Long-Term Portfolio Management Using Attribute Selection and Combinatorial Fusion

Jason Irukulapati
Laboratory of Informatics and Data Mining, Department of Computer and Information Sciences
Fordham University
New York, New York USA
jirukulapati@fordham.edu

D. Frank Hsu
Laboratory of Informatics and Data Mining, Department of Computer and Information Science
Fordham University
New York, New York USA
hsu@cis.fordham.edu

Christina Schweikert
Division of Computer Science, Mathematics and Science
St. John's University
Queens, New York USA
schweikc@stjohns.edu

Abstract—**Stock portfolios can be constructed based on information provided by financial attributes, such as earnings per share, profit margin, and profit to earnings ratio. Statistical models are often employed to maximize a portfolio's expected return using a combination of those financial indicators. Portfolios are often diversified to include stocks from different industries to minimize risk. In this work, we study the long-term portfolio management problem using attribute selection and combinatorial fusion. Attribute selection is performed using two criteria: diversity strength of an attribute and a sliding rule method. Selected attributes are fused using four combination methods: average combination, weighted combination based on diversity strength or performance strength, and mixed group rank. Analysis based on a historical stock training set and a market test case demonstrate that results from the combinations not only perform better than each individual attribute but also exceed the performance of the Russell 2000 Index.**

Keywords—combinatorial fusion, portfolio management, cognitive diversity, cognitive information fusion

I. INTRODUCTION

Traditional portfolio management often relies on statistical methods to create optimal portfolios. The Capital Asset Pricing Model (CAPM), developed by William Sharpe and John Lintner, builds off of the mean variance model proposed by Markowitz [1]. In the mean variance model, investors choose portfolios that minimize the variance of portfolio return given an expected return and maximize expected return given variance. Based on this optimization, the mean-variance model algebraically constrains the weights of all the assets in this portfolio to satisfy these restrictions. CAPM builds off of this by turning this algebraic relation into a prediction about the relationship between expected return and risk that one can test. CAPM also makes two assumptions about the market: 1) complete agreement, in which investors all agree that returns are sampled from the same distribution and 2) there is borrowing and lending at a universal risk-free rate for investors. The Arbitrage Pricing Theory (APT), developed by Stephen Ross, asserts a linear relationship between the expected return of assets in a portfolio and their variance with respect to other random variables (the covariance), similar to CAPM [2]. However, while CAPM simply looks at variance with the overall market portfolio, APT looks at the variance between returns and multiple factors underlying the price of each asset. Despite the popularity of these methods, they still suffer from a shortcoming since they assume the distribution of returns follows a normal distribution and is identically and independently distributed. However, the distribution of returns is neither normal nor identically and independently distributed, but typically demonstrates skewness and changing variance. This suggests that some conclusions obtained from tests of CAPM tend to be over fitted [3].

In this paper, we propose an algorithmic approach to portfolio management using attribute selection and combinatorial fusion. Algorithmic approaches to portfolio management have been done (e.g. by [4] and [5]) in the past. We will select and combine features directly to form predictors of stock performance and pick out a well performing portfolio. Feature selection will be based on two criteria: diversity strength and a sliding rule method. The algorithm's final performance will be measured by calculating the chosen portfolio's return over a one year period.

II. MULTIPLE ATTRIBUTE SYSTEMS

A. Dataset

We sourced historical data from the Bloomberg Terminal on over 3000 stocks from February 2007 to December 2016. Stocks from the dataset that were missing any values were removed. After cleaning the data, 1903 stocks remained in our dataset.

B. Financial Attributes

Let A = $\{A_1, A_2,..., A_t\}$ be a list of attributes, where each attribute is a feature of a stock (see Table 1 below, where t=8). We chose these features since they were used and justified to be fundamental indicators in [5]. To measure the performance of each individual stock, we use the Return on Equity (ROE) of that particular stock, as done by [5]. Let D = $\{d_1, d_2,..., d_n\}$ be a list of stocks, where each d_i represents a stock. Each attribute assigns a numerical value to the stock for that attribute. The data is pre-processed in the following way:

(i) Min-Max normalization of x in $\{A_j(d_i): i = 1, 2, ..., n\}$ to x', where $0 \leq x' \leq 1$ using the transformation [5]:

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

(ii) We assume the higher the score, the better the variable. For attributes favoring a lower score, the value is simply reversed by 1-x.

