**Introduction**

Ensemble learning have been widely used in the field of machine learning, plenty of successful classifier was created based on it. ensemble method refers to the structure which built by several individual classifier and integrate all decisions into a specific result. in this method, there are two different ensemble types which are homogeneous and heterogeneous. in homogeneous method the ensemble only include individual classifier in similar type such as decision tree ensemble only include individual classifier which is decision tree, the heterogeneous method is built by different types of classifier such as decision tree and neural network. this essay will explore several common topics under the field of ensemble learning, it will focus on how to created diverse classifiers, how to fuse the decisions from individual classifiers and how to establish the weights that individual classifiers contribute to the ensemble’s answer.

**Diverse classifiers**

diverse classifiers refers to the individual classifiers which can show different performance. a successful ensemble model include different individual classifiers which all showing high performance. for example, in a true false classification task, assuming in three classifiers there would be three different result produced for each classifier when there are three testing samples, the final result will be produced by voting.

|  |  |  |  |
| --- | --- | --- | --- |
|  | test sample 1 | test sample 2 | test sample 2 |
| c1 | True | False | True |
| c2 | False | True | True |
| c3 | True | True | False |

1. performance increased

|  |  |  |  |
| --- | --- | --- | --- |
|  | test sample 1 | test sample 2 | test sample 2 |
| c1 | True | False | True |
| c2 | True | False | True |
| c3 | True | False | True |

1. no increasing

|  |  |  |  |
| --- | --- | --- | --- |
|  | test sample 1 | test sample 2 | test sample 2 |
| c1 | True | False | False |
| c2 | False | True | False |
| c3 | False | False | True |

(c) performance decreased

table 1, example to ensemble method

table 1 showing the example, 1.a showing a successful classifier, all three classifier showing 66.66% accuracy, the whole ensemble model showing 100% accuracy. 1.b showing there is no changes after ensembling. 1.c showing there is a decreasing performance since three classifier showing only have 33.33% accuracy. in this example, a good ensemble have high performance in each classifiers and each classifiers are also unique.

based on the example, the individual classifiers should be diverse and showing high performance, thus, for creating diverse individual classifiers, two main method will be presented which are serial and parallel. serial method is used when the individual classifiers have strong relationship between each other and must be generate in a specific order. parallel method is used for the individual classifiers which do not have strong relationship between each other and all them can be generated at same time. boosting is a example of serial and bagging(Breiman, 1996) is a example of parallel. boosting algorithm is allowed to increase a weak classifier to strong classifier, its created by training a basic classifier first, then try to change the distribution of samples based on the performance of basic classifier. then training the new basic classifier based on changed samples. then whole process required repeat those process for several times. in the whole process, for each specific data sample distribution two method can be used for training specific classifier which are re-weighting and re-sampling. bagging algorithm normally based on bootstrap sampling to training each individual classifier, the bootstrap sampling is working by random selecting specific number of samples for training individual classifier each time, the selected samples also used for next random select. In summary, all those methods are available to created a diverse classifier based on ensemble methods.

**Increase diversity**

For increasing the diversity of model, it will focus on two parts which are diversity measure and diversity increase. the diversity measure is used to assess the diversity of individual classifier, the common method is to consider the similarity between each two classifiers, for a binary classification task, training by data set D = {(x1,y1),(x2,y2)....(xm,ym)}, the contingency table in two classifier should be

|  |  |  |
| --- | --- | --- |
|  | hi = +1 | hi = -1 |
| hj = +1 | a | c |
| hj = -1 | b | d |

a showing the number when both classifier predict positive value, b,c,d is following by similar rule, based on this case, following would be some common method for measure its diversity.

disagreement measure(figure 2) is one which can be used for measure it, its value is between 0 and 1, when the value higher, the diversity is stronger.

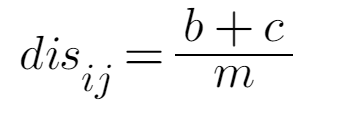


figure 2, disagreement measure

correlation coefficient(figure 3) is another method, its value is between -1 and 1, when there is no relation between both classifier, the value would be 0, if there is positive relationship, the value would be 1, if navigate, it would be -1.

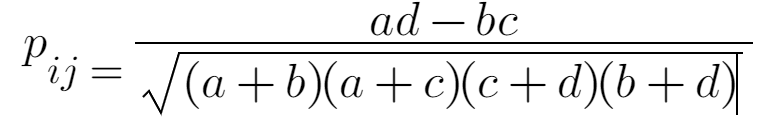


figure 3, correlation coefficient

Q-statistic(figure 4) is another method to assess the diversity. value Qij is similar with Pij, and |Qij| > =|Pij|.

