Automatic trading method based on piecewise aggregate approximation and multi-swarm of improved self-adaptive particle swarm optimization with validation

Rodrigo C. Brasileiro, Victor L.F. Souza, Adriano L.I. Oliveira*

Universidade Federal de Pernambuco, Centro de Informática, Av. Jornalista Aníbal Fernandes, s/n, Recife 50740-560, PE, Brazil

1. Introduction

A time series is a set of observations ordered according to some parameter [1]. Financial time series, for example, represent stock prices in the financial market and follow a chronological order [2]. This kind of data can be used for analyzing and understanding the past and forecasting the future, allowing information users to improve their decision-making. Besides, time series analysis also aims to identify patterns in the data and distinguish them from random variations. These tasks of time series analysis combined with the high computing power available nowadays, have made it a tool widely used in several economic sectors, such as government, industry and commerce [3].

Financial time series represent the stock prices over time and exhibit behavior similar to a data stream. Many works report on the use of data mining techniques to predict the future direction of stock prices and to discover patterns in the time series data to provide decision support for trading operations. Traditional optimization methods do not take into account the possibility that the function to be optimized, namely, the final financial balance for operations considering some stock, may have multiple peaks, i.e., be represented by multimodal functions. However, multimodality is a known feature of real-world financial time series optimization problems. To deal with this issue, this article proposes the PAA-MS-IDPSO-V approach (Piece-wise Aggregate Approximation - Multi-Swarm of Improved Self-adaptive Particle Swarm Optimization with Validation). The proposed method aims to find patterns in financial time series to support investment decisions. The approach uses multi-swarms to obtain a better particle initialization for the final optimization phase since it aims to tackle multimodal problems. Furthermore, it uses a validation set with early stopping to avoid overfitting. The patterns discovered by the method are used together with investment rules to support decisions and thus help investors to maximize the profit in their operations in the stock market. The experiments reported in this paper compare the results obtained by the proposed model with the Buy-and-Hold, PAA-IDPSO approaches and another approach found in the literature. We report on experiments conducted with S&P100 index stocks and using the Friedman Non-Parametric Test with the Nemenyi post-hoc Test both with 95% confidence level. The results show that the proposed model outperformed the competing methods and was able to considerably reduce the variance for all stocks.

1. Introduction

Data mining is the process of “making better use of data” and is founded on the theory that historical data store information that can be used to predict future behavior [4]. In the financial market scenario, data mining can be used to discover hidden patterns in historical data time series, which can represent their behavior, including trends, seasonality and other information. Thus, these patterns can be used to advise on purchase and sale of stocks and thereby assist investors in their decision-making [2].

High dimensionality (a large number of data points) and continuous updating are among the main features of financial time series data. As a consequence of the high dimensionality, the increase in runtime and storage space are some of the main problems that data mining techniques must tackle. One of the most common approaches to dimensionality reduction is to transform the time series to another domain. This transformation may enable a faster computation of the similarity among processed data as if they were the original time series data, yet at a lower computational cost [5].

We have recently proposed PAA-PSO, a data mining technique applied to time series for stock trading [6]. This approach is a model
that combines Piecewise Aggregate Approximation (PAA) and Particle Swarm Optimization (PSO) to discover the best representative pattern of time series, which is, the pattern that obtains the best financial results in trading. This pattern is used in conjunction with trade rules to automate the decision on buying, selling or keeping the stock. We have shown that PAA-PSO obtains equal or better results with a lower computational cost when compared to SAX-GA, which combines Symbolic Aggregate Approximation (SAX) with Genetic Algorithm (GA) for trading stocks [7]. We have also shown that PAA-PSO outperformed the Buy-and-Hold strategy, in which an investor buys the stock and keeps it in his investment portfolio for a long period. This strategy is widely used as a reference for comparison with proposed models in studies applied to the financial market [7–11].

However, both PAA-PSO and SAX-GA produce a high variance in their results, i.e., when running n times the same experiment, the results vary significantly. In practice, this generates poor results for a decision support system since such variability in the results represents a high risk in the transactions carried out automatically by the system or indicated to the investors. Thus, the primary objective of this paper is to introduce a method that reduces the variance, in addition to providing profits better than those obtained by PAA-PSO [6].

