

AdaBoost

维基百科，自由的百科全书

AdaBoost為英文"Adaptive Boosting"（自适应增强）的缩写，是一种机器学习方法，由約阿夫·弗羅因德和羅伯特·沙皮爾提出。^[1]AdaBoost方法的自适应在于：前一个分类器分错的样本会被用来训练下一个分类器。AdaBoost方法对于噪声数据和异常数据很敏感。但在一些问题中，AdaBoost方法相对于大多数其它学习算法而言，不会很容易出现过拟合现象。AdaBoost方法中使用的分类器可能很弱（比如出现很大错误率），但只要它的分类效果比随机好一点（比如两类问题分类错误率略小于0.5），就能够改善最终得到的模型。而错误率高于随机分类器的弱分类器也是有用的，因为在最终得到的多个分类器的线性组合中，可以给它们赋予负系数，同样也能提升分类效果。

AdaBoost方法是一种迭代算法，在每一轮中加入一个新的弱分类器，直到达到某个预定的足够小的错误率。每一个训练样本都被赋予一个权重，表明它被某个分类器选入训练集的概率。如果某个样本点已经被准确地分类，那么在构造下一个训练集中，它被选中的概率就被降低；相反，如果某个样本点没有被准确地分类，那么它的权重就得到提高。通过这样的方式，AdaBoost方法能“聚焦于”那些较难分（更富信息）的样本上。在具体实现上，最初令每个样本的权重都相等，对于第k次迭代操作，我们就根据这些权重来选取样本点，进而训练分类器C_k。然后就根据这个分类器，来提高被它分错的样本的权重，并降低被正确分类的样本权重。然后，权重更新过的样本集被用于训练下一个分类器C_k^[2]。整个训练过程如此迭代地进行下去。

AdaBoost算法

用xⁱ和y_i表示原始样本集D的样本点和它们的类标。用W_k(i)表示第k次迭代时全体样本的权重分布。这样就有如下所示的AdaBoost算法：

- 初始化：输入参数为训练集D={x¹, y₁, ..., xⁿ, y_n}，最大循环次数k_{max}，采样权重W_k(i)=1/n, i=1, ..., n；
- 迭代计数器k赋值为0；
- 计数器k自增1；
- 使用W_k(i)采样权重对弱学习器C_k进行训练；
- 对弱学习器C_k的训练结果进行评估并记录进误差矩阵E_k中；
- α

k

←

1
2

ln⁡

1−

E

k

E

k

{\displaystyle \alpha _{k}\leftarrow {\frac {1-E_{k}}{2}}\ln {\frac {1-E_{k}}{E_{k}}}}
- W

k
+
1

(
i
)
←

W

k

(
i
)

Z

k

×
⎧

e

−

α

k

,

if

h

k

(

x

i

)
=

y

i

e

α

k

,

if

h

k

(

x

i

)
≠

y

i

⎫

{\displaystyle W_{k+1}(i)\leftarrow {\frac {W_{k}(i)}{Z_{k}}}\times \left\{{\begin{matrix}e^{-\alpha _{k}},&{\text{ if }}h_{k}(x^{i})=y_{i}\\e^{\alpha _{k}},&{\text{ if }}h_{k}(x^{i})\neq y_{i}\end{matrix}}\right.}
- 当k=k_{max}时停止训练
- 返回结果 C_k和α_k, k=1, ..., k_{max}（带权值分类器的总体）
- 结束**

注意第5行中，当前权重分布必须考虑到分类器 C_k 的误差率。在第7行中， z_k 只是一个归一化系数，使得 $W_k(i)$ 能够代表一个真正的分布，而 $h_k(x^i)$ 是分量分类器 C_k 给出的对任一样本点 x^i 的标记（+1或-1）， $h_k(x^i) = y_i$ 时，样本被正确分类。第8行中的迭代停止条件可以被换为判断当前误差率是否小于一个阈值。

最后的总体分类的判决可以使用各个分量分类器加权平均来得到：

$$g(x) = [\sum_{k=1}^{k_{max}} \alpha_k h_k(x)]$$

这样，最后对分类结果的判定规则是：


$$H(x) = \text{sign}(g(x))$$

软件实现

- AdaBoost and the Super Bowl of Classifiers - A Tutorial on AdaBoost. (<http://www.inf.fu-berlin.de/inst/ag-ki/adaboost4.pdf>) (页面存档备份 (<https://web.archive.org/web/20131101020946/http://www.inf.fu-berlin.de/inst/ag-ki/adaboost4.pdf>), 存于互联网档案馆)
- Adaboost in C++ (<http://codingplayground.blogspot.com/2009/03/adaboost-improve-your-performance.html>) (页面存档备份 (<https://web.archive.org/web/20111006163224/http://codingplayground.blogspot.com/2009/03/adaboost-improve-your-performance.html>), 存于互联网档案馆), an implementation of Adaboost in C++ and boost by Antonio Gulli
- icsiboost (<https://code.google.com/p/icsiboost/>) (页面存档备份 (<https://web.archive.org/web/20130601234409/http://code.google.com/p/icsiboost/>), 存于互联网档案馆), an open source implementation of Boostexter
- JBoost (<http://jboost.sourceforge.net>) (页面存档备份 (<https://web.archive.org/web/20180603131348/http://jboost.sourceforge.net>), 存于互联网档案馆), a site offering a classification and visualization package, implementing AdaBoost among other boosting algorithms.
- MATLAB AdaBoost toolbox. Includes Real AdaBoost, Gentle AdaBoost and Modest AdaBoost implementations. (<https://web.archive.org/web/20110817114237/http://graphics.cs.msu.ru/en/science/research/machinelearning/adaboosttoolbox>)
- A Matlab Implementation of AdaBoost (<http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=21317&objectType=file>) (页面存档备份 (<https://web.archive.org/web/20190919023724/http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=21317&objectType=file>), 存于互联网档案馆)
- Multi-threaded MATLAB-compatible implementation of Boosted Trees (<https://sites.google.com/site/carlosbecker/resources/gradient-boosting-boosted-trees>) (页面存档备份 (<https://web.archive.org/web/20140522125421/https://sites.google.com/site/carlosbecker/resources/gradient-boosting-boosted-trees>), 存于互联网档案馆)
- milk (<http://luispedro.org/software/milk>) (页面存档备份 (<https://web.archive.org/web/20130917063455/http://luispedro.org/software/milk>), 存于互联网档案馆) for Python implements AdaBoost (<https://web.archive.org/web/20120711210335/http://packages.python.org/milk/adaboost.html>).

- [MPBoost++ \(http://www.esuli.it/mpboost\)](http://www.esuli.it/mpboost) (页面存档备份 (<https://web.archive.org/web/20110604124807/http://www.esuli.it/mpboost>), 存于互联网档案馆), a C++ implementation of the original AdaBoost.MH algorithm and of an improved variant, the MPBoost algorithm.
- [multiboost \(https://web.archive.org/web/20150419050429/http://www.multiboost.org/\)](https://web.archive.org/web/20150419050429/http://www.multiboost.org/), a fast C++ implementation of multi-class/multi-label/multi-task boosting algorithms. It is based on AdaBoost.MH but also implements popular cascade classifiers and FilterBoost along with a batch of common multi-class base learners (stumps, trees, products, Haar filters)。
- [NPatternRecognizer \(http://npatternrecognizer.codeplex.com/\)](http://npatternrecognizer.codeplex.com/) (页面存档备份 (<https://web.archive.org/web/20130820072633/http://npatternrecognizer.codeplex.com/>), 存于互联网档案馆), a fast machine learning algorithm library written in C#. It contains support vector machine, neural networks, bayes, boost, k-nearest neighbor, decision tree, ..., etc.
- OpenCV implementation of several boosting variants (<https://web.archive.org/web/20120924165410/http://opencv.willowgarage.com/documentation/cpp/boosting.html>)
- Into (<https://web.archive.org/web/20100709025652/http://intopii.com/into/>) contains open source implementations of many AdaBoost and FloatBoost variants in C++.
- Mallet (<http://mallet.cs.umass.edu/>) (页面存档备份 (<https://web.archive.org/web/20120426044516/http://mallet.cs.umass.edu/>), 存于互联网档案馆) Java implementation.
- [adabag \(https://web.archive.org/web/20150505023754/http://cran.r-project.org/web/packages/adabag/\)](https://web.archive.org/web/20150505023754/http://cran.r-project.org/web/packages/adabag/) adabag: An R package for binary and multiclass Boosting and Bagging.
- Scikit-learn (<https://web.archive.org/web/20150426104718/http://scikit-learn.org/dev/modules/ensemble.html#adaboost>) Python implementation.

参考书目

1. Freund, Yoav; Schapire, Robert E. A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting. 1995. [CiteSeerX: 10.1.1.56.9855](#).
 2. O. Duda, Peter E. Hart, David G. Stork, *Pattern Classification*, 2nd Edition, Wiley, 2000, [ISBN 978-0-471-05669-0](#)
-

取自“<https://zh.wikipedia.org/w/index.php?title=AdaBoost&oldid=78676111>”

■