Fall 2013 CS 412 Assignment 4 Report

Author: Weijie Liu, NetId: wliu43

Section 1. Introduction

In this assignment, I use Naive Bayes as basic method, and use AdaBoost as ensemble method.

Section 2. Classification Result and Evaluation Measures

Table 2.1 The classification results of adult and breast_cancer
(NB means Naïve Bayes; Boost means AdaBoost)

Dataset	adult				breast_cancer			
	training		test		training		test	
	NB	Boost	NB	Boost	NB	Boost	NB	Boost
TP	313	307	5905	5767	29	29	17	13
FN	82	88	1541	1679	27	27	12	16
FP	282	310	5060	5745	28	32	20	13
TN	928	900	18450	17765	96	92	57	64
Accuracy	0.773	0.752	0.787	0.760	0.694	0.672	0.698	0.726
Error Rate	0.227	0.248	0.231	0.240	0.306	0.328	0.302	0.274
Sensitivity	0.792	0.777	0.793	0.775	0.518	0.518	0.586	0.448
Specificity	0.767	0.743	0.785	0.756	0.774	0.742	0.740	0.831
Precision	0.526	0.498	0.539	0.501	0.509	0.475	0.459	0.5
F-1 Score	0.632	0.607	0.641	0.608	0.513	0.496	0.515	0.473
F-.5 Score	0.564	0.536	0.575	0.539	0.511	0.483	0.480	0.489
F-2 Score	0.720	0.699	0.725	0.698	0.516	0.509	0.556	0.458

Table 2.2 The classification results of led and poker
(NB means Naïve Bayes; Boost means AdaBoost)

Dataset	led				poker			
	training		test		training		test	
	NB	Boost	NB	Boost	NB	Boost	NB	Boost
TP	324	399	162	207	740	678	448	398
FN	314	239	189	144	7	69	11	61
FP	65	92	27	41	284	242	217	196
TN	1384	1357	756	742	10	52	2	23
Accuracy	0.818	0.841	0.810	0.837	0.720	0.701	0.664	0.621
Error Rate	0.182	0.159	0.190	0.163	0.280	0.299	0.336	0.379
Sensitivity	0.508	0.625	0.462	0.59	0.991	0.908	0.976	0.867
Specificity	0.955	0.937	0.966	0.948	0.034	0.177	0.009	0.105
Precision	0.833	0.813	0.857	0.835	0.723	0.737	0.674	0.670
F-1 Score	0.631	0.707	0.600	0.691	0.836	0.813	0.797	0.756
F-.5 Score	0.738	0.767	0.732	0.771	0.764	0.766	0.718	0.702
F-2 Score	0.551	0.656	0.508	0.627	0.922	0.867	0.896	0.819

Section 3. Does your framework perform equally good on training and test datasets? Why or why not?

Answer:

Overall speaking, the results in the above table tell that my framework perform equally good on training and test datasets because the metrics of the classification results of the training and testing datasets are similar.

However, we have to notice that the results of the training dataset of Poker are slightly better than those of the test datasets.

Why? In my opinion, the reason is that the training and the testing datasets of some of the datasets (adult, breast_cancer and led) are from the same data source, which means that the training datasets are of the same structure of the test datasets. However, in terms of Poker, the training and test datasets may not come from the same source, which leads to the different structure of training and test datasets. As a result, the classifier from the training dataset may give a different result if the input is test dataset.

Section 4. Parameters you chose during implementation and why you chose these parameters.

Answer:

No parameters are chosen in Naïve Bayes.

There are three parameters in AdaBoost.

(1) The first parameter is T, the number of rounds of Iteration and the second parameter is R, the sample rate in one round. In my implementation, T=10 and R=0.5

First, we have notice that AdaBoost trains several weak classifiers and let them vote for the final classification result according to the weights of each classifier. T should be reasonably large enough for the framework to get enough weak classifiers to get the final classification result. From previous academic works and my own experiments, I can tell that T=10 is good enough. Of course, other parameter may be good.

Second, in each round, it uses samples of the total tuples to generate the classification in that round. The goal for me to choose the sample rate R is to let each of the tuple be chosen at least once in the AdaBoost. Since the weight of a tuple chosen will increase and decrease in the process of AdaBoost, now, for simplification, I just assume that the weight of a tuple remains the same. If the total number of training tuples is N, we can have P_{never}, the probability of a certain tuple never chosen in T rounds is $P_{never} = P^T = (\frac{N-1}{N})^{RNT}$, if N=100, R=0.5, T=10, we have 0.0066. Then the expectation of the number of tuples never chosen is 100*0.0066=0.66<1. Thus, R=0.5 is a good choice of the sample rate.

(2) The third parameter in AdaBoost is E, the error rate threshold for repeating generating a weak classifier. I choose E=0.5.

The reason is: I use this following equation to modify the weight of each tuples:

$$D_{i+1}(i) = \frac{D_i(i)}{Z} \times \begin{cases} \frac{Error}{1-Error}, & if\ classification\ corrects \\ 1, & otherwise \end{cases}$$

We can see that, if Error>E=0.5, $\frac{Error}{1-Error} = \frac{1}{1-Error} - 1 > 1$, this contradicts the principle that we should decrease the weight of the tuple which is classified correctly in this round. Thus, E=0.5.

Section 5. Conclusion on whether the ensemble method improves the performance of the basic classification method you chose, why or why not.

Answer:

The ensemble method, AdaBoost, doesn't improve the performance of the Naïve Bayes classification much. I think the main reason for this is the randomness significantly influences the performance of the AdaBoost. Sometimes the results of AdaBoost are better than Naïve Bayes, while sometimes AdaBoost fails to improve the performance. A more important reason is that the Naïve Bayes is not ensemble compatible. The detailed analysis can be found in Section 6.

Section 6. Verify your guess on whether your basic classification method is ensemble compatible.

Answer:

I guess the Naïve Bayes is not ensemble compatible.

The effectiveness of boosting on Naïve Bayes is not a new topic. From the previous analysis in the academic works [1], we can see that

Naïve Bayes is a stable classifier with a strong bias. More detailed analysis points out that the final boosted model share the same form of the base learner, thus imposing a strong uncorrectable bias which prevents the bias reduction mechanism in boosting from taking effect.

Reference:

[1] Kai Ming Ting, Zijian Zheng. A study of AdaBoost with Naïve Bayesian Classifiers. Computational Intelligence, Volume 19, Number 2, 2003