Expert Systems With Applications 000 (2016) 1–17

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Clustering stock price time series data to generate stock trading recommendations: An empirical study

Binoy B. Nair[a,*], P.K. Saravana Kumar[a], N.R. Sakthivel[b], U. Vipin[a]

[a] Department of Electronics and Communication Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, India

[b] Department of Mechanical Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, India

ARTICLE INFO

Article history:
Received 30 May 2015
Revised 17 August 2015
Accepted 1 November 2016
Available online xxx

Keywords:
Stock
Trading
Recommender
Clustering
Time-series

ABSTRACT

Predicting the stock market is considered to be a very difficult task due to its non-linear and dynamic nature. Our proposed system is designed in such a way that even a layman can use it. It reduces the burden on the user. The user's job is to give only the recent closing prices of a stock as input and the proposed Recommender system will instruct him when to buy and when to sell if it is profitable or not to buy share in case if it is not profitable to do trading. Using soft computing based techniques is considered to be more suitable for predicting trends in stock market where the data is chaotic and large in number. The soft computing based systems are capable of extracting relevant information from large sets of data by discovering hidden patterns in the data. Here regression trees are used for dimensionality reduction and clustering is done with the help of Self Organizing Maps (SOM). The proposed system is designed to assist stock market investors identify possible profit-making opportunities and also help in developing a better understanding on how to extract the relevant information from stock price data.

1. Introduction

The price of a stock keeps changing depending on the supply and demand of the stocks. Due to the extremely nonlinear nature of stock price movements, forecasting the stock prices and timing the buy/sell decisions becomes an extremely challenging task. It is this risk that tends to keep a vast majority of people from trading in stocks. In this study, a stock trading recommender system that learns patterns from the historical stock price data and recommends when to buy/sell stocks and thus, can help the laypeople invest in equity markets profitably is proposed.

Initial studies carried out in the past century questioned the very predictability of stock markets (Cowles, 1933), (Cowles, 1944). Several other studies claimed the stock price movements to be random (Cootner, 1964), (Fama, 1965), with the efficient market hypothesis (Fama, 1970) ruling out any possibility of making excess returns from the market. However, recent studies, e.g. (Atsalakis & Valavanis, 2009), (Nair & Mohandas, 2015a), (Brabazon, O' Neill, & Dempsey, 2008), (Brabazon & O' Neill, 2006) and (Nair & Mohandas, 2015), have indicated that it is in fact possible to forecast the stock price movements and use the forecasts to generate excess returns. It is also observed that application of data mining techniques tend to generate good forecasting accuracy, as in (Nair et al., 2011), (Atsalakis & Valavanis, 2009), (Nair, Mohandas, & Sakthivel, 2011) and (Nair, Minuvarthini, Sujithra, & Mohandas, 2010). Different aspects of data mining have been explored in the earlier attempts at design of stock trading recommender, e.g. in (Nair & Mohandas, 2015b), a classifier based recommender system is proposed while (Nair et al., 2015) proposes a recommender based on mining of temporal association rules. However, it was observed that clustering based stock trading recommender systems have not been explored. In the present study, an attempt has been made to design a clustering based stock trading recommender system that can employ the historical stock price data to generate buy/sell recommendations.

Technical analysis has traditionally been very popular with stock traders and is still widely used to forecast stock price movements. Technical analysis involves identification of future trends in stock price movements based on the historical stock price values. There are a large variety of technical indicators in use today (Eng, 1988), however, selection of the optimal set of technical indicators for the given market conditions and identification of the optimal technical indicator parameters is quite challenging, thus limiting their utility. There have been a few studies that attempt to identify the optimal technical indicator parameters, such as in (De Brito &

* Corresponding author.
E-mail addresses: b_binoy@cb.amrita.edu (B.B. Nair), pksaravanakumar@gmail.com (P.K.S. Kumar), nr_sakthivel@cb.amrita.edu (N.R. Sakthivel), vipunne@gmail.com (U. Vipin).

http://dx.doi.org/10.1016/j.eswa.2016.11.002
0957-4174/

Oliveira, 2012), (Bodas Sagi, Soltero, Hidalgo, Fernández, & Fernández, 2012), (Soltero, Bodas, Hidalgo, Fernández, & De Vega, 2012) and (Briza & Naval, 2011), however, selection of the relevant indicators themselves is an issue. There are very few studies, e.g. (Nair & Mohandas, 2015a), which take this aspect into account. The second most common approach followed, is to employ statistical techniques such as Autoregressive (AR) and Generalized Autoregressive Conditional Heteroskedasticity models or simple linear regression (Clare & Miffre, 1995), (McKibben, 1972) models for forecasting the stock prices or stock price returns. These techniques have been found to offer poor forecast accuracy when compared to more recent soft computing based forecasting techniques. Soft computing based techniques are now being widely used for forecasting of nonlinear time series, as seen from (Atsalakis & Valavanis, 2009) and (Nair & Mohandas, 2015a). Hence, soft computing based techniques appear to be highly suitable for the proposed stock trading recommender system.

In the present study, separation of trend and cyclic components of the stock price time series is carried out as a preprocessing step. As suggested in (Hodrick & Prescott, 1997), the stock price time series can be considered to be made up of a (linear) trend component and a (nonlinear) cyclic component. Since the cyclic component governs the short-term price fluctuations, an accurate forecast of the cyclic component is essential for a trader wishing to profit by executing short-term trades. It has also been proven based on the theory of dynamical systems that a system exhibiting chaotic behaviour in time domain tends to behave deterministically in its phase-space representation (Huffake, 2010). This property has been successfully utilized in (Wan & Chai, 2014), (Nair et al., 2011), (Zhang & Li, 2010) etc. and it was reported to generate better results. In the present study as well, rather than using the stock price time series as-it-is, the phase space representation of the cyclic component (of the stock price time series) is used with a view to improving the overall accuracy of the system.

