Stacking generalization via machine learning for trend detection in financial time series

Vittorio Carlei[1], Gaetano Adamo[2], Oksana Ustenko[3], and Valeriya Barybina[4]

[1] University of Chieti-Pescara, Pescara 65122, Italy
carlei@unich.it,
[2] Qi4M, Rome 00144, Italy
adamo@qi4m.com,
[3] Qi4M, Rome 00144, Italy
ustenko@qi4m.com
[4] Qi4M, Rome 00144, Italy
barybina@qi4m.com

Abstract. The task of understanding and modeling the dynamics of financial data has a significant practical value. In particular, it can help intercept trend inversion signals, providing an accurate future forecast that is important for asset allocation, investment planning, portfolio risk hedging and so on. Yet, the irregular fluctuations, chaotic dynamics and constantly changing patterns of financial data make time series modeling a challenging task in this domain. In this paper, we propose a classifier ensemble operator based on stacking generalization, which is applied to a pool of individual signals generated by a Poisson process-based model. The forecasting ability of the methodology is tested on a set of price time series. The results of the ensemble model application demonstrate the increased accuracy of prediction and a mitigated sensitivity of the model to parameters, outperforming the output of individual model components.

Keywords: Poisson process; Classifier ensemble; Stacking generalization; Neural networks; Trend detection

1 Introduction

Time series forecasting and its application across various domains are among the main topics of interest for the machine learning community. Financial time series refer to a range of economic and business data, such as exchange rate, stock market, price index, national income and so on. The financial market itself is a non-stationery and regime-switching environment, where high volatility and noise are intrinsic features. Besides, financial time series are often affected by several complex economic variables that include growth, interest rate, inflation, deflation, political and psychological factors, etc. [1, 2]. Due to the irregular fluctuations, chaotic dynamics and constantly changing patterns of financial time series it is quite challenging for traditional statistical models to produce an accurate prediction in this domain [3]. It is, however, important to overcome the challenge and adopt a method for intercepting changes in relevant trend dynamics, as this can significantly contribute to the decision process support, resulting into a more efficient asset allocation, investment planning, portfolio risk hedging etc.

When looking specifically at the stock market dynamics, we infer that the persistently high price volatility implies high risks for shareholders [4]. By diversifying the investment portfolio, it is possible to overcome the issue of company specific risk, yet the returns will continue to be exposed to the systematic market risk [4]. With effective ways of intercepting changes in stock market trend dynamics, asset managers and investors are able to exercise more informed and accurate decision-making.

The increase in computing power and availability of data have caused the shift from applying linear statistical models (exponential smoothing, ARIMA) to extensively using artificial Neural networks (NNs) for financial data forecasting purposes [5, 6]. Empirical results [7] have shown that the latter outperform linear regression models in the complex and chaotic market environment and are more efficient in producing accurate market signals. In order to deal with sampling and modeling uncertainties that can affect forecasting accuracy and robustness, NNs are typically used as ensembles of several network models [8, 9, 10, 11]. Ensembles are a single classification architecture with a defined set of features, employed in order to combine forecasts from the multiple and diverse classifiers that comprise them. This approach is based on the studies, which confirm that a variety of classifiers make uncorrelated errors, and by using a set of classifiers that operate on diverse features it is possible to improve the overall classification accuracy, as opposed to using the results of a single best one [11, 12, 13, 14].

The ensemble methods used are meta-algorithms built on various machine learning techniques that in combination construct a single predictive model that can decrease variance (bagging), bias (boosting) and predictions (stacking) [15, 16, 17]. However, more recent research suggests that instead of creating a set of classifiers and then combining it, one should rather build a pool with the available set of classifiers and adopt the ones that perform better than others, reducing the initial size of the pool. Zhou et al. [18] explicitly describe the conditions where it would be appropriate to create an ensemble of many but not all the available classifiers.

Our motivation behind applying the stacking generalization approach is that different learning models have different areas of expertise across the input space [12]. Aiolfi and Timmermann [19] suggest there is evidence that forecasting models have a varying relative performance over time. In addition, the process behind time series generation often has recurrent structures due to factors such as seasonality [20]. Therefore, we assume that the metalearning strategy allows the ensemble to better intercept changes in relative performance of models and adapt to the dynamic environment. Besides, an important reason for the chosen approach is the features of the base learner, employed to produce the signals that compose the ensemble. We assume that the price trend inversion follows an inhomogenous Poisson process. This is useful for detecting changes in price movement but exhibits a high sensitivity to parameters, depending on observation time (e.g. probability threshold) or the method used for estimating the intensity of the Poisson process observed. Such parameter sensitivity of the model is effectively mitigated during further ensemble construction.