TABLE 1. FINANCIAL ATTRIBUTES AS FEATURES

Feature Labels	Financial Attributes
A_1	*Earnings per Share (EPS)*
A_2	*Earnings before interest, tax, depreciation, and amortization (EBITDA)*
A_3	*Profit Margin*
A_4	*Price to Earnings (P/E Ratio)*
A_5	*Cash Flow per Share*
A_6	*Price to Book Ratio*
A_7	*Current Enterprise Value*
A_8	*Volume*

The EPS represents the return on investment to the owner of a share of stock. It is essentially a portion of the entire future stream of earnings the owner of the stock is entitled to upon buying it. As such, it serves as an indicator of the future growth and prospects of a company. If a company is poised to have higher earnings per share, that means it performs better and would facilitate an increase in the price [3]. The earnings before interest, taxes, depreciation, and amortization (EBITDA) measures a company's earnings from its current operations, before the net income is calculated. As a metric for valuation, it can give investors an idea of how much money a company just starting out generates before paying taxes, interests on debt, and accounting for noncash changes. Essentially, the investor gains insight into direct cash flow from business operations. The growth of EBITDA over time can also suggest long-term profitability. Thus, EBITDA serves as a gauge of a company's fair market value [6].

Profit margin is calculated as the ratio of net income to sales. It typically increases as expenditures decrease or as revenues increase. Profit margin can inform investors as to whether or not a company has the ability to gain earnings from its operations. Thus, a higher profit margin suggests that a firm earns favorable revenue given its costs of production and its operating expenses [7]. The P\E Ratio or price to earnings ratio is the ratio of the price of a stock to its earnings per share. The price to earnings ratio serves as a measure of bias towards particular stocks for an individual period of time. In particular, the price to earnings ratio encodes exaggerated investor expectation with respect to growth in earnings and dividends. When investors are optimistic, stocks throughout the market tend to have higher P/E ratios. Conversely, when investors are pessimistic, stocks throughout the market tend to have lower P/E ratios [8].

The cash flow per share is the ratio of the free cash flow of the entire company to the number of outstanding shares. Cash flow represents the amount of actual money flowing into or out of a business. Therefore, a positive cash flow means the company is directly generating cash without spending too much, and is thus managing its money efficiently for sustainable long term performance. A negative cash flow on the other hand, suggests cash is flowing outside of a company [9].

The price to book ratio is the ratio of a stock's market price to its book value. The book value can give multiple insights into the valuation of a company's stock. Book value provides two important valuations of a company: expected future normal earnings and liquidation value. The latter especially holds true for firms experiencing losses; book value can serve as a proxy for the value of abandoning the stock of such a company [10].

The current enterprise value of a company measures how much a company's current operations are worth to financial markets. It measures company value by incorporating equity, debt, and cash. It represents the market capitalization of a company's equity, while also taking into account the market value of the company's debt. It basically serves as an approximation for what it would cost to buy the company. For an investor, the current enterprise value gives them a concrete notion of the underlying value of their stake in the company [11]. The volume refers to the amount of trading sustained in a security or in the market for a given duration of time. Large volumes can demonstrate that significant news or events are expected or have occurred [12].

C. Rank-Score Characteristic Function and Cognitive Diversity

Previous works have explored the use of combinatorial fusion at the systems level to combine the predictions of various algorithms as scoring systems for portfolio management [4]. In this paper, we consider each attribute (feature) as a scoring system.

