Improving bagging performance through multi-learner ensemble

Kuo-Wei (David) Hsu

Advisor: Dr. Jaideep Srivastava

Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota 55455

Introduction

This research proposes the use of multi-learner ensemble, or ensemble of heterogeneous algorithms, to improve the prediction performance of bagging (bootstrap aggregation) for classification problems.

Idea flow

Ultimate goal:
Effective and efficient modeling.

Thoughts:
Ensemble is effective. Two algorithms are popular: boosting and bagging. Boosting is effective but sequential (and hence less efficient); bagging is parallel (and hence more efficient) but less effective.

Research problem:
How to improve bagging prediction performance without changing its parallel structure?

Idea:
Bootstrap is the exclusive source of diversity in bagging. Since diversity is important, we might improve bagging by introducing another source of diversity.

Question #1:
What is the relationship between diversity and overall accuracy of an ensemble?

Answer #1:
A non-linear function describing a relationship between diversity and correlation, which is related to overall diversity by others.

Question #2:
Where is diversity from?

Answer #2:
Diversity could come from using different datasets, feature sets, and algorithms. We call the first one T-diversity and the last one A-diversity, which could be viewed as heterogeneity.

Question #3:
Can heterogeneity (diversity from using different algorithms) improve diversity in bagging?

Answer #3:
When heterogeneity is up to certain degree (i.e. when two algorithms are sufficiently different), the gain of diversity would be larger than a certain quantity with a lower bound of probability.

Multi-learner ensemble:
Using different algorithms could improve diversity and diversity, up to certain degree, could improve prediction performance, we have theoretical support to the use of multiple heterogeneous learning algorithms in ensemble.

Solution – bagging with heterogeneous algorithms:
We use different algorithms as base algorithms in bagging and design various plans to select algorithms.

Diversity

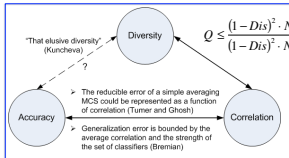
The intuition behind ensemble learning is that "*two heads are better than one*" or "*crowds are smarter than individuals*". In order to have ensemble working, member classifiers in ensemble are expected to be different. The degree of difference could be described by diversity. Several diversity measures are proposed, while none of them are standard.

Question and answer #1:

Number of samples		Classifier 1	
		incorrect	correct
Classifier 2	incorrect	N^{00}	N^{10}
	correct	N^{01}	N^{11}

$$Q = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}}$$

$$DIS = \frac{N^{01} + N^{10}}{N^{11} + N^{10} + N^{01} + N^{00}}$$



$$Q \le \frac{(1-Dis)^2 \cdot N^2 - 4 \cdot Dis \cdot N}{(1-Dis)^2 \cdot N^2 + 4 \cdot Dis \cdot N}$$

Figure 1. Relationships among diversity, correlation, and overall accuracy.

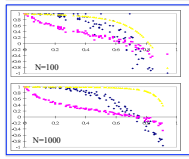


Figure 2. Experimental results. X-axis is the value of DIS and y-axis corresponds to values of Q or correlation. A (blue) diamond and a (pink) square represent respectively an observed and an observed correlation, while a (yellow) triangle gives an upper bound of the corresponding value of Q.

Question and answer #2:

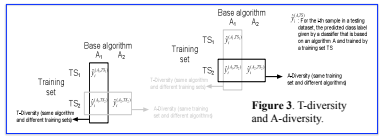


Figure 3. T-diversity and A-diversity.

- Definitions
  - A is (α, β)-stable if $\Pr[E[I(\hat{y}^{(A,TS_1)} \neq \hat{y}^{(A,TS_2)})] \le \alpha] \ge \beta$
  - A_1 and A_2 are (δ, γ)-differentiable if $\Pr[E[I(\hat{y}^{(A_1,TS)} \neq \hat{y}^{(A_2,TS)})] \ge \delta] \ge \gamma$
- Assumptions
  - A_1 is (α_1, β_1)-stable, A_2 is (α_2, β_2)-stable, they are (δ_2, γ_2)-differentiable
- Proposition
  - $\Pr[(E[I(\hat{y}^{(A_1,TS_1)} \neq \hat{y}^{(A_1,TS_2)})] - E[I(\hat{y}^{(A_1,TS_1)} \neq \hat{y}^{(A_2,TS_2)})]) \ge \delta_2 - 2 \cdot \alpha_1] \ge \gamma_2 \cdot \beta_1$
- Conclusions
  - Combination of different classifiers would provide better diversity, when diversity from using different algorithms is sufficiently larger than diversity from applying an algorithm on different training sets
  - Heterogeneity of algorithms could serve as another source of diversity

Table 1. Experimental results with respect to diversity in various measures.

Averages of diversity values		10 synthetic data sets				14 benchmark data sets			
		Bagging		Boosting		Bagging		Boosting	
		HO	HE	HO	HE	HO	HE	HO	HE
Measures where a lower value indicates more diversity	Q	1	0.58	-0.84	-0.74	1	0.78	-0.94	-0.93
	ρ	1	0.26	-0.26	-0.18	1	0.42	-0.24	-0.19
	DF	0.24	0.11	0.04	0.03	0.12	0.06	0	0
	κ	1	0.48	-0.08	0.04	1	0.75	0.04	0.15
	θ	0.12	0.07	0.06	0.06	0.16	0.13	0.08	0.08
	EEC	157	30	28	11	7	2	2	1
Measures where a higher value indicates more diversity	DIS	0	0.26	0.54	0.48	0	0.12	0.48	0.43
	E	0	0.26	0.54	0.48	0	0.12	0.48	0.43
	KW	0	0.06	0.14	0.12	0	0.03	0.12	0.11
	GD	0	0.61	0.91	0.93	0	0.56	0.98	0.99
	CFD	0	0.74	0.95	0.96	0	0.69	0.99	0.99
	$WCEC$	-429	212	-495	597	215	322	326	346
	DFD	0	0.01	0.48	0.42	0.02	0.06	0.75	0.79
	M	0	0.29	0.09	0.21	0	0.11	0.07	0.08

HE/HO presents hetero-/homogeneous ensemble. HE uses pairwise combinations from algorithms DT, NB, kNN, ANN, and SMO; HO uses each of them individually

Bagging with heterogeneous algorithms

Question and answer #3:

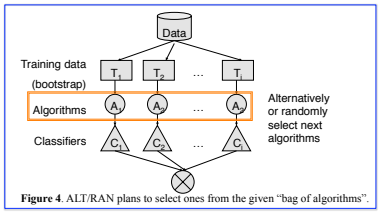


Figure 4. ALT/RAN plans to select ones from the given "bag of algorithms".

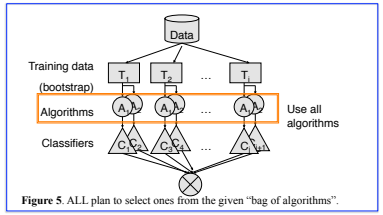


Figure 5. ALL plan to select ones from the given "bag of algorithms".

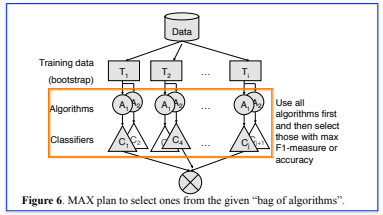


Figure 6. MAX plan to select ones from the given "bag of algorithms".

Table 2. Experimental results from using RAN on 40 benchmark datasets.

The number of winning/tying/losing data sets. Row vs. column in mean error rate.

Win/Tie/Lose (a mean error rate)		Classic Bagging		AdaBoost		Random Forests	DECORATE		Bagging?
		DT	NB	DT	NB	Forests	DT	NB	DT+NB
Random Forests		19/3/18	35/0/5	37/0/3	29/3/8	0/2/38	18/5/17	33/2/5	24/2/14
Classic Bagging	DT	--	23/3/14	19/2/19	15/3/22	11/2/27	12/2/26	22/0/18	19/3/14
	NB	--	--	17/3/22	13/5/22	11/1/28	15/2/23	22/4/14	11/2/26
AdaBoost	DT	--	--	--	20/3/18	13/3/24	13/3/22	24/1/15	17/3/20
	NB	--	--	--	--	10/1/29	19/2/22	25/1/14	14/2/23
Random Forests		--	--	--	--	--	23/3/14	30/8/12	23/2/15
DECORATE	DT	--	--	--	--	--	--	30/2/21	18/2/20
	NB	--	--	--	--	--	--	--	8/3/29

The number of winning/tying/losing data sets. Row vs. column in mean F1-measure for minority.

Win/Tie/Lose (a mean F-measure)		Classic Bagging		AdaBoost		Random Forests	DECORATE		Bagging?
		DT	NB	DT	NB	Forests	DT	NB	DT+NB
Random Forests		25/2/12	36/0/4	32/2/6	26/2/11	0/2/38	20/1/18	36/0/4	20/3/5
Classic Bagging	DT	--	20/2/18	11/4/25	15/3/22	11/3/26	21/2/17	22/2/16	19/0/15
	NB	--	--	14/3/23	13/2/22	13/2/25	18/1/21	24/1/15	19/2/19
AdaBoost	DT	--	--	--	19/2/19	20/1/19	28/0/14	24/1/15	23/2/15
	NB	--	--	--	--	14/2/24	18/1/21	22/2/16	19/3/18
Random Forests		--	--	--	--	--	28/4/7	28/1/11	20/3/12
DECORATE	DT	--	--	--	--	--	--	24/1/15	20/2/18
	NB	--	--	--	--	--	--	--	13/4/23

Conclusions

- There exists a non-linear relationship between diversity and correlation, which could be related to overall accuracy.
- Let's consider two algorithms used in heterogeneous ensemble. If one is stable enough and the other is different enough, then there is a lower bound of the probability that the diversity for such a mini ensemble would be larger than a certain quantity.
- Bagging with heterogeneous algorithms could outperform classic bagging, and provide comparable performance to some others.

Future work

- More insights into the dynamics
  - Different combinations of base algorithms
  - Prediction performance vs. number of iterations
- Different ways to combine predictions from classifiers
- More plans to select algorithms
  - For example, self-adjustment or "context-aware" plans
- Parallel implementation
  - From algorithm perspective, e.g. multi-core hardware
  - From data perspective, e.g. cloud computing platform

Related publications

- K.-W. Hsu: **Applying Bagging with Heterogeneous Algorithms to Health Care Data**, in Workshop on Data Mining for Healthcare Management, 2010
- K.-W. Hsu and J. Srivastava: **Relationship between Diversity and Correlation in Multi-Classifier Systems**, in Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2010
- K.-W. Hsu and J. Srivastava: **Diversity in Combinations of Heterogeneous Classifiers**, in Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2009
- K.-W. Hsu and J. Srivastava: **An Empirical Study of Applying Ensembles of Heterogeneous Classifiers on Imperfect Data**, in Workshop on Data Mining When Classes are Imbalanced and Errors Have Costs, 2009

Acknowledgments

I would like to express my sincere gratitude to Dr. Jaideep Srivastava for his encouragement and continuous support. My sincere thanks also goes to people who gave me comments, suggestions, and critiques.

For further information

Please contact *kuowei@cs.umn.edu*.