Lazy Bayesian Rules: A Lazy Semi-Naïve Bayesian Learning Technique Competitive to Boosting Decision Trees

Zijian Zheng, Geoffrey I. Webb, Kai Ming Ting
Deakin University
Victoria Australia
Appeared in ICML ‘99

Paper Overview

- Description of LBR, Adaboost and Bagging
- Experimental Comparison of algorithms

Naïve Bayesian Tree

- Each tree node is a naïve bayes classifier

Lazy Bayesian Rules

- Build a special purpose bayesian classifier based on the example to classify
- greedily choose which attributes to remain constant and which should vary

Table 1: The Lazy Bayesian Rule learning algorithm

LBR(Att, D, E_{test})
  INPUT: *Att*: a set of attributes,
    D: a set of training cases described using *Att* and classes,
    E_{test}: a test case described using *Att*.
  OUTPUT: a predicted class for E_{test}.
LocalNB = a NB classifier trained using *Att* on D
Errors = errors of *LocalNB* estimated using N-CV on D
Cond = *true*
REPEAT
  $TempErrors_{best}$ = the number of cases in D + 1
  FOR each attribute A in *Att* whose value v_A on E_{test} is not missing *DO*
    D_{subset} = cases in D with $A = v_A$
    TempNB = a NB classifier trained using $Att - \{A\}$ on D_{subset}
    TempErrors = errors of *TempNB* estimated using N-CV on D_{subset} + the portion of *Errors* in $D - D_{subset}$
    IF ((*TempErrors* < $TempErrors_{best}$) *AND* (*TempErrors* is significantly lower than *Errors*))
    THEN
      $TempNB_{best}$ = *TempNB*
      $TempErrors_{best}$ = *TempErrors*
      A_{best} = A
  IF (an A_{best} is found)
  THEN
    Cond = *Cond* $\wedge$ ($A_{best} = v_{Abest}$)
    LocalNB = $TempNB_{best}$
    $D = D_{subset}$ corresponding to A_{best}
    Att = *Att* $- \{A_{best}\}$
    Errors = errors of *LocalNB* estimated using N-CV on D
UNTIL (no A_{best} is found)
classify E_{test} using *LocalNB*
RETURN the class

Boosting / Bagging

- Adaboost
  - train on examples
  - evaluate performance
  - re-train new classifier with weighted examples
  - repeat
  - when classifying, vote according to weights
- Bagging
  - train many times on samples drawn with replacement
  - when classifying, vote equally

Table 2: Error rates (%) of LBR, NB, C4.5, AdaBoost, and bagging

Domain	LBR	NB	C4.5	AdaBoost	bagging
Annealing	2.8	3.0	10.4	7.8	8.9
Audiology	28.0	28.4	23.7	19.4	21.5
Breast(W)	2.6	2.6	5.1	3.8	4.4
KR-KP	2.7	12.5	0.8	0.5	0.7
Credit(A)	14.4	14.5	14.5	15.9	14.3
Echocar.	30.8	30.8	36.2	33.9	35.1
Glass	34.1	34.0	35.8	33.1	32.8
Heart(C)	17.6	17.6	24.0	22.3	20.7
Hepatitis	14.7	14.7	21.0	17.7	18.9
Horse	20.3	21.0	17.0	21.3	16.8
HouseVote	6.7	9.9	5.1	4.9	4.6
Hypo	1.5	1.8	1.3	1.2	1.2
Iris	6.6	6.6	7.5	7.5	7.1
Labor	10.7	10.7	19.5	18.2	18.6
LED 24	39.7	39.5	36.9	36.4	31.8
Liver	36.2	36.2	35.7	36.0	36.0
LungC.	49.7	49.4	59.4	55.0	61.3
Lympho	18.5	18.3	23.6	17.7	20.7
Pima	25.1	25.2	25.5	27.0	25.0
Postop.	35.9	36.0	29.9	39.7	31.1
P.Tumor	52.4	52.4	61.0	58.4	57.8
Promoters	10.7	10.7	26.3	8.4	17.2
Solar	17.1	19.1	15.7	17.0	15.9
Sonar	25.4	26.5	32.1	22.1	24.9
Soybean	7.0	9.8	10.3	6.9	8.2
Splice	4.3	4.5	6.5	4.7	5.5
TicTacToe	19.1	29.5	16.5	1.3	9.3
Wine	2.5	2.5	11.8	4.7	8.9
Zoology	6.2	6.2	8.5	6.2	8.4
mean	18.7	19.8	21.4	18.9	19.6
e.r. mean		1.12	1.21	.93	1.07
w/t/l		4/11/14	10/0/19	13/1/15	14/0/15
p. of wtl		.0304	.1360	.8506	1.0000

e.r. mean: geometric mean of error ratios.

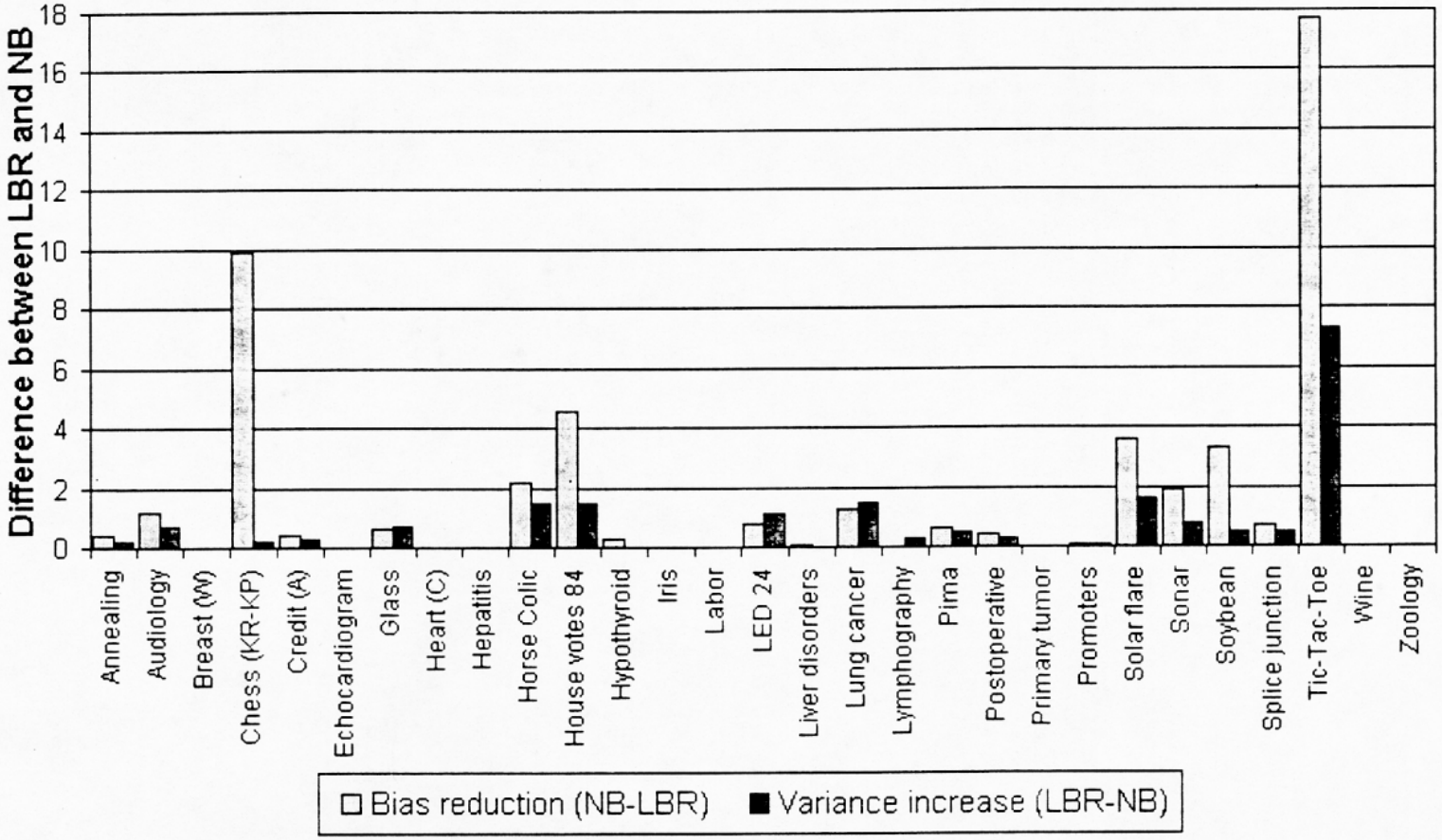

Figure 1: Bias and variance comparison between LBR and NB

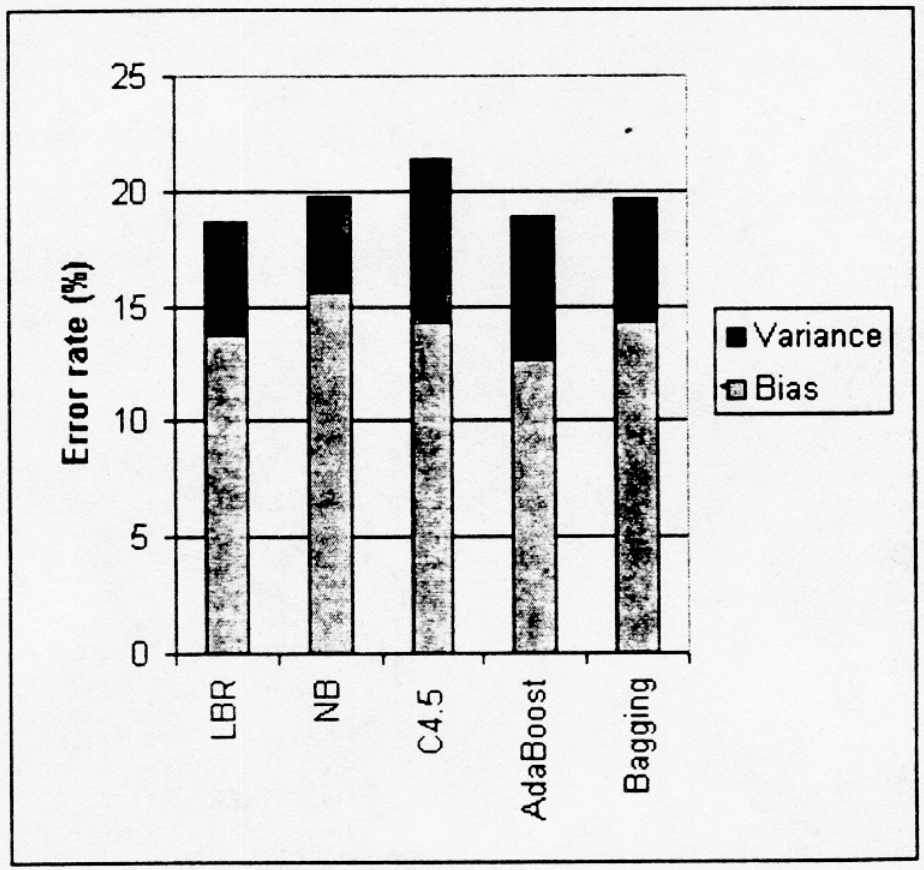

Figure 2: Error rate with bias and variance decomposition

Table 3: Compute time (in seconds) of LBR, and AdaBoost

Domain	AdaBoost	LBR
Annealing	6.74	3.84
Audiology	8.50	21.75
Breast(W)	1.65	0.35
KR-KP	31.10	388.74
Credit(A)	3.67	1.30
Echocar.	0.41	0.02
Glass	1.22	0.16
Heart(C)	1.44	0.20
Hepatitis	1.09	0.09
Horse	4.25	0.77
HouseVote	1.44	0.86
Hypo	13.29	54.14
Iris	0.22	0.01
Labor	0.31	0.01
LED 24	2.98	0.97
Liver	0.64	0.16
LungC.	0.61	0.15
Lympho	1.02	0.12
Pima	2.90	0.64
Postop.	0.48	0.03
P.Tumor	6.43	1.42
Promoters	1.13	0.10
Solar	5.06	2.92
Sonar	2.88	1.30
Soybean	8.95	55.50
Splice	40.20	266.30
TicTacToe	3.52	1.94
Wine	0.54	0.04
Zoology	0.45	0.03
mean	5.28	27.72
e.r. mean		1.34
w/t/l		24/0/5
p. of wtl		.0004