Boosting (machine learning)

From Wikipedia, the free encyclopedia
(Redirected from Boosting (meta-algorithm))

Boosting is a machine learning ensemble meta-algorithm for reducing bias primarily and also variance[1] in supervised learning, and a family of machine learning algorithms which convert weak learners to strong ones.[2] Boosting is based on the question posed by Kearns and Valiant (1988, 1989):[3][4] Can a set of **weak learners** create a single **strong learner**? A weak learner is defined to be a classifier which is only slightly correlated with the true classification (it can label examples better than random guessing). In contrast, a strong learner is a classifier that is arbitrarily well-correlated with the true classification.

Robert Schapire's affirmative answer in a 1990 paper[5] to the question of Kearns and Valiant has had significant ramifications in machine learning and statistics, most notably leading to the development of boosting.[6]

When first introduced, the *hypothesis boosting problem* simply referred to the process of turning a weak learner into a strong learner. "Informally, [the hypothesis boosting] problem asks whether an efficient learning algorithm […] that outputs a hypothesis whose performance is only slightly better than random guessing [i.e. a weak learner] implies the existence of an efficient algorithm that outputs a hypothesis of arbitrary accuracy [i.e. a strong learner]."[3] Algorithms that achieve hypothesis boosting quickly became simply known as "boosting". Freund and Schapire's arcing (Adapt[at]ive Resampling and Combining),[7] as a general technique, is more or less synonymous with boosting.[8]

Contents [hide]

Boosting algorithms [edit]

While boosting is not algorithmically constrained, most boosting algorithms consist of iteratively learning weak classifiers with respect to a distribution and adding them to a final strong classifier. When they are added, they are typically weighted in some way that is usually related to the weak learners' accuracy. After a weak learner is added, the data is reweighted: examples that are misclassified gain weight and examples that are classified correctly lose weight (some boosting algorithms actually decrease the weight of repeatedly misclassified examples, e.g., boost by majority and BrownBoost). Thus, future weak learners focus more on the examples that previous weak learners misclassified.

There are many boosting algorithms. The original ones, proposed by Robert Schapire (a recursive majority gate formulation[5]) and Yoav Freund (boost by majority[9]), were not adaptive and could not take full advantage of the weak learners. However, Schapire and Freund then developed AdaBoost, an adaptive boosting algorithm that won the prestigious Gödel Prize. Only algorithms that are provable boosting algorithms in the probably approximately correct learning formulation are called boosting algorithms. Other algorithms that are similar in spirit to boosting algorithms are sometimes called "leveraging algorithms", although they are also sometimes incorrectly called boosting algorithms.[9]

Examples of boosting algorithms [edit]

This section requires expansion.
(December 2009)

The main variation between many boosting algorithms is their method of weighting training data points and hypotheses. AdaBoost is very popular and perhaps the most significant historically as it was the first algorithm that could adapt to the weak learners. However, there are many more recent algorithms such as LPBoost, TotalBoost, BrownBoost, MadaBoost, LogitBoost, and others. Many boosting algorithms fit into the AnyBoost framework,[9] which shows that boosting performs gradient descent in function space using a convex cost function.

Boosting algorithms are used in Computer Vision, where individual classifiers detecting contrast changes can be combined to identify Facial Features.[10]

Criticism [edit]

In 2008 Phillip Long (at Google) and Rocco A. Servedio (Columbia University) published a paper[11] at the 25th International Conference for Machine Learning suggesting that many of these algorithms are probably flawed. They conclude that "convex potential boosters cannot withstand random classification noise," thus making the applicability of such algorithms for real world, noisy data sets questionable. The paper shows that if any non-zero fraction of the training data is mis-labeled, the boosting algorithm tries extremely hard to correctly classify these training examples, and fails to produce a model with accuracy better than 1/2. This result does not apply to branching program based boosters but does apply to AdaBoost, LogitBoost, and others.[12][11]

See also [edit]

- AdaBoost
- Random forest
- Alternating decision tree
- Bootstrap aggregating (bagging)
- Cascading
- BrownBoost
- CoBoosting
- GentleBoost
- LPBoost
- logistic regression
- maximum entropy methods
- neural networks
- support vector machines
- Gradient boosting
- RankBoost
- margin classifiers
- cross-validation
- machine learning
- Boosting methods for object categorization

Implementations [edit]

- Orange, a free data mining software suite, module Orange.ensemble
- Weka is a machine learning set of tools that offers variate implementations of boosting algorithms like AdaBoost and LogitBoost
- R package GBM (Generalized Boosted Regression Models) implements extensions to Freund and Schapire's AdaBoost algorithm and Friedman's gradient boosting machine.
- jboost; AdaBoost, LogitBoost, RobustBoost, Boostexter and alternating decision trees

References [edit]

Footnotes [edit]

1. ^ Leo Breiman (1996). "BIAS, VARIANCE, AND ARCING CLASSIFIERS" (PDF). TECHNICAL REPORT. Retrieved 19 January 2015. "Arcing [Boosting] is more successful than bagging in variance reduction"
2. ^ Zhou Zhi-Hua (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall/CRC. p. 23. ISBN 978-1439830031. "The term boosting refers to a family of algorithms that are able to convert weak learners to strong learners"
3. ^ *a* *b* Michael Kearns(1988); *Thoughts on Hypothesis Boosting*, Unpublished manuscript (Machine Learning class project, December 1988)
4. ^ Michael Kearns; Leslie Valiant (1989). "Crytographic limitations on learning Boolean formulae and finite automata". *Symposium on Theory of computing* (ACM) **21**: 433–444. doi:10.1145/73007.73049. Retrieved 18 January 2015.
5. ^ *a* *b* Schapire, Robert E. (1990). "The Strength of Weak Learnability" (PDF). *Machine Learning* (Boston, MA: Kluwer Academic Publishers) **5** (2): 197–227. doi:10.1007/bf00116037. CiteSeerX: 10.1.1.20.723.
6. ^ Leo Breiman (1998). "Arcing classifier (with discussion and a rejoinder by the author)". *Ann. Statist.* **26** (3): 801–849. Retrieved 18 January 2015. "Schapire (1990) proved that boosting is possible. (Page 823)"
7. ^ Yoav Freund and Robert E. Schapire (1997); *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences, 55(1):119-139
8. ^ Leo Breiman (1998); *Arcing Classifier (with Discussion and a Rejoinder by the Author)*, Annals of Statistics, vol. 26, no. 3, pp. 801-849: "The concept of weak learning was introduced by Kearns and Valiant (1988, 1989), who left open the question of whether weak and strong learnability are equivalent. The question was termed the *boosting problem* since [a solution must] boost the low accuracy of a weak learner to the high accuracy of a strong learner. Schapire (1990) proved that boosting is possible. A *boosting algorithm* is a method that takes a

weak learner and converts it into a strong learner. Freund and Schapire (1997) proved that an algorithm similar to arc-fs is boosting.

9. ^ *a* *b* *c* Llew Mason, Jonathan Baxter, Peter Bartlett, and Marcus Frean (2000); *Boosting Algorithms as Gradient Descent*, in S. A. Solla, T. K. Leen, and K.-R. Muller, editors, *Advances in Neural Information Processing Systems* 12, pp. 512-518, MIT Press
10. ^ OpenCV CascadeClassifier http://docs.opencv.org/modules/objdetect/doc/cascade_classification.html
11. ^ *a* *b* Random Classification Noise Defeats All Convex Potential Boosters
12. ^ Long version published as Phillip M. Long and Rocco A. Servedio (2010); *Random Classification Noise Defeats All Convex Potential Boosters*, Machine Learning 78(3), pp. 287-304

Notations [edit]

- Yoav Freund and Robert E. Schapire (1997); *A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting*, Journal of Computer and System Sciences, 55(1):119-139
- Robert E. Schapire and Yoram Singer (1999); *Improved Boosting Algorithms Using Confidence-Rated Predictors*, Machine Learning, 37(3):297-336

External links [edit]

- Robert E. Schapire (2003); *The Boosting Approach to Machine Learning: An Overview*, MSRI (Mathematical Sciences Research Institute) Workshop on Nonlinear Estimation and Classification
- Zhou Zhi-Hua (2014) *Boosting 25 years*, CCL 2014 Keynote.

Categories: Classification algorithms | Ensemble learning