Collaborative Bagging of Boosting Ensembles

Nino Arsov[1], Martin Pavlovski[1], Lasko Basnarkov[1], and Ljupco Kocarev[1,2,3]

[1]Macedonian Academy of Sciences and Arts, Skopje, Republic of Macedonia

[2]Ss. Cyril and Methodius University, Skopje, Republic of Macedonia

[3]BioCircuits Institute, UC San Diego, La Jolla, CA 92093-0402, USA

Introduction

The main challenge of this research is to nurture the idea of ensemble-based classification by combining well-known ensemble generation methods bagging and boosting. Since the former reduces variance, while the latter ameliorates overfitting, the outcome of a multi-model that combines both strives towards a comprehensive net-balancing of the Bias-Variance trade-off (or dilemma). To further improve this, we alter the bagged-boosting scheme by introducing collaboration between the multi-model's constituent learners at various logical levels. The ultimate outcome represents collaborative bagging of boosting ensembles. The aforementioned collaboration is delivered in two flavors/variants: during or after the boosting process. These collaborative approaches to combining ensembles present a novel complex interactive classification scheme. Applied among a crowd of classification ensembles, they introduce a new bi-level interactive multi-model. In our approach, the underlying base ensembles can improve their quality by collaborating with each other, or through their constituent components (weak learning models). In addition, through voting, they can reach a joint compromise in order to generalize, i.e. infer on unobserved data. These complex ensemble principles characterize and encompass the basis of our proposed model.

Methods

Suppose a single boosting [1] ensemble consisting of ST weak learners is reorganized as a subbagged ensemble of S boosting ensembles, each containing T weak learners. Subbagging, which is a variant of Bootstrap Aggregation (Bagging) [2], aggregates models trained on randomly drawn sub-samples without replacement (thus producing smaller subsets that the training dataset itself), and paves the road to collaborative information exchange. We introduce two collaborative schemes between boosting ensembles: (1) during generation/training of boosters and (2) after boosters are trained.

There are two major benefits of collaboration. First, it refreshes the data fed to each individual booster and allows it to observe and learn instances that were previously unobserved. Second, it potentially aids the process of finding the optimal training subsets which is considered to be an NP-Hard combinatorial problem. Although, the collaboration itself seems promising and the benefits above are assumptive, it ought to be supported by computational learning theory. For this purpose, we link it to existing stability theory, most notably established by Bousquet and Elisseeff [3,4], henceforth resulting in two collaboration flavors: M-CLB and C-CLB. We have tested our approach for collaborative bagging of boosting ensembles with different boosters, but here we report our findings only for Gentle Boost. The purpose of this is the fact that Gentle Boost is intended to terminate the training process with a lower risk of overfitting to the training data and to reduce the noise susceptibility. Moreover, it has been observed that it outperforms other boosters [1], thus eliminating the need of comparison with them.

Margin-Based Collaboration (M-CLB)
Operates among weak learners in the Gentle Boost ensembles, during boosting.

Cumulative Collaboration (C-CLB)
Operates among prediction-ready (trained) Gentle Boost ensembles by repetitive training.

It is worth noting that both strive to improve the upper bounds of the overall generalization error. M-CLB and C-CLB provide a tighter, and eventually lower upper bound based on the exponential and classification loss functions respectively.

Theorem 1 (Generalization error upper bound for Subbagged Gentle Boost). Assume that the loss function ℓ is B-Lipschitzian, and $0 \leq \ell(\Phi_{\mathcal{X}}, z) \leq M$, for all $z \in \mathcal{Z}$, where $\Phi_{\mathcal{X}}$ is the output of a subbagging algorithm whose base machine is Gentle Boost. Moreover, assume that subbagging is done by sampling S sets of size $p < N$ from some $\mathcal{X} \in \mathcal{Z}^N$ uniformly and without replacement.

Methods

Let the weak learning algorithm A have (pointwise) hypothesis stability β_w with respect to ℓ and let $\epsilon_* = \text{Weak}_D(A)/2 > 0$. Then, for sufficiently large p, for all T, for Subbagged Gentle Boost in T rounds with probability at least $1-\delta$ over the random draw of $\mathcal{X} \sim D^N$,

$$R(\Phi_{\mathcal{X}}) \leq R_{emp}(\Phi_{\mathcal{X}}) + 2B\beta_p \frac{p}{N} + (4Bp\beta_p + M)\sqrt{\frac{\ln(1/\delta)}{2N}},$$

where β_p is the stability of the base machine (a Gentle Boost ensemble) and is evaluated as a function of the weak learner's stability β_w as in [5], i.e.

$$\beta_p = \frac{2}{N}\Sigma_{t=1}^{T} \frac{2^{t^2+1}(\frac{N\beta_w}{2}+4)^t}{\epsilon_*^{2t-1}}, \delta = e^{-N\epsilon_*^2/2}.$$

Theorem 2 (Classification-loss-oriented upper generalization error bound of Subbagged Gentle Boost). Let $\ell_1(\Phi_{\mathcal{X}}, z)$ be a 1-Lipschitzian classification loss function, for all $z \in \mathcal{Z}$, where $\Phi_{\mathcal{X}}$ is the outcome of a real-valued Subbagged Gentle Boost consisted of S base boosting ensembles. Then, for any $N \geq 1$, and any $\delta \in (0,1)$, with probability at least $1-\delta$ over the random draw of a training set $\mathcal{X}$,

$$R(\Phi_{\mathcal{X}}) \leq R^1_{emp}(\Phi_{\mathcal{X}}) + 4\eta\beta_p + (8\eta N\beta_p + 1)\sqrt{\frac{ln(1/\delta)}{2N}},$$

where $R^1_{emp}(\Phi_{\mathcal{X}})$ is the empirical error w.r.t. $\ell_1(\Phi_{\mathcal{X}}, z)$ and $\eta = p/N$.

Why CLB Approaches Work

Lower upper bound

Theorem 3 (Monotonic minimization of the exponential loss by Gentle Boost). Let t be the current round of boosting and let $F(\mathbf{x})$ be the outcome of a Gentle Boost algorithm from the previous $t-1$ rounds of training on a dataset $\mathcal{X} \in \mathcal{Z}^N$. Assume that $f_t(\mathbf{x})$ is the outcome of a real-valued weak learning algorithm, added to the ensemble, then with respect to the exponential loss,

$$\Sigma_{i=1}^{p} e^{-y_i(F(\mathbf{x}_i)+f_t(\mathbf{x}_i))} \leq \Sigma_{i=1}^{p} e^{-y_i F(\mathbf{x}_i)}.$$

Theorem 4 (M-CLB yields almost-everywhere lower empirical exponential loss of Gentle Boost). Let t be the current round of Gentle Boost with an outcome $F_{t,\mathcal{X}}(\mathbf{x}) = \Sigma_{s=1}^{t} f_{s,\mathcal{X}}(\mathbf{x})$ and assume that M-CLB is injected after training $f_{t,\mathcal{X}}(\mathbf{x})$, i.e. between rounds t and $t+1$, yielding $f^h_{t,\mathcal{X}'}$ and $F^h_{t,\mathcal{X}'}$, respectively. Then, with a high probability of ω, M-CLB yields a lower empirical Gentle Boost error $R^{M-CLB}_{emp}(F_{t+1,\mathcal{X}'})$ at round $t+1$ with respect to the exponential loss $\ell(F_{t+1,\mathcal{X}'}, z_i), z_i \in \mathcal{X}'$, or

$$R^{M-CLB}_{emp}(F^h_{t,\mathcal{X}'} + f_{t+1,\mathcal{X}'}, \mathcal{X}') \leq R_{emp}(F_{t,\mathcal{X}} + f_{t+1,\mathcal{X}}, \mathcal{X}),$$

$$\omega = \mathbb{P}\Big[\Sigma_{\mathbf{z}\in\mathcal{X}'}\, y f_{t+1,\mathcal{X}'}(\mathbf{x}) \geq \Sigma_{\mathbf{z}\in\mathcal{X}}\, y f_{t+1,\mathcal{X}}(\mathbf{x}) \vee$$
$$\Sigma_{\mathbf{z}\in\mathcal{X}'}\, y F^h_{t,\mathcal{X}'}(\mathbf{x}) \geq \Sigma_{\mathbf{z}\in\mathcal{X}}\, y (F_{t,\mathcal{X}}(\mathbf{x}) + f_{t+1,\mathcal{X}}(\mathbf{x}))\Big].$$

Theorem 5 (C-CLB provides a lower empirical error w.r.t. the classification loss of Subbagged Gentle Boost). Let $\Phi_{\mathcal{X}}$ be the outcome of a real-valued collaborative Subbagged Gentle Boost trained on $\mathcal{X}$. If C-CLB is used as a method for collaboration between its constituent boosting ensembles, then

$$R^1_{emp}(\Phi^{(\tau+1)}_{\mathcal{X}}) \leq R^1_{emp}(\Phi^{(\tau)}_{\mathcal{X}}),$$

where $\tau = 1, \ldots, \mathcal{T}-1$.

Tighter upper bound

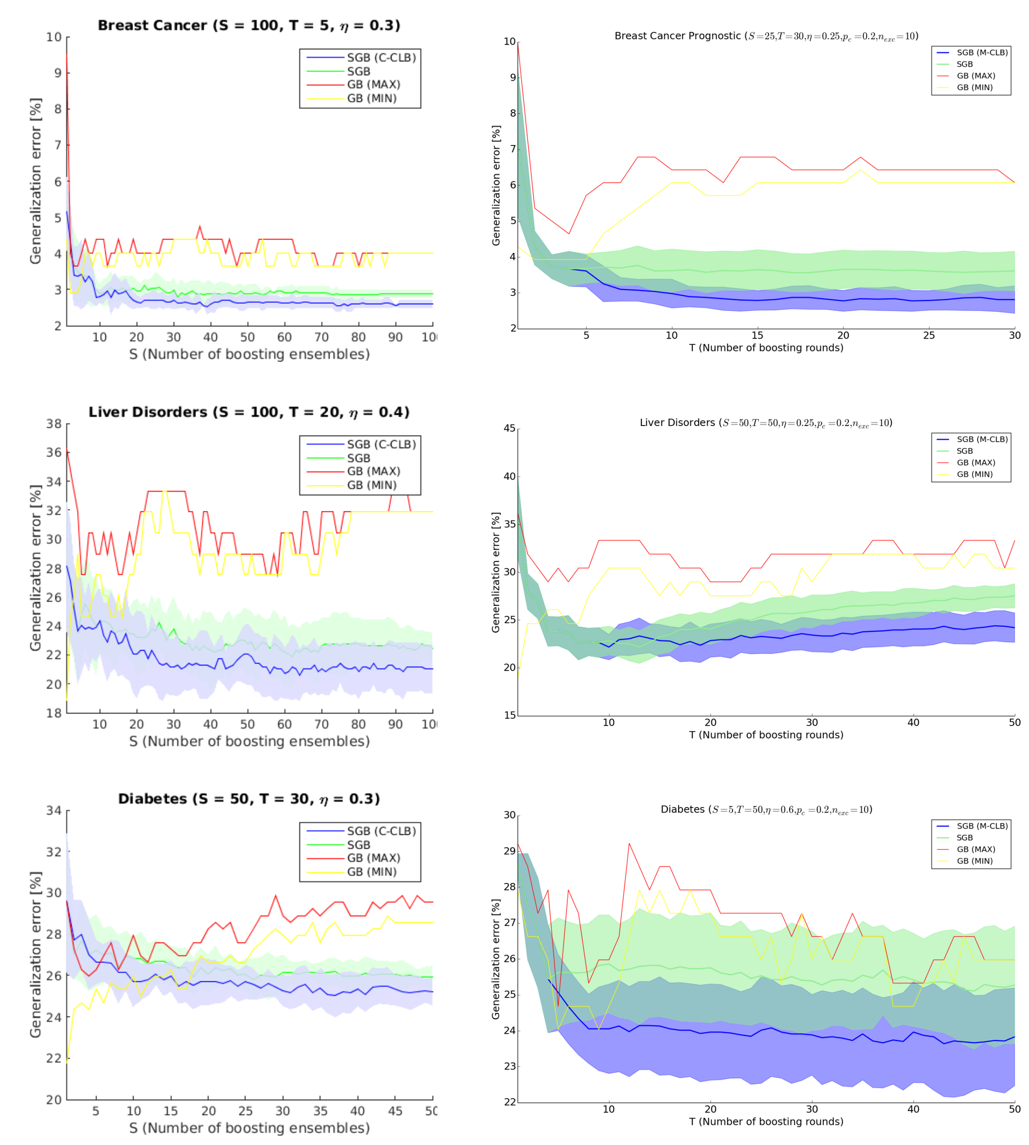
Proposition 1. Let f be the outcome of a real-valued classification algorithm, trained on a dataset $\mathcal{X}$ and let ℓ be the exponential or classification loss. Then for any two correctly classified training