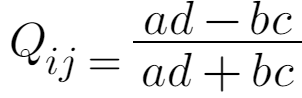


figure 4, Q-statistic

k-statistic(figure 5) is also used for measure the diversity, when k = 1, it means both classifier is basically same, when both classifiers only same in specific value, k=0.

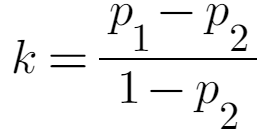


figure 5, k-statistic

in figure 5, p1(figure 6) is the possibility when both classifier total same, p2(figure 7) show the possibility when both classifier only same in specific value.

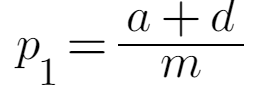


figure 6, calculation of P1

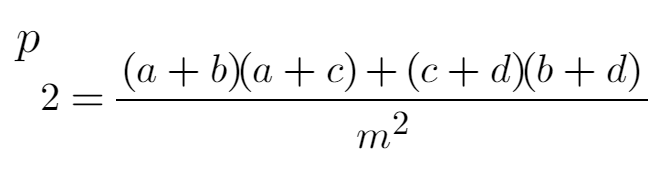


figure 7, calculation of P2

diversity increase is focus on four parts which are data sample changes, input attribute changes, output changes and permater changes.

first, the data sample changes refers to training the classifier based on different data samples from original data set such as bootstrap sampling which was shown before in bagging method and sequence sampling in adaBoost (Freund & Schapire, 1997). this method already been used in bootstrap sampling method by random selecting and re-sampling to improving the model,

it is helpful in increasing the diversity of classifier and model performance, but this method is not working well if there is some classifiers are not sensitive to the changes of data such as linear model and SVM.

second, the input attributes changes is working by changing the attributes of samples, random subspace(figure 8) is a example of it(Ho, 1998), in this method, select a specific number of attributes for each classifier, it saves time for training classifier and reduce the redundancy of useless attributes when predict results, its diversity also increase for individual classifier at same time, but this method also take longer time for exploring data, if there is less redundancy data, this method may not working well.

third, output changes is also helpful for increasing the diversity of individual classifier, flipping output(Breiman, 2000) is a example of it, in this method, it changes the label of data sample to increase diversity, output smearing(Breiman, 2000) is another example of it, in the method it change the classifier output to regression output for increasing the diversity.

final, parameter changes is also available to increase the diversity of individual classifier, this method is working by changing the parameter of individual classifiers such as the number of nodes, initializing weights in neural network, it is useful to increase the diversity of model but also takes longer time. a example in negative correlation it make individual classifier use different parameter by regularization.

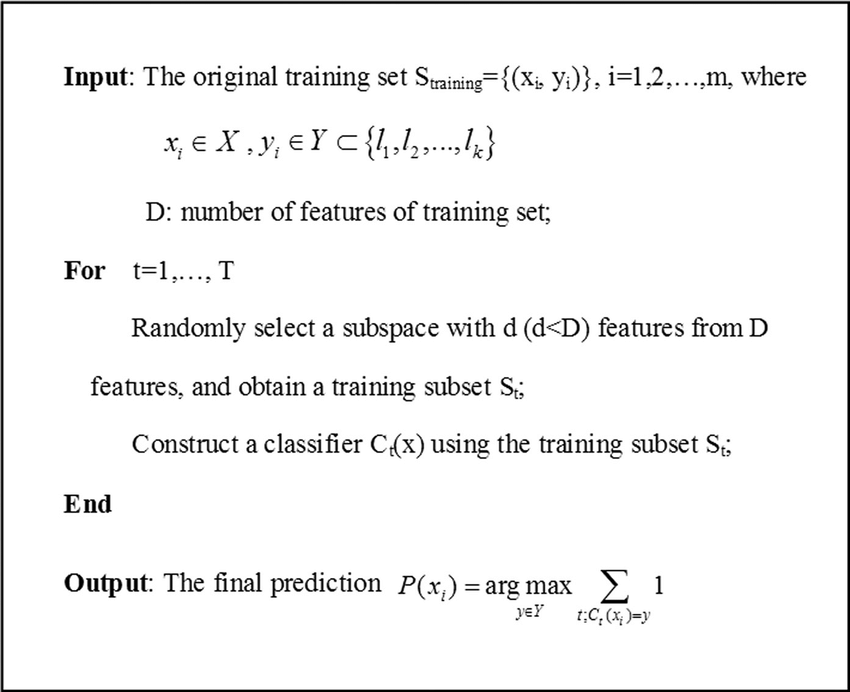


figure 8, random subspace algorithm

**Fuse decision**

there are three methods can be used to fuse the decisions from individual classifier (Dietterich 2000).

the first one is averaging, both simple averaging and weighted averaging are used for fusing decision from different classifiers, simple averaging(figure 9) is working by averaging all results from different classifiers, the decision would be the average result of the whole model, this method is more suitable when there is less difference in the results between individual classifiers.

