DATA MINING TECHNIQUES AND STOCK MARKET

Mr. Rahul Thakkar,
Lecturer and HOD,
Naran Lala College of Professional & Applied Sciences, Navsari

ABSTRACT

Without trading in a stock market we can’t understand the full potential of trading and the rewards that come from the trading of stocks. We can earn far greater returns from individual stocks than we will ever find from a diversified mutual fund, if we pick the right stocks for trading on day to day basis. But along with greater returns come greater risk. Thus, we need better direction through which we can minimize the risk involved during trading.

Keywords: Data Mining, Intraday Trading, Stock Market, Association Rule Mining, Neural Network, Clustering and Decision Tree

1. Association rule mining in stock market

In data mining, association rule is a popular and well researched method for discovering interesting relations between variables in large databases.

Association rule mining finds interesting associations or correlation relationships among large set of data items. Association rules are useful for determining correlations between attributes of a relation and have applications in stock market investment, marketing, financial, and retail sectors. Furthermore, optimized association rules are an effective way to focus on the most interesting characteristics involving certain attributes. Optimized association rules are allowed to contain uninstantiated attributes and the problem is to determine instantiations such that either the support or confidence of the rule is maximized.

Association mining, which is widely used for finding association rules in single and multidimensional databases, can be classified into intra and inter transaction association mining. Intra-transaction association refers to association in the same transaction; inter-transaction association
indicates association among different transactions. Most contributions in association mining focus on intra-transaction association also referred to traditional association mining.

Stock Prices are considered to be very dynamic and susceptible to quick changes because of the underlying nature of its domain and in part because of the mix of known parameters (open, high, low and close price) and unknown factors (government stability, results of election, and rumors).

2. Decision tree in stock market

Decision trees are powerful and popular tools for classification and prediction. In data mining, a decision tree is a predictive model which can be used to represent both classifiers and regression models. Decision trees are outstanding paraphernalia for making monetary or number based decisions where a lot of complex information needs to be taken into account. It provides an effective structure in which different decisions and the propositions of taking those decisions can be laid down and assessed. It also helps you to form a precise, fair picture of the risks and rewards that can result from a particular choice. In a stock market, how to find right stock and right timing to buy has been of great interest to investors.

3. Clustering in stock market

Clustering is a tool for data investigation, which solves categorization problems. Its objective is to assign cases (community, items, actions etc.) into groups, so that the amount of relationship can be strong between members of the same cluster and weak between members of different clusters. In clustering, there is no preclassified data and no distinction between independent and dependent variables. Instead, clustering algorithms search for groups of records (the clusters composed of records similar to each other). The algorithms discover these similarities. This way each cluster describes, in terms of data collected, the class to which its members belong. Clustering may reveal associations and structure in data which, though not previously evident, nevertheless are sensible and useful once found.

As part of a stock market analysis and prediction system consisting of an expert system and clustering of stock prices, data are needed. Stock markets are recently triggering a growing interest. In particular, it would be useful to find, inside a given stock market index, groups of companies sharing a similar temporal behavior. To this purpose, a clustering approach to the problem may represent a good strategy.

4. Neural network in stock market

Neural networks have been successfully applied in a wide range of verified and unverified learning applications. Neural network methods are commonly used for data mining tasks, because they often produce logical models. A neural network is a computational technique that benefits from techniques similar to ones employed in the human brain. The main advantage of neural networks is that they can estimate any nonlinear function to a random degree of accuracy with a suitable number of hidden units. The development of powerful communication and trading facilities has enlarged the scope of selection for investors.

Information gain technique used in machine learning for data mining can be applied to assess the analytical relationships of numerous economic and monetary variables. Neural network models for level estimation and classification are then examined for their ability to provide an effective forecast of future values. A cross validation technique is also employed to improve the simplification ability of several models.
5. Data Mining Algorithms

There are various data mining algorithms which can be used to build the mining model. But choosing the right algorithm for the right business task is critical. Different algorithms can be used to do the same business tasks but each algorithm produces different results.

The various types of algorithms are as follows:

- **Association algorithm:** It discovers the association between different elements in a database. It creates the association rules; common application of association rule algorithm is to discover the association rules for market basket analysis.

- **Classification algorithm:** It forecasts one or more distinct variables, based on the other attributes in the database.

- **Regression algorithm:** It forecasts one or more continuous variables, such as gain or loss, based on other attributes in the database.

- **Segmentation algorithm:** It divides data into set, or groups, of items that have similar characteristics.

- **Sequence analysis algorithm:** It reviews frequent sequences or occurrences in database, Web path flow is an example of sequence analysis algorithm.

An application can use any of the algorithms from the aforementioned algorithms of the data mining; segmentation algorithm can be used to explore data, regression algorithm can be used to predict continuous variables and association algorithm can be used to find the association between two or more elements from the database.

6. Association Rule Mining

Association rule knowledge is an accepted and well researched method for discovering interesting relationships between elements in large databases. It is projected to recognize robust rules discovered in databases using different measures which are interesting. For example, the rule \{chicken, clothes\} => \{burger\} found in the sales data of a supermarket would indicate that if a customer buys chicken and clothes together, he is likely to also buy burger. Such information can be used as the basis for decisions about marketing activities such as, promotional pricing or product placement. Association rule mining algorithms can be applied in many application areas such as Web mining, intrusion discovery, Continuous production, stock market and bioinformatics.

Sequence mining considers the order of the items either within a transaction or across the transaction, while it is not so in association rule mining. Association rules are if/then statements that help discover associations between apparently distinct data in a database. An example of an association rule would be "If a customer buys a chicken and clothes, he is 85% likely to also buy burger."

First, minimum support is applied to find all frequent itemsets in a database. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

While the second step is straightforward, the first step needs more attention. Apriori is a classic algorithm for frequent itemset mining and association rule learning occurs over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database.

Finding all frequent itemsets in a database is difficult since it involves searching all possible itemsets. The set of possible itemsets is the power set over I and has size \(2^n - 1\). Although the size of the powerset grows exponentially in the number of items \(n\) in \(I\), efficient search is possible using the
downward-closure property of support, which guarantees that for a frequent itemset, all its subsets are also frequent and thus for an infrequent itemset, all its supersets must also be infrequent. Exploiting this property, Apriori can find all frequent itemsets.

7. **Support Confidence & Accuracy of Association Rules**

The potential usefulness of a pattern is a factor defining its interesting. It can be estimated by a utility function, such as support. The support of a pattern refers to the percentage of task-relevant data tuples for which the pattern is true.

Each discovered pattern should have a measure of certainty associated with it that assesses the validity or “trustworthiness” of the pattern. A certainty measure for rules of the form “A=> B” is confidence.

Estimating classifier accuracy is important in that it allows one to evaluate how accurately a given classifier will label future data, that is, data on which the classifier has not been trained. Accuracy estimates also help in the comparison of different classifier.

**CONCLUSION**

Trading, like most other things, requires that we have a general philosophy about how to do things in order to avoid careless errors. Would we make furniture without a wood? Would we play cricket without bat and ball? We hope not. So before we dig deeper into trading, we certainly need a considered plan by using certain techniques of data mining before trading our hard-earned savings.

**REFERENCES**

3. L. Bing and H. Wynne, “Using General Impressions to Analyze Discovered Classification Rules,” Proc. 3rd International Conference on Knowledge Discovery and Data Mining (Kdd-97), Newport Beach, California, USA, pp. 31-36, 1997.
14. Jiawei Han & Micheline Kamber: Data Mining Concept and Techniques, Morgan Kaufmann