Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 12, December 2013, pg.412 – 417

SURVEY ARTICLE

Survey on Lazy Ensemble Methods for Improving Accuracy of Lazy Learner

Mansi Y. Parmar[1], Prof. Sonal P.Rami [2]

[1]IT Department, Gujarat Technical University, India

[2]IT Department, Gujarat Technical University, India

[1] mansi.parmar18@gmail.com; [2] sonalrami88@gmail.com

Abstract— In classification, to handle test instances, supervise learning is divided in to two parts Eagar and Lazy learning. Eagar learning build decision theory form training instance and that theory is applied on all test instances. Where Lazy learning is focus on each test instance and provides local optimal solution for each test instance. Diversity is the one issue of lazy learning, it suffers from reduce diversity. It is key issue in combination approach. It is possible to merge lazy and ensemble approach for build lazy learner. Lazy ensemble is build using ensemble approach that uses multiple classifiers and combines its predictions using combination methods and find optimal label. Many approaches for ensemble like lazy stacking, lazy bagging and other methods. Stacking with Lazy learning is used for building Lazy ensemble learner provides desire accuracy. Lazy Stacking Ensemble uses different individual "type" of classifiers as base classifiers for labelling new instance. Survey can be done only on numeric datasets and Lazy stacking is outperforming then other ensemble methods when it compared with other alternative methods in terms of classification accuracy that concern with diversity

Keywords—Lazy learning; Ensemble classification; Lazy stacking; Diversity; Accuracy

I. INTRODUCTION

A classification task begins with training data for which the target values are known. The discovered knowledge is often represented in the form of IF (conditions) THEN (class) classification rules, One of its main sub-domain is supervised learning that form decision theories or functions to accurately assign unlabeled (test) instances into different predefined classes Depending on training data how the new test instance is analysed can be defined as supervise learning. Supervise learning can be divided in Eagar learning and Lazy learning. Eager learning methods construct a general [3], explicit description of the target function when training examples are provided. Instance-based learning methods simply store the training examples, and generalizing beyond these examples is postponed until a new instance must be classified. Each time a new query instance is encountered, its relationship to the previously stored examples is examined in order to assign a label to the new instance.

Ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a vote of their predictions. A set of classifiers with similar training performances may have different generalization performances, combining outputs of several classifiers reduces the risk of selecting a poorly performing classifier. It has been discovered that a classifier ensemble can often outperform a single classifier [10].

A large body of research exists on classifier ensembles and why ensemble techniques are effective [1]. K-nearest neighbour's algorithm is a key element in lazy learning. The kNN is one of the most thoroughly analysed algorithms in machine learning, due in part to its age and in part to its simplicity [9] One of the best approaches that work under instance base- learning in ensemble classifiers is lazy bagging (LB) that builds bootstrap replicate bags based on the characteristics of test instances have great success to build lazy learner but have reduce diversity at some point that reduce accuracy of classifiers [7]. Another ensemble approach is Lazy stacking that is Meta combination method that apply on lazy learning that increase diversity and uses more diverse base classifiers. To make effective accuracy need to take more diverse base classifiers that uses by Lazy stacking.

II. RELATED WORKS

Lazy Learning is basically instance based learning that focus on particular instance. Lazy stacking is one of the effective ensemble approach use to construct lazy learner. By surveying of multiple ensemble methods like mean aggregation which combine by aggregate simple manner like voting, cluster stacking which combines different clusters, Intra-cluster stacking refer to the method of stacking within clusters and taking the mean of level 1 output as intra-cluster stacking Inter-cluster stacking is, averages the outputs of classifiers within a cluster then performs stacking on the averaged level 0 outputs and stacking a Meta combination, analysis define that stacking is more effective than others[10].

Method	Performance GI	PF1	PF2	PF3
Best Base Classifier	0.79	0.68	0.72	0.78
Mean Aggregation	0.763	0.669	0.732	0.773
Greedy Selection	0.792	0.684	0.734	0.779
CES	0.802	0.686	0.741	0.785
Stacking (Aggregated)	**0.812**	**0.687**	**0.742**	**0.788**
Stacking (All)	0.809*	0.684	0.726	0.773
Intra-Cluster Stacking	0.799*	0.684	0.725	0.775
Inter-Cluster Stacking	0.786*	0.683	0.735	0.783

Fig 1: Comparison of different ensemble methods stacking is more effective define as bold letter [10]

Lazy learners are suffering from reducing diversity; because they form a decision theory that is especially tailored for the test instance it is recovered by multiple methods that construct Lazy lerner.In ensemble methods stacking formwork is more effective to build Lazy Lerner.

A. *Lazy Bagging*

Lazy Bagging is one of Lazy ensemble approach used to build Lazy learner that build replicate bootstrap bags and uses same “type” of classifiers for prediction so it have good experience to build lazy learner but at some point it lakes diversity. Accuracy of classifier model in form of diversity is depending on type of base classifiers. If same type of base classifier is used then it contains similar prediction that reduces diversity, It. it is better to use different “type” of classifiers as base classifiers [7].

III. IMPORTANCE OF THE STUDY PROBLEM OBJECTIVES

A. *Lazy Stacking*

The above observations motivate our research to find a method that annihilates reducing diversity. Stacking is an ensemble that uses different “type” of base classifiers for labelling new instance. So we have expected that by using stacking along with lazy learners, we can provide the desire diversity [4].

Lazy Stacking is basically Meta level combination method. It applies lazy local learning to the nearest neighbours of a test instance (base part), which produces more accurate base classifies than applying to another part for build the global learner. Lazy learners are suffering from reducing diversity, so by choosing different classifier in stacking to the whole training set, the performance of the joint lazy and stacked learners can be increased accuracy. The increase in performance of LS can mainly attributed to the diversity of our model to be outlined in the section.

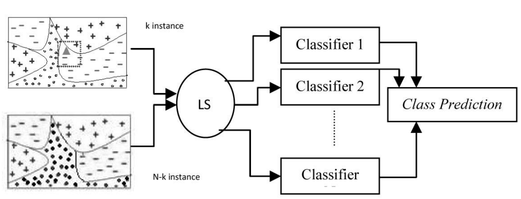

Fig 2: LS Diagram; LS waits until a test instance arrived and make a set with k instances out of the NN subset and also another set with N− k instances out of the original dataset. Afterwards the class label of the test instance is defined by the majority vote from the output of the base learners [4].

IV. METHODOLOGY

Lazy Stacking is different from Traditional Stacking. In Lazy Stacking the learning process delayed until the arrival of a test instance, Because of its lazy learning nature. When a test instance xi needs to be classified, LS will try to divide whole dataset into two partition one is used to build base classifiers and that base classifiers predictions are combined with second partition using cross validation and select the best prediction.LS first try to find the kNN of xi from the training set T, and uses the discovered kNN, along with the original training set T, to build bootstrap bags for stacking prediction. We will propose a β-similar concept to automatically determine the value of K for each dataset T, because kNN of xi play a crucial role for LS to classify xi [4].

LS sample K instances from the kNN subset (S) and N-K instances from the original learning set (T). The first N-K instances sampled from T, such that LS's base classifiers can be as independent as possible. Instead of directly putting all xi's kNN into each bag, LS applies bootstrap sampling on xi's kNN subset as well. It is expected that our procedure will ensure xi's kNN have a better chance to appear in each bootstrap bag, with No decrease of the bag independency.xi's kNN to have a better chance to appear in each bag and thus help LS build base classifiers with less variance when classifying xi.

After the construction of each bootstrap bag B´i, LS builds a classifier Ci from B´i, applies Ci to classify xi and generates a prediction Ci (xi). LS repeats the same process for L times, by different classifiers and eventually produces L predictions for xi, C1(xi), C2(xi),..,CL(xi). After that, the class y that wins the majority votes among the L base classifiers is selected as the class label for xi.

Lazy Stacking Procedure [4]

Input: 1) T: Training datasets
2) Xi: A test instance with unknown label
3) L: Number of bootstrap bags
4) β: The number of k nearest neighbour for xi

Output: yi: A new class label.

1) K ← Determine K value for T.
2) Calculate attribute weight
3) S ← Find K nearest neighbour of xi instance form T.
4) Divide training dataset into N-k and kNN
5) For i form 1 to L
   A) P ←Apply bootstrap sampling on kNN instances
   B) Bi ←Apply bootstrap sampling on N-instances
   C) Bi' ← Bi ∪ P
   D) Build Classifier Ci by using instance of Bi and apply Ci on xi prediction, Denoting predicted class label by Ci (xi).
6) End For
7) Y ← max yi (Y with majority votes)
8) Return Y.

A. *Attribute Selection*

K value defines automatically partition for more accuracy instead of kNN. The value of K decides the region surrounding a test instance xi which improves the certainty of the base classifiers in classifying xi. The accuracy of decision can be increase by adding reference parameters like 1NN, 3NN, 5NN as shown below.

Attributes	1-NN		3-NN		5-NN	
Selected	Eager	Lazy	Eager	Lazy	Eager	Lazy
10% (1)	82.6	**95.5**	82.6	**95.5**	82.6	**95.5**
20% (3)	95.5	**97.8**	94.4	**100.0**	94.4	**100.0**
30% (4)	95.5	**98.3**	93.3	**98.9**	93.3	**98.9**
40% (5)	97.2	**98.9**	93.3	**98.9**	93.3	**98.9**
50% (7)	97.8	**98.3**	96.6	**98.3**	96.6	**98.3**
60% (8)	98.9	98.9	96.6	**98.3**	96.6	**98.3**
70% (9)	97.2	**97.8**	94.9	**97.8**	94.9	**97.8**
80% (10)	97.2	**98.3**	94.4	**97.2**	94.4	**97.2**
90% (12)	96.6	**97.8**	96.1	**96.6**	96.1	**96.6**
100% (13)	98.3		96.1		96.1	

Fig 3: Improving kNN selection by adding K values [6]

In this propose a β- similar concept [2] to automatically determine the value of K for each dataset T based on entropy. The large value of k leads to the increase of bag dependency and very small value of k leads to decrease in accuracy so we should find an appropriate value of k. Derive sampling entropy based approach to automatically determine the value of K for each data set.

In Lazy attribute selection, attribute is selected on new query is arrived, similar concept it by default define as 0.99 for analysis where, to help an instance xi, find similar neighbours, we need to find the weight of each Attribute so that the weighted distance function can indeed capture instances similar to xi Information-gain Ratio (IR) as a weight measure for each attribute. The attribute weights is for finding the kNN subsets i.e. attributes with larger weights have stronger effects in finding the nearest neighbours of an instance. A new parameter like Minimum Description Length (MDL) measure which is least bias toward multi valued attribute, Multivariate split – based on combination of attributes [6]. The weighted Euclidean distance for finding kNN of an instance, it effect lazy attribute selection in kNN that improves performance. IR0(Ai) is the Normalized information gain for the i'th attribute and R is the total number of attributes:

$$Dis(x_i, x_j) = \frac{1}{R}\sqrt{\sum_{i=1}^{R} IR'(A_i).(x_j^{A_j} - x_i^{A})}$$

B. Diversity between Classifiers

Diversity among the members of a team of classifiers is deemed to be a key issue in classifier combination, and has been recognized as a very important characteristic. An ensemble combines their outputs to improve the performance of a single classifier. The strategic combination can reduce the total error. Therefore we need base classifiers whose decision boundaries are adequately different from those of others, such a set of classifiers is said to be diverse [8]. Diversity means at any point two classifiers make different errors can say that they are diverse. To improve the performance of classification must catch diverse base classifiers that they get accurate results in the form of reduce errors LS improve the result by taking more diverse different "type" of classifiers.

C. Statistical Methods for Performance Testing

Different statistical methods can be applied on 10 real world datasets to cheek the performance of the performance of lazy stacking, methods like algorithm correlation, win/loss/tie comparison, absolute average accuracy, t-test. Q statistic, Friedman test and Nemenyi test. The correlation between any two algorithms can be observed by calculating the correlation between those two random variables. K4.5 has much stronger correlation with kNN than LB does [8].

LS's wins (8) versus losses (2) compared with LB is also statistically significant. The results indicate that out of the 10 numeric datasets, LS wins other methods in 8 numeric datasets (a probability of 8/10=80%).

In Q-Statistic, Two classifier Ci & Ck [3]

$$Q_{i,k} = \frac{ad - bc}{ad + bc}$$

	D_k correct (1)	D_k wrong (0)
D_i correct (1)	a	b
D_i wrong (0)	c	d

Total, $a + b + c + d = 1$

Fig 4: The 2 × 2 relationship table with probabilities

It defines accuracy of learner that define that how many instances are correctly classify form whole dataset in percentage. The Friedman test: first determines if there are statistically significant differences between any pair of Methods over all datasets, followed by a posthoc. Nemenyi test to calculate a p-value for each pair of methods [7]

V. RESULT & DISCUSSION ANALYSIS

A. Classification Accuracy Comparison

Dataset	*KNN*	C4.5	TB	Lazy Bagging	Lazy Stacking
Ionosphere	84.18±0.89	89.12±1.18	91.34±0.83	**92.97±0.74**	91.76±0.83
Bupa	63.47±2.21	64.20±2.90	69.17±1.88	70.23±2.04	**72.28±1.38**
ImageSeg.	87.24±0.37	96.45±0.24	96.94±0.47	**97.74±0.49**	97.62±0.38
Ecoli	**86.01±1.07**	82.56±1.17	83.72±1.06	83.30±1.11	85.97±0.69
Wine	96.06±0.93	92.09±1.28	94.47±0.87	95.62±0.89	**96.33±1.09**
Glass	64.95±1.87	66.92±2.65	72.62±1.88	74.31±1.62	**74.55±1.77**
Pima	75.00±1.05	73.65±1.21	75.17±0.81	76.23±0.68	**76.89±1.08**
Sonar	73.18±6.75	73.08±3.63	75.91±2.21	80.05±1.78	**84.66±1.87**
vowel	40.51±2.02	79.19±1.57	86.37±0.99	90.19±1.12	**92.45±0.98**
Vehicle	65.06±1.03	71.56±2.05	73.41±0.98	74.94±1.02	**75.59±1.15**

Fig 5: Classification Accuracy on 10 Numeric Datasets Selected From the UCI Data Repository. For each dataset. The accuracy of the method with the high accuracy is marked with bold face [3].

Survey can be done on Numeric datasets by taking In order to measure the performance of the proposed algorithms in this work, we employed 10-time 5-fold cross validation for each dataset, and assess their performance, based on the average accuracy over 10 trials. Here, the survey of different learning algorithms like kNN, TB, LB those apply to build Lerner, on different 10 real numeric datasets form UCI repository

As per the comparison to other statistical methods and survey analysis define that as the mean accuracy in 10 UCI numeric datasets by lazy bagging classification is 83.558 and by lazy stacking is 84.81[1]

B. Different Learning Algorithms

Different Learning algorithms are applied like C4.5, kNN, TB and LB [9].

C4.5 developed and uses gain ratio for selection of attribute for splitting. It provides an improvement over ID3 as it deals with nominal and numerical attributes as well as able to handle missing and noisy data

The choice of k also affects the performance of k-nearest neighbour algorithm. If value of k is small, and noise is present in the pattern space, then noisy samples may win the majority votes, which results into misclassification error. Nearest neighbour classifier are very slow in classifying a new sample. TB is direct sampling training data and builds bootstrap bags on that. LB is lazy ensemble approach different form TB and outperforms then TB but lake of diversity problem. This all algorithms performances are compared with LS.

VI. CONCLUSION AND FURTHER RESEARCH

This survey defines that Lazy ensemble is the best approach to build lazy learner in that lazy stacking is outperforms then other ensemble methods as per the analysis. Lazy Ensembles is applied on only numeric datasets till now. Current work can be extended on applying on mixed values that have more complexity contains values of numeric, real and categorical values that performance will compare with existing work.

REFERENCES

[1] Jiawei Han and Micheline Kamber,” Data Mining: Concepts and Techniques”, ISBN 13: 978- 1-55860-901-3, ISBN 10: 1-55860-901-6.

[2] Lior Rokach,” ENSEMBLE METHODS FOR CLASSIFIERS”chepter-45

[3] Rojar rope xemia, “Data Mining: A Closer Look: chepter 2s”, http://www.pearson.ch/1471/9780201741285/Data Mining-A- Tutorial-Based-Primer-.aspx

[4] Haleh Homayouni, Sattar Hashemi, and Ali Hamzeh,” Instance-based Ensemble Learning Algorithm with Stacking Framework”, , 2nd International Conference on Softwar Technology and Engineering(ICSTE),IEEE-2010

[5] Haleh Homayouni1, Sattar Hashemi2 and Ali Hamzeh,” A Lazy Ensemble Learning Method to Classification”, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010

[6] Rafael B. Pereira, Alexandre Plastino, Bianca Zadrozny Luiz Henrique de C. Merschmanny, Alex A. Freitas,” Lazy Attribute Selection Choosing Attributes at Classification Time”, July- 3, 2010

[7] Xingquan Zhua,* Ying Yangb," A lazy bagging approach to classification" Pattern Recognition 41,2980 2992,2008

[8] S.A. Gilpin, D.M Dunlavy," Relationships between Accuracy and Diversity in Heterogeneous Ensemble Classifiers", Department of Energy's Nuclear Security Administration under Contract DE-AC04- 94AL85000, , SAND2009,694OC

[9] Hetal Bhavsar, Amit Ganatra," A Comparative Study of Training Algorithms for Supervised Machine Learning", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012

[10] Sean Whalen and Gaurav Pandey," A Comparative Analysis of Ensemble Classifiers: Case Studies in Genomics", ArXiv: 1309.5047v1 [cs.LG] 19 Sep 2013

[11] SASO DˇZEROSKI, BERNARD ZENKO," Is Combining Classifiers with Stacking Better than Selecting the Best One?", Machine Learning, 54, 255–273, 2004

[12] In´es M. Galv´an, Jos´e M. Valls, Miguel Garc´ıa and Pedro Isasi," A lazy learning approach for building classification models", march 201

[13] Mohamed Bibimoune1;2, Haytham Elghazel1, Alex Aussem," An Empirical Comparison of Supervised Ensemble Learning Approaches", Solving Complex Machine Learning Problems with Ensemble Methods, Prague, Czech Republic, 27 September 2013

[14] Niklas Lavesson," Evaluation and analysis of supervised learning algorithms and classifiers", Blekinge Institute of Technology Licentiate Dissertation No 2006:04, ISSN 1650- 2140, ISBN 91-7295-083-8.

