

# Applying Moving Objects Patterns towards Estimating Future Stocks Direction

Dahab Galal<sup>1,a</sup>, Nada Hassan<sup>1</sup> and Doaa S. Elzanfaly<sup>1</sup>

<sup>1</sup>*Information Systems, Informatics and Computer Science, British University in Egypt*

**Abstract.** Stock is gaining vast popularity as a strategic investment tool not just by investor bankers, but also by the average worker. Large capitals are being traded within the stock market all around the world, making its impact not only macro economically focused, but also greatly valued taking into consideration its direct social impact. As a result, almost 66% of all American citizens are striving in their respective fields every day, trying to come up with better ways to predict and find patterns in stocks that could enhance their estimation and visualization so as to have the opportunity to take better investment decisions. Given the amount of effort that has been put into enhancing stock prediction techniques, there is still a factor that is almost completely neglected when handling stocks. The factor that has been obsolete for so long is in fact the effect of a correlation existing between stocks of the same index or parent company. This paper proposes a distinct approach for studying the correlation between stocks that belong to the same index by modelling stocks as moving objects to be able to track their movements while considering their relationships. Furthermore, it studies one of the movement techniques applied to moving objects to predict stock movement. The results yielded that both the movement technique and correlation coefficient technique are consistent in directions, with minor variations in values. The variations are attributed to the fact that the movement technique takes into consideration the sibling relationship

## 1 Introduction

Capitalizing in the stock market has always been a powerful investment tool that is gaining popularity fast. However, the one characteristic all stock markets have in common is uncertainty; an effect of the stock's short and long term states. Such a trait is undesirable yet unavoidable if people are to participate in the stock market; this is when prediction techniques for the stock market become a major point of interest.

There already exists a number of computational techniques which aim to enhance the stock prediction process including but not limited to; Linear Regression, Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Another approach to stock prediction involves the use of Database Management Systems such as real time database characteristics. Databases in which the stock data is stored. Given the amount of techniques that are used to enhance stock prediction, the effect of a correlation that exists between stocks of the same index or mother company is a factor that is almost completely neglected in most of the aforementioned techniques.

This paper aims to study a new perspective of stock modelling that focuses on the correlation between stocks within the same index or parent company as a key factor to draw better predictions. The proposed approach considers stocks as objects that constantly move in

accordance with their values, thus becoming moving objects. The idea is that by applying prediction patterns of moving objects such as "Following Pattern" and "Attraction & Avoidance Pattern", all the stocks involved in a relationship would be used to predict each stock's future direction relative to the rest of the 'sibling' stocks. If a stock is said to have an attraction/avoidance relationship with another stock, it means that when the first stock rises/falls in value, then the second stock would follow accordingly.

This paper is organized as follows; section two provides a brief literature review of the various techniques that were applied for stock prediction. Section three discusses the proposed model and how the movement technique is applied to the model. Section four evaluates the test results and experiments, lastly, section five provides a conclusion and discussing the future work.

## 2 Literature Review

Current prediction techniques concerning stocks vary between psychological, financial and technical models. These techniques are applied to a stock's collected values over a period of time to predict its future value. Most of these techniques are in fact based on data mining techniques.

<sup>a</sup> Corresponding author: Dahab.Galal@bue.edu.eg

The most commonly used technique is that of linear regression. Linear regression allows analysts to predict the volume of a given stock taking into consideration the fluctuations in its values over a large period of time [12]. It is further used to detect patterns that occur with respect to actual and target stock values. The identified patterns allow for further prediction of stock values. One significant limitation of this technique is the fact that the dataset needs to be large to produce meaningful results. Furthermore, it projects the data onto a straight line, disregarding outliers and their impact on the determined patterns [7].

Artificial Neural Networks (ANN) is another technique that is employed in stock prediction. This technique has the added benefit of predicting the direction in which the stock value would move, meaning whether it would increase or decrease [11]. One issue with this technique is that it does not provide accurate results with respect to prediction of stock values. Another issue lies in its susceptibility to noisy data. An attempt to handle those limitations was to integrate ANN with Support Vector Machines (SVMs) that were introduced in stock prediction to tackle stock market movement as well as the variation in price indices [7]. The work done by [2] at al. shows that SVMs outperformed various other techniques including case based reasoning and ANNs with respect to prediction of stock direction.

In order to model the time aspect of stocks, real time databases were considered. Transactions stored within real time databases have strict time constraints [5]. Such transactions are referred to as tardy constraints and later overload the system. The issues with this representation include its improper handling of abortion and recovery of transactions. Furthermore, there is a constant trade-off between security and performance [1]

A common technique for measuring correlation between stocks the correlation coefficient. It is defined as follows:

$$Correlation = p = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad (1)$$

Where Cov (x, y) is the covariance between X and Y,  $\sigma_x$  is the standard deviation of X and  $\sigma_y$  is the standard deviation of Y [3]. As with previous techniques, the correlation coefficient technique only takes into consideration the stocks' closing price, which is a single dimension. It doesn't investigate how incorporating more than one dimension would affect the correlation between the two objects.

All of the above mentioned techniques have an important aspect in common, which is the fact of drawing assumptions about the future estimate of a stock, they use the history of the same individual stock as well as the possible fluctuations of the market. This discards any correlation between stocks of the same index or mother-company to be a significant factor. This leads to the main problem, which is the lack of techniques that take into consideration the correlation between stocks and how such a relation can aid the enhancement of stock prediction.

### 3 Modelling Stocks as Moving Objects

To overcome the limitations of the aforementioned stock prediction techniques, we propose a distinct approach for studying the correlation between stocks that belong to the same index by modelling them as a set of interrelated moving objects where the index is treated as a parent and stocks within it to be siblings to be able to track their movements while considering their relationship. Modelling stocks as moving objects has a threefold aim: first, getting benefits of the enriched capabilities of the object-oriented data model especially the inheritance property to model the parent-child and sibling relationships. Second, keeping track of the temporal and spatial aspects of the moving objects along with maintaining their hierarchy and relationships. Third, applying different movement patterns on moving stocks and determine its impact on stock value predictions.

A stock can be abstracted as <point, instance> where the point represents the stock and the instance represents its value in a certain time. The trajectory, that is the path a moving object follows through space as a function of time, represents the motion of a moving point in a discrete form [6]. Thus, trajectories of moving points are often defined as a sequence of (x, y, t) rows:

$$T = \{(x1, y1, t1), (x2, y2, t2), \dots, (xn, yn, tn)\} \quad (2)$$

Where x, y ∈ R and t1 < t2 < ... < tn, [13]. The trajectory of each object represents a sequence of locations each coupled with a certain timestamp. In our case, the object is a stock that is represented by its closing value coupled with the volume sold at that value.

As it has been briefly discussed above, if stocks are to be modelled and handled as points of moving objects, both the temporal and spatial factors would be taken into consideration where each stock would be managed as a moving entity over a period of time. As a result the proposed model could offer a significance advantage over Real-Time databases as stocks get to keep their relationships, their timestamp factor and adds to that their fluctuation of prices can be modelled as movement, this movement can be analysed by using different Moving Objects data model techniques.

#### 3.1 Stock Movement Patterns

In the context of moving object, a number of movement patterns have recently been identified [9], most notable of which are the "Following" pattern and the "Attraction/Avoidance" pattern. The "Following" pattern shows the dynamic relationship between objects following the idea of how a predator follows its prey [10]. This pattern is limited by the fact that it defines a one-sided relationship. This implies that if one object follows the other, the reverse is not guaranteed. Furthermore, it does not specify a movement direction for the object in question.

On the other hand, the "Attraction/Avoidance" pattern is a technique that identifies whether two objects attract or avoid each other. This relationship would result from their dependence on each other in some way [8, 10]

The Attraction/Avoidance technique takes into consideration the two moving objects and the distance

between them in order to determine when they [9]. A threshold is defined, dictating the maximum allowed distance between the two objects so as to consider that they have met. For this work, the designated threshold is in fact a particular time frame. The two stocks are considered to have met when the closing value and volume pair for each of the stocks occur within the specified time frame. The number of times that they occur within that time frame is considered the meeting frequency of those two stocks.

As an example, given the threshold of four days, stock A has closing price of 2.00\$ with an associated volume of 5000 stocks; while stock B has closing price of 4.00\$ with an associated volume of 3000 stocks. As such, stock A and stock B will meet if these same values occur again within the predetermined four days. The number of times that that condition evaluates to true would in turn represent the meeting frequency of those two stocks. It is however infeasible to directly measure the relationship between two trajectories using meeting frequency. This is because two objects may seem to be frequently meeting when in fact there are other issues that cause the excessive meetings (i.e. Outliers or noisy data).

To further achieve this notion, a permutation test is performed with the null hypothesis that the two movement sequences of the two stocks are independent. The null hypothesis is in fact the baseline for comparison. The permutation test involves randomly shuffling the movement sequences of the one of the two stocks and then checking whether the null hypothesis holds or not. If the hypothesis holds, then it is concluded that the two stocks are independent. Otherwise, it is proven that the stocks have a certain relationships which needs to be further classified as either attraction or avoidance.

In order to determine whether two stocks attract or avoid each other, the significance value is calculated as indicated by the algorithm. The calculated value is based on the actual meeting frequency as well as the permuted frequency. The resulting significance value lies in one of three ranges; either the resulting value is greater than 0.5, is equal to 0.5, or is less than 0.5, thus emphasizing that only two ranges are of interest. If the significance value is greater than 0.5 then the two stocks are attract. Conversely, as the value decreases, the stocks avoid each other.

#### 4 Model Evaluation

To evaluate the proposed model, an application has been built as an adaptation of the algorithms of a tool called MoveMine that was created by a team in Pennsylvania university to measure the movement of animals and their preys to then be displayed using Google Earth [8 - 10].

The application has been developed to track a certain pattern which is the “Attraction and Avoidance” pattern in the S&P 500 stock market index data set [4]. The S&P 500 Index is managed by S&P Dow Jones Indices. The S&P 500 stock market index represent 502 common stocks that are issued by 500 large-cap companies and are traded on the American Stock Exchange Market; some of which include IBM, Apple, and EMC. The S&P 500 companies takes up 75 percent of the American Equity market by Capitalization. Although the index is called S&P “500”, it

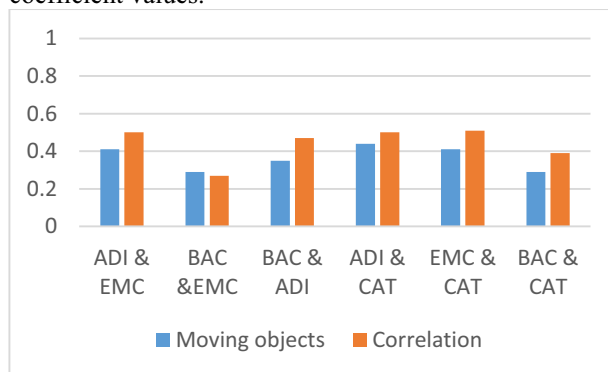
actually contains 502 stocks. This particular data set was chosen for the existence of siblings’ relationship to validate the research hypothesis.

The dataset contained the Ticker, the opening stock value, the highest stock value, the lowest stock value, the closing stock value, and the volume (number of stocks sold). Each record of the above values was recorded for a single day. The dataset represents each of the 502 stocks over the period of 365 days.

Those values were chosen for two reasons; firstly because they are the most used values when it comes to stock prediction and secondly they are the most representative of a stocks state given a certain point in time. The values were further uniformly normalized (values lie between 0 and 1) so as to apply the attraction/avoidance technique. It was taken into consideration that the stocks were from the same industry, which is the information technology industry.

A number of pairs of stocks were selected to investigate the accuracy of modelling stocks as moving objects. In order to determine the quality of results, the resulting significance values were compared to correlation coefficient values [14]. Correlation coefficient values are between -1 and 1, where 0 indicates independence. Positive values indicate that the two stocks move in the same direction, while negative values indicate that they move in opposite directions. As such, this technique was used as a standard of comparison. The correlation coefficient values were normalized to values between 0 and 1.

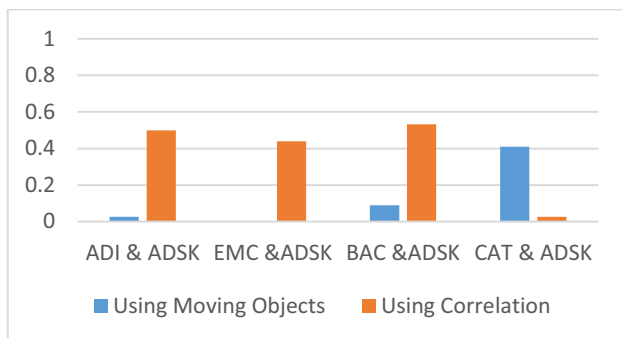
In Figure 1, it is clear that both the Attract/Avoid technique and the correlation coefficient technique yield the same values with respect to their avoidance. The discrepancy between the values produced by each technique are attributed to the fact that the attract/avoid technique takes into consideration aspects that were otherwise ignored. This shows that the siblings’ relationship does in fact add value to the prediction as it does not deviate drastically from the correlation coefficient values.



**Figure 1.** Stock Movement Prediction showing the proposed approach as comparison with the correlation equation

Among the set of stocks that were chosen, there was one particular stock that when compared with others, always yields different results than those produced by the correlation coefficient, as shown in Figure 2. This implies that there are certain characteristics provided by the “Attraction/Avoidance” technique that may cause skew for certain objects. This is consistent with the permutation step that is applied to ensure that two stock objects are truly

attracted/avoid each other. This further proves that for this stock, the correlation coefficient does not take into consideration possible outliers that may affect the identified movement pattern.



**Figure 2.** Stock movement prediction showing anomalist

By taking a closer look at the resulting values, stocks BAC and EMC have very close values with a lesser correlation value. This shows that the Attract/Avoid pattern identifies the avoidance for what it is and yet takes the value one step further to include the two stocks' history as siblings. This aspect was unaccounted for in the correlation coefficient technique

## 5 Conclusion and Future Work

In conclusion, it can be established that the use of the object oriented data concepts to model related stocks has positively affected stock prediction. The limitations of previous stock prediction techniques did not take into consideration the possible dependencies between different stocks. The dependencies are a result of the fact that a number of stocks may all belong to the same index, making those stocks siblings. Stocks were modelled as moving objects and were represented by their closing value and volume. One of the movement pattern techniques was selected to assess whether applying the moving object model holds for stock. The Attract/Avoid technique was selected due to its significance with respect to the model in question. The results shows that treating stocks as moving objects took into consideration aspects that were otherwise ignored by the correlation coefficient. Furthermore, the results prove that with respect to predicting movement direction, both techniques yield the same direction.

It is intended to study further the anomaly presented by some of the stocks, to determine the cause of such behaviour. Furthermore, more movement patterns can be applied and investigated to further enhance stock prediction.

## References

1. Akdere, M., Cetintemel, u., & Upfal, E. (2009, August). Database-Support for Continuous Prediction Queries over Streaming Data. Providence, Rhode Island, USA.
2. Chen, A., Leung, M., & Daouk, H. (2003). Application of Neural Networks to an Emerging

3. Financial Market: Forecasting and Trading the Taiwan Stock Index. *Computers & Operations Research*, 901-923.
3. Cherewyk, P. (2015). Calculating Covariance For Stocks. Retrieved from Investopedia: <http://www.investopedia.com/articles/financial-theory/11/calculating-covariance.asp>
4. Economic Research Federal Reserve Bank of St. Louis. 2000. S&P 500© (SP500). [ONLINE] Available at: <https://research.stlouisfed.org/fred2/series/SP500/downloaddata>. [Accessed 15 July 15].
5. Hajari, H., & Hakimpour, F. (February 2014). A Spatial Data Model For Moving Objects Database. *International Journal of Database Management Systems ( IJDM )* Vol.6, No.1.
6. Imfeld, S., & P., L. (2002). Analyzing Relative Motion within Groups of rackable Moving Point Objects. *Information Systems*.
7. Kara, Y., Boyaciglu, M. A., & Baykan, O. K. (2011). Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines. *Expert Systems with Applications*, 5311-5320.
8. Kays, R., Li, Z., Ji, M., Lee, J.-G., Tang, L.-A., Yu, Y., & Han, J. (2011). MoveMine: Mining Moving Objects Database. Illinois.
9. Kin, T., Wu, F., Li, Z., & Han, J. (2013). MoveMine 2.0: Mining Object Relationships From Movement Data. Pennsylvania & Illinois.
10. Li, Z., Ding, B., Wu, F., Lei, T. K., Kays, R., & Crofoot, M. C. (2014). Attraction and Avoidance Detection From Movements. Pennsylvania, USA.
11. Mahajan, K. S., & Kulkarni, R. (2013). A Review: Application of Datamining Tools For Stock Market. Kolhapur, India: Institute of Business Education & Research Center.
12. Olaniyi, A. S., Adewole, K. S., & Jimoh, R. G. (2011). Stock Trend Prediction Using Regression Analysis: A Data Mining Approach. *ARPN Journal of Systems and Software*, 4.
13. Pena, J. H., & Santos, M. Y. (2011). Representing, Storing and Mining Moving Objects Data. *World Congress on Engineering* . London: WCE.
14. Stock Charts. (2015). Technical Indicators and Overlays. Retrieved from Stock Charts: [http://stockcharts.com/school/doku.php?id=chart\\_school:technical\\_indicators:correlation\\_coeffici](http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:correlation_coeffici)