Combining Multiple Algorithms for Portfolio Management using Combinatorial Fusion

Conference Paper · July 2017

DOI: 10.1109/ICCI-CC.2017.8109774

4 authors:

Yuxiao Luo
Fordham University

Bruce S Kristal
Brigham and Women's Hospital

Christina Schweikert
St. John's University

D. Frank Hsu
Fordham University

Y. Luo, B.S. Kristal, C. Schweikert, D.F. Hsu. Combining Multiple Algorithms for Portfolio Management using Combinatorial Fusion. Proc. 2017 IEEE 16th Int'l Conf. on Cognitive Informatics & Cognitive Computing (ICCI*CC'17), 2017, 361-366.

Combining Multiple Algorithms for Portfolio Management using Combinatorial Fusion

Yuxiao Luo[1], Bruce S. Kristal[2], Christina Schweikert[3], and D. Frank Hsu[1]

[1] Laboratory of Informatics and Data Mining, Department of Computer and Information Sciences, Fordham University New York, NY 10023, USA

[2]Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital and Division of Sleep Medicine, Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

[3] Division of Computer Science, Mathematics and Science, St. John's University, Queens, NY 11439, USA

Abstract— **Several financial indicators or attributes are used to evaluate the performance of stocks when constructing and managing a portfolio. It is advantageous to utilize multiple algorithms for analyzing and combining these financial indicators, instead of using and optimizing a single algorithm. In this paper, we use the recently developed Combinatorial Fusion Analysis (CFA) to improve portfolio performance at both the attribute level and at the algorithm level. The first phase employs the following algorithms for attribute selection and combination: multiple regression, random forest, support vector machines, and neural networks, and combinatorial fusion according to either diversity strength or performance strength. The second phase involves combining the outputs from these multiple algorithms, using both score and rank combination. Our results suggest that different systems may be preferable for different portfolio sizes. Our results also directly demonstrate that combinatorial fusion can improve portfolio performance.**

Keywords—information fusion, portfolio management

I. Introduction

Investors typically combine diverse assets when building a portfolio to minimize the unsystematic risk, such as can occur in a specific industry or company [22]. Since there are more underlying uncertainties in the financial market, assets for a portfolio can be selected by using various attributes and measurements based on historical stock data. The Sharpe ratio, developed by William F. Sharpe, is the industry standard for measuring risk-adjusted return and determining the expected reward for investing in a risky asset versus a risk-free asset [13]. The following financial indicators have been considered widely in academic and professional communities [1]: Sharpe ratio [13], Price to earnings ratio (P/E) [2], Earnings per share (EPS) [5], Net profit margin (NM) [3], Cash flow per share (CFS) [20], Price to book ratio (P/B ratio) [20], Net income to common margin [20], Price to earnings ratio to growth ratio (PEG ratio) [23], Dividend payout ratio (DPR) [7], Dividend yield [20], Return on common equity (RETURN.COM.EQY) [20], Beta (EQY BETA) [6], Standard deviation [6], Earnings before interest, taxes, depreciation, and amortization (EBITDA) [20], Return on equity (ROE) [16, 20], Sustainable growth rate (SGR) [12], and Free cash flow (FCF) [18].

There are pros and cons for each of the listed attributes. Not surprisingly, none of them work well under all different markets and economic conditions. If the market uncertainty could be modeled by the bell-shaped normal distribution, then the mean-variance models, such as the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT) can produce optimal portfolios [25]. The inability to determine what the distribution characteristics based on limited historical data are subject to several types of errors, including probability and measurement. This situation becomes more complicated as the number of model choices is increased.

In this paper, we abandon the single algorithm (model) optimization method and view portfolio management as a problem in attribute selection and combination by multiple algorithms. Given the number of financial attributes available for stock performance analysis, we use different algorithms to find attributes that are most significant for stock selection. In this study, attribute selection and combination is performed using the following six algorithms: (A) multiple regression, (B) random forest, (C) support vector machines, (D) neural networks, along with (E_1) diversity or (E_2) performance strength. Five of these algorithms are then combined (A, B, C, D, and E_1; and A, B, C, D, and E_2) using combinatorial fusion techniques, where $2^5-1-5 = 26$ rank combinations and 26 score combinations are performed. Our method is quite different from other combination methods (e.g.: [8]) in many aspects. For example, we use a rank-score characteristic (RSC) function to measure the diversity between two attributes or algorithms. In addition, we consider all the 2^n-1-n cases of possible combinations, where n is the number of attributes or algorithms. The Sharpe ratio is used as the metric for performance evaluation in this paper.

In the remainder of this paper, we describe the attributes and stock dataset in Section II, Combinatorial Fusion Analysis in Section III, attribute combinations in Section IV, combination of multiple algorithms based on performance strength or diversity strength in Section V, and conclusion and final remarks in Section VI.

Proc. 2017 IEEE 16th Int'l Conf. on Cognitive Informatics & Cognitive Computing (ICCI*CC'17)
N. Howard, Y. Wang, A. Hussain, F. Hamdy, B. Widrow & L.A. Zadeh (Eds.)
978-1-5386-0771-8/17/$31.00 ©2017 IEEE

II. FINANCIAL INDICATORS AND STOCK DATASET

A. Financial indicators as attributes

Let A = {a_1, a_2, ... , a_{13}} be a list of attributes where each a_i corresponds to a financial indicator (see Table I). Here we use Sharpe Ratio as the performance measure of individual stocks and the overall portfolio. For each d_i in D, the set of stocks being evaluated, there is a numerical value given by each attribute a_i in A. The attribute values are pre-processed according to the following:

(i) Min-max normalization of x in $\{A_i(d_j): j = 1, 2, \dots, n\}$ to x', where $0 \leq x' \leq 1$, using the transformation

$$x' = \frac{x - \min\{A_i(d_j): d_j \text{ in D}\}}{\max\{A_i(d_j): d_j \text{ in D}\}}$$

(ii) The evaluation assumes the higher the score, the more desirable the variable. Accordingly, the order of scores where a lower value is more desirable, is simply reversed.

TABLE I. FINANCIAL INDICATORS AS ATTRIBUTES

	A (a_1, a_2, …, a_{13})
a1	*P/E ratio*
a2	*Earnings per share (EPS)*
a3	*5yr. average net margin*
a4	*Cash flow per share*
a5	*Price-To-Book Ratio*
a6	*Net income margin*
a7	*Dividend yield*
a8	*Return on common stockholder's equity*
a9	*Equity beta*
a10	*Earnings before interest, tax, depreciation and amortization (EBITDA)*
a11	*Return on Equity (ROE)*
a12	*Sustainable growth rate*
a13	*Free Cash Flow (FCF)*
	Performance measure:
	Sharpe ratio

B. Stock dataset

We obtained the dataset from a Bloomberg Terminal which includes financial indicators and the closing prices of 525 stocks from the time period 06/20/2006 to 06/20/2016. Any stocks with missing data, such as values for closing price or financial indicators across the 10 years or indicator values, were removed. After this process, we were left with 257 stocks. In order to match up the federal reserve rates with the closing price on the same date, any records with missing federal reserve rates were filled with the closest rates (20 records).

More specifically, let r_t be the return and y_t be the federal funds interest rate for days t = 1, ..., T . Sharpe Ratios of these 257 stocks are calculated based on the following steps:

Step 1. Compute excess return for each day (250 trading days per year): $e_t = r_t - \frac{y_t - 1}{250}$

Step 2. Convert excess return into an excess returns index: $g_1 = 100$ $g_t = g_{t-1} \times (1 + e_t)$

Step 3. Compute numer of years of data n by taking 250 trading days of a year

Step 4. Compute compounded annual growth rate: $CAGR = \left(\frac{g_t}{g_1}\right)^{1/n} - 1$

Step 5. Compute the annualized volatility: $v = \sqrt{250} SD[e_t]$

Step 6. Compute Sharpe Ratio: $SR = \frac{CAGR}{v}$

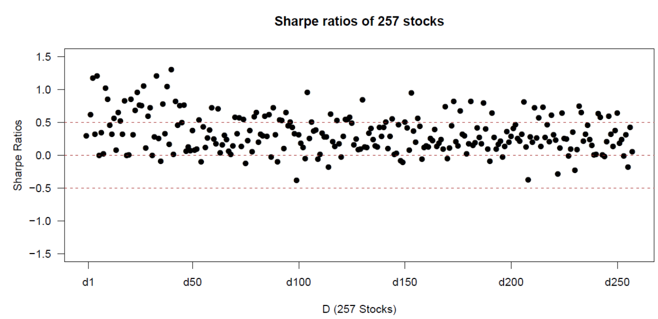
Let D = {d_1, d_2, …, d_n}, where n = 257, be the final dataset consisting of 257 stocks. The 13 financial indicators are represented as attributes in A = {a_1, a_2, …, a_m}, where m = 13. The 257 stocks are grouped into 12 different sectors according to the Bloomberg database system, with four not being defined (Table II).

TABLE II. SECTOR CLASSIFICATION OF STOCKS

Sector Types	**# of Stocks**
Basic Industries	21
Capital Goods	32
Consumer Durables	8
Consumer Non-Durables	27
Consumer Services	42
Energy	8
Finance	16
Health Care	38
Miscellaneous	11
Public Utilities	10
Technology	31
Transportation	9
Not Defined	4

The plot of the 257 stocks on their 10-year average annual Sharpe ratios is shown in Figure 1. This figure shows a variation among this portfolio of 12 industry sections.

Fig. 1. Plot of Shape ratios for the 257 stocks {d_1, d_2, …, d_{257}}



III. COMBINATORIAL FUSION ANALYSIS

A dataset may be analyzed using various descriptive methods, such as regression and forecasting or predictive algorithms, such as classification, neural network, and support vector machine (SVM). Each of which is expected to yield varied results and performance under different market situations. Combination or ensemble methods have been developed with the goal of improving overall performance by combining the results of several methods (e.g.: [25]). The issues related to combination methods involve "when" and "how" to combine these methods or algorithms. Our framework uses the recently developed Combinatorial Fusion Analysis (CFA), which entails the combination of multiple scoring systems [9, 10, 11]. Each scoring system can exist at either of the two different contexts (levels): as an indicator / attribute or as an algorithm / method. A distinctive advantage of CFA over other combination methods is the rank-score characteristic (RSC) function and its corresponding notion of "cognitive diversity" between two attributes or algorithms [9, 10].

A. Multiple scoring systems

A scoring system A, on a set of stocks D = {d_1, d_2, …, d_n}, consists of a score function $s_A(d_i)$, which maps d_i to a set of real numbers in R. There is a corresponding rank function, $r_A(d_i)$, which is generated by assigning ranks to the stocks by sorting their score values in descending order. The Rank-Score Characteristic (RSC) function $f_A(i)$, is a mapping from a rank to its corresponding score value, and is defined as the composite function of the score function s_A and the inverse rank function r_A^{-1} [9, 10]:

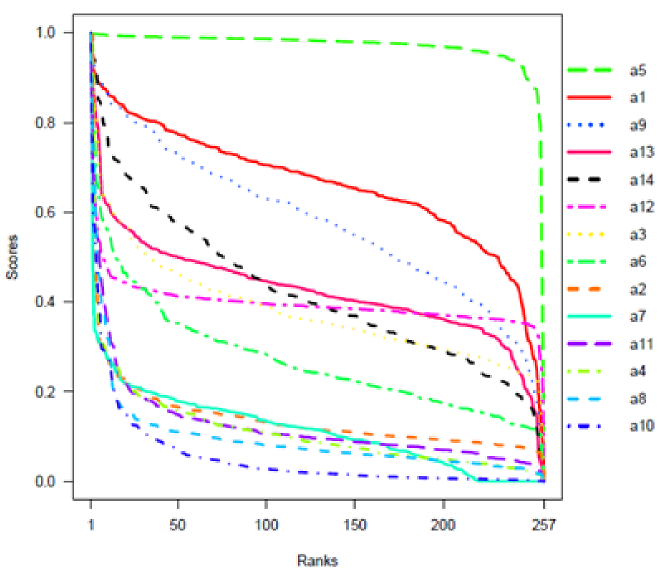
$$f_A(i) = s_A\left(r_A^{-1}(i)\right) = (s_A \cdot r_A^{-1})(i),$$

where i is in N={1,2,…,n} and $f_A(i)$ is in R = the set of real numbers.

B. RSC graph and cognitive diversity

A combination of two systems A and B is considered positive if the performance of the combined system exceeds or equals the best of the individual systems. It has been proposed and demonstrated that the variation between the rank-score functions of two systems, f_A and f_B, can be used as a predictor of a positive combination of systems A and B in several application domains [4, 9, 11, 14, 15, 17, 19, 20, 24]. Figure 2 shows an example of the RSC graph for each of the 13 RSC functions f_{a_i}, where i=1, 2, …, 13 at attribute level, where the x-coordinate represents the rank, the y-coordinate represents the score, and each curve represents the RSC function of the 13 systems (attributes) used in this study.

Fig. 2. RSC graph of 13 attributes a_1 to a_{13}



In statistics, diversity between two scoring systems A and B can be defined as the correlation between score functions s_A and s_B (e.g. Pearson's z correlation) or the correlation between rank functions r_A and r_B (e.g. Kendall's τ or Spearman's ρ). In this paper, the diversity between two scoring systems A and B is defined as the Euclidean distance between two RSC functions, f_{A_i} and f_{A_j}, as follows:

$$d(A_i, A_j) = \sqrt{\sum_{k=1}^{257} (f_{A_i}(k) - f_{A_j}(k))^2}$$

C. Diversity strength and performance strength

For each individual scoring system, we would like to know how it contributes to the overall diversity between the system and all of the other systems. In this case, we define diversity strength of the system A_i to be the average diversity between A_i and the 12 other systems as follows:

$$ds(A_i) = \left(\sum_{j \neq i} d(A_i, A_j)^2\right) \Big/ 12, \qquad where\ i, j \in [1,13]\ and\ i \neq j.$$

Table III(A) gives a ranking of the 13 attributes according to their diversity strength.

TABLE III. DIVERSITY STRENGTH AND PERFORMANCE STRENGTH FOR THE 13 ATTRIBUTES

3(A) DIVERSITY STRENGTH RANKING 3(B) PERFORMANCE STRENGTH RANKING

	Diversity Strength	Rank		Performance.Strength	Rank
a1	0.417567952	2	a1	-0.426610855	12
a2	0.254201281	11	a2	0.288567911	3
a3	0.251501395	12	a3	0.187624423	7
a4	0.263329911	9	a4	0.191868242	6
a5	0.70081557	1	a5	-0.390579755	11
a6	0.240905066	13	a6	0.157855483	9
a7	0.265655087	7	a7	-0.509701811	13
a8	0.275793169	5	a8	0.253440724	5
a9	0.362480118	3	a9	0.181553847	8
a10	0.311137058	4	a10	-0.028854153	10
a11	0.257987763	10	a11	0.272811143	4
a12	0.26382821	8	a12	0.349269156	1
a13	0.267438828	6	a13	0.294127794	2

In this paper, the performance of a system (attribute) a_i is determined by its relation to the Sharpe ratio. As such, the performance strength of the system (attribute) a_i is defined to be the Spearman's rho rank distance between the rank function of a_i, r_{a_i}, and the rank function of the Sharpe ratio (SR), r_{SR}. Therefore, the greater performance strength represents the more significant attribute. Table III(B) gives a ranking of the 13 attributes according to their performance strength.

IV. COMBINING MULTIPLE ATTRIBUTES

In order to select the attributes that contribute the most to forecasting stock performance, we employ the following algorithms for attribute selection: multiple regression (mechanical variable screening), random forest, support vector machines, neural networks, and diversity strength or performance strength. Each algorithm ranks the attributes based on their significance. We then pick the top 6 attributes for combination by each of these algorithms.

For each of the six algorithms, a combination of the six selected attributes is constructed using that algorithm. The new six scoring systems: A, B, C, D, E_1, and E_2 are obtained using average score combination, where the six algorithms are: multiple regression (A), random forest (B), support vector machine (C), neural network (D), diversity strength (E_1), and performance strength (E_2). The six scoring systems are shown in Table IV.

TABLE IV. TABLE OF THE SIX SCORING SYSTEMS

Rank order	MR	RF	SVM	NN	Diversity Strength	Performance Strength
1	a1	a7	a1	a5	a5	a12
2	a2	a1	a7	a4	a1	a13
3	a7	a5	a2	a3	a9	a2
4	a11	a2	a4	a7	a10	a11
5	a13	a11	a3	a1	a8	a8
6	a9	a13	a13	a6	a13	a4
Combined scoring system ->	A	B	C	D	E_1	E_2

The diversity strength, as defined previously, is computed for each scoring system, according to the RSC functions resulting from score combination and rank combination, shown in Tables V(A) and V(B), respectively.

TABLE V. DIVERSITY STRENGTH FOR RSC FUNCTIONS
(A)SCORE COMBINATION

	Diversity.Strength	Rank
A	0.071457817	6
B	0.07378261	5
C	0.138616029	2
D	0.075854054	4
E1	0.070614691	7
E2	0.152765258	1

(B) RANK COMBINATION

	Diversity.Strength	Rank
A	0.037336652	2
B	0.035614728	3
C	0.043204879	1
D	0.027568899	6
E1	0.027845819	5
E2	0.027021736	7

V. COMBINING MULTIPLE ALGORITHMS

Let A, B, C, D, E_1 and E_2 be the six scoring systems (algorithms) obtained in Section IV. In this section, we investigate combinatorial fusion within two different groups of five algorithms, each consisting of A, B, C, D, and either E_1 or E_2. Performance of each algorithm A as a scoring systems on the set of stocks in in D = {d_1, d_2, ..., d_n} is evaluated by the Spearman's rho rank correlation between rank function r_{A_i} and r_{SR}. Moreover, we examine the performance at different portfolio sizes. In each combination, the average score combination and average rank combination of n number, n = 2, 3, 4, or 5 of these 5 algorithms is used.

A. Results

The performance for the score and rank combination of the scoring systems A, B, C, D, E_1 is shown in Table VI. At each portfolio size, the top-performing method of combination is shown. According to this analysis, smaller portfolio sizes result in better performance, since they include the top stocks identified by a method. As the portfolio size increases, the overall performance decreases; however, further analysis would need to be conducted to test if larger portfolios are more stable, or less risky, over time. The graph for portfolio size of 5 stocks is illustrated in more detail in Figure 3. Similarly, the results for the five algorithms A, B, C, D, E_2 are shown in Table VII and Figure 4.

B. Discussion

Based on the results from Tables VI and VII as well as Figures 3 and 4, the combination case including performance strength E_2 generally outperforms the case using the diversity strength E_1. More specifically, E_2 can add power for portfolios sizes of 5, 10, 20, 30 and 40 for rank combination (see right column of Tables VI and VII) and for portfolio sizes of 10, 20, 30 and 40 for score combination (see left column of Tables VI and VII).

We further observe from Figures 3 and 4 (portfolio size of 5 stocks) that combination of all five algorithms do not produce the best result. This phenomena exists also for portfolios sizes of 10, 20, 30, and 40 (see Table VI and VII). In fact, the best performance happens for portfolio sizes of 5 when combining A and E_2 for both score and rank combinations, while it happens when combining B, C, E_1 and B, C, D for score combination and rank combination respectively.

TABLE VI. PERFORMANCE OF THE BEST COMBINATION FOR 5 SCORING SYSTEMS: A, B, C, D, AND E_1 AT EACH PORTFOLIO SIZE

	Score combination of 5 systems		**Rank combination of 5 systems**	
Portfolio size	***Best method***	***Performance***	***Best method***	***Performance***
5	BCE$_1$	0.9361496	BCD	0.8817705
10	CD	0.6669999	C	0.650772
20	ACD	0.5060687	ABD	0.4573593
30	ACD	0.4106033	AC	0.4071315
40	AC	0.3851184	C	0.3847278

Fig.3. Portfolio of size 5 in terms of average return for score and rank combinations of all 31 possible combinations of scoring systems: A, B, C, D, E_1

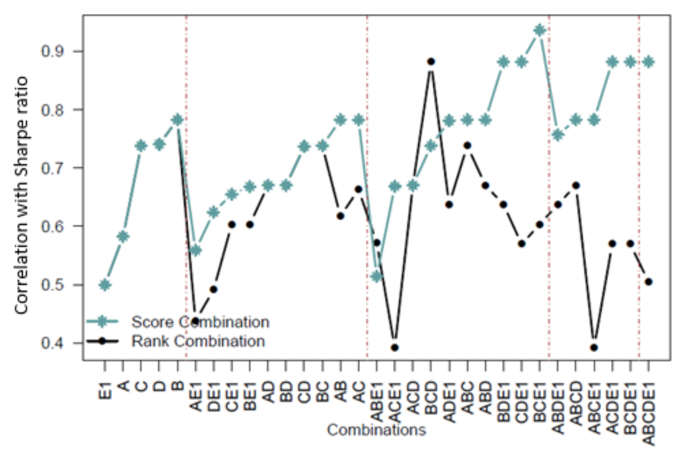
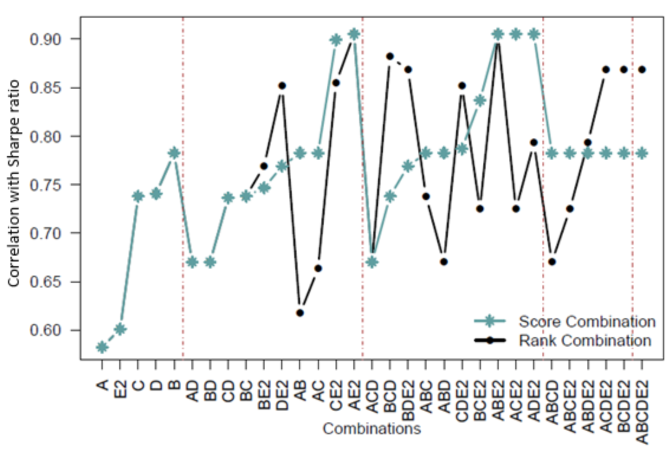


TABLE VII. PERFORMANCE OF THE BEST COMBINATION FOR 5 SCORING SYSTEMS: A, B, C, D, AND E_2 AT EACH PORTFOLIO SIZE

	Score combination of 5 systems		**Rank combination of 5 systems**	
Portfolio size	***Best method***	***Performance***	***Best method***	***Performance***
5	AE_2	0.9050895	AE_2	0.9050895
10	DE_2	0.6910167	$BCDE_2$	0.6910167
20	E_2	0.5975445	E_2	0.5975445
30	E_2	0.5438441	E_2	0.5438441
40	E_2	0.4948047	E_2	0.4948047

Fig.4. Portfolio of size 5 in terms of average return for score and rank combinations of all 31 possible combinations of scoring systems: A, B, C, D, E_2



VI. CONCLUSION AND FINAL REMARKS

Instead of treating market uncertainty as modelled by the mean variance bell-shaped normal distribution, we view portfolio management as a problem in attribute selection and combination by multiple algorithms. A recently developed combinatorial fusion analysis is used not only to select attributes for each algorithm but also to combine multiple algorithms. Our framework demonstrates that different algorithms may be preferred for different portfolio size and that combination of algorithms can indeed improve portfolio performance. In addition, it is also demonstrated that the algorithm using performance strength (E_2) to perform weighted rank combination plays an important role in improving portfolio performance at all sizes from 5, 10, 20, 30 to 40.

Our study provides a novel approach to portfolio management. We call special attention to the use of rank-score characteristic (RSC) functions f_A and f_B to measure the diversity (called cognitive diversity in Section III(B)) between two scoring systems (attributes in this paper) A and B [9, 10]. As noted before, since the RSC function f_A is defined from a rank in N to a score in R, the notion of a cognitive diversity between scoring systems A and B is independent from each of the data items d_i in D (Section III). As such, cognitive diversity and diversity strength are useful in big data analytics to measure the diversity between attributes or algorithms [11, 15, 17, 24, 26].

Our study suggests the following items for further investigation:

(1) Selection of attributes: In this paper, we select the top 6 attributes for each algorithm to perform combination. In general, each algorithm could select a different number of attributes using a distinctive threshold appropriate for that specific algorithm.

(2) Cognitive Diversity between two algorithms: In this paper, we use cognitive diversity to compute diversity strength. In the future, we will plot the RSC graphs for scoring systems A, B, C, D, and E_1 or E_2, respectively. We will test to see if the conventional wisdom will still hold. Specifically, we will test whether the combination of two (or more) systems can be better than each individual system only if they are relatively good and diverse.

(3) Criteria for positive combinations: a positive combination typically requires both good performance and good diversity. But are the criteria for performance and diversity related to each other, and, if so, how?

(4) Performance evaluation: In this paper, we use the Spearman's rank distance between the outcome of the algorithm and the Sharpe ratio to measure the performance of the algorithm. We will explore other means of performance evaluation using real performance of the portfolio on Dow Jones or Standard and Poor across a temporal span, e.g., 3 months, 6 months, or 1 year.

(5) As a simulation and performance test, we will construct stock portfolios based on the system combinations. We will then test the performance of these portfolios in the market for a different time frame.

REFERENCES

[1] P. Barnes. The analysis and use of financial ratios: A review article. Journal of Business Finance and Accounting, 14(4):449-461, 1987.

[2] S. Basu. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. The Journal of Finance, 32(3):663-682, 1977.

[3] K. Berman, J. Knight, and J. Case. Financial Intelligence for Entrepreneurs: What You Really Need to Know About the Numbers. Harvard Business Press, Sep 2008. ISBN 1422119157.

[4] Y.S. Chung, D.F. Hsu and C.Y. Tang. On the Diversity-Performance Relationship for Majority Voting in Classifier Ensembles. In: Haindl M., Kittler J., Roli F. (eds) Multiple Classifier Systems. MCS 2007. Lecture Notes in Computer Science, vol 4472. Springer, Berlin, Heidelberg, 2007.

[5] M. Davies, R. Paterson, A. Wilson. UK GAAP, chapter 23, pages 1321-1375. Palgrave Macmillan UK, 1997.

[6] K.R. French, G.W. Schwert, and R.F. Stambaugh. Expected stock returns and volatility. Journal of Financial Economics, 19(1):3-29, 1987.

[7] A. Gill, N. Biger, and R. Tibrewala. Determinants of dividend payout ratios: Evidence from United States. The Open Business Journal, 3(1):08-14, 2010.

[8] N. Hazarika and J.G. Taylor. Combining models. In Proceedings of the 2001 International Joint Conference on Neural Networks (IJNN-2001), 2001, 1847-1851.

[9] D.F. Hsu, Y.S. Chung and B. S. Kristal. Combinatorial fusion analysis: methods and practices of combining multiple scoring systems. Advanced Data Mining Technologies in Bioinformatics Idea Group Inc, pages 32-62, 2006.

[10] D.F. Hsu, B.S. Kristal, and C. Schweikert. Rank-score characteristics (RSC) function and cognitive diversity. Brain Informatics 2010, Lecture Notes In Artificial Intelligence, Yiyu Yao, Ron Sun, Tomaso Poggio, Jiming Liu, and Ning Zhong (Eds.), Springer-Verlag Berlin Heidelberg, 2010, pp. 42-54.

[11] D.F. Hsu and I. Taksa. Comparing rank and score combination methods for data fusion in information retrieval. Information Retrieval, 8:449-480, 2005.

[12] R. Johnson and L. Soenen. Indicators of Successful Companies. European Management Journal, 21(3):364-369, 2003.

[13] D. Kidd. The Sharpe Ratio and the Information Ratio. In Investment Performance Measurement. CFA INSTITUTE, 2011. URL http://www.cfapubs.org/doi/pdf/10.2469/ipmn.v2011.n1.7

[14] Y. Li, D.F. Hsu, and S.M. Chung. Combination of multiple feature selection methods for text categorization by using combinatorial fusion analysis and rank-score characteristic. Int. J. Artif. Intell. Tools 22, 1350001 (2013), 25 pp.

[15] K.-L. Lin, et al. Feature Selection and Combination Criteria for Improving Accuracy in Protein Structure Prediction. IEEE Transactions on Nanobioscience, 6(2):186-196, 2007.

[16] D. Martani and R. Khairurizka. The effect of financial ratios, firm size, and cash flow from operating activities in the interim report to the stock return. Chinese Business Review, 8(6): 44-55, 2009.

[17] C. Mesterharm and D. F. Hsu. Combinatorial Fusion with On-line Learning Algorithms. In The 11th International Conference on Information Fusion, pages 1117-1124, 2008.

[18] J. Mills and J.H. Yamamura. The power of cash ow ratios. Journal of Accountancy, 186(4):53-61, 1998.

[19] C. Schweikert, Y. Li, D. Dayya, D. Yens, M. Torrents, and D. F. Hsu. Analysis of Autism Prevalence and Neurotoxins Using Combinatorial Fusion and Association Rule Mining. In Bioinformatics and BioEngineering, pages 400-404, 2009.

[20] D.L. Scott. Wall Street Words: An Essential A to Z Guide for Today's Investor. Houghton Mifflin Company, Boston, Massachusetts, revised edition, 2003. ISBN 0-395-85392-3.

[21] C. Spearman. The proof and measurement of association between two things. American Journal of Psychology, 15:72-101, 1904.

[22] M. Statman. How many stocks make a diversified portfolio? The Journal of Financial and Quantitative Analysis, 22(3):353-363, Sep 1987.

[23] M.A. Trombley. Understanding the PEG ratio. The Journal of Investing, 17(1):22-25, 2008.

[24] H.D. Vinod, D.F. Hsu, and Y. Tian. Combinatorial fusion for improving portfolio performance. Advances in Social Science Research Using R, 196:95-105, 2010. Springer New York.

[25] H.D. Vinod and D.P. Reagle. Preparing for the worst: Incorporating downside risk in stock market investments. Wiley Interscience (2005).

[26] J. M. Yang, Y. F. Chan, T. W. Shon, B. S. Kristal, and D. F. Hsu. Consensus scoring criteria for improving enrichment in virtual screening, Journal of Chemical Information and Modeling 45(4), (2005) 1134-1146.