Two parameters to be identified for generating the phase space representation of a time series are the delay (d) and the embedding dimension (M). The most common approach to determining d is the technique proposed by Fraser and Swinney in (Fraser & Swinney, 1986), while False Nearest Neighbours (FNN) technique, proposed in (Kennel, Brown, & Abarbanel, 1992) is widely used for identifying M. In this study, however, a novel technique for identification of d and M using regression trees (described in detail in Section 2.2.2), has been presented and empirically validated. The proposed technique has also been compared to the traditional method proposed in (Fraser & Swinney, 1986).

The second novelty in the proposed stock trading recommender systems is to employ clustering algorithms to identify similarity short term price movements. Traditionally, time series clustering algorithms like Symbolic Aggrigate Approximation (SAX) (Lin, Keogh, Wei, & Lonardi, 2007) have been used for the purpose, e.g. in (Nair, Xavier, Mohandas, Anusree, & Kumar, 2014) and (Canelas, Neves, & Horta, 2013), however, there appears to be very little available literature that explores the possibility of employing clustering algorithms to time series data. In this study, a novel technique of transforming the time series data and then clustering it in a higher dimensional space to find similarities in the stock price movements over time, is employed for generating trading recommendations.

Clustering algorithms are very widely used in data mining with k-means clustering algorithm and its variants being the most popular of the clustering algorithms. Though better than the simple hierarchical clustering algorithms such as the single and complete linkage algorithms, k-means clustering technique is unable to yield good quality clusters with high dimensional data (Xu & Wunsch II, 2005). Neural network based clustering algorithms such as Self Organising Maps (SOM) appear to be better suited in this regard. In (Bação, 2005), performance of k-means clustering and SOM was compared. It was observed that the mean quadratic error, the standard deviation of the quadratic error and the structural error were lesser for SOM when compared to k-means clustering. SOMs have been used for predicting stock prices (Afolabi & Olude, 2007), (Sugunsil & Somhom, 2009), stock picking (Khan, Bandopadhyaya, & Sharma, 2008), identification of liquid stocks (Widiputra & Christianto, 2012), trading preference discovery (Tsai, Lin, & Wang, 2009) and study of stock price bubbles (Gao & Xu, 2009). SOMs have also been combined with SVMs (Ismail, Shabri, & Samsudin, 2011), (Hsu, Hsieh, Chih, & Hsu, 2009), Genetic programming (Hsu, 2011), wavelets (Li & Kuo, 2008) for better performance. It was observed from the literature that SOM based systems can be successfully used for mining information from financial data. Hence, in the present study, SOMs have been employed for clustering the stock price data. The technique is described in Sections 2.4 and 2.5 below.

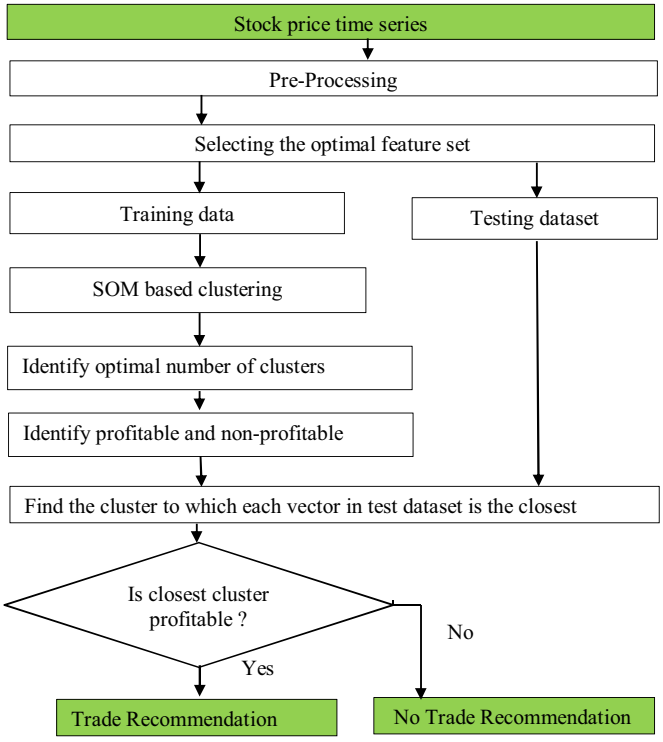

Fig. 1. Top level block diagram of the proposed clustering based stock trading recommender system.

Once the time series data points are clustered, the clusters are identified as profitable or non-profitable, based on the technique as detailed in Section 2.6. For every new data point obtained, the closest cluster to the point is identified and the proposed recommender system generates 'trade' (buy/sell) or 'no-trade' (hold) recommendations depending on whether the point belongs to profitable or non-profitable cluster, as explained in Section 2.7.

The top level block diagram of the proposed system is given in Fig. 1.

Based on the techniques employed for pre-processing, optimal feature set selection and the training and testing techniques used, a total of sixteen different variants of the proposed recommender system are generated. In order to establish the efficacy of the proposed system, all sixteen variants are validated on stocks drawn from US, UK, Brazilian and Indian stock markets (stocks are selected based on the additional criterion that they must belong to different sectors of industry) using eight different performance metrics.

Please cite this article as: B.B. Nair et al., Clustering stock price time series data to generate stock trading recommendations: An empirical study, Expert Systems With Applications (2016), http://dx.doi.org/10.1016/j.eswa.2016.11.002