

Forecasting Daily Stock Market Return Using Dimensionality Reduction

Xiao Zhong

Department of Engineering Management and Systems Engineering
Missouri University of Science and Technology
204 Engineering Management, 600 W. 14th Street, Rolla, MO, 65409-0370, USA
xz4y4@mst.edu

David Enke*

Laboratory for Investment and Financial Engineering
Department of Engineering Management and Systems Engineering
Missouri University of Science and Technology
221 Engineering Management, 600 W. 14th Street, Rolla, MO, 65409-0370, USA
enke@mst.edu

*Corresponding author: 573-341-4749, enke@mst.edu

Abstract

In financial markets, it is both important and challenging to forecast the daily direction of the stock market return. Among the few studies that focus on predicting daily stock market returns, the data mining procedures utilized are either incomplete or inefficient, especially when a large amount of features are involved. This paper presents a complete and efficient data mining process to forecast the daily direction of the S&P 500 Index ETF (SPY) return based on 60 financial and economic features. Three mature dimensionality reduction techniques, including principal component analysis (*PCA*), fuzzy robust principal component analysis (*FRPCA*), and kernel-based principal component analysis (*KPCA*) are applied to the whole data set to simplify and rearrange the original data structure. Corresponding to different levels of the dimensionality reduction, twelve new data sets are generated from the entire cleaned data using each of the three different dimensionality reduction methods. Artificial neural networks (*ANNs*) are then used with the thirty-six transformed data sets for classification to forecast the daily direction of future market returns. Moreover, the three different dimensionality reduction methods are compared with respect to the natural data set. A group of hypothesis tests are then performed over the classification and simulation results to show that combining the *ANNs* with the *PCA* gives slightly higher classification accuracy than the other two combinations, and that the trading strategies guided by the comprehensive classification mining procedures based on *PCA* and *ANNs* gain significantly higher risk-adjusted profits than the comparison benchmarks, while also being slightly higher than those strategies guided by the forecasts based on the *FRPCA* and *KPCA* models.

Key words: Daily stock return forecasting; Principal component analysis (*PCA*); Fuzzy robust principal component analysis (*FRPCA*); Kernel-based principal component analysis (*KPCA*); Artificial neural networks (*ANNs*); Trading strategies