To this end, this paper proposes the PAA-MS-IDPSO-V technique. PAA-MS-IDPSO-V combines (i) the Piecewise Aggregate Approximation (PAA), for time series representation and dimensionality reduction, (ii) a Multi-Swarm of Improved Self-adaptive Particle Swarm Optimization algorithm (IDPSO) with the early stopping criterion [12] (iii) and the use of validation set [13].

The PAA-MS-IDPSO-V is responsible for discovering patterns that are used by the proposed business rules as investment strategies serving as a decision support system to automate operations in the stock market (decisions about the best moments for buying and selling stocks) and help investors maximize their profits with controlled risk. The proposed method utilizes (i) multi-swarms and (ii) validation set aiming to control the risk (reduce the variance of the results). Multi-swarm is a technique that uses various swarms simultaneously to deal with multiple peaks (often used in multimodal functions, typical of real-world problems) [14]. The validation set and early stopping criterion are used to avoid the overfitting of the model [15].

We report on experiments carried out using stocks from the S&P100 aimed at evaluating the effectiveness of the proposed model. Friedman Non-Parametric Tests with the Nemenyi post-hoc Test with 95% confidence level were performed on the results to compare the proposed method to the PAA-IDPSO [6] and the approach introduced by Teixeira and Oliveira [9] regarding profitability.

This paper is organized as follows: Section 2 discusses decision support systems, time series, optimization algorithms - specifically PSO, IDPSO and multi-swarms approaches - validation and the early stopping criterion. Section 3 presents the proposed method. In Section 4 the experiments are reported, and the results are analyzed. Section 5 presents the conclusions.

2. Fundamentals

This section briefly reviews the main areas of research of this paper (i) decision support systems, (ii) representation technique used in time series data mining, i.e., the PAA approach, (iii) optimization based on PSO algorithms, and also (iv) use of validation set and early stopping criterion.

2.1. Decision Support Systems for stock trading and data streams

Decision Support Systems (DSSs) are computer systems that analyze and compile a significant amount of data and are used to assist users in their decision-making [16]. DSSs are used in many areas, including financial and stock trading systems [16–21].

Zhang et al. [17] proposed a DSS based on an aggregate ensemble learning framework used for mining noisy data streams without preprocessing the data. Kao et al. [18] proposed a DSS for forecasting stock prices based on a hybrid approach integrated by wavelet-based feature extraction with Multivariate Adaptive Regression Splines (MARS) and Support Vector Regression (SVR). Geva and Zahavi [19] proposed an automated intraday stock recommendation system that incorporates both market data and textual news. Oliveira et al. [20] proposed an automatic approach, called Lexicon acquisition, to perform sentiment analysis in financial market through microblogging messages. According to Shynkevich et al. [21], the market changes when new information is disclosed, e.g., information derived from news articles, which affect the decisions made by investors. In this context, Shynkevich et al. proposed a decision support system capable of reading these news articles simultaneously, providing different degrees of relevance to the information based on the sector of the financial market that one wishes to operate.

2.2. Time series data mining

2.2.1. Piecewise Aggregate Approximation (PAA)

Piecewise Aggregate Approximation (PAA), proposed by Keogh et al. [22], is an approach used for data representation and dimensionality reduction in time series data mining. In this method, a time series window of size n is divided into k segments of equal length, and the average value of the data of the segments is then used as the representative value of each segment. Hence, a time series PAA representation will be a k-dimensional vector of the means values of each segment. Fig. 1 depicts an example of a PAA representation.

PAA is performed in two steps [22]. Initially, the original time series window data must be standardized. The purpose of this step is to convert the data to the same relative amplitude, keeping the original form of the data. The statistical standardization is computed via Eq. (1).

\[ x'_i = \frac{x_i - \mu_k}{\sigma_k} \]  

Fig. 1. PAA representation of a time series Q. In this example, PAA parameters are n = 15, k = 3.
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات