REMOVAL OF CONFUSING TRAINING SAMPLES AS A GENERIC MECHANISM TO IMPROVE AND DIVERSIFY TRADING STRATEGIES DISCOVERED BY BOOSTING-BASED OPTIMIZATION

OLGA BARINOVA

Moscow State University, Dept. of Computational Mathematics and Cybernetics, Graphics and Media Lab, Moscow, Russian Federation

VALERIY GAVRISHCHAKA

Head of Quantitative Research, Alexandra Investment Management, 767 Third Avenue New York, NY 10017, USA

Recently proposed boosting-based optimization offers a generic framework for the discovery of portfolios of complimentary base trading strategies with stable (non-resonant) performance over wide range of market regimes and robust generalization abilities. The framework can be based on many different types of boosting algorithms developed for classification problems. Stable portfolios with fixed capital allocations (weights) for base strategies can be discovered even with the classical AdaBoost. However, wide variety of market regimes and existence of very noisy periods with insufficient determinism reduces universe of financial instruments and achievable performance ranges for which such portfolio strategies can be found. One of the practically useful boosting generalizations is input-dependent boosting. This approach not only provides natural regularization but also introduces adaptive regime adjustment through weight time-dependence for additional strategy improvement and diversification. Here we introduce an alternative approach for regularization of boosting-based optimization that is based on confusing (noisy) sample removal in the training phase. Besides being able to improve and diversify discovered strategies on its own, this method can also contribute to the combined regularization when used with other generalized types of boosting. The generic algorithm for confusing sample removal is outlined and illustrated using real market example of mid-frequency intraday trading.

1. Introduction

Recently proposed boosting-based optimization offers a generic framework for the discovery of compact and interpretable portfolios of complimentary base trading strategies with stable performance over wide range of market regimes and robust generalization abilities [1,2]. Several types of stable portfolios with fixed capital allocations (weights) for base strategies can be discovered even

with the classical AdaBoost [3]. However, wide variety of market micro-regimes and existence of very noisy periods with insufficient determinism limits the universe of the financial instruments and desired performance characteristics for which such portfolio strategies with fixed weights can be found.

Many formal regularization methods for dealing with high-noise cases are often of limited value in practice. One of the practically useful boosting generalizations is input-dependent boosting which moderates emphasis on the difficult and/or noisy periods during boosting iterations [4]. This approach not only provides natural regularization but also introduces adaptive regime adjustment through weight time dependence [2]. However, even more important is the generic ability of the input-dependent boosting to find much greater variety of the portfolio strategies. This would allow building stable portfolios of portfolios using boosting frameworks or just simple equal-weight combination.

In this paper we outline an alternative approach for regularization of boosting-based optimization that is based on confusing (noisy) sample removal in the training phase. This method was originally proposed for the classification problems [5]. Besides being able to improve and diversify discovered strategies on its own, this method can also contribute to the combined regularization when used with other generalized types of boosting. The criteria for the training sample removal are based on the employed boosting framework itself without any additional assumptions. In the following the proposed algorithm is described and illustrated using real market example of mid-frequency intraday trading.

2. Algorithm for Confusing Sample Removal as Boosting Regularization

2.1. *Boosting-based optimization framework*

Portfolio strategy discovery is a direct optimization rather than classification problem. However, several formal arguments presented in [2] demonstrate that boosting for classification can be naturally generalized for the framework of boosting-based optimization originally proposed in [1]. Boosting framework can be extended to optimization tasks, where instead of minimizing classification error we are trying to minimize a number of intervals where return of a combined strategy is lower than a given threshold. Similarly, one

can minimize violations of an arbitrary complex condition. Therefore, even though the algorithm for confusing sample removal was originally proposed for classification problems [5], it can be readily accommodated for the boosting-based optimization.

2.2. *Problem of overfitting and possible solutions*

One of the key problems concerning practical use of machine learning is overfitting. Boosting was first believed to be immune to overfitting. It was reported to eventually lower its test error even after the training error reaches zero. Later it was found that boosting could also be sensitive to noise and overfits in some applications [6,7].

Many works are devoted to explaining how to avoid boosting overfitting. Several authors reported that boosting tends to increase the weights of few hard-to-learn samples. Several modifications of the re-weighting scheme were proposed that make weights change more smoothly [6]. The property of concentrating on few hard-to-learn patterns can be interpreted in terms of margin maximization and leads to regularized modifications [8]. Another source of boosting overfitting is unboundness of loss function used in standard boosting algorithms. Viewing boosting as a gradient descent search for a good fit in function space allows modifying loss functions [9]. Presence of label noise in data is widely considered as the reason of boosting overfitting [7]. Other authors see reasons of boosting overfitting in general concept of average loss minimization exploited by boosting algorithm which was shown to be inadequate for classification problems with overlapping classes distribution [5].

