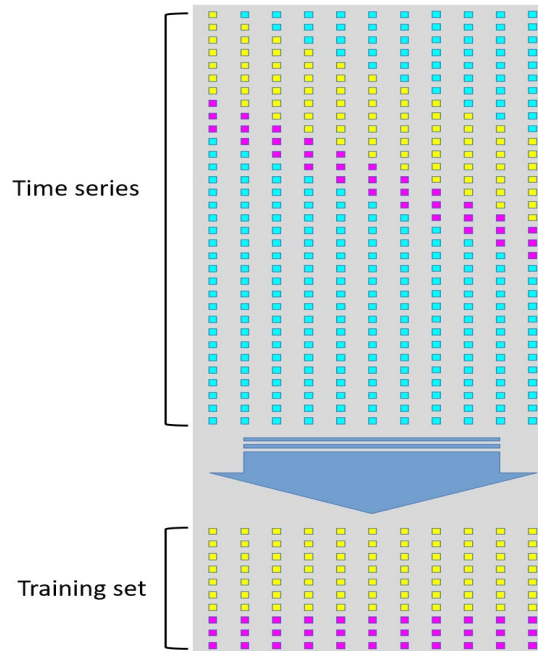
Fig. 1. Time series splitting to training examples

Financial time series usually have trend, seasonality and high-frequency oscillations [11]. The base of the technical analysis in financial time series is an appropriate curve fitting to the points which are already known.

3. Distributed Computing in Financial Forecasting

When a particular computation is a too much time consuming it is rational such calculation to be done on many computational machines in parallel [12]. When artificial neural networks are used for financial time series forecasting, the training can be organized with genetic algorithms [13] or differential evolution [14]. Both global optimization techniques are perfect for parallel computing and of course for distributed computing because global population can be separated in many local populations and the process of optimization can go simultaneously.

In the technical analysis time series are conditionally divided (Fig. 2) into past (lag) and future (lead) periods [15]. Based on past events (financial time series) it is possible to predict the feature trend if some kind of training is applied in advance. The process of training of the examples could be done by using of artificial neural network [15] to predict the output as shown in Fig. 3.

The training process on the local devices goes into two common phases, which are related to the way the population individual are represented. Artificial neural network topology is predefined and its optimization is beyond the scope of

this approach, but the weights of the artificial neural network are subject to optimization. The population of the evolutionary algorithm is organized as sets of artificial neural networks weights matrices.

Fig. 2. Financial time series

Fig. 3. Artificial neural network training
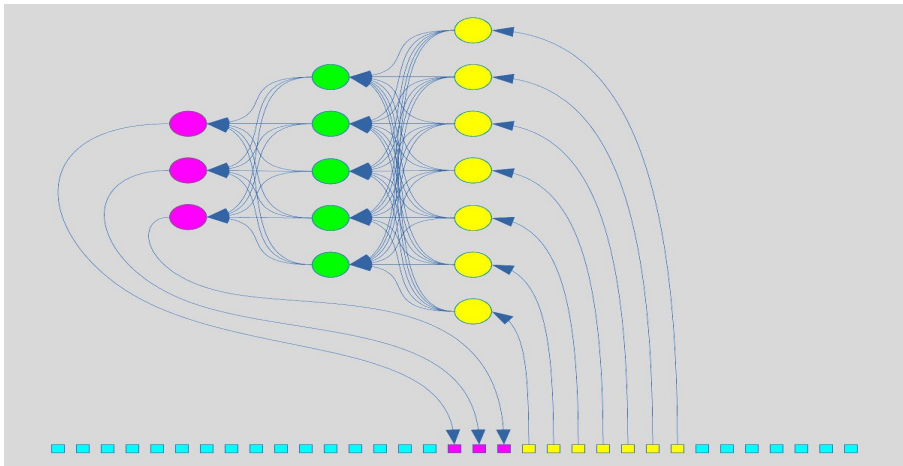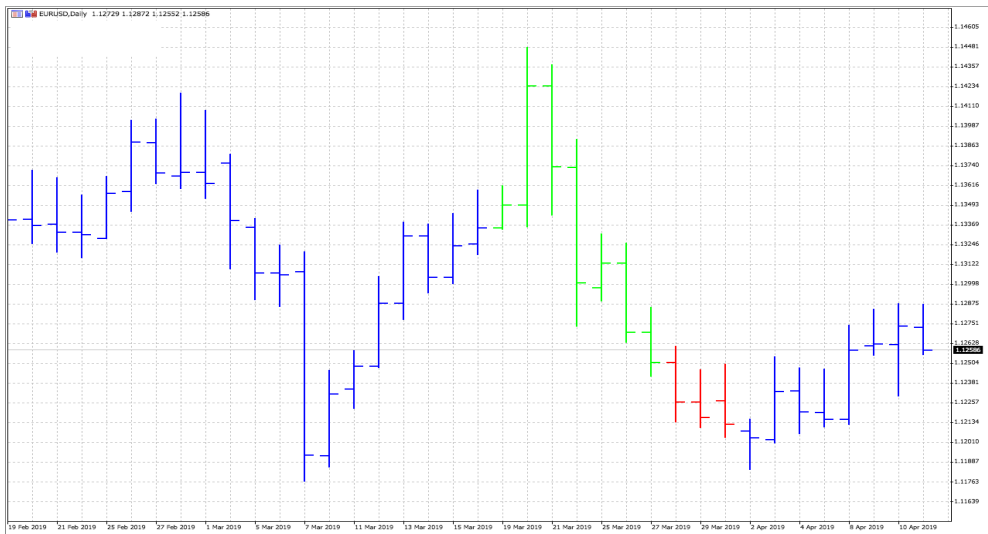
Each set of weights (population individual) is provided to the back-propagation training procedure until there is no significant improvement of the forecasting capabilities. After that crossover and mutation are applied and a new generation is added to the population. Uniform crossover is used as the best way to prevent total information destruction in an artificial neural network. Newly

created individuals are also trained with back-propagation algorithm and the evolution algorithm goes continuously.

Such symbolic restarting of the back-propagation training has wide similarity with simulated annealing. After the crossover and the mutation operators' weights of the artificial neural network are partially distressed and disordered. This helps the back-propagation training to proceed from a previous point of the training process and to escape possible local optimums. When this hybrid training is done on mobile devices the training process goes 24/7. The best-found weights of an artificial neural network are reported to the remote server. In this way, best-found individuals form a global evolutionary algorithm population.

4. Android OS Crowdsensing Solution

Distributed computing forecasting described in the previous section is very attractive and very cost-efficient, but it has a common disadvantage - subjective human opinion is not taken in any form. It is very well known that financial markets are extremely volatile. Global disasters like the COVID-19 pandemic are capable to put financial markets in a highly-unpredictable state. In such extreme situations, pure technical analysis is useless.

Subjective human opinion in the shape of human intuition can improve forecasting results significantly in situations where methods of pure applied mathematics are not applicable. Human intuition is not completely understood yet, but it is based on rich life experiences and subconscious information processing in the human brain. This research purposes aggregation of the information collected from a mobile device voting system and forecasting improvement with the aggregated data in use. A screenshot of the Android OS voting interface is shown in Fig. 4.
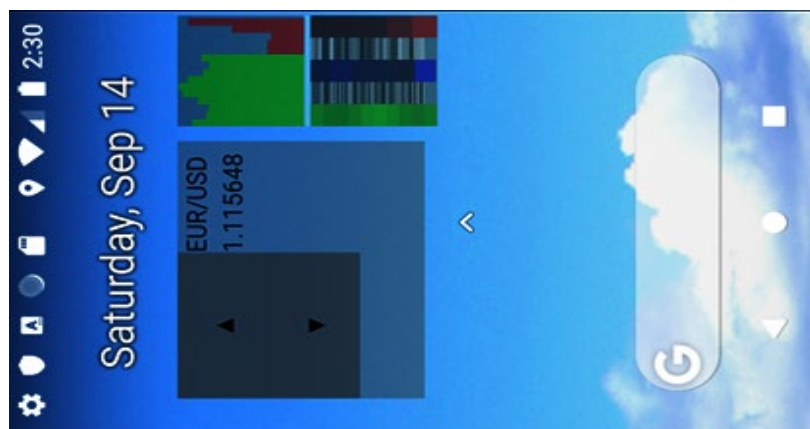