The main contribution of this paper is the presentation and experimental analysis of the ability of a stacking generalization metamodel, that employs the Poisson process for initial signals' pool generation, to improve the diversity and the predictive accuracy of ensembles in financial time series forecasting. In Section 2 we describe the model behind the signal generation and subsequent classifier ensemble construction. The experimental evaluation of the output of our model in comparison with its individual components is given in Section 3, which is followed by the main conclusions of this paper and an outline of possible future research directions in Section 4.

2 Classifier ensemble construction

In finance we adopt the term "trend" to describe the direction of asset prices. In other words, the term "trend" captures the ascending or descending price fluctuations over a period of time, with the latter not being subject to any limitations. Our goal is to estimate the changes of trend direction, predicting its future behavior through inferences from historical analysis of financial time series data.

We begin by defining the strategy; given a discrete time series of prices[5] $\{p_t\}_{t=1,..,n}$, and a rule $\Re$, which depends on the length of the investment horizon, we define the direction of the trend $\forall\ t = 1, ..., n$ as a sequence:

$$d_t \doteq \begin{cases} +1 & \text{if the trend is ascending under } \Re \\ -1 & \text{if otherwise} \end{cases} \tag{1}$$

We then use $\{d_t\}_{t=1,...n}$, which represents the target we want to predict, to build three sequences Nt. Let[6] $L : t \longrightarrow L(t) \in \mathbb{N}$ be such that $L(t) <= t$ and

$$\begin{aligned} N_t^+ &\doteq \textstyle\sum_{1=t-L(t)}^{t} \mathbb{I}_{[d_i=1]} \\ N_t^- &\doteq \textstyle\sum_{1=t-L(t)}^{t} \mathbb{I}_{[d_i=-1]} \\ N_t &\doteq \textstyle\sum_{i=t-L(t)}^{t} \mathbb{I}_{[d_i \neq d_{i-1}]} \end{aligned} \tag{2}$$

For the sake of simplicity, we will keep $L(t)$ constant for every $t = 1, ..., n$. These three sequences represent the realization of the three counting processes[7] at time t: for each $t = 1, ..., n$, N_t^+ (N_t^-) counts the number of times the price

[5] Temporal indexing t refers to an arbitrary observation frequency.

[6] The L function is the time window on which the parameters will be estimated. This function could also depend on $\Re$, or on the relevant state of the market [21].

[7] i.e. The increments over a time frame of L steps.

has increased (decreased) over the last L observations. N_t, instead, counts the number of times the trend has switched between N_t^+ and N_t^-. Let τ_t^+, τ_t^- and τ_t be the sequences that represent the inter-arrival time between successive events, represented respectively by N_t^+ , N_t^- and N_t .

We interpret the sequence $\left\{N_t^+\right\}$ as a non-homogeneous Poisson process[8] $\mathcal{N}_t^+$, whose intensity is $\lambda^+(t)$. Consequently, the sequence $\left\{\tau_t^+\right\}$ represents the realization of an event at a particular arrival time, that is a trajectory of the process $\mathfrak{T}_t^+$ of exponential random variables with parameter $\lambda^+(t)$. The same approach is used for the successions $\left\{N_t^-\right\}$ and $\{N_t\}$.

The above-stated is confirmed if empirical observations satisfy the standard assumptions [22]. The choice of $\Re$ and L and the techniques used to estimate the parameter[9] $\lambda^*(t)$, the examples of which will be provided in the following section, are determinant to the outcome.

We proceed with the notion that a random variable $X \sim exp(\lambda)$ has $E[X] = \frac{1}{\lambda}$ and $\mathrm{Var}[X] = \dfrac{1}{\lambda^2}$, while in the case of a variable $Y \sim \mathrm{Poisson}(\lambda)$ we observe the following: $\mathrm{E}[X] = \mathrm{Var}[X] = \lambda$. As shown in Figure 1, a multitude of techniques may be applied in order to estimate the intensity of the process.