Why CLB Approaches Work

instances $z_i, z_k \in \mathcal{X}$, such that $0 \leq y_i f_{\mathcal{X}}(\mathbf{x}_i) \leq y_k f_{\mathcal{X}}(\mathbf{x}_k)$,

$$|\ell(f_{\mathcal{X}}, z_i) - \ell(f_{\mathcal{X}^{\setminus i \cup z}}, z_i)| \geq |\ell(f_{\mathcal{X}}, z_k) - \ell(f_{\mathcal{X}^{\setminus k \cup z}}, z_k)|, \quad z \in \mathcal{Z}.$$

Results



The experiments shown above indicate that both collaborative approaches, M-CLB and C-CLB, manifest significant improvements of the overall generalization error, thus outperforming both Gentle Boost and Subbagged Gentle Boost. Note that, a collaborative Subbagged Gentle Boost consisted of S Gentle Boost ensembles is compared with ST weak learners organized into a single Gentle Boost ensemble. C-CLB displays lower variance because it gradually increases the number of ensembles used, hence the mean error is stabilized, while M-CLB uses a fixed number of ensembles which implies a higher variance, but has a negligible computational overhead.

Conclusion

In this research, we propose a method for improving the generalization error of bagged (or subbagged) algorithms, particularly for Subbagged Gentle Boost. First, we provide an upper-bound-based theoretical background for this specific class of algorithms; we construct upper bounds on the generalization error and use collaboration to improve them afterwards. Our two approaches manifest a visible improvement of the error. We believe that the same approach is going to have an improving effect on a broader class of bagged/subbagged learning algorithms.

References

1. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 'Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors).' The annals of statistics 28.2 (2000): 337-407.
2. Breiman, Leo. "Bagging predictors." Machine learning 24.2 (1996): 123-140.
3. Bousquet, Olivier, and André Elisseeff. 'Stability and generalization.' The Journal of Machine Learning Research 2 (2002): 499-526.
4. Elisseeff, Andre, Theodoros Evgeniou, and Massimiliano Pontil. 'Stability of randomized learning algorithms.' Journal of Machine Learning Research. 2005.
5. Kutin, Samuel, and Partha Niyogi. The interaction of stability and weakness in AdaBoost. Technical Report TR-2001-30, Computer Science Department, University of Chicago, 2001.