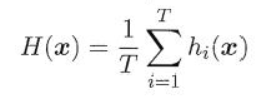


figure 9, simple averaging

weighted averaging(figure 10) is working by adding weight when averaging the results of different classifier, the weight normally learnt by training sample, but sometimes it may caused over fitting if there is much weight should be learnt, this method is suitable for the result produced from classifiers have strong difference. in discuss the benefit and drawback of averaging method, it is easy to understand and clear to use, it is suit for most regression problem, but it is not able to resolve classification problem(Perrone & Cooper, 1993).

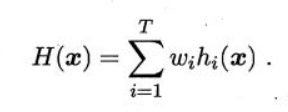


figure 10, weighted averaging

the second one is voting, it also can be divided to three different types which are majority voting, plurality voting and weighted voting. the majority voting(figure 11) is a method which make decision by gaining more than half votes. the plurality voting(figure 12) is a method which select the decision from highest voting, if there are more than one, then random select. the weighted voting(figure 13) is the method when calculating the votes, adding specific weight to each result, then select the highest voting. in discuss its advantage and disadvantage, this method is easy to understand and operate, it is working well in fuse decision from different classifier, but it is not suitable for a large complex data set.

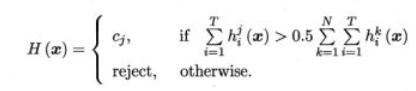


figure 11, majority voting

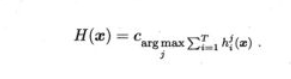


figure 12, plurality voting

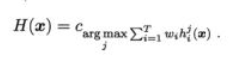


figure 13, weighted voting

the third one is learning, the learning method integrate different classifier based on another classifier, stacking(figure 14) is a example of this method. in this method it required to training several basic classifier first, then use the output of them as input of meta-learner, use the label of original data set as true label of training set which used for meta-learner, then repeat the process several times. in discuss its benefit and drawback, this method is suitable for large data set, splitting data, improve performance and get better results for ensemble model, however, it must have enough data to learn,the segmentation may reduce the height of data level, it is not able to provide sufficient guarantee for training(Wolpert, 1992; Breiman, 1996).

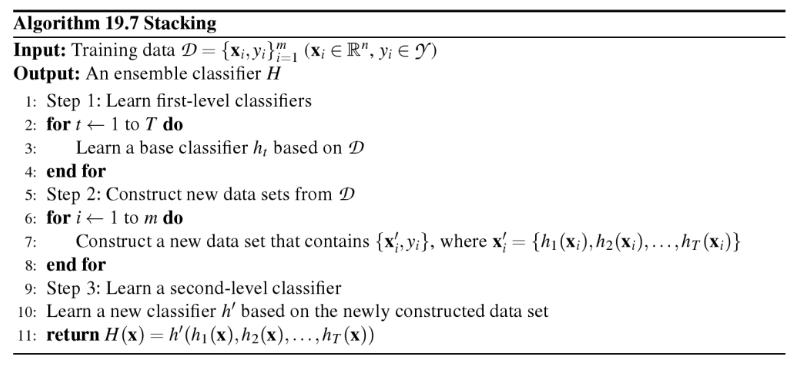


figure 14, Stacking algorithm

**Establish weights**

For establishing weights of ensemble model, the two types of structure in ensemble model should be considered which are serial and parallel, in the serial structure, the weights should be changed for increasing the performance of model, the example boosting is presented before for introducing this structure. a example of boosting which are adaboost(figure 15) will be used for presenting how the weight be established for contributing to the whole model.

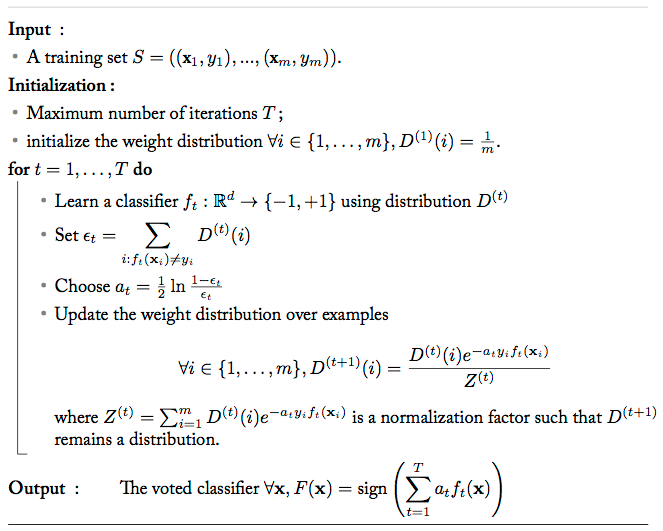


figure 15, Adaboost

Adaboost use re-weighting method to establish the weight for increasing the performance of classifier. in this method, it first initialize the distribution of weights(D1) in training sample to produce basic classifiers(G1), its weight was defined by figure 16, then update the new weight to wrong classificated samples by time exp(-a1yiG1(Xi)), then use the new weight distribution training next classifier and repeat the process to T. then ensemble(figure 17) all the classifier into a whole model. in discuss its benefit and drawback, this method is working well in update weights, but also easy to cause overfitting.

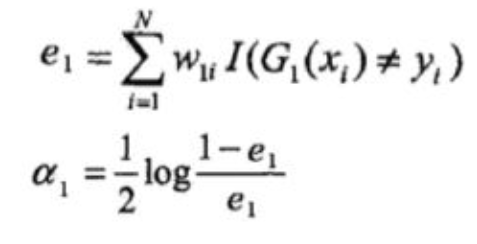


figure 16, initlized weigth

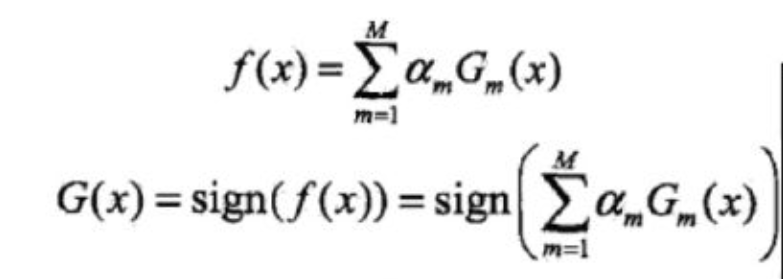


figure 17, ensemble the model

in parallel structure, bagging which was showed before will be used to present the establish of weight, random forest(Breiman, 2001) algorithm(figure 18) which as a example of bagging will be used to present the establish of weights. according to figure 18, there is not a specific weight changed in the whole process, which means there is not a requirement for changing the weight for each decision tree, all individual classifier should have same weight when predict the results.

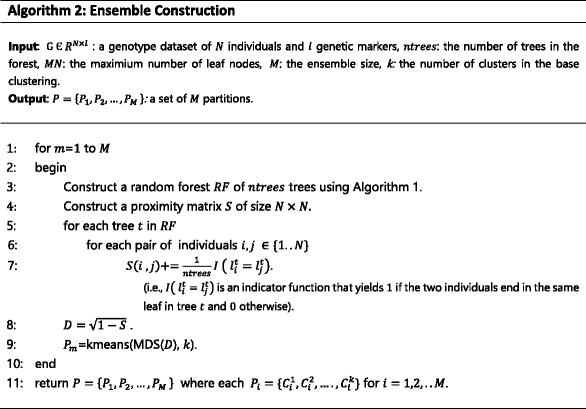


figure 18, Random forest

**Exploration**

for exploring the difference of both bagging and boosting method, it can be divided into several fields, in the select of samples, bagging use bootstrap sampling which is random select and back all selected samples for other classifier, boosting normally re-weighting the samples. in the changes of weights, bagging classifier have equal amount of samples and equal weights each other, boosting required update the weight for each classifiers. for predict the value, bagging have equal weight for each predicted value and boosting have specific weights for each value, for the structure, bagging method is parallel running and boosting method is serial running, for increase the diversity, bagging method more use sample changes and attributes changes, boosting more use weights changes. for discussion their benefit and drawback, bagging is more suitable for independent dataset, it is working well in reduce the variance, boosting is suitable for the dataset which have strong relationship to each other, it is working well for reduce bias.

**New method**

The new method will focus on resolving the problem how to create diversity individual classifiers under ensemble method. it already been discussed that why diversity is important for individual classifier, how to estimate the diversity of individual and how to increase the diversity of individual classifier. based on those finding, a new method for creating diversity individual classifier will be introduced.

In the new method the individual classifier is allowed to created first, the diversity measure method will working as a updater, the average value A will be produced for looping each two specific individual classifier based on k-statistic, the method is parallel, the individual classifiers is suitable for the classifier which are sensitive to the changes of sample, the number of samples in bootstrap sampling will be updated for each loop.

------------------------------------------------------------------------------------------------------------------------

**algorithm 1: diversity updater**

------------------------------------------------------------------------------------------------------------------------

**Input**: a training set S = {(x1,y1),(x2,y2).....(xn,yn)}

**output**: a diversity auto-updater based on threashood value T, Valve T in [0,1]

**initialized** value A to 1, bootstrap sampling number N and samples update frequence F in [0, 1]

**while** A > than T **do**

**initialized** ensemble model by training individual classifier on S

**for** C in individual classifiers **do**

**for** M in individual classifiers **do**

**calculate** K-statistic

**add** K-value to a specific list L

**end for**

**end for**

**Averaging**  the value of list L to A

N = N - S\*F

**end**

-----------------------------------------------------------------------------------------------------------------------

In this method, the diversity updater will update the diversity rate into a specific value T based on bootstrap sampling updater. it used k-statistic to calculate the diversity rate of model. the method is useful for individual classifier which is sensitive to sample changes only, which means it should be showing less efficient to the individual classifiers such as linear model and SVM.

for improving this algorithm, the method to calculate diversity value is allowed to select from all four diversity estimate algorithm, the method for increase the diversity is also available to select from four method based on data sample and the model which used for individual classifier.

**Conclusion**

in discussion, ensemble method is working well for increasing weak classifier to strong classifier, it include both homogeneous and heterogeneous, there are also two different kinds of structure which are serial and parallel, diversity is a significant part for create ensemble method, in addition, it is also important for fusing decision from different individual classifier, establish is also a significant part for increasing the performance of weak classifiers to strong classifiers. in this essay, it presented some exist solution for solving those problem and also provide a new method for resolving the problem create diversity individual classifier under ensemble method which are diversity updater. it is important to provide specific method for different case types, there is not a method are available to resolve all the problems once.

**Reference**

Adams, D.C., 2014. ‘A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data’. *Systematic Biology*, vol.*63, no.*5, pp.685-697.

Bomberger, W.A., 1996. ‘Disagreement as a Measure of Uncertainty’. *Journal of Money, Credit and Banking*, vol.*28, no.*3, pp.381-392.

Breiman, L., 1996. ‘Bagging predictors’. *Machine learning*, vol.*24, no.*2, pp.123-140.

Breiman, L., 1996. ‘Stacked regressions’. *Machine learning*, vol.*24, no.1*, pp.49-64.

Breiman, L., 2000. ‘Randomizing outputs to increase prediction accuracy’. *Machine Learning*, vol.*40, no.*3, pp.229-242.

Breiman, L., 2001. ‘Random forests’. *Machine learning*, vol.*45, no.*1, pp.5-32.

Breiman, L., 2001. ‘Using iterated bagging to debias regressions’. *Machine Learning*, vol.*45, no.*3, pp.261-277.

Dietterich, T.G., 2000, June. ‘Ensemble methods in machine learning’. In *International workshop on multiple classifier systems*, pp. 1-15.

Dietterich, T.G. and Bakiri, G., 1994. ‘Solving multiclass learning problems via error-correcting output codes’. *Journal of artificial intelligence research*, vol.*2*, pp.263-286.

Freund, Y. and Schapire, R.E., 1997. ‘A decision-theoretic generalization of on-line learning and an application to boosting’. *Journal of computer and system sciences*, vol.*55, no.*1, pp.119-139.

Mujica, L.E., Rodellar, J., Fernandez, A. and Güemes, A., 2011. ‘Q-statistic and T2-statistic PCA-based measures for damage assessment in structures’. *Structural Health Monitoring*, vol.*10, no.*5, pp.539-553.

Perrone, M.P. and Cooper, L.N., 1992. ‘When networks disagree: Ensemble methods for hybrid neural networks*’*. *BROWN UNIV PROVIDENCE RI INST FOR BRAIN AND NEURAL SYSTEMS.*

Wolpert, D.H., 1992. ‘Stacked generalization’. *Neural networks*, vol.*5, no.*2, pp.241-259.

Yang, J.H. and Yang, M.S., 2005. ‘A control chart pattern recognition system using a statistical correlation coefficient method’. *Computers & Industrial Engineering*, vol.*48, no.*2, pp.205-221.