Abstract

Trading profitably in stock markets consistently is considered to be an extremely challenging task. The Efficient Market Hypothesis (EMH) even suggests that it is not possible to generate excess returns from the market, in short, that it is not possible to obtain returns higher than that generated by the passive Buy and Hold (B&H) strategy. Though there is empirical evidence to the contrary, traders who make the highest gains from stock trading have a very deep understanding of the market dynamics and are able to ‘time’ the purchase and sale of stocks in a manner that generates returns much higher than any other fixed-income instrument. For a layman who is not well versed in the way stock markets behave, making profits in stock market could be a daunting task. To address this issue, a stock trading recommender system that can take over the role of an ‘expert’ by learning patterns in the stock price movements and recommend buy/sell/hold trading decisions, is highly desirable.

A large body of literature dealing with various aspects of financial data, its analysis and applications has been published over the past century (1930-2013). A short survey of literature in the area is initially carried out, with over a hundred articles considered, starting from the work on predictability of Dow Jones Industrial Average (DJIA) carried out by Cowles in 1933, and right up to the recent work in the field. The survey focused on major aspects such as the forecasting task (forecasting price/index, identifying buy/sell points, predicting stock/index movement direction, etc.), the target stocks/indices selected for the purpose, the technique selected (statistical, soft computing etc.), the input features selected (closing prices, daily returns, financial ratios etc.) and the results reported. It is seen from the survey that the two traditional techniques employed by stock traders are fundamental analysis and technical analysis. Fundamental analysis utilizes the information about the overall performance of the firm under consideration, for eg. its balance sheet, annual reports, market capitalization etc. and arrives at conclusions on the long-term performance of the firm. The stock is then bought or sold based on the conclusions. However, it was observed that to trade profitably using fundamental analysis, stocks have to be held for months or even years before being sold, making it unsuitable for a layman who has limited funds and is thus interested in short term profits. Technical analysis, on
the other hand, is quite popular in the stock trading community, and attempts to analyze the behavior of stock price over time with the help of technical indicators and identify patterns from it, which could then be used for trading. This approach is preferred by traders who are interested in making profits over a short term (eg. a few days or weeks). However, there are hundreds of technical indicators in use today and the specific indicators applicable under the given market condition as well as the parameters of each indicator are chosen subjectively, based on the expertise of the trader. A more recent and promising approach is employing soft computing based techniques to ‘learn’ the underlying patterns in stock price movements and then use the information to recommend the appropriate stock trading decision. From the literature survey, soft computing based recommender systems are seen to be highly effective in uncovering patterns in stock price movements and are seen to significantly outperform conventional techniques such as fundamental/technical analysis and regression based models. On the other hand, such systems are highly complex requiring fine tuning of system parameters to arrive at the optimal results, putting them beyond the reach of a layman. An attempt has been made in this study to design and empirically validate the efficacy of soft computing based stock trading recommender systems that can operate at a high level of abstraction, thereby insulating the lay-user from the underlying system complexities while at the same time allowing him/her to benefit from the superior performance that soft computing based systems offer.

In this study, four different soft computing based approaches to designing stock trading recommender systems are explored and empirically validated. The traditional approach of forecasting the stock prices and then recommending trading decisions based on the forecast has been explored first. Three different soft computing techniques, namely Artificial Neural Networks (ANNs), Adaptive Network based Fuzzy Inference System (ANFIS) and decision tree are evaluated for their efficacy in forecasting 1-25 days-ahead stock index values for S&P-500, BSE-Sensex and FTSE-100. A novel technique using Hodrick-Prescott (HP) filter and Discrete Fourier Transforms (DFT) for identification of the optimal number of training samples for the recommender system has also been presented. It is observed from the results that proposed training sample set selection technique improves the forecasting performance of all the three techniques with the combination of decision tree with the proposed training sample set selection technique being the most accurate. The combination of decision tree based forecasting system with a trading rule results in the forecasting based recommender system. Three different variants
of the recommender system realized by modifying the trading rule are evaluated on stocks drawn from the Indian and UK stock markets using eight different performance measures: total profits, average profits, profit per successful trade (P/ST), loss per loss making trade, total trades, maximum drawdown, profit factor (PF) and win ratio. Results indicate that the trading rule has a significant impact on the performance of forecasting based recommender system and its ability to beat the traditional benchmark B&H strategy.

Dynamical systems theory suggests time series that demonstrate chaotic behavior in time domain tend to behave deterministically in their phase space. Since stock price time series data are inherently nonlinear in nature, the second approach explored in this study involves designing a recommender system that integrates the phase space representation of stock price time series with clustering techniques to forecast buy/hold/sell recommendations. First minimum of mutual information and false nearest neighbors techniques are used to identify the delay and embedding dimension respectively, based on which the phase space representation is obtained. Clustering is then used to group the state-space vectors into similar clusters. Profitable clusters are identified from the cluster centroids and the vectors belonging to the profitable clusters are used to train an Adaptive Resonance Theory MAP (ARTMAP) neural network to generate ‘Trade’ or ‘No trade’ recommendations. Three different clustering techniques are evaluated, namely, the k-means, Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and the Fuzzy C-Means (FCM), resulting in three variants of the proposed recommender system. Three additional types of recommender systems are also evaluated viz., (a) ARTMAP based trading recommender trained using the entire set of state space vectors (b) clustering based recommender (with three variants based on the three clustering techniques discussed above) and (c) ARTMAP based recommender trained using the stock price time series data. Hence a total of eight different recommender systems are considered. Trading performance of these systems is empirically validated on stocks drawn from the Indian, UK and US stock markets using eight different performance measures. It is observed that phase-space- k-means clustering-ARTMAP based stock trading recommender system is able to generate the highest returns and is able to consistently beat the B&H strategy for all the stocks considered. The other two variants employing UPGMA and FCM too were able to beat the B&H strategy but generated returns lower than the k-means based variant. Clustering based recommender systems were also able to
generate returns higher than B&H for 89% of the datasets considered but the returns were not as high as the phase-space-k-means clustering-ARTMAP based recommender system.