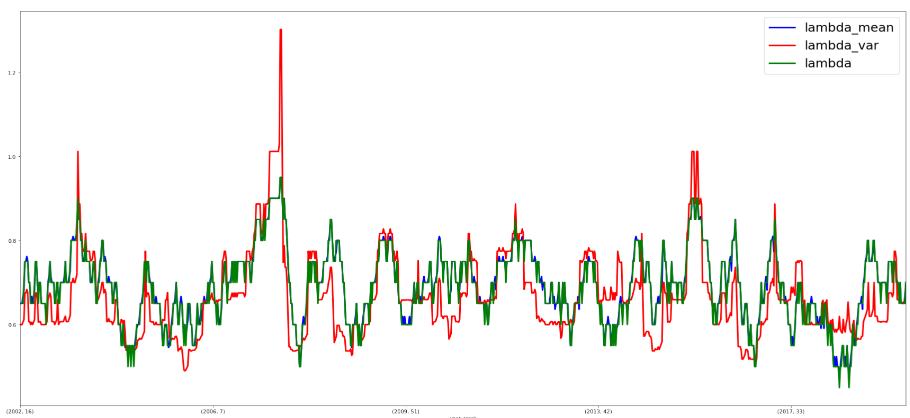

Fig. 1. The graph above represents a plot of three different intensity $\lambda(t)$ estimates of the process $\mathcal{N}_t$, built on the historical data of the examined price series, with weekly observation frequency, $L = 20$, $\Re = \{trend\ is\ ascending\ if\ today's\ return\ is\ greater\ than\ yesterday's\ return\}$. Intensity estimation was performed through the calculation of variance and mean of the inter-arrival time observations (red and blue respectively), and the average of the counting process observations (green) on L observations.

[8] Homogeneity is avoided to pursue a general path.
[9] Refers to one of the three processes we built in 2.

There is no universality regarding the intensity estimation techniques. Therefore, the intensity calculation varies depending on the specific application case. The results presented in this study are obtained via counting process average.

Having obtained the intensities of the processes built in 2, and knowing the parameters $\lambda^*(t)$, we are able to calculate event probabilities for the trend at step $n+1$.

For example,using inter-arrival times, we can estimate the probability that at the next observation the price will exhibit an ascending $\Re$-trend

$$\mathbb{P}(\mathfrak{T}^+_{n+1} < \mathfrak{T}^-_{n+1}) = \mathbb{P}(d_{n+1} = 1) = \frac{\lambda^+(n+1)}{(\lambda^+(n+1) + \lambda^-(\text{n+1}))} \tag{3}$$

and can, consequently, open a long position if the output 3 exceeds a Θ threshold[10]. Similarly, we can calculate

$$\mathbb{P}(\mathfrak{T}_{n+1} < \mathfrak{T}^+_{n+1}) = \frac{\lambda(n+1)}{(\lambda(n+1) + \lambda^+(\text{n+1}))}, \tag{4}$$

which is the probability that the trend switch will occur prior to the arrival of an event characterized by a positive trend below $\Re$. Therefore, in the ascending trend phase, a short position can be opened if the relevant output exceeds a Θ threshold.

Another example of applying the counting process is to calculate the probability that the next price observed is going to be in a descending phase under $\Re$:

$$\mathbb{P}(\mathcal{N}^+_{(n+1)} = 0) = e^{-\lambda^+(n+1)}; \tag{5}$$

or in an ascending phase

$$\mathbb{P}(\mathcal{N}^-_{(n+1)} = 0) = e^{-\lambda^-(n+1)}, \tag{6}$$

suggesting a long position if $\mathbb{P}(\mathcal{N}^-_{(n+1)} = 0) > \mathbb{P}(\mathcal{N}^+_{(n+1)} = 0)$.

By doing so, at each event X_i, $i = 1, ..., K$ that describes the trend, we associate a predictor R_i, $i = 1, ..., K$, whose output is a binary sequence (long, short) $P_{i,t}$ with $t = 1, ..., T$, which represents one of our basic classifiers.

Note that once fixed a rule $\Re$, it is possible to create multiple predictors by changing the definition of events whose probability we want to estimate, the Θ threshold and the defining function L(t). The accuracy of these predictors depends on the distance between the lambda estimates, whereas the sensitivity to parameters implies that the *a posteriori* choice of the parameters can lead to overfitting. To overcome this limitation, we have created several hundred classifiers based on a multitude of parameters and have left the ensemble creation task to a Neural network algorithm.

Ensemble methods are based on the idea of combining a set of individual predictors R_1, R_2, ... , R_m, in order to build a decision function f by means

[10] For example, if this value is greater than its historical average plus c times its historical standard deviation.

of an aggregation operator that combines the individual forecasts instead of the popular keep-the-best (KTB) model [23, 24].

Our base learning model generates eight predictive signals. We subsequently employ a deep learning algorithm to construct a classifier ensemble that will improve the accuracy of the final signal, which we will use to obtain the prediction P_f. The precise process flow is shown in figure 2.

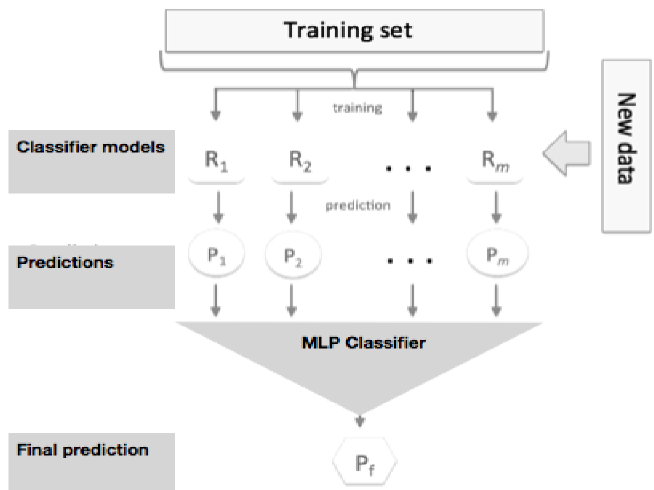

Fig. 2. The diagram shows each step of the process flow for the generation of the final prediction.

We employ the feedforward multilayer perceptron as our metalearning model to build the function f. The forecast for $y_{(t+1)}$ is obtained using single predictors, which were prior built employing the Poisson process, as explanatory variables. The target function is identical to the one used for individual predictors construction. Our final output is a binary indicator, used as the decision support signal in relation to a particular asset:

- 0 for a cash/short operation to be executed on the day t+1;
- 1 for a long positions to be executed on the day t+1.

3 Experimental evaluation

In the table below we provide a comparison between the classification accuracy of the stacking ensemble and of the individual predictors. The backtesting results are obtained using the walk-forward validation method on four time series, selected arbitrarily [25].

Table 1. Output comparison between the stacking ensemble and the individual predictors.

Classifier	Nvidia	Microsoft	DAX	S&P500
Classifier R_0	54%	55%	53%	51%
Classifier R_1	53%	57%	55%	55%
Classifier R_2	58%	58%	57%	56%
Classifier R_3	53%	54%	52%	54%
Classifier R_4	57%	58%	57%	55%
Classifier R_5	54%	55%	53%	53%
Classifier R_6	53%	55%	52%	53%
Classifier R_7	48%	50%	48%	46%
Stacking ensemble	73%	73%	75%	74%

The accuracy of the classifiers is calculated on an out-of-sample ten-year period. Due to the application of walk-forward validation, the behavior and performance of individual classifiers vary over time. The stacking ensemble that uses different classifiers as input, is employed to determine the best performing one on a relevantly small time window.

4 Conclusion

Forecasting financial time series is a challenging task for analysts and researchers, due to the irregular fluctuations, chaotic dynamics and constantly changing patterns that characterize financial data. In this paper we have proposed a Neural network metamodeling technique, which is based on stacking generalization classifier ensemble, applied to a pool of individual signals from a Poisson process-based model.

Empirical results of practical data experiments on a set of price time series demonstrated that the Neural network-based metamodel can be used as an effective approach to financial time series forecasting as it outperforms the individual components of the underlying model. It shows substantially improved predictional accuracy, which was the main focus of our application, when modeling stock price trend dynamics, as well as a mitigated sensitivity to parameters. Our subsequent objective is exploring the ability of increased accuracy signals to enhance the financial performance of a relevant investment product. Further research may also be directed to exploring the scalability of the proposed method in relation to other price series (e.g. testing multi-index or multi-asset class application).