In this paper we propose a regularization which is based on removing so-called "confusing samples" [5] from the training set. The algorithm does not require any prior information about the data like noise ratio and noise distribution model.

2.3. *Confusing samples in boosting-based optimization framework*

In case of binary classification problems with overlapping class distributions both samples $(x,+1)$ and $(x,-1)$ can occur with positive probabilities. Then "confusing samples" are those samples that have posterior probability of their own label lower than of the opposite one. Forcing classifier to fit confusing samples was shown to cause overfitting.

We can formally define "confusing samples" for boosting-based optimization framework as follows:

Definition 1. "Confusing sample" is an interval x_i from training data, for which holds one of the two following conditions:

$$\left(P[r\left(S(x_i)\right) < r_c] > 0.5\right) \text{ but } \left(r\left(S(x_i)\right) > r_c\right) \quad (1)$$

$$\left(P[r\left(S(x_i)\right) > r_c] > 0.5\right) \text{ but } \left(r\left(S(x_i)\right) < r_c\right) \quad (2)$$

Here $r(S(x))$ stands for return of a combined strategy S and r_c denotes the required minimal value of the return on a chosen horizon. If necessary, more complex return/risk conditions can be used without any changes in the general formalism. However, we will refer to these particular conditions for simplicity.

Informally, "confusing samples" are those intervals, for which the observed value of return differs from the one we expect to see. These are hard or noisy examples. It was shown that in case of overlapping classes, training set always contains a number of confusing samples [5].

Removing all confusing samples from the training set reduces overlapping classes' task to a deterministic classification task with the same Bayesian separating surface. This statement provides a procedure for avoiding a possible boosting overfitting. We estimate the posterior probabilities of training set samples labels. Those samples that have average posterior probability estimate of their label lower than of its opposite are considered to be "confusing". After removing these samples boosting can be applied to the pruned training set.

2.4. *Confusing samples detection*

Our goal is to estimate the posterior probabilities of return being greater or lesser than a given threshold on training intervals and to exclude samples for which either (1) or (2) is held. In this work posterior probability estimates are extracted from boosting itself. Friedman et al. [10] provide a view on boosting as an additive logistic regression model. To obtain the probabilities, the logit transformation must be inverted as follows:

$$P(y \mid x) \approx \frac{1}{1+\exp\left(-2 \cdot F(x)\right)} \quad (3)$$

Other methods for converting boosting predictions to probabilities such as Platt's Scaling or Isotonic Regression [11] could be more accurate. However, they require significantly more data for proper estimation. Therefore, Friedman et al. method (3) is more practical for most applications and will be used here.

The overall algorithm consists of iterative process where the data is randomly divided into 2 parts: a boosted committee is trained on the first part, and posterior for samples in the second part are estimated. At each iteration we acquire an estimate of class probabilities for training instances and later

posterior estimates are averaged. Finally, "confusing" samples are determined according to (1) and (2), and removed.

3. Application Example

In some cases the regularization approach based on confusing sample removal may improve multiple performance measures simultaneously. However, more important is that it provides an efficient generic mechanism for the diversification of the strategies discovered by boosting-based optimization which allows meeting more complex multi-objective performance criteria.

For the illustration we apply boosting-based optimization to the single financial instrument and the pool of the typical momentum strategies operating on 5min intraday bars with no overnight positions. Each base strategy is either long-only or short-only and based on one of the two types of entry indicators. The entry indicators are double exponential (two exponential moving averages with different scales) and relative strength index (RSI) [12]. When price crosses up both EMAs, entry signal for long position is generated (cross-down signals short position). Similar rules are for RSI indicator. Exits are controlled by volatility-based trailing stops with Parkinson volatility measure [13].

In the presented example we considered 820 days of historical intraday data (from ESignal) for the mid-cap index (ticker: MID) ending on 2008/07/16, where 500 days ending on 2007/03/31 are used for training. The search is restricted to the strategies operating on the last 30 5min bars of each trading day with the same availability of the max capital at the start of the day (no reinvestment). Realistic fees of 0.5c per share and stop loss slippage of 5c for MDY (MID-based tradable ETF) were assumed and scaled to MID accordingly.

Both types of boosting optimization runs (with and without confusing sample removal) used the following parameters: τ=15days, r_c=0.15%, and T=7. Runs with more complex condition that additionally included limitation on max daily loss (< 0.5%) have also been considered. Confusing-sample removing procedure was based on the posterior probability estimation given by (3). In simpler terms, the confusing interval is that where condition ($r > r_c$) or the chosen complex condition is violated. The boosted portfolio was obtained by applying AdaBoost to the pool of base strategies on the training set except shifting 100 day window (i.e., 400 days for training). Posterior probabilities have been estimated for these 100 days. Window was shifted with a step of 50 days, i.e., every sample in the training data (except boundary region) received two estimations of the posterior probability. A sample was considered to be

confusing when at least one of the two estimations indicated that. Finally, the boosting algorithm was applied to the full training set with removed confusing samples.

