Analysis and Design Insights for an E-Finance Platform Using Parallel Processing

DARIE MOLDOVAN¹, MIRCEA MOCA¹, MIRCEA RUSU²

¹Business Information Systems Department
Babeș-Bolyai University of Cluj-Napoca
58-60, Teodor Mihali Street
²IT Research Department
Company for Applied Informatics Cluj-Napoca
107, Republicii Street
ROMÂNIA
{darie.moldovan; mircea.moca}@econ.ubbcluj.ro, dmrusu@cianet.ro

Abstract: In this paper we provide the analysis for an e-finance platform and propose a design solution using state-of-the-art models in finance and parallel processing. The problem that we address lays in the context of financial markets that are rapidly changing and the majority of investors are facing important problems such as: the lack of integrated, advanced tools to analyze data, high-level services that automatically react to market changes and propose investment alternatives in near real-time and the lack of mechanisms to integrate clean data from multiple sources. These issues lead, in most cases, to poor investment decisions. Hence, in this paper we proposed the design of an e-finance framework to provide a set of services that tackle these issues. The proposed services rely on computing-intensive methods like: text-mining, Neural Networks and Genetic Algorithms, enhanced by applying relevant findings from the efficient-market theory study. In order to improve the execution of these services we propose their parallelization on distributed computing infrastructures. In our discussion we highlighted the importance of such a framework for the financial market investors.

Key-Words: E-finance platform, financial models, finance, e-services in finance, computing-intensive methods, parallel processing, system analysis, system design.

1 Introduction

Nowadays, while the conditions in the financial markets are rapidly changing, the majority of investors are facing important problems such as: the lack of integrated, advanced tools to analyze data, high-level services that automatically react to market changes and propose investment alternatives in near real-time and the lack of mechanisms to integrate clean data from multiple sources. These issues lead, in most cases, to poor investment decisions.

The topic of this paper is based on the objectives of the European strategy for the time period 2014/2020 [1] that has the following three priorities:

- Intelligent and sustainable growth by developing a knowledge and initiative-based economy
- Sustain a competitive and efficient economy that relies on its own resources
- Social and territorial cohesion support

In this context, the proposed framework design comprises a collection of infrastructure modules and financial models exposed as e-services. The financial, knowledge-based models are intended to leverage on the parallel processing power, in order to improve their execution. Virtually, a percentage of 80-90% of the investors are losing money when entering into risky financial transactions – including stocks, futures, options, FOREX [2,3]. Therefore we consider this area of optimizing trading and portfolios as being in need of independent, professional and consistent resources of information in order to improve the odds of not losing money in investments.

Not only the individual, retail investors are taking poor decisions when trading, but the proportion applies also to corporate investors [3]. Even though, as a whole, the corporate sector is a winner. It means that a condition for succeeding is the access to the latest novelties in the financial research area, for capturing the fast changing markets and acting promptly. The access to this knowledge is usually reserved for the large capitalized banks, which are capable to support large research departments and invest in their training and development. Even in some of these cases, adverse conditions of the markets are able to cause high damages to top companies (like
difficulties encountered by Lehmann Brothers and SocieteGenerale in recent years).

The financial markets are some of the most competitive markets in the world, and for the smaller players that do not have access to the above described resources (strong research and/or large capital) it is very difficult to perform reasonably. Our aim is to identify the main research challenges in the aim of building an e-finance framework to integrate sophisticated services for the financial markets investors to facilitate their investment decisions. For this, we provide elements of the system analysis and briefly present the sketch of a proposed design solution.

The remaining of this paper is organized as follows. Section 2 provides some insights regarding the financial markets issues on which we base our system proposal, while Section 3 presents the analysis of the system from both financial and technical perspectives. In Section 4, we design the proposed solution. Finally we conclude and present future work.

2 Background

2.1 Informational efficiency
Considering the efficient-market hypothesis built by Fama [4] a market is fully efficient when it reflects all the information available. The prices on the market will remain in equilibrium until new information will be available, having in this way a random walk behavior. It becomes very difficult for an investor to build a successful strategy under these conditions, as one can consider the price prediction as impossible since the real price of an asset is already the one in the market.

But the efficiency of the markets can also be scaled on different levels of efficiency, making a more precise differentiation between the various markets and providing useful information to the investors (either retail or corporate investors) [5]. The strong-form efficiency asserts that all the information is comprised in the price and no excess return can be obtained on the long term. The semi-strong form of efficiency implies that the prices are rapidly adjusting to the new information and on the long term neither the technical analysis or the fundamental analysis are useful for obtaining excess returns. In weak form efficiency the historical prices are not considered as reliable grounds for predicting the future prices, but fundamental analysis could be applied in order to obtain excess returns on the long term.

Different studies [6, 7, 8], suggest the different intensity in time of the efficiency in the markets. Strong correlations periods are alternating with random walk periods. This behavior in the markets needs different approaches from the investing strategies point of view and intensive monitoring and computation. The investors can take profit of the inefficiencies in the markets only if they are able to quickly design trading strategies. For those adopting the technical analysis for the trading strategies, the rapid data analysis and models calibrations are issues needing to be solved in a time efficient manner. This proves to be a very difficult task for the individual investor, as he or she does not possess the infrastructure for complex calculations.

2.2 Financial forecasting algorithms
In this section we give a background on financial forecasting algorithms and approaches for their parallelization as found in literature. Based on the following discussion we will argue the proposed design solution in Section 4.

As we discussed above, our proposed services rely on computing-intensive data mining methods like GAs, text-mining and Neural Networks (NNs) [9]. In order to meet business requirements (in terms of responsiveness to financial markets changes and reliability of the service) we need to minimize the execution time of the services. That is, we need to distribute their execution and, therefore, the challenge is to parallelize the execution of the proposed algorithms.

The literature presents two common approaches for parallelizing such applications, based on their specifics: MapReduce and Bag-of-Tasks (BoT). While MapReduce computations are characterized by handling large volumes of input and intermediate data, the BoT applications have low requirements in terms of communication bandwidth or disk storage and tasks are independent of one another [10].

In [9] the authors characterize the data mining applications as relying on computing-intensive processing tasks running on huge datasets. Hence, these applications are natural candidates for execution on high performance and throughput infrastructures like clusters and computational grids. Many data mining algorithms can be implemented as BoT applications. Thus, text-mining and NNs are plausible scenarios for BoT [9, 11]. On the other hand, the GAs can be parallelized using the MRPGA model, an extension of MapReduce for Genetic Algorithms [12]. We consider these approaches suitable for our proposed framework, since we could rely on previous experience [10] in working with MapReduce.
3. System analysis
3.1 Financial perspective

We identified several categories of shortcomings that lead to the problem that a significant proportion of the financial market investors make poor and costly decisions. These categories concern the decisional process, data, software tools and action on financial market. During the decisional process for investing, some difficulties arise, some of them having as factors of influence psychological reasons [13, 14] with sentiments like greed or fear, or some technical factors such as very general software tools, not adapted to the investor needs and the lack of knowledge in the field of financial mathematics and statistics. All these factors are making the decisional process hard to control. The knowledge needed in order to enter in trading as an informed investor cover areas like financial economics, financial mathematics, experimental finance, behavioral finance. All these areas are difficult to be covered by individuals and even by some companies, leading to a lack of solid basis, which will ultimately affect their decisions. The financial markets environment is always changing, and trading strategies need to adapt in order to perform in a consistent manner. This need can be satisfied only by strong computational mechanisms.

Software tools such as AIQ Systems [15], Amibroker[16], BestDirect[17], eSignal[18], invesTools[19], Metastock[20], Tradestation[21] were built in order to facilitate the decision making based especially on technical analysis, but it takes time to understand and learn their mechanisms, especially if we consider the above mentioned factors that are disturbing the decision process. More, most of the software tools represent an unsorted on performance group of indicators, leaving the uninitiated investor (which represents 80-90% of the total number of investors) to chose from he’s own perspective what to consider for decision. That leads to a lack of advanced customer oriented services. One can hardly find an integrated analysis system, as most of them are built considering just a few branches of the domain, their integration needing to be done by every investor, in relation with his own knowledge about the importance or usefulness of the tools.

The data used in analysis consists on multiple sources available to the public, but some of them difficult to access, especially due to high costs of access for the individual investor (specialized financial news portals very useful in data aggregation). With the abundance of news and information, the investor faces a lack of clean information to fundament he’s decisions. Filtering the useful information out of noise and correlating the information from different sources of data in order to obtain a clean image for decision taking becomes a very complicated task, which can be done using specific data mining techniques [22, 23, 24]. The large datasets involved imply powerful tools for aggregating, processing and delivering, like distributed systems that can easily multiply the power of computation.

3.2 Technical perspective

To address the previously described problems, our goal is to design a framework that integrates a set of financial services. This could be accessed by investors in real time, in order to get investment alternatives based on three sources of data: financial markets, the Internet and the investor’s portfolio. In Figure 1 we present a high level data-flow representation between four main types of entities: the financial services company, the Internet, the Investors and the financial markets. On one hand, the company aggregates financial market information from the Internet as unstructured meta-data related to financial markets specific behaviors.

Figure 1 High-level data-flow diagram for the proposed solution.