References

1. Y. Sun, Exchange rate forecasting with an artificial neural network model: can we beat a random walk model? *Master of Commerce and Management (MCM) thesis*, Lincoln University, New Zealand, 2005.

2. E. W. Tyree, J.A. Long, Forecasting currency exchange rates: neural networks and the random walk model. *Proceedings of the third international conference on artificial intelligence applications*, Wall Street, New York, 1995.
3. A. Timmermann, C. W. J. Granger, Efficient market hypothesis and forecasting, *International Journal of Forecasting*, vol. 20, pp. 15-27, 2004.
4. B. Krollner, B. Vanstone, and G. Finnie, Financial time series forecasting with machine learning techniques: A survey, *Paper presented at the European symposium on artificial neural networks: Computational and machine learning*, Bruges, Belgium, 2010.
5. C. Bellgard, P. Goldschmidt, Forecasting across frequencies: linearity and non-linearity, *Proceedings of the conference on advanced investment technology, Gold Coast, Australia*, pp. 41-48, 1999.
6. N. Kourentzes, D. K. Barrow, S. F. Crone, Neural network ensemble operators for time series forecasting, *Expert Systems with Applications*, vol. 41, pp. 4235-4244, 2014.
7. M.-C. Chan, C.-C. Wong, C.-C. Lam, Financial Time Series Forecasting by Neural Network Using Conjugate Gradient Learning Algorithm and Multiple Linear Regression Weight Initialization, *Computing in Economics and Finance*, vol. 61, 2000.
8. A. Krogh, J. Vedelsby, Neural Network Ensembles, Cross Validation and Active Learning, *Advances in Neural Information Processing Systems*, vol. 7, pp. 231-238, 1995.
9. M. Khashei, M. Bijari, An artificial neural network (p, d, q) model for time series forecasting, *Expert Systems with Applications*, vol. 37, pp. 479-489, 2010.
10. T. G. Dietterich. Ensemble methods in machine learning, *Multiple Classifier Systems*, pp. 1-15, 2000.
11. L. I. Kuncheva, Combining pattern classifiers: Methods and algorithms, *Wiley-Interscience*, 2004.
12. C. Chatfield, The analysis of time series: An introduction, ed. 6, CRC Press, 2013.
13. G. Brown, L. I. Kuncheva, "Good" and "bad" diversity in majority vote ensembles, *Proceeding of Multiple Classifier Systems*, pp. 1-15, 2010.
14. L.K. Hansen, P Salamon, Neural network ensembles, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12(10), pp. 993-1001, 1990.
15. L. Breiman, Bagging predictors, *Machine Learning*, vol. 24(2), pp. 123-140, 1996.
16. Y. Freund, R. E. Schapire, Experiments with a new boosting algorithm, *Proceedings of the 13th International Conference on Machine Learning*, pp. 148-156, 1996.
17. D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, *Journal of Artificial Intelligence Research*, vol. 11, pp. 169-198, 1999.
18. Z.-H. Zhou, J. Wu, W. Tang, Ensembling neural networks: Many could be better than all, *Artificial Intelligence*, vol. 137(1-2), pp. 239-263, 2002.
19. M. Aiolfi, A. Timmermann, Persistence in forecasting performance and conditional combination strategies, *Journal of Econometrics*, vol. 135(1), pp. 31-53, 2006.
20. J. Gama, P. Kosina, Tracking recurring concepts with meta-learners, *Portuguese Conference on Artificial Intelligence*, pp. 423-434, 2009.
21. D. Ilalan, A Poisson process with random intensity for modeling financial stability, *The Spanish Review of Financial Economics*, vol. 14, pp. 43-50, 2016.
22. J. Norris, Markov Chains (Cambridge Series in Statistical and Probabilistic Mathematics), Cambridge: Cambridge University Press, 1997.
23. E. Graham, C. Granger, A. Timmermann, *Handbook of Economic Forecasting*, vol. 1, 2006.

24. P. G. Zhang, L. V. Berardi. Time series forecasting with neural network ensembles: an application for exchange rate prediction, *Journal of the Operational Research Society*, vol. 52(6), pp. 652-664, 2001.
25. R. M. Stein, Benchmarking default prediction models: pitfalls and remedies in model validation, *Journal of Risk Model Validation*, vol. 1, pp. 77-113, 2007.