In this case we found that the best results for the runs without regularization based on confusing sample removal are obtained for the objective function that includes additional condition on max daily loss (< 0.5%). When regularization is used, the original single condition ($r > r_c$) works better, since additional condition on max daily loss causes excessive removal of training data points. In the figure below we present Profit&Loss(%) time series of these best runs with and without regularization as line 1 and 2 respectively. Buy&hold benchmark is represented as line 0. Line 3 presents equal-weight combination of the two portfolios obtained with and without sample removal.

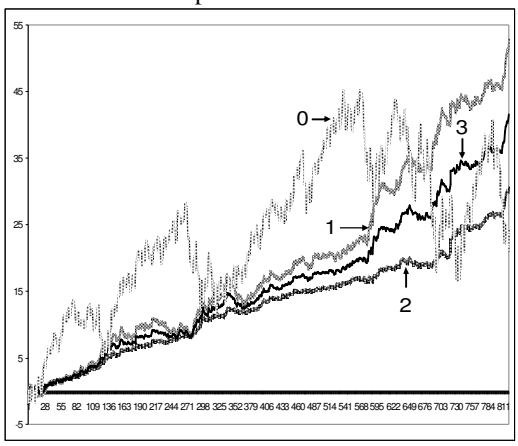

Figure 1. Profit&Loss(%) vs days for buy&hold (0) and boosted strategies (1, 2 – with and without regularization, 3 – equal weight combination of the two).

Both boosting strategies show stable out-of-sample performance and are clearly superior to the buy&hold strategy. Average annual return and Sharpe ratio of monthly returns is better for the portfolio obtained with regularization compared to the one obtained without it: 16.6% vs 9.3% and 2.18 vs 2.15 (3% risk-free rate is assumed). However, the max daily loss of the regularized portfolio (1.1%) is significantly higher than that obtained without regularization (0.5%). Thus, for the multi-objective goal (max return, max Sharpe ratio, and min daily loss), regularization in the form of confusing sample removal provides efficient mechanism for the additional diversity of the trading strategies. Simple equal weight combination of the two portfolios

allows obtaining very attractive portfolio with average annual return of 13.0%, Sharpe ratio of 2.4, and max daily loss of 0.7% which would be significantly more difficult (and time consuming) to obtain without combination of the two approaches.

4. Conclusions

The novel regularization method for the boosting-based optimization is outlined and illustrated using real market example. Regularization is based on the removal of confusing samples corresponding to noisy and/or difficult-to-model regimes from the training data. The introduced approach provides efficient mechanism for the improvement and diversification of the portfolio strategies discovered by boosting-based optimization.

References

1. V.V. Gavrishchaka, *New Mathematics and Natural Computation*, 2, 315 (2006).
2. V.V. Gavrishchaka, O.V. Barinova, A.P. Vezhnevets and M.A. Monina, Discovery of multi-component portfolio strategies with continuous tuning to the changing market micro-regimes using input-dependent boosting, in *Computational Finance and its Applications* III (2008).
3. R.E. Schapire, The Design and Analysis of Efficient Learning Algorithms, *PhD Thesis, MIT Press (1992).*
4. R. Jin, Y. Liu, L. Si, J. Carbonell and A.G. Hauptmann, A new boosting algorithm using input-dependent regularizer*, in Proc. of ICML-2003, Washington DC (2003).*
5. A. Vezhnevets and O. Barinova, Avoiding Boosting Overfitting by Removing Confusing Samples, in Proc. of ECML (2007).
6. C. Domingo and O. Watanabe: Madaboost: A modication of adaboost. In *13th Annual Conference on Comp. Learning Theory* (2000).
7. Ph. M. Long and R.A. Servedio, Random Classification Noise Defeats All Convex Potential Boosters, in *Proc. of ICML* (2008).
8. G. Ratsch: Robust Boosting and Convex Optimization. *Doctoral dissertation, University of Potsdam (2001).*
9. N. Krause and Y. Singer, Leveraging the Margin More Carefully, *ACM International Conference Proceeding Series*; Vol. 69 (2004).
10. J. Friedman, T. Hastie and R. Tibshirani, *Annals of Statistics*, 28, 2, 337-407 (2000).
11. A. Niculescu-Mizil and R. Caruana: Obtaining Calibrated Probabilities from Boosting, *Proc. 21st Conference on Uncertainty in Artificial Intelligence* (2005).
12. J.O. Katz and D.L. McCormick, *The Encyclopedia of Trading Strategies*, McGraw-Hill (2000).
13. M. Parkinson, *Journal of Business*, 53, 61-68 (1980).