A feature A_j, $\{j=1...8\}$ for a set of stocks $D = \{d_1, d_2,..., d_n\}$, consists of a score function $s_A(d_i)$ which maps every stock in D to R, the set of real numbers, where each real number denotes a score (Fig. 1). We derive a rank function $r_A(d_i)$ by sorting the score values in descending order. The rank function maps every stock in D to N, the set of natural numbers where each natural number represents the rank of that stock. The Rank-Score Characteristic (RSC) function assigns a score value to a corresponding rank. One defines it as the composite of the score function s_A and the inverse of the rank function r_A [13, 14].

Fig.1. Score function s_A, rank function r_A, and rank-score characteristic (RSC) function f_A of the scoring system A. The relationship among these three functions is also depicted in the formula.

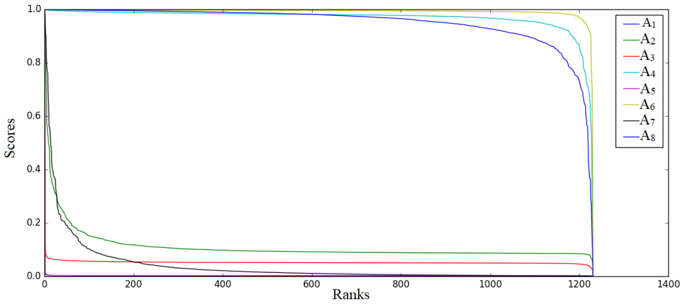
$$\begin{tikzcd} D \arrow[r, "r_A"] \arrow[dr, "s_A"'] & N \arrow[d, "f_A"] \\ & R \end{tikzcd}$$

$$f_A(i) = (s_A \circ r_A^{-1})(i) = s_A(r_A^{-1}(i))$$

$$i \epsilon N = \{1, 2, \dots, n\}$$

The graphs of the RSC functions for each of our eight attributes are shown in Figure 2 below:

Fig. 2. RSC Graph of 8 Features A_1 to A_8

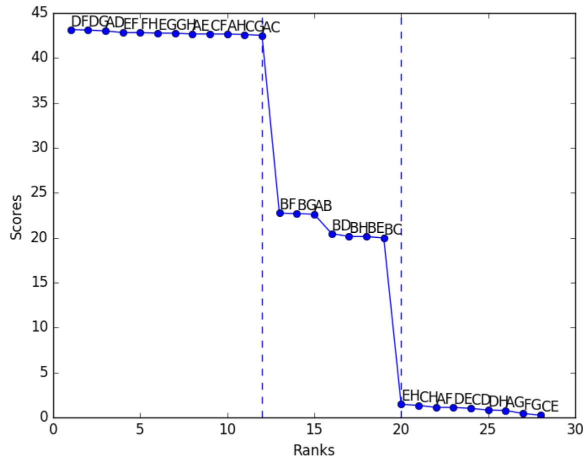


In statistics, diversity between two scoring systems is defined as the correlation between score function s_A and s_B (eg. Pearson's z correlation between s_A and s_B) or as the correlation between rank functions r_A and r_B (Kendall's tau or Spearman's rho between r_A and r_B). In this paper, we use the **cognitive diversity** between attributes A_j and A_k as the Euclidean distance between two rank-score functions for A_j and A_k [14, 15, 16]:

$$d(A_j, A_k) = \left(\sum_{i=1}^{1903} (f_{A_j}(i) - f_{A_k}(i))^2\right)^{1/2}$$

Since cognitive diversity is the Euclidean distance between two rank-score functions, it maps the pair of attributes to a real number. We derive a score function, called the diversity score function, which maps the set of all possible pairs of attributes to the set of real numbers, where each real number score is the cognitive diversity of each pair of attributes. From this score function, we derive a diversity rank function by sorting the pairs of attributes in descending order according to their score. Thus, each of the pairs is ranked by their diversity. Using this diversity score function and diversity rank function, we derive a diversity RSC function that maps the ranks of attribute pairs to the scores, each of which represents the cognitive diversity of each pair. Figure 3 below displays the diversity RSC function for all possible pairs of our attributes.

Fig. 3. Diversity RSC graph for 28 pairs of attributes



III. ATTRIBUTE SELECTION

In this section, we define two methods of attribute selection: diversity strength and a sliding rule method. Both methods utilize the concept of cognitive diversity, as defined in section II, between two financial attributes in terms of their scoring behaviors.

A. *Diversity Strength*

We define the diversity strength of an attribute as the frequency with which it appears in the top 20 pairs of features with the highest diversity in the Diversity RSC function. We choose the 5 features with the highest diversity strength out of our 8 total features. The diversity strength method chose EPS, Profit Margin, P/E Ratio, Price to Book Ratio, and Current Enterprise Value as features for future combination.

B. *Sliding Rule Method*

Performance Strength of a feature is simply its measured performance as an individual scoring system by taking the average of the ROE of what the scoring system decides are the 100 best stocks. We incorporate both diversity strength and performance strength into our attribute selection through a "sliding rule" method. We created two ranking orders of attributes in terms of diversity strength and in terms of performance strength. We move a sliding ruler from the first rank down to the second, then the third, until we obtain six common attributes. This method chose EPS, Profit Margin, P/E Ratio, Cash Flow Per Share, Current Enterprise Value, and Volume.

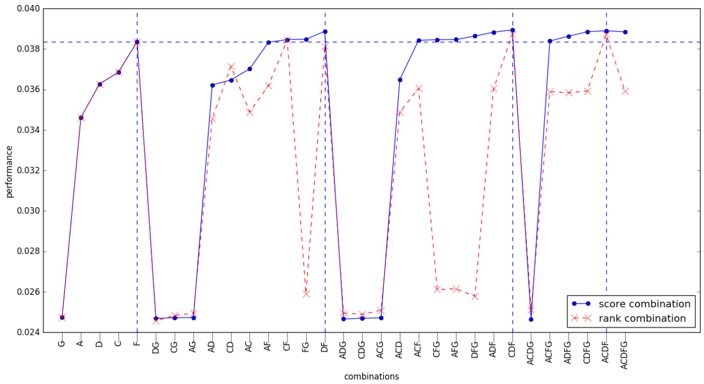
IV. COMBINING MULTIPLE FINANCIAL ATTRIBUTES CHOSEN BY DIVERSITY STRENGTH

We explore four different ways of fusing these five attributes selected by diversity strength: average combinations, diversity weighted combinations, performance weighted combinations, and combinations by mixed group rank. For average combinations, diversity weighted combinations, and performance weighted combinations, the performance of individual features remains the same since applying a weight to an individual feature does not fundamentally change it.

A. *Average Combinations*

For average combinations, fusion is done by simply taking the average of multiple features. The performance for average score combinations and average rank combinations are displayed below in Figure 4.

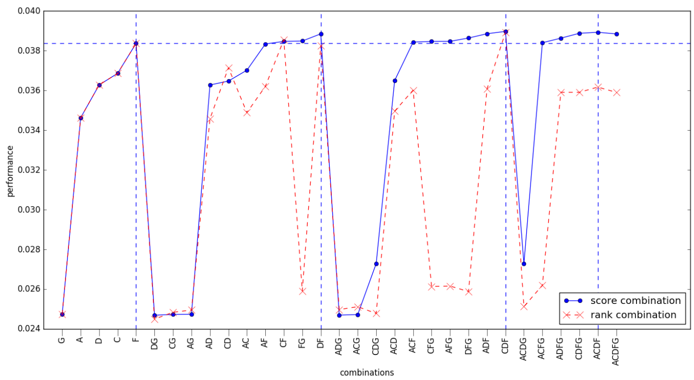
Fig. 4. Average combinations of features selected by diversity strength



B. *Weighted Combinations Based on Diversity*

Here, fusion is done by taking a weighted sum of features, where the weights assigned to each feature are calculated by normalizing the individual diversity strengths of our features intended for fusion. The performance of the diversity weighted score combinations and diversity weighted rank combinations are displayed below in Figure 5:

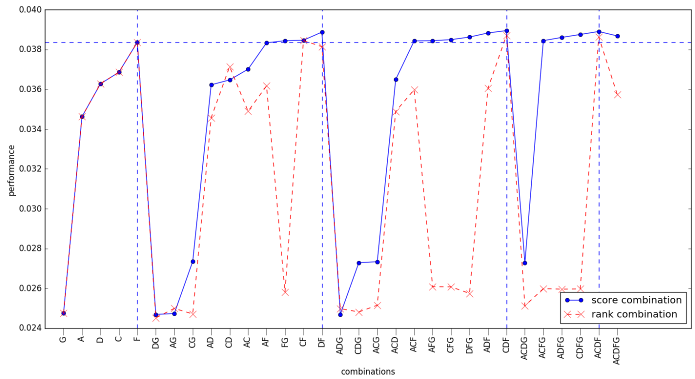
Fig. 5. Diversity Weighted combinations of features selected by diversity strength



C. *Weighted Combinations Based on Performance*

Fusion is again done by taking a weighted sum of features; however, the weights are calculated by normalizing the individual performance strengths of the features we wish to fuse. The performance of the performance weighted score combinations and rank combinations are displayed below in Figure 6:

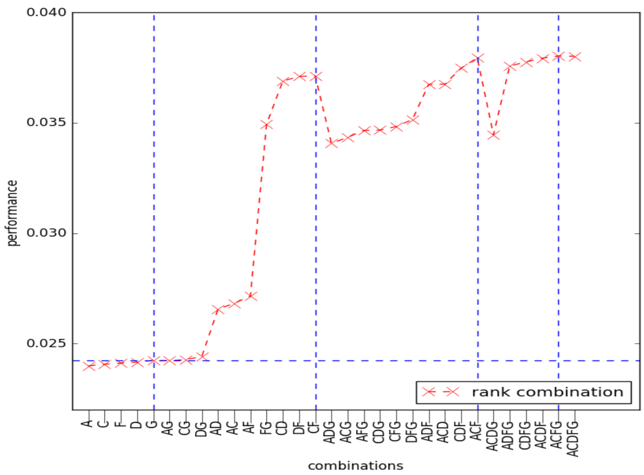
Fig. 6. Performance weighted combinations of features selected by diversity strength



D. *Mixed Group Rank*

Mixed Group Rank (MGR) is a method for fusing ranking systems developed by [17]. The advantage of this method is that it possesses the characteristics of monotonicity and convex sublevel sets. For each combination of ranking systems, MGR linearly combines the lowest ranks of the sub groups of the combination. The performance of rank combinations by mixed group rank is displayed below in Figure 7:

Fig.7. MGR combinations of features selected by diversity strength



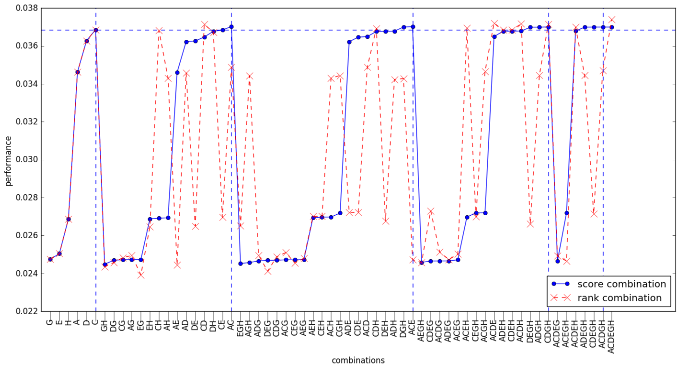
V. COMBINING MULTIPLE FINANCIAL ATTRIBUTES CHOSEN BY THE SLIDING RULE METHOD

Here, we perform the same fusions listed in the previous section, but on the set of six attributes chosen by the sliding rule method. Again, the performance of individual features for average combinations, diversity weighted combinations, and performance weighted combinations remain the same since applying a weight to just an individual feature does not change its predictions.

A. *Average combinations*

The performance of average combinations (both by rank and by score) is displayed below in Figure 8:

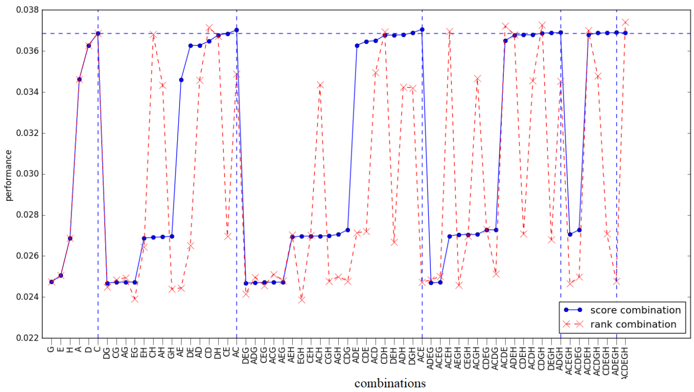
Fig. 8. Average combinations of features selected by the sliding rule method



B. *Weighted Combinations Based on Diversity*

The performance of diversity weighted combinations of features selected by the sliding rule is displayed in Figure 9 below:

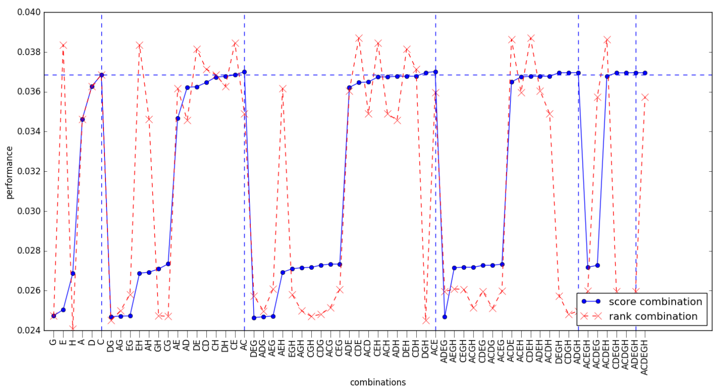
Fig. 9. Diversity weighted combinations of features selected by the sliding rule method



C. *Weighted Combinations Based on Performance*

The performance of performance weighted combinations of features selected by the sliding rule is displayed in Figure 10 below:

Fig.10. Performance weighted combinations of features selected by the sliding rule method

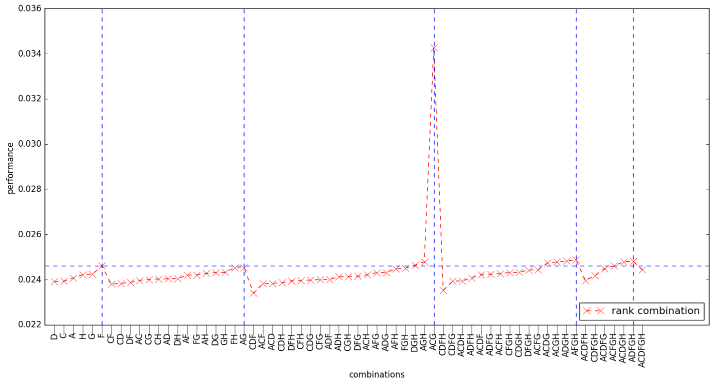


Although Figure 9 looks similar to Figure 8, the performance of some of the rank combinations in Figure 9 exceeds that of the rank combinations in Figure 8. Also, although the performance of the score combinations in Figure 9 looks similar to that of the score combinations in Figure 8, one can see that the performance of many score combinations changes depending on whether they were made using performance weighted combinations or diversity weighted combinations.

D. *Mixed Group Rank Combinations*

The performance of MGR combinations of features selected by the sliding rule is displayed in Figure 11 below:

Fig.11. MGR combinations of features selected by the sliding rule method



VI. MARKET TEST CASE

A. *Testing*

We use the models obtained through fusion in Sections IV and V to create portfolios of 10, 20, 30, and 40 stocks. For each algorithm in sections IV and V, the top 3 combinations are chosen and we select stocks from the intersection of the top 100 stocks recommended by each result. Within each portfolio, stock allocations are determined by the score for score combinations and by the rank for rank combinations. For score combinations, allocations are calculated by normalizing the scores of each of the stocks in the portfolio. For rank combinations, allocations are calculated by normalizing the ranks of each stock in the portfolio. However, the allocations are then reversed, so that lower ranked stocks receive higher allocations. We test each portfolio by simulating an investment of 1 million dollars using market test data from January 03, 2017 to March 19, 2018. The returns of each portfolio over that time period were calculated using a Bloomberg Terminal. The results are displayed in Table 2 for features selected by diversity strength and Table 3 for features selected by the sliding rule method. We use the return of the Russell 2000 Index as our benchmark. From January 03, 2017 to March 19, 2018, the Russell 2000 Index saw a return of 14.13%. Many of our portfolios exceed this performance.

TABLE 2. RETURN ON PORTFOLIO WITH FEATURES CHOSEN BY DIVERSITY STRENGTH

Method	**Portfolio Returns**			
	10 Stocks	***20 Stocks***	***30 Stocks***	***40 Stocks***
Average Score Combination	20.01%	33.79%	37.04%	35.23%
Average Rank Combination	24.9%	8.60%	20.28%	23.40%
Performance Weighted Score Combination	-0.08%	8.66%	-2.27%	15.76%
Performance Weighted Rank Combination	8.40%	13.22%	20.13%	21.91%
Diversity Weighted Score Combination	27.37%	41.79%	37.37%	36.13%
Diversity Weighted Rank Combination	24.94%	8.7%	19.66%	22.38%
Mixed Group Rank	58.29%	22.93%	37.65%	32.25%

TABLE 3. RETURN ON PORTFOLIO WITH FEATURES CHOSEN BY SLIDE RULE METHOD

Method	**Portfolio Returns**			
	10 Stocks	***20 Stocks***	***30 Stocks***	***40 Stocks***
Average Score Combination	29.88%	22.61%	25.38%	25.17%
Average Rank Combination	19.4%	12.95%	13.9%	15.5%
Performance Weighted Score Combination	24.45%	24.46%	24.46%	24.46%
Performance Weighted Rank Combination	12.51%	20.12%	20.13%	20.56%
Diversity Weighted Score Combination	27.06%	11.95%	13.44%	15.12%
Diversity Weighted Rank Combination	26.87%	24.43%	24.41%	24.36%
Mixed Group Rank	25.03%	24.94%	24.91%	25.02%

B. Discussion

Combinatorial Fusion has been applied in many areas, such as virtual screening [16], mobile network selection [18], bioinformatics [19], and visual decision combination [20]. In this paper, we demonstrate how Combinatorial Fusion can be applied to financial indicators for portfolio management. We choose features using two different methods: diversity strength of individual features and the sliding rule method. The diversity strength method chose EPS, Profit Margin, P/E Ratio, Price to Book Value, and Current Enterprise Value as features for fusion. On the other hand, the sliding rule method chose EPS, Profit Margin, P/E Ratio, Cash Flow Per Share, Current Enterprise Value, and Volume. From Tables 2 and 3, we note that the MGR method performs quite well because it is a convex function. In addition, we also see that the diversity weighted based rank or score combinations have relatively good results, in particular diversity weighted score combination in Table 2 and diversity weighted rank combination in Table 3.

VII. SUMMARY AND REMARKS

In this paper, we test each financial attribute as a scoring system which consists of a score function, a rank function, and a rank-score characteristic (RSC) function. Cognitive diversity between the two attributes A_j and A_k was used to define the concept of diversity strength of an attribute. In section III, we use diversity strength and a sliding rule method to select five attributes and six attributes, respectively. Then in sections IV and V, we used four combination methods to combine these five and six selected attributes, respectively: average combination, diversity weighted combination, performance weighted combination, and the mixed group rank (MGR). Each of the three methods of combination, other than MGR, entails both rank and score combinations. The market test case demonstrates that results from our approach not only outperform individual attribute cases, but also exceed that of the Russell 2000 Index, which had a return of 14.13% during the same period of January 03, 2017 to March 19, 2018.

Further investigation needs to be conducted to determine the best attribute selection and attribute combination method for different portfolio sizes. In addition, it would be advantageous to identify which methods perform best for long- or short-term investments.

REFERENCES

[1] Fama, E.F. and French, K.R., The Capital Asset Pricing Model: Theory and Evidence (August 2003). CRSP Working Paper No. 550; Tuck Business School Working Paper No. 03-26.

[2] Huberman, G. and Wang, Z., "Arbitrage Pricing Theory," in Steven N. Durlauf and Lawrence E. Blume, eds., The New Palgrave Dictionary of Economics, 2nd ed. (2008). London: Palgrave Macmillan.

[3] Subramaniam, V.A., Tharshiga, P. Impact of Earning Per Share (EPS) on Share Price (Listed Manufacturing Companies in Sri Lanka) (December 3, 2013). International Journal of Innovative Research and Studies, Vol. 2(12), (2013) pp. 250-258.

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

[5] Vinod, H.D., Hsu, D.F., Tian, Y. Combinatorial Fusion Analysis for Improving Portfolio Performance. In: Vinod H. (eds) Advances in Social Science Research Using R. Lecture Notes in Statistics, vol 196. Springer, New York, NY, (2010) pp. 95-105.

[6] Ebitda. (2014). In Qatar Financial Center, & Qatar Financial Center (Eds.), *QFinance: the ultimate resource* (5th ed.). London, UK: A&C Black.

[7] Profit margin. (1995). In J. K. Shim, & J. G. Siegel, *Dictionary of economics, wiley*. Los Alamitos, CA: Shim and Siegel.

[8] Basu, S. (1977), Investment Performance of Common Stock in Relation to the Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. The Journal of Finance, 32, 32: pp. 663-682. Doi: 10.111/j.1540-6261.1977.tb01979.x

[9] Berman, K., Knight, J., Financial Intelligence for Entrepeneurs: What you really need to about the numbers. Boston, Mass: Harvard Business Press. p. 184.

[10] Subramanyam, K.R. and Venkatachalam, M. The Role of Book Value in Equity Valuation: Does the Stock Variable Merely Proxy for Relevant Past Flows? (June 1998). Available at SSRN: https://ssrn.com/abstract=113388

[11] Enterprise Value. (2014). In Qatar Financial Center, & Qatar Financial Center (Eds.), *QFinance: the ultimate resource* (5th ed.). London, UK: A&C Black.

[12] Volume. (2003). In D. L. Scott, *Wall Street words* (3rd ed.). Boston, MA: Houghton Mifflin.

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

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

[15] Lyons, D,M,, and Hsu, D,F. Method of Combining Multiple Scoring Systems for Target Tracking using Rank-Score Characteristics. Information Fusion 10(2), (2009) pp. 124-136.

[16] Yang, J.M., Chen, Y.F., Shen, T.W., Kristal, B., Hsu, D.F. Consensus scoring for improving enrichment in virtual screening. J Chem Inform Model, 45, (2005) pp. 1134–1146.

[17] Zhang, C., Vardi, Y., Melnik, O. "Mixed Group Ranks: Preference and Confidence in Classifier Combination," in *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 26, no. 8, (2004) pp. 973-981. doi:10.1109/TPAMI.2004.48

[18] Kustiawan, I., Liu, C.Y., Hsu, D.F. Vertical Handoff Decision Using Fuzzification and Combinatorial Fusion, IEEE Communications Letters, Volume: 21, Issue: 9, (2017) pp. 2089 – 2092.

[19] Schweikert, C., Brown, S., Tang, Z., Smith, P., Hsu, D.F. Combining multiple ChIP-seq peak detection systems using combinatorial fusion. BMC Genomics, 13(Suppl 8):S12 (2012).

[20] Schweikert, C., Mulia, D., Sanchez, K., Hsu, D.F. The diversity rank-score function for combining human visual perception systems, Brain Informatics, 3, (2016) pp. 63–72.