Time Series Forecasting using Boosting Techniques With Correlation Coefficient

Luzia Vidal de Souza
Aurora T. R. Pozo
Anselmo Chaves Neto
Joel M. C. Rosa

06 de março de 2007

Motivation

- An essential element for many management decisions is an accurate forecasting.
- New Machine Learning techniques for forecasting
- Combination of Genetic Programming (GP) and Boosting Techniques to Obtain Accuracy in Forescasting Tasks

Boosting

- Boosting emerged as a way of combining many weak classifiers to produce a powerful ”committee”.
- It is a iterative process that gives more and more importance to bad classification
- Simple strategy that results in dramatic improvements in classification performance.
- There is a straight relation between Boosting and Theory of Optimization
- It was adapted to Regression Problems

Boosting Elements

- A training dataset $(x_1, y_1), \ldots (x_m, y_m)$
- A number of iterations $t = 1, \ldots, T$
- A base classifier $h_t : X \rightarrow \mathbf{R}$
- Weights $D_t(i)$ for the elements of the training dataset
- Weights α_t for the base classifiers

Adaboost Algorithm for Boosting (Binary Classification Problems)

Given: $(x_1, y_1), \ldots (x_m, y_m)$**,**$Y \in \{-1, 1\}$

Initialize $D_1(i) = 1/m$ For $t = 1 \ldots T$:

- Train base learner using distribution D_t
- Get base classifier $h_t : X \rightarrow$ R
- Choose $\alpha_t \in$ R
- Update:

$$D_{t+1}(i) = \frac{D_t(i)exp(-\alpha_t y_i h_t(x_i))}{Z_t} \quad (1)$$

Z_t is a normalization factor

Output the final classifier

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \quad (2)$$

Classifiers Weights α_t were left unspecified

- The base classifier h_t works to minimize:

$$\epsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i] \tag{3}$$

- A possible choice for α_t in binary classification is:

$$\alpha_t = \frac{1}{2} ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \tag{4}$$

Generalization Error

- The great interest in building boosting classifiers lies on getting good classification out-of-sample.

- Freund & Schapire (1997) showed that the generalization error is bounded by something like:

$$\widehat{Pr}[H(x) \neq y] + \tilde{O}\left(\sqrt{\frac{Td}{m}}\right) \tag{5}$$

Boosting for Regression Problems

Given: $(x_1, y_1), \ldots (x_m, y_m)$,$Y \in R$

Initialize $D_1(i) = 1/m$ For $t = 1 \ldots T$:

- Train base learner using distribution D_t
- Get base hypothesis h_t : $X \rightarrow$ R
- Evaluate Loss Function

$$L_i(t) = 1 - exp\left(\frac{|h_t(x_i) - y_i|}{max_{i=1,\ldots,m}|h_t(x_i) - y_i|}\right) \tag{6}$$

- Compute $\beta_t = \dfrac{\bar{L}_t}{1 - \bar{L}_t}$
- Update:

$$D_{t+1}(i) = \frac{D_t(i)^{1-L_i}}{Z_t} \tag{7}$$

  Z_t is a normalization factor

Output the final classifier (geometric median of $h_t()$)

$$F(x) = min\{y \in R : \sum_{h_t(x) \leq y} log\left(\frac{1}{\beta_t}\right) \geq \frac{1}{2}\sum_{t=1}^{T} log\left(\frac{1}{\beta_t}\right)\} \tag{8}$$

Boosting Coefficient Correlation for Regression Problems (Ad Hoc)

Given: $(x_1, y_1), \ldots (x_m, y_m)$,$Y \in R$

Initialize $D_1(i) = 1/m$ For $t = 1 \ldots T$:

- Train base learner using distribution D_t
- Get base hypothesis $h_t : X \to$ R
- Evaluate Loss Function

$$L_i(t) = 1 - exp\left(\frac{|h_t(x_i) - y_i|}{max_{i=1,\ldots,m}|h_t(x_i) - y_i|}\right) \quad (9)$$

- Update:

$$D_{t+1}(i) = \frac{\rho(h_t(x), y) D_t(i)^{1-L_i}}{Z_t} \quad (10)$$

  $\rho_t = \rho(h_t(x), y)$ is the Pearson Correlation Coefficient between $h_t(x)$ and y evaluated in the training sample. Z_t is a normalization factor

Output the final classifier

$$F(x) = \sum_{t=1}^{T} \frac{\rho_t h_t(x)}{\sum_{t=1}^{T} \rho_t} \quad (11)$$

GPBoost

- Genetic Programming can be used to build the rule of prediction $h_t(.)$, the pseudo-code for its algorithm is given bellow

  1. Randomly create an initial population

  2. Repeat until a good solution or a stop criterion is reached

     (a) Evaluate of each program by means of the fitness function (example:$fit = \sum_{i=1}^{m} |h(x_i) - y_i| * D_t(i)$)

     (b) Select a subgroup of individuals onto apply the genetic operators

     (c) Apply the genetic operators

     (d) Replace the current population by this new population

  3. End

- 3 time series with 1100 observations: Returns of Djiad, Ibovd and Nasdaq stock markets indexes.
- Separate 110 days to test the models performances.
- Implement a Trading Strategy to sell or buy according to the forecasts of returns. If return tomorrow is said to be positive, buy it today, otherwise sell it.

Table - Financial Returns in 110 days

Method	BCC	ARMA	GP	GPBoost
Djiad	2.0%	0.5 %	-1.0%	-3.5%
Ibovd	16.9%	-1.7%	-8.7 %	-6.9%
Nasdaq	7.3 %	-5.9%	-8.8%	-3.7%

Table - Financial annualized returns

Method	BCC	ARMA	GP	GPBoost
Djiad	4.6%	1.0 %	-2.3%	-8.0%
Ibovd	76.6%	-3.9%	-19.9 %	-15.7%
Nasdaq	16.7%	-13.5%	-20.2%	-8.5%