2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

AI Robo-Advisor with Big Data Analytics for Financial Services

Min-Yuh Day,
Department of Information Management,
Tamkang University,
Taipei, Taiwan,
myday@mail.tku.edu.tw

Tun-Kung Cheng,
Department of Economics,
Tamkang University,
Taipei, Taiwan,
ctk@mail.tku.edu.tw

Jheng-Gang Li,
Index and Quantitative Investment Department,
Yuanta Securities Investment Trust,
Taipei, Taiwan,
jervis.li@yuanta.com

***Abstract*—Robo-Advisors has been growing attraction from the financial industry for offering financial services by using algorithms and acting as like human advisors to support investors making investment decisions. During the investment planning stage, portfolio optimization plays a crucial role, especially for the medium and long-term investors, in determining the allocation weight of assets to achieve the balance between investors expectation return and risk tolerance. The literature on the topic of portfolio optimization has been offering plenty of theoretical and practical guidance for implementing the theory; however, there is a paucity of studies focusing on the applications which are designed for Robo-Advisors. In this research, we proposed a modular system and focused on integrating big data analysis, deep learning method and the Black-Litterman model to generate asset allocation weight. We developed a portfolio optimization module which takes the information from a variety of sources, such as stocks prices, investor profile and the other alternative data, and used them as input to calculate optimal weights of assets in the portfolio. The module we developed could be used as a sub-system for Robo-Advisors, which offers a customized optimal portfolio based on investors preference.**

Keywords— Robo-Advisors, Investment Management, Financial Technology, Deep Learning, Black-Litterman, Portfolio Optimization, Big Data Analysis

I. INTRODUCTION

Financial technology (Fintech) has gotten a lot of traction from the financial institutions and start-ups since recent years. World Economic Forum in 2015 identified financial technology as a disruptive innovation which will reshape the future of the financial industry and make a significant impact on the functions of financial services, which including payment, insurance, deposits and lending, capital raising, and investment management. As an innovation in investment management sector, robo-advisors is the digital platform of financial service that provides automated investment management services to investors by using algorithms and acting as like human advisors to support investors making investment decisions. During the investment planning stage, portfolio optimization plays a crucial role, especially for the medium and long-term investors, in determining the allocation weight of assets to achieve the balance between investors expectation return and risk tolerance. The literature on the topic of portfolio optimization has been offering plenty of theoretical and practical guidance for implementing the theory; however, there is a paucity of studies focusing on the applications which are designed for Robo-Advisors. To fill this research gap, this research aims to develop the system framework for robo-advisor which provides portfolio optimization to support investors during the process of making investment decisions. We proposed a modular system and focused on integrating big data analysis, deep learning method and the Black-Litterman model to generate asset allocation weight. We developed a portfolio optimization module which takes the information from a variety of sources, such as stocks prices, investor profile and the other alternative data, and used them as input to calculate optimal weights of assets in the portfolio. The module we developed could be used as a sub-system for Robo-Advisors, which offers investors an optimal portfolio based on their preference and expectations.

II. RELATED WORKS

A. Robo-Advisors

Robo-advisors is a digital platform which provides investment advisory and financial services for investors by using algorithms with the characteristics of low costs, availability, and ease-of-use. Traditional wealth management is both expensive and exclusive. Automated investment management in the form of robo-advisors seeks to change this and brings wealth management to an affordable price [1]. Robo-advisors is a disruptive innovation in financial services driven by big data [1, 2]. By applying big data and artificial intelligence, it has changed the way of providing wealth management services and forcing traditional advisory to evolve new business model [2, 3].

Robo-advisors make dynamic adjustments and asset allocation to form an optimal portfolio base on the given conditions and financial market status [2]. By collecting customers data and conducting customer segment, robo-advisors allocate funds into assets and re-balance the portfolio based on algorithms, and preparing a report for investors to review their investment performance. The algorithm used by robo-advisors is based on data science and statistical model, following the changing financial market condition, which seeking an optimal portfolio to fit the client's preference and expectation.

B. Deep Learning

Deep learning is a branch of machine learning method in artificial intelligence research field. Composing of multiple processing layers, deep learning allows computational models to extract higher levels of abstraction. In recent years, deep artificial neural networks have proved itself by its success in pattern recognition, speech recognition, and natural language processing [4]. Aside from the application of picture recognition, the sequence to sequence translation is also a trending topic. Most of the former are done with

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

convolutional neural networks, and most of the latter is done with recurrent neural networks, particularly Long Short-Term Memory (LSTM).

C. Long Short-Term Memories for Time Series Forecasting

LSTM, a type of recurrent neural network used in deep learning, has the unique structure enables the model to make a prediction with earlier memories. It has gates to help it decides which information to remember, ignore, forget and select when new information flows in. LSTM can be not only applying on the sequence to sequence translation but also on financial prediction problems, such as pricing securities, constructing portfolio and risk management. Financial prediction problems involve large data sets with complex data interactions that are difficult to specify in standard methods in finance. It may detect these interactions in the data and produce a useful result. Literature [5] also stated that time-series problem could be solved by deep learning approach and unsupervised feature learning. Time series prediction problems are a difficult because of the complexity of a sequence dependence among the input variables added by time series. LSTM can handle these sort of sequence dependences because large architectures can be successfully trained. In addition, it has several advantages, such as expands and includes all possible relevance items to the input data, captures non-linearities and complex interactions among input data more easily to avoid over-fitting and increase in-sample fit versus traditional models.

D. The Black-Litterman Model

The Black-Litterman model [6-10] was developed by Fischer Black and Robert Litterman in 1992 at Goldman Sachs. It solved the issue of input sensitive when using the Markowitz Mean-Variance model [11]. The Black-Litterman model allows the investors to combine their views about certain assets with the market equilibrium returns to formulate the posterior expected return. This model has been widely accepted by the Wall Street mainstream for asset allocation.

III. METHODOLOGY

In this research, the System Development Research Methodology in information systems research [12] was implemented as our method.

A. Research Design

The development has divided into five stages for our robo-advisor system which is shown as follows:

1) Construct a Conceptual Framework: Describe and analyze the research problems for building a robo-advisor. Next, we investigate system requirement and functions, understand the system process and procedure, and survey system criteria and methods.

2) Develop a System Architecture: Develop a modularizable and expandable system architecture. Define the core functions of the system and describe the relationship between each function.

3) Analyze & Design the System: Design and implement the database and knowledge base for the robo-advisor system and evaluate the potential solutions.

4) Build the System: Build the robo-advisor system and explore the solutions by progressively understanding the difficulties and complexity while implementing the system.

5) Observe & Evaluate the System: Observe and evaluate the performance of the robo-advisor system. Improve the system based on the experiences learned from the result of experiments.

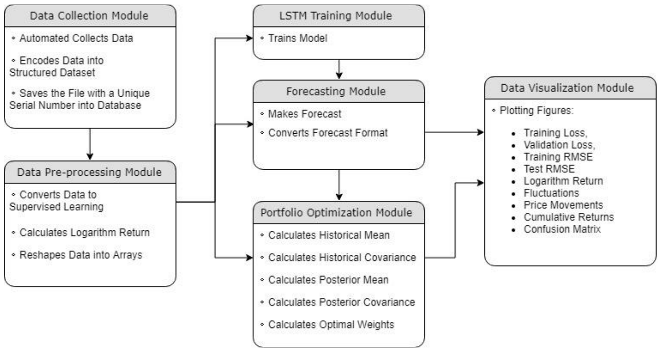

Fig. 1. System Architecture

B. System Architecture

To perform the task of building a customized optimal portfolio, the system is composing by several modules written in Python language (shown in figure1). The modules in our system include:

1) Data Collection Module

It executes automated data collecting, extracting the daily adjusted closing price, encoding the data into a structured dataset, and saving the file with a unique serial number into a database for future access and retrieval.

2) Data Preprocessing Module

It converts and transforms the time series data into a dataset for supervised learning. It calculates the logarithm return at each time point, and reshape the data into arrays that can be processed in the LSTM module. Each array contains a specified number of lag periods, the past n-days returns of adjusted closing prices, as input variables; and a specified number of predict period, the next n-days returns of adjusted closing prices, as the target variable.

3) LSTM Training Module

It performs the function of training the model to capture the long-term tendencies and pattern for making forecasts. It processes the input data and compares its forecasts with the given true values. It will pass the weight to next epoch after each training epochs. It returns the training history and trained model as outputs.

4) Forecasting Module

This module is specially designed to make forecasting with a batch size of 1 to manually solve the error occurring when using the Stateful LSTM. This module uses the model trained by LSTM module to make a forecast and converts the forecast values into the formats that can be used in the various form of applications.

5) Data Visualization Module

It visualizes the data, including records of training loss, validation loss, training RMSE (root mean square errors) and test RMSE, plots figures of logarithm return fluctuations and price movements, cumulative returns, and confusion matrix for portfolios and ETF components.

6) Portfolio Optimization Module

In this module, we perform the portfolio optimization. It generates investor views about assets by using the result obtained from LSTM forecasting module. It calculates the historical mean and covariance, the posterior mean and covariance, and optimal portfolio weights.

C. Subjects and Data Collection

The daily adjusted closing price of ETFs (Exchange Traded Funds) is chosen as our input and target variable. ETF is marketable security that tracks an index, bonds, commodities, or a basket of assets. The benefits of trading ETFs is that it has higher liquidity and lower trading fees comparing with standard mutual funds.

Our data collecting period is 10 years long from 2007/01/03 to 2016/12/30, totally 2518 trading days. 322 ETFs were collected and covered 44 categories of assets (shown in table 1). All of them were incepted before 2007 and still circulating in the U.S. market after 2016. The data was automatically collected by our data collecting module used the data source providing by Yahoo Finance.

D. The Black-Litterman Formula

We used the Black-Litterman model and Markowitz Mean-Variance Optimization to perform the portfolio optimization. The formula for the new Combined Return Vector (E[R]) is

$$E[R] = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1} \, [(\tau\Sigma)^{-1}\Pi + P'\Omega^{-1}Q]$$

where

- E[R] is the new (posterior) combined return Vector (N x 1 column vector);
- τ is a scalar; Σ is the covariance matrix of excess returns (N x N matrix);
- P is a matrix that identifies the assets involved in the views (K x N matrix or 1 x N row vector in the special case of 1 view);
- Ω is a diagonal covariance matrix of error terms from the expressed views representing the uncertainty in each view (K x K matrix);
- Π is the implied equilibrium return vector (N x 1 column vector); and,
- Q is the view vector (K x 1 column vector).

E. Steps to implement the Black-Litterman Model

To implement the Black-Litterman Model, the process can be divided into following steps:

1) *Calculate historical mean and covariance.*
2) *Use historical mean as a priori implied equilibrium return.*
3) *Combine the investor views.*
4) *Calculate the posterior mean and covariance.*
5) *Calculate the portfolio weight through the mean-variance optimization using posterior mean and covariance.*

IV. RESULTS

In this section, we implemented the Black-Litterman model to perform portfolio optimization by integrating with the investor views generated by the LSTM forecasting module as an input for the Black-Litterman model. Six assets were chosen to construct our portfolio (shown in Table I). To show the process of portfolio optimization, we considered a scenario that requires a fund manager to rebalance the portfolio weights at the beginning of each quarter (date is shown in Table II) to reach a higher cumulative return compared with the S&P 500 as a benchmark in 2016. This requires using our LSTM forecasting module to generate the investor views about the performance of chosen assets in each quarter.

TABLE I. SIX ASSETS PORTFOLIO

Symbol	Name	Category
IVW	S&P 500 Growth ETF	U.S. Large Cap Growth Equities
IVE	S&P 500 Value ETF	U.S. Large Cap Value Equities
IWN	Russell 2000 Value ETF	U.S. Small Cap Value Equities
IWO	Russell 2000 Growth ETF	U.S. Small Cap Growth Equities
EFA	MSCI EAFE ETF	Foreign Developed Equities
EEM	MSCI Emerging Markets ETF	Emerging Markets Equities

TABLE II. DATE OF PORTFOLIO REBALANCING FOR EACH QUARTER

	Q1	Q2	Q3	Q4
Date	2016/1/4	2016/4/1	2016/7/1	2016/10/3

A. The LSTM Based Investor Views

We first prepared the investor views by using the LSTM model to predict the quarterly returns for each quarter. We used the past 120 daily returns as training data and forecasts the daily return for the next 60 trading days, which is exactly a quarter long. With the characteristics of the logarithm returns, we can simply obtain the quarterly returns by adding up the forecasted daily returns for next 60 days from the outputs of our LSTM forecasting module. With the predictions of returns and prices for each asset at the end of each quarter (shown in Table III), we can then formulate our investor views of each quarter (shown in table IV).

TABLE III. FORECASTED CUMULATIVE RETURNS AND PRICES

	IVW	IVE	IWO	IWN	EFA	EEM
Initial Price	113.64	85.96	137.26	89.32	55.18	30.57
Predicted Return	4.99%	5.57%	3.10%	3.12%	2.84%	-0.34%
Actual Return	0.73%	2.41%	-5.05%	1.77%	-1.14%	9.39%
Predicted Price	119.30	90.75	141.51	92.10	56.75	30.47
Actual Price	114.47	88.04	130.33	90.90	54.55	33.44
Difference	4.26%	3.15%	8.14%	1.34%	3.98%	-9.73%

The default investor views for each quarter will look like this:

- View 1: U.S. Large Cap Growth Equities will outperform, or underperform, U.S. Large Cap Value Equities by X %.
- View 2: U.S. Small Cap Growth Equities will outperform, or underperform, U.S. Small Cap Value Equities by X %.
- View 3: Foreign Developed Equities will outperform, or underperform, Emerging Markets Equities by X %.

TABLE IV. PREDICTED INVESTOR VIEWS

	View 1	View 2	View 3
Predicted	-0.58%	-0.02%	3.17%
Actual	-1.68%	-6.82%	-10.53%
Difference	1.10%	6.80%	13.71%

B. The Black-Litterman Model

Following the steps to implement the Black-Litterman model, we prepared the quarterly historical data, including historical average returns and the covariance matrix (shown in table V and VI) for each asset in our portfolio as the input for the Black-Litterman model. Next, we combined the investor views with the historical returns using the Black-Litterman formula to generate the posterior mean and posterior covariance matrix. And finally, using the same technique developed by Markowitz, we put the posterior mean and posterior covariance matrix back to the mean-variance optimization to obtain the portfolio weights. In Table VII, it presented the Black-Litterman portfolio weights, which were generated by cooperating with the LSTM investor views, and the Markowitz portfolio weights.

TABLE V. HISTORICAL COVARIANCE MATRIX

Historical Covariance Matrix						
	040_IVW	**041_IVE**	**058_IWO**	**057_IWN**	**086_EFA**	**108_EEM**
040_IVW	0.004767	0.004202	0.005206	0.004212	0.004095	0.004547
041_IVE	0.004202	0.004218	0.004477	0.004013	0.003924	0.004432
058_IWO	0.005206	0.004477	0.007846	0.005978	0.004442	0.004929
057_IWN	0.004212	0.004013	0.005978	0.0053	0.003857	0.004373
086_EFA	0.004095	0.003924	0.004442	0.003857	0.005194	0.005326
108_EEM	0.004547	0.004432	0.004929	0.004373	0.005326	0.008446

TABLE VI. POSTERIOR COVARIANCE MATRIX

Posterior Covariance Matrix						
	040_IVW	**041_IVE**	**058_IWO**	**057_IWN**	**086_EFA**	**108_EEM**
040_IVW	0.004809	0.004242	0.005249	0.004251	0.004133	0.004588
041_IVE	0.004242	0.004259	0.004518	0.004051	0.003963	0.004473
058_IWO	0.005249	0.004518	0.007909	0.006032	0.004482	0.004971
057_IWN	0.004251	0.004051	0.006032	0.00535	0.003893	0.004413
086_EFA	0.004133	0.003963	0.004482	0.003893	0.005244	0.005377
108_EEM	0.004588	0.004473	0.004971	0.004413	0.005377	0.008513

TABLE VII. RETURN VECTORS AND RESULTING PORTFOLIO WEIGHTS

Return Vectors and Resulting Portfolio Weights				
Asset Class	**Posterior Return**	**Historical Return**	**Difference**	**Black-Litterman Weight**
040_IVW	1.35%	2.52%	-1.17%	2.31%
041_IVE	1.00%	1.19%	-0.18%	83.65%
058_IWO	-0.64%	0.91%	-1.55%	0.00%
057_IWN	-0.77%	-0.21%	-0.56%	4.58%
086_EFA	-1.00%	-0.62%	-0.39%	9.46%
108_EEM	-3.30%	-2.13%	-1.17%	0.00%

C. Portfolio Performance

We used the weights to rebalance the portfolio at the beginning of each quarter and presented the performance of the portfolio (shown in table VIII and figure 2). As a comparison, we using S&P 500 index as a benchmark.

TABLE VIII. ANNUAL PORTFOLIO STATISTICS

	Black-Litterman Portfolio - the LSTM Investor Views	**S&P 500 Index**
Annual return	16.151%	9.643%
Annual volatility	13.897%	13.169%
Sharpe ratio	1.14697	0.76492
Stability	0.82500	0.78754
Max drawdown	-10.105%	-10.306%
Skew	-0.35652	-0.36795
Kurtosis	2.49845	2.21958
Daily value at risk	-1.688%	-1.619%
Alpha	0.06445	0.00000
Beta	1.01485	1.00000
Information ratio	0.10935	NaN

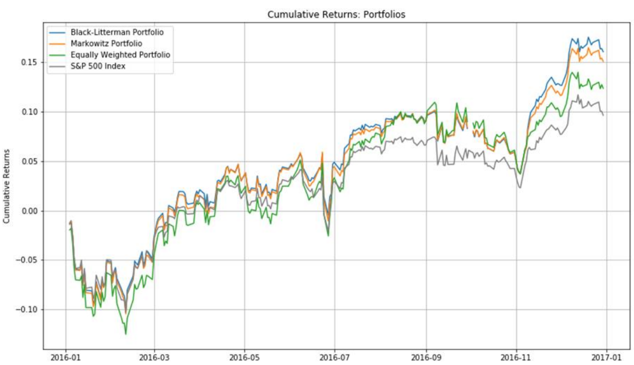

Fig. 2. Cumulated return of the LSTM based portfolio

V. Conclusion

We proposed a modular system and focused on integrating big data analysis, deep learning method and the Black-Litterman model to generate asset allocation weight. We developed a portfolio optimization module which takes the information from a variety of sources, such as stocks prices, investor profile and the other alternative data, then used them as input to calculate optimal weights of assets in the portfolio. For the practitioner implications, the module we developed could be used as a sub-system for Robo-Advisors, which offers investors a customized optimal portfolio based on their preference and the goal of the financial plan.

References

[1] J. Rättyä, "March of the Robo-advisors: The potential for global expansion of digital asset management platforms," 2016.

[2] P. Sironi, *FinTech Innovation From Robo-Advisors to Goal Based Investing and Gamification*. Somerset: Somerset : Wiley, 2016.

[3] S. Chishti and J. Barberis, *The FINTECH Book The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries*. New York: New York : Wiley, 2016.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, pp. 436-444, 05/28/print 2015.

[5] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters,* vol. 42, pp. 11-24, 6/1/ 2014.

[6] F. Black and R. Litterman, "Global portfolio optimization," *Financial analysts journal,* vol. 48, pp. 28-43, 1992.

[7] T. M. Idzorek, "A step-by-step guide to the Black-Litterman model," *Forecasting expected returns in the financial markets,* p. 17, 2002.

[8] A. Meucci, "Beyond Black-Litterman in practice: A five-step recipe to input views on non-normal markets," *Available at SSRN 872577,* 2006.

[9] G. Zhou, "Beyond Black-Litterman: letting the data speak," *Journal of Portfolio Management,* vol. 36, p. 36, 2009.

[10] C. Walters, "The Black-Litterman model in detail," *The Black-Litterman Model in Detail (June 20, 2014),* 2014.

[11] H. Markowitz, "Portfolio selection," *The journal of finance,* vol. 7, pp. 77-91, 1952.

[12] J. F. Nunamaker, M. Chen, and T. D. M. Purdin, "Systems Development in Information Systems Research," *Journal of Management Information Systems,* vol. 7, pp. 89-106, 1990.