The third category of recommender systems evaluated, takes an alternate approach to the design of stock trading recommender systems by viewing the issue of generating trading recommendations as a classification problem rather than a time series forecasting problem. Classifier based recommender systems are designed that can classify the one-day-ahead trading recommendations into one of the two classes: ‘Trade’ or ‘No Trade’. Classifier based recommender systems proposed in this study are also designed to overcome the issues associated with technical analysis discussed above. The classifier based recommender systems considered in the present study initially start with a set of thirty-one technical indices drawn from the three broad categories of technical indicators, namely, overlays, price-based and volume-based indicators. Training and testing time frames are then selected based on the HP filter- DFT based technique proposed in this study. Since one of the major issues with technical indicators is the selection of optimal technical indicator parameters, Genetic Algorithm (GA) – decision tree based optimization of parameters is carried out with the objective function being the maximization of profit factor for the training data. Once the optimal parameters are identified, the second issue associated with technical indicators is addressed: the selection of relevant features (technical indicators). This is accomplished in the present study with the help of two different techniques. First technique involves selecting those features (technical indicators) that form the nodes of the optimal decision tree obtained using the above technique. The technical indicators that do not form the part of the tree are considered to be irrelevant and are discarded for the next stage. The second feature selection technique proposed in this study involves the selection of the relevant features with the help of a linear Support Vector Machine (SVM). Features that contribute less than 2% to the sum of feature weights are considered to be insignificant and discarded. Once the features have been selected, the reduced input feature set is used to train a classifier to generate one of the two classes ‘Trade’ or ‘No Trade’ as the one-day-ahead trading recommendation. Five different classifiers are considered: SVM, ANN, cANTMiner, decision tree and Naïve Bayes. Recommender systems combining each of the feature selection techniques with classifiers are designed. Variants of the recommender systems with GA-optimized technical indicators but without feature selection were also evaluated to empirically validate the effectiveness of feature selection on recommender trading performance.
Trading performance of all the recommender systems are evaluated on eight performance measures listed above on stocks drawn from the Indian and UK markets. Trading performance of traditional technical indicator based trading strategies (i.e. with fixed technical indicator parameters and no soft computing techniques involved) is also evaluated. Results indicate that overall performance of purely technical indicator based trading is rather poor. It is observed that trading recommender systems are able to generate much higher returns. It is also seen that feature selection tends to improve the performance of classifier based recommender systems considered in the study. Of all the recommender systems considered, the GA-technical indicator-decision tree-SVM based stock trading recommender system with decision tree based feature selection outperforms other recommender systems as well as the B&H strategy.

Stock trading recommender systems capable of generating stock trading recommendations by mining temporal association rules from stock price data are the fourth category of recommender systems explored in this study. An extension to the Apriori Association Rule Mining (ARM) algorithm to enable mining of temporal association rules is proposed in this chapter. Two temporal ARM based recommender systems are proposed. The first recommender system converts the stock price time series data (training data) to its symbolic representation using Symbolic Aggregate Approximation (SAX). Day-of-the-Week (DoW) information is then added to each symbol. Each such pair is an itemset. The symbol-DoW pairs are then grouped into weekly transactions with each transaction composed of the symbol-DoW pair that represents the stock price and the day of that particular week. The stock price dataset is thus converted into a transaction database from which the frequent two-itemsets are identified and association rules mined. Relevant rules that can be used for trading are then identified from the mined rules. For every new incoming stock price sample, the price is converted into an itemset using the above technique and is matched with the antecedent part of the trading rules. The matching trading rule with the highest support is selected and the consequent decision is recommended. It must be noted that the optimal recommender system parameters: the number of symbols and the minimum support value are identified using evolutionary algorithms. Two algorithms, GA and Artificial Bee Colony (ABC) algorithms are considered. Two variants are considered, one with PF and the other with P/ST as the performance measures to be maximized. The second temporal ARM based recommender system proposed in this study generates one-day-ahead trading recommendations and also begins by converting the stock price time series (training) data to its
symbolic representation. The symbols are then grouped in pairs to form a transaction database with each transaction composed of two items. Association rules are then mined from the transaction database. Rules relevant for the trading process are identified from the set of mined rules. The optimal values for the SAX breakpoints is identified using evolutionary algorithms. As in the first temporal ARM based recommender, GA and ABC are considered with maximization of PF and P/ST as objective functions. For every new sample, the price is transformed into an itemset using the above technique and is matched with the antecedent part of the trading rules. The matching trading rule with the highest support is selected and the consequent one-day-ahead trading decision is recommended. Trading performance of both the categories of temporal ARM based recommender systems is evaluated on stocks selected from the Indian and UK markets with performance being quantified in terms of the eight performance measures listed above. It is observed from the results that both categories of temporal ARM based stock trading recommender systems proposed in this study are able to outperform the traditional B&H strategy. It is also observed that for both categories of temporal ARM based recommender systems, optimization with P/ST as the objective function generates higher out-of-sample profits. Temporal ARM based one-day-ahead trading recommender is seen to generate the highest out-of-sample profits at the cost of an increase in the number of trades that need to be executed in the same time-frame.

From the trading performance, it is clearly observed that all four soft computing based approaches can be successfully used to implement stock trading recommender systems that can learn patterns in stock price movements and generate trading recommendations that are easy for a layman to understand. The results also indicate that simply following the trading recommendations suggested, results in profits which are much higher than that generated by the B&H strategy thereby offering an empirical negation of the EMH.