As particular sources of data in this category, we can imagine news historical databases and social network flows. On the other hand, the company would aggregate historical data from financial markets. Based on large data-sets aggregated from these sources, the company would use the e-finance platform to perform advanced financial analysis in order to obtain valuable information concerning the predictability of the markets. Further, for particular portfolios descriptions, the company would provide its investors (service consumers) with valuable investment alternatives. Consequently, investors make decisions and interact with the financial markets.
The main challenge from the technical perspective is the design and development of a scalable, open system that is able to gather, store and process large amounts of heterogeneous data in near real-time. From this perspective, we identified research challenges in the following areas:

- **Data gathering**, which requires the design of efficient, low-cost and flexible means of collecting both product and process data from the earliest stages of the production (i.e., feed quality); also collecting context information (i.e., time, localization, etc.) in order to perform context-aware data analysis will be considered.
- **Data integration**, which requires the design of efficient data pre-processes like data cleansing, de-duplication, normalization in order to obtain high-quality integrated data; Data storage.
- **Data analysis**, which requires the design and development of domain-specific multidimensional analysis models (OLAP);

### 3.2.2 Distributed Computing Infrastructures

The envisioned solutions for the proposed financial services are computing-intensive, since they employ genetic algorithms (GAs), neural networks or data-mining (DM) algorithms. For these algorithms, the literature provides us with models of parallelization on distributed computing infrastructures. Hence, [12] presents the MRPGA model (MapReduce for Parallelizing Genetic Algorithms) used to parallelize GAs, [9] proposes a bag-of-tasks (BoT) approach to run DM applications on computing grids and N2Grid is a framework that allow neural nets to use grid resources [25]. Thus, we propose to build distributed solutions for the services, so we can leverage on Distributed Computing Infrastructures (DCI) like clusters, grids, clouds or desktop grids.

### 4. Design solution

**Powerful algorithms** are needed into an ever changing environment where the individuals involved are competing for profit. The **adaptability of the algorithms** to the new conditions of the markets is a must in order to obtain consistent profits. A certain strategy working in a certain period of time will surely become useless relatively shortly.

Our approach is to select from well known but also state of the art technical indicators combinations that prove to work in a consistent manner during certain market conditions. We will choose a variety of combinations between the parameters of the indicators in order to determine a satisfactory solution from a business point of view. The artificial neural nets will be trained using these parameters as inputs and their scope will be to obtain estimations for the future price of a certain asset.

Since the training of a neural net can be very time consuming due to the high complexity, **powerful resources** for computation are needed. One possibility is to use the **GRID distributed capabilities** in order to obtain the results in a reasonable manner for business purposes.

In the last decades, a constant preoccupation for the portfolio managers was to build up indexes in order to **increase their performances** in the market [26]. The indexes can be used in portfolio management or (creating a portfolio structure and performance benchmark) or for the estimation of the trend of a certain component. Our aim is to **obtain a clusterization of the financial markets** by identifying the securities with a resembling evolution during certain periods of time and by comparing them with the evolution of indexes built by experts in the finance field, to **analyze how advisable is to decide to invest** based on these indexes.

A first application based on this idea was realized on the American index Dow Jones Composite Average, and its sub-indexes using a method for clusterization named Gene Trajectory Clustering (GTC), first used in the gene trajectory study [27, 28]. The results demonstrated the **validity of using the method in the financial area**, by obtaining valid clusters considering the objectives proposed the identification of homogeneous and stable clusters, useful for the investors [29, 30]. Financial news plays an important role in the market trends generating tremendous volatility when unexpected events occur. The investment decision has always had a component of “What other people think” [31], investors paying attention to advices and rumors from all sorts of sources: from friends, specialists, newspapers, television and more recently from the internet. But not always the markets have a rational behavior, to react positively to good news and negatively to bad news. It happens for the markets to react completely irrational, causing more uncertainty for the investors [32]. Important questions arise: is this happening because already settled sentiment of the investors? Is it possible for bad news not to influence the trend because a positive sentiment was already established in the markets?

The objective of our platform is to **determine the market sentiment** from the financial news
during certain periods, to determine its intensity and to overlap it to the market trend, in order to find anomalies or indicators of trend reversals.

The methods for achieving the objective would imply specialized text mining tools to extract keywords from the news that can be associated to market sentiments. The mining algorithms will parse the financial news, whether the news has or not specific keywords associated. The developing of intelligent algorithms for auto learning will be an important task, in order to be able to obtain results in a time efficient manner, a critical factor in the financial markets. This process can be very resource-consuming, possibly needing enhanced processing capabilities, like distributed computing [33].

**Acknowledgments**

This work was partly supported by CNCSIS IDEI II, 2596/2009 and TE_316 Grants.

**References:**