Fig. 4. Android OS voting interface

Users have the opportunity to vote for future changes in the particular financial time series. The vote is collected on a remote server. Voters are classified in four common groups: High voting frequency and high guess rate users (group a); Low voting frequency and high guess rate users (group b); High voting frequency and low guess rate users (group c); Low voting frequency and low guess rate users (group d). This classification is the base of the aggregated forecasting proposed in this research.

Each group of the voters participates with a different coefficient in the final forecast calculation and mathematically is expressed as follows:

$$f(x) = \frac{w_a}{n_a}\sum_{a=1}^{n_a} x_a + \frac{w_b}{n_b}\sum_{b=1}^{n_b} x_b + \frac{w_c}{n_c}\sum_{c=1}^{n_c} x_c + \frac{w_d}{n_d}\sum_{d=1}^{n_d} x_d \qquad (1)$$

where

$$N = n_a + n_b + n_c + n_d \qquad (2)$$

$$w_a + w_b + w_c + w_d = 1 \qquad (3)$$

The relation (2) represents the number of overall voting users from four groups (where the sum N is the number of total votes), while relation (3) expresses weights given to each of the four groups. The sum of the weights is chosen to be one and each single weight is between 0 and 1.

When a particular forecast should be issued single votes for the particular situation are taken and they are aggregate by groups to form the forecast. The forecast is in the form of chance the price to go up (value between 0.0 and 1.0, Eq. 1) and the opposite chance for the price to go down (also between 0.0 and 1.0). The sum of the two chances gives 1. The vector x in Eq. 1 is a binary vector. It has 1 for each vote of a user who thinks that the price will go up and it has 0 for each vote of the user who thinks that the price will go down. The total number of available votes for a particular forecasted event is N. The size of each group is given by the small letter n with an index.

Each group participate in the aggregated forecast with a given coefficient presented by the small letter w with and index. The values of the coefficients are subject to a multi-criteria problem and are chosen by the decision-makers.

5. Experiments and Results

For the experiments, EUR/USD pair is taken and ten separate events are selected for forecasting. For each event, 30 separate attempts for prediction are done. In Fig. 5 the obtained results are shown where each dot represents the average success.

The output of the forecasting model is a real number between zero and one. As close, the value is to one as stronger is the prediction that the price will go up.
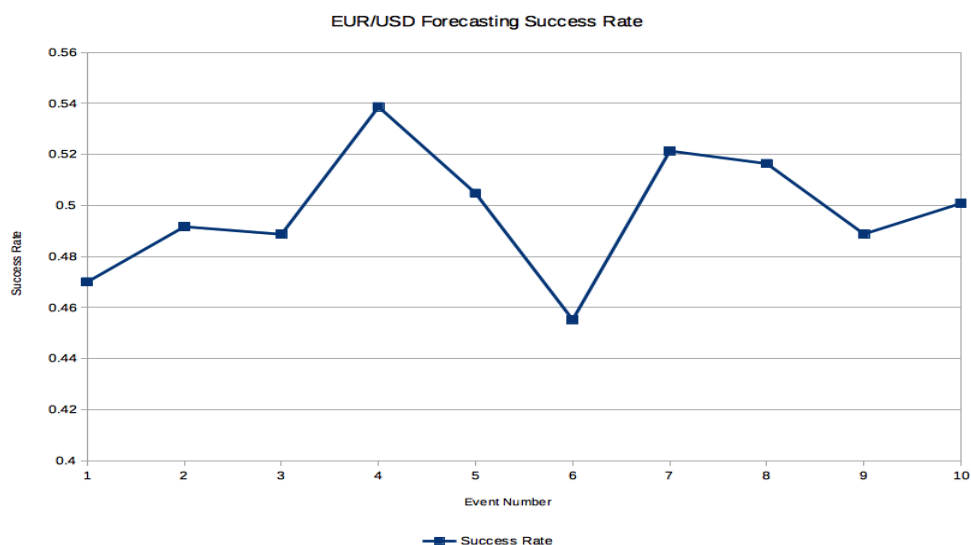
Fig. 5. Experimental results

As close is the value to zero as stronger is the prediction that the price will go down. As it is shown in Fig. 5 in half of the cases forecast was in the right direction. The experiments were done in a small closed group and that is why results are in an early stage of system testing. In order for better results to be achieved much bigger group of participants should be involved. It is essential people with high interest in trading to share their votes for a longer period.

