Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 55 (2015) 1165 – 1173

Procedia Computer Science

Information Technology and Quantitative Management (ITQM 2015)

Double layer interval graph model: the universal tool for data driven market analysis and forecasting

Nikolay V. Markov*, Andrey V. Dmitriev, Svetlana V. Maltseva

National Research University Higher School of Economics, 33 Kirpichnaya st., Moscow,105187, Russia

Abstract

This scientific work is dedicated to the development, improvement and application of double layer interval weighted graphs (DLIG) for non-stationary time series forecasting. This model appears to be the universal and easy-to-use tool for modeling the non-stationary time series and forecasting. We observe the double layer version of the model because it's the most representative way in the sense of main idea though you can add several layers more for different purposes. The first layer of the graph is based on empirical fluctuations of system and displays the most potential fluctuations of the system at the time of system training. The second layer of the graph as a superstructure of the first layer displays the degree of modeling error and it's connected with the first layer nodes by edges. The second layer is the way of supervised training implementation with the aim of error minimization.

Peer-review under responsibility of the Organizing Committee of ITQM 2015

Keywords: market analysis, machine learning, DLIG, non-linear dynamics, forecasting

1. Introduction

Today market specialists need the universal tool for market analysis especially in time series forecasting. The reason for this is the presence of numerous factors of probabilistic nature in time series, which are often not possible to account because of ignorance of the laws of random variables distribution or cumbersome computational procedures. On the other hand, the risk factor is a strong incentive for saving money and resources and optimization of all business processes.

Market risks are associated with fluctuations in prices for different kinds of goods, services, financial instruments, and more. In this regard, in the conditions of market crisis, characterized by sharp price

* Corresponding author. Tel.: +7-985-306-26-83;.
E-mail address: nmarkov@hse.ru.

Peer-review under responsibility of the Organizing Committee of ITQM 2015
doi:10.1016/j.procs.2015.07.090

fluctuations, the major problem is the choice of the mathematical model for the evaluation, analysis and prediction of non-stationary time series.

There are lots of statistical and structural models of analysis and forecasting of time series [1], however, due to the low efficiency of most of them in conditions of unstable markets there are some interesting and innovative approaches to forecasting such as, for example, models based on fuzzy logic and Markov chains [2], [3], [4]. The main purpose of both statistical and structural prediction models is the minimization of the prediction error [5]. We are interested in dynamic models (as a part of structural models) based on the differential equations. We are also interested in dynamic chaos modeling [6] and in adaptive models, that are capable, based on the training set, independently configure themselves to identify explicit and implicit "preconditions" to sharp fluctuations inside of the time series.

For example, let's overview a mathematical model based on fuzzy logic. It involves the use of nonstandard methods of approximate calculations. In this model, the universal set, formed from the values of the time series, can be divided into intervals. The resulting intervals are fuzzy sets (A_i) with the certain logical relations for them. The disadvantage of this approach is the lack of well-defined, well-grounded algorithm for estimating the optimal number of intervals for a specific task. The user model defines a number of intervals divided oscillation amplitude values of the time series by himself. In addition, the length of intervals of the plurality of universal set affects the prediction accuracy [7], [8].

During the research of the model described below, it was decided to borrow some effective techniques of modern machine learning models, including neural networks and fuzzy logic models. Thus, for example, to break the universal set, we used the scale of "1-9", proposed in [9]. The reasoned usage of scale "1-9" in the conditions of our problem helped us to break the universal set into 18 intervals; later (after the implementation of a software product) that gave us the acceptable ratio of CPU time used, as well as the accuracy of modeling and forecasting. It should be noticed that there are no clear recommendations on the conditions of division into intervals. T. Saaty applied the nine-level ranking to the hierarchy analysis process in decision-making, not for time series analysis. The next step in the modeling of fuzzy time series is the formation of fuzzy logical relationships ($A_i \to A_j$) and their grouping scheme.

$$\left\{\begin{matrix} A_i \to A_{j1} \\ A_i \to A_{j2} \end{matrix}\right\} \Rightarrow A_i \to A_{j1}, A_{j2} \tag{1}$$

In continuation it should be said that the goal of this research is not only to design a new approach to time series forecasting, but also to design its practical implementation.

2. Algorithm of double layer interval graph training

The basis of the proposed prediction model is system training based on structural features of the time series (training set).

2.1. Primary training

Consider the non-stationary time series $\{F(t_0), F(t_1), ..., F(t_n)\}$. Nonstationarity of time series is due to the high financial indicators fluctuation amplitude on financial market, characterized by rapid changes and price shocks. We form the universal set $\{\Delta F_r(t_0), \Delta F_r(t_1), ..., \Delta F_r(t_n)\}$ from the set of relative variations of adjacent values of time series, that can be counted using formula: