Asian Journal of Pharmaceutical and Clinical Research

Innovare Academic Sciences
Knowledge to Innovation

Special Issue (April)

Online - 2455-3891
Print - 0974-2441

Full Proceeding Paper

SURVEY ON ADVISOR INTELLIGENCE THROUGH PURCHASE PATTERNS AND SALES ANALYTICS

GAYATRI K PRADHAN[1], SARATH GOLLAPALLI[2], JANAKI MEENA M[1], SYEDIBRAHIM SP[1]*

[1]School of Computing Sciences and Engineering, VIT University, Chennai, Tamil Nadu, India. [2]Broadridge Financial Solutions, Hyderabad, India. Email: Syedibrahim.sp@vit.ac.in

Received: 19 January 2017, Revised and Accepted: 20 February 2017

ABSTRACT

In mutual fund, an individual or a firm that is in the business of giving advice about securities to clients is an investment advisor. Investment advisers are individuals or firms that receive compensation for giving advice on investing in stocks, bonds, mutual funds, or exchange-traded funds. Investment advisors manage portfolios of securities. Advisors can use new cognitive and analytics capabilities to better understand their clients and needs and have a stronger ability to deepen relationships with a better portfolio. In this paper, we analyze data points for each advisor, and distinguish the best prospects, obtain insight into their experience and credentials, and learn about their portfolio, in other words, to recognize the pattern of portfolio of the advisors. Such analysis helps the sales people to sell the fund company products to the suitable advisors based on the nature of the product they want to sell. This is done by investigating what kind of products advisors have been buying, and what kind of products they might be looking for. This helps to increase the sales of the products as sales people will be reaching the appropriate advisors.

Keywords: Big data analytics, Investment advisor, Pattern recognition, Evolutionary algorithm, Correlation, Leader-follower.

 DOI: http://dx.doi.org/10.22159/ajpcr.2017.v10s1.19743

INTRODUCTION

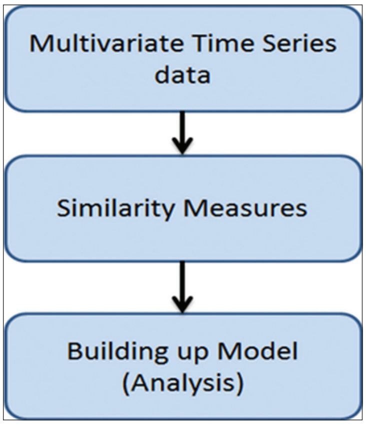
Among investors owning mutual fund shares, most of it holds the fund shares through an intermediary such as a broker-dealer, bank, fund supermarket or platform, insurance company, investment advisor. Investors choose intermediary which best suits their needs.

Investors use intermediaries to obtain a number of benefits. Investors typically show interests to the intermediary for recommendations on how to invest their money. With the help of the investment advisor, the investor may decide to include mutual funds as part of a portfolio of investments. Regardless of how assets are allocated between types of securities, with the choice to purchase mutual funds, an investor becomes a mutual fund shareholder. Due to the large number of shareholders who prefer to use intermediaries, as shown in Fig. 1 intermediary bridges the gap between mutual funds and fund shareholders.

Big data analytics helps to learn the portfolios of the investment advisors and recognize the behavioral pattern of an investment advisor. Such analysis will help the sales person or the fund companies to find relevant advisor which will be beneficial for the sales person to increase the sales of his products.

Further, this recognized pattern can be mapped into 10 global broad Morningstar category groups (equity, allocation, convertibles, alternative, commodities, fixed income, money market, tax preferred, property, and miscellaneous).

The Morningstar Global Category Assignments were introduced in 2010 to help investors search for similar investments entertained across the globe. There are different flavors of funds, and the investment advisors can be categorized into different categories. Financial time series analysis is concerned with theory and practice of asset valuation over time. It is a highly empirical discipline, but like other scientific fields theory forms the foundation for making inference. There is, however, a key feature that distinguishes financial time series analysis from other time series analysis. Both financial theory and its empirical time series contain an element of uncertainty. For example, there are various definitions of asset volatility, and for a stock return series, the volatility is not directly observable. Statistical theory and methods play an important role in financial time series analysis. Our financial time series data are a multivariate time series data set. A Multivariate time series data analysis is used when one wants to model and explain the interaction and comovements among a group of time series variables. Hence, multivariate time series data can be handled by converting it into univariate data using appropriate similarity measures and then building model on it. This can be pictured in following way:



This similarity measure can vary from problem to problem. For example, as author [1] proposed a distance function based on the assumed independent Gaussian models and used a hierarchical clustering method to group seasonality sequences into a desirable number of clusters. Here, an independent Gaussian model is a distance function which is a similarity measure and model is build using hierarchical clustering. Similarly, n number of analysis can be done on multivariate time series data. In this paper, few of the analysis have been done on multivariate time series data. The first section of the paper explains an approach toward clustering multivariate time series data. This will cluster advisors with similar behavior into same clusters. The second section of the paper explains detecting leaders from correlated advisors among n number of advisors. An approach of analyzing lead-lag relation among a set of time series data is explained.

LITERATURE SURVEY

In recent past, there is an increased interest in time series clustering research, particularly for finding useful similar trends in multivariate time series in various applied areas such as environmental research, finance, and crime [2]. Clustering multivariate time series have potential for analyzing large volume of finance data at different time points as investors are interested in finding market trends of various funds such as equity, allocation, convertibles, alternative, commodities, fixed income, money market, tax preferred, property, and miscellaneous so that it will help them to invest. Most of the traditional time series clustering algorithms deal with only univariate time series data and for clustering high dimensional data; it has to be transformed into single dimension using a dimension reduction technique. An approach of finding similar pattern author [3] has explained a tradition search and classification technique.

A novel approach based on cosine similarity with hierarchical clustering model has been proposed to find similar advisors. The proposed methodology goes as follows:

Euclidean distance measure is commonly used for non-time series data clustering. However, it is not suitable for multivariate time series clustering. Instead of Euclidean distance measure in stand-alone mode, cosine similarity provides better results. The problem of finding similar advisor can be solved in two steps:

1. Computing cosine similarity.
2. Hierarchical clustering.

We now discuss each step in detail.

Cosine similarity

Cosine similarity is a measure of similarity between two non-zero vector. Computing cosine can be explained with following example.

Consider an example to find the cosine similarity between two time series t_1 and t_2

t_1 = (5, 0, 3, 0, 2, 0, 0, 2, 0, 0)

t_2 = (3, 0, 2, 0, 1, 1, 0, 1, 0, 1)

Calculate cosine similarity using

$$\text{Cos}\left(t_1, t_2\right) = \frac{t_1 . t_2}{|t_1||t_2|} = 0.94$$

Similarly, find cosine similarity between each pair of time series data. This cosine score is then stored into a database table.

	t_1	t_2	t_3	t_4
t_1	Cos (t_1, t_1)	Cos (t_1, t_2)	Cos (t_1, t_3)	Cos (t_1, t_4)
t_2	Cos (t_2, t_1)	Cos (t_2, t_2)	Cos (t_2, t_3)	Cos (t_2, t_4)
t_3	Cos (t_3, t_1)	Cos (t_3, t_2)	Cos (t_3, t_3)	Cos (t_3, t_4)
t_4	Cos (t_4, t_1)	Cos (t_4, t_2)	Cos (t_4, t_3)	Cos (t_4, t_4)
Here t_1, t_2… are time series				

This result of cosine similarity computation can be used input for hierarchical clustering.

Hierarchical clustering

A hierarchical clustering algorithm creates a hierarchical decomposition of the given data set objects. Depending on the decomposition approach, hierarchical algorithms are classified as agglomerative (merging) or divisive (splitting). The agglomerative approach starts with each data point in a separate cluster or with a certain large number of clusters. Each step of this approach merges the two clusters that are the most similar. Thus after each step, the total number of clusters decreases. This is repeated until the desired number of clusters is obtained or only one cluster remains. By contrast, the divisive approach starts with all data objects in the same cluster. In each step, one cluster is split into smaller clusters, until a termination condition holds. Agglomerative algorithms are more widely used in practice. In this paper, we are using agglomerative algorithm. Agglomerative approach can be done in two different ways: Top-down or bottom-up approach. This will result into cluster with advisors of similar type in the same cluster. Hence, it will result into discovery of similar type of advisors.

An approach toward analyzing advisors intelligence led to discover leaders among a set of time series. This is done by analyzing lead-lag relations among the time series data. An efficient algorithm has been explained by authors [4] which are able to track the lagged correlation and compute the leaders incrementally is been proposed for climate science data. The problem of leadership discovery can be solved in three main steps:

1. Compute the lagged correlation between each pair of time series.
2. Construct an edge-weighted directed graph based on lagged correlations to analyze the lead-lag relation among the set of time series.
3. Detect the leaders by analyzing the leadership transmission in the graph.

We now discuss each step in detail.

Lagged correlation computation

The first step is to compute the lagged correlation between each pair of time series.

We propose to aggregate the effects of various lags and define an aggregated lagged correlation. How to compute the aggregated lagged correlation can be explained by the following example. Fig. 2a shows two-time series X (top) and Y (bottom) with a length of 150. The window length is set to be 120 and we consider the window marked by the dotted rectangle. Fig. 2b shows the lagged correlation at each lag l computed by Equation (1) over the two windows. The maximum lag m=60, i.e., $|l| \leq 60$. When l<0 (i.e., Y is delayed), the positive correlation only exists for $l \in [-60, -39]$ (the shadowed area). When l≥0 (i.e., X is delayed), starting from l=1, we observe a strong increase in positive correlation and it achieves a peak value of 0.81 at l=32. To identify the leadership (X leads Y or Y leads X), we need to aggregate all the observed correlation values over the entire lag span and take the expected correlation value given the two cases of l. The aggregated lagged correlation between two-time series S_i and S_j, denoted as $E_{ij}^{®}$, is then defined as the larger expected correlation value:

$$Eij® = \max\left(E_{ij}(r|l \geq 0), E_{ij}(r|l < 0)\right)$$

We say that S_i leads S_j if $E_{ij}^{®} = E_{ij}(r|l < 0)$, and S_i is led by S_j otherwise if E_i $j^{®} = E_{ij}(r|l \geq 0)$. Such leadership ($S_i$ leads S_j or *vice versa*) is also called the lead-lag relation between S_i and S_j.

Next step is to construct an edge-weighted directed graph based on lagged correlations to analyze the lead-lag relation among the set of time series.

Graph construction

To model the leadership relationships among a set of time series, constructing an edge-weighted graph will help to learn lead-lag

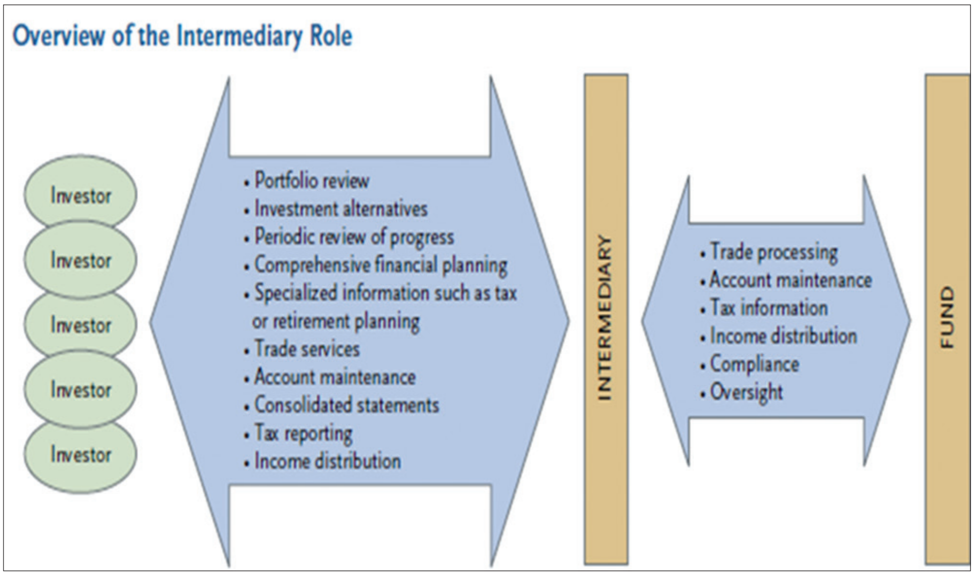


Fig. 1: Overview of the intermediary role

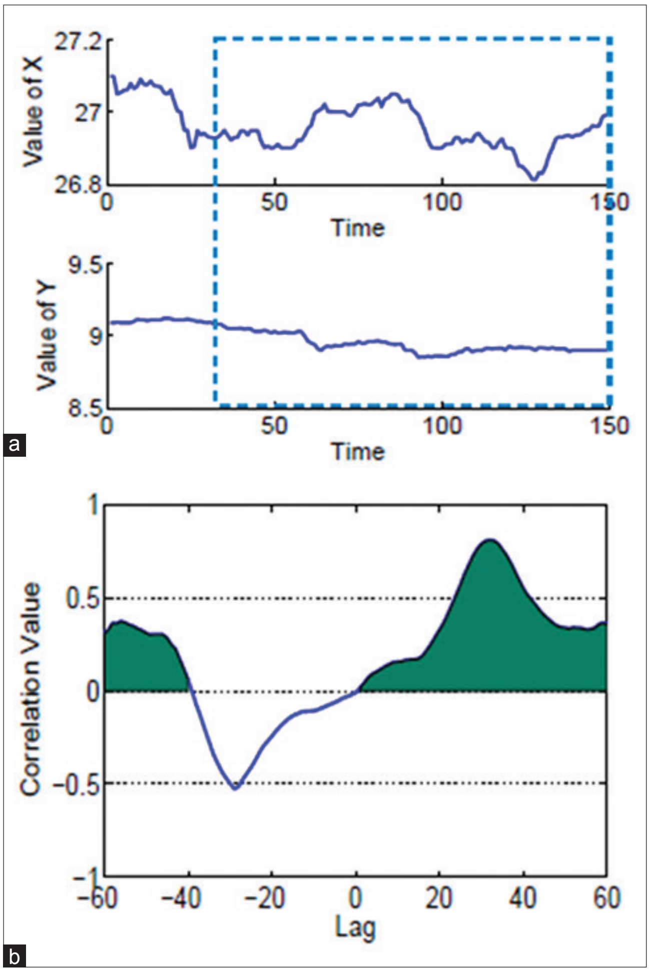


Fig. 2: (a) Two time series, (b) lagged correlation

relationship among the set of time series data. A simple edge-weighted directed graph, G (V, E), where the nodes V={S_1, S_2.... S_N} represents N time series, and the directed edges E represents lead-lag relations between pair of time series. An edge (S_i, S_j) indicates that S_i is led by Sj and its weight is set as $E_{ij}(r)$. Since we are interested in significant lead-lag relations, a correlation threshold g is set such that only those pairs Si and Sj with Eij(r)>g have edges in G.

Leader extraction

Based on the structure of G and the PageRank values of time series, extract the leaders by eliminating redundant leaderships. The basic idea is to first sort the time series by the descending order of their PageRank values and then to remove iteratively the time series that is led either by previously found leaders or by the descendant of previously found leaders. This will result in number of leaders.

CONCLUSION

This paper basically talks about analysis of investment advisors intelligence. A novel problem of discovering leaders from set of time series data based on lagged correlation has been proposed. A time series is learned as leader time series. The movement of leader time series triggers many other time series which are called as followers. Behavior of leader time series helps in learning the behavior of the follower's time series. Proceeding with further analysis of investment advisors intelligence led to pattern recognition of the investment advisor behavior. An approach toward this problem showed great interest in using EAs for pattern recognition tasks and also came with other possible use of EAs combined with other approaches for the development of fully automated pattern recognition systems.

REFERENCES

1. Chandra B, Gupta M, Gupta MP. A Multivariate Time Series Clustering Approach for Crime Trends Prediction, Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference, 12-15 October; 2008.
2. Kumar M, Patel NR, Woo J. Clustering seasonality patterns in the presence of errors. Proceedings of KDD '02. Edmonton, Alberta, Canada.
3. de la Fraga LG, Coello CA. A Review of Applications of Evolutionary Algorithms in Pattern Recognition. Berlin, Heidelberg: Springer-Verlag; 2011. p. 3-28.
4. Wu D, Ke Y, Yu JX, Yu PS, Chen LS. Detecting Leaders from Correlated Time Series. The Chinese University of Hong Kong, University of Illinois at Chicago, The Hong Kong University of Science and Technology.