6. Conclusion

The proposed aggregated financial forecasting based on human-computer mobile distributed computing shows to be very promising in situations where pure technical analysis has weaknesses. The main disadvantage is related to the lack of a strict scientific explanation of how human intuition works. Even with such a disadvantage, the proposed forecasting approach may be applied successfully in many practical situations.

As further directions of research, it will be interesting nonlinearities to be investigated in the way in which values of the four groups are calculated to a single prediction. Another interesting direction of further research is related to the estimation of multi-criteria coefficients according decision-makers preferences.

Acknowledgment

Communication Technologies for a Single Digital Market in Science, Education and Security (ICTinSES)", approved by DCM # 577/17.08.2018.

References

1. Stoyanova, K., Guliashki, V.: MOEAs for portfolio optimization applications. Lambert Academic Publishing, 42 pages (2018).
2. Bohn, A., Guting, T., Mansmann, T.: MoneyBee aktienkursprognose mit kunstlicher intelligenz bei hoher rechenleistung. Wirtschaftsinf 45, 325–333, (2003).
3. Chokesatean, P.: Credibility-based Binary Feedback Model for Grid Resource Planning. University of Pittsburgh, (2008).
4. Chang, C., Narayana Srirama S., Buyya, R.: Indie fog an efficient fog-computing infrastructure for the Internet of Things. Computer 50(9), 92–98, (2017).
5. Zankinski, I., Barova, M., Tomov P.: Hybrid approach based on combination of backpropagation and evolutionary algorithms for artificial neural networks training by using mobile devices in distributed computing environment. In: Proc. of Large-Scale Scientific Computing, pp. 425–432, (2018).
6. Leppanen, T., Alvarez Lacasia, J., Tobe, Y., Sezaki, K., Riekki, J.: Mobile crowdsensing with mobile agents. Autonomous Agents and Multi-Agent Systems 31, 1–35, (2017).
7. Laukkanen, S., Kangas, A., Kangas, J.: Applying voting theory in natural resource management A case of multiple-criteria group decision support. Journal of Environmental Management 64(2), 127–137, (2002).
8. Bandt, C., Pompe, B.: Permutation entropy a natural complexity measure for time series. Physical Review Letters 88(17), 174102, (2002).
9. Salvador, S., Chan, P.: Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis 11(5), 561–580, (2007).
10. Lan, Y.: Computational approaches for time series analysis and prediction data-driven methods for pseudo-periodical sequences. University of Bradford, (2009).
11. Kolev, K., Sevova, J., Ketipov, R., Blagoev, I., Petrov, P., Kostadinov, G., Zankinski, I.: Removal of linear component and sinusoidal harmonics from time series with differential evolution. In: Proc. Annual University Scientific Conference of Vasil Levski National Military University, pp. 1586–1594, (2019).
12. Megiddo, N.: Applying parallel computation algorithms in the design of serial algorithms. Journal of the Association for Computing Machinery 30(4), 852–865, (1983).
13. Donate, J.P., Li, X., Sanchez, G.G., Miguel, A.S.: Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm. Neural Computing and Applications 22, 11–20, (2013).
14. Balabanov, T. Zankinski, I., Dobrinkova, N.: Time series prediction by artificial neural networks and differential evolution in distributed environment. In: Proc. of Large-Scale Scientific Computing Conference, vol. 7116, pp. 198–205, (2012).
15. Tomov, P., Zankinski, I., Balabanov, T.: Training of artificial neural networks for financial time series forecasting in android service and widgets. Problems of Engineering Cybernetics and Robotics 71, 50–56, (2019).