International Journal of Computer Theory and Engineering, Vol.2, No.6, December, 2010
1793-8201

Trend Spotting Technological advancement, Issues, Challenges and Opportunities in Information Fusion Methods specific to Multiple Classifier Systems - A Survey Paper

Amit P Ganatra*, Yogesh P Kosta**

Abstract—**This paper is a precipitate of handpicked research papers in the field of Information fusion specific to multiple classifier systems, following systematic consolidation, clearly spotlighting technology trends, issues, opportunities and challenges, apart from inking authors' views and critical suggestions in the form of questions at the end of this paper. Large amounts of data are nowadays available because gathering data is easy and inexpensive. Most data is raw and to be useful relevant knowledge has to be extracted from it. Technology alone does not deliver a solution. Data mining is a process that must be reliable, repeatable with even with a little knowledge about data mining by the people it provides the solution. Data analytics is the day in and day out activity for almost all the organizations. People with business problem and with the data on that issue, expect more sophisticated and actionable information with higher accuracy and time (cost) effective solution to make decisions from the data. Data mining functionalities provide the organization with some form of intelligence. If we have several sets of data obtained from various sources, where the nature of features are different (heterogeneous features), a single classifier cannot be used to learn the information contained in all of the data. And if we have to perform different tests then each test generates data with a different number and type of features, which cannot be used collectively to train a single classifier. In such cases, data from each testing modality can be used to train a different classifier, whose outputs can then be combined. Applications in which data from different sources are combined to make a more informed decision are referred to as Information fusion applications, and ensemble based approaches like bagging and all the variants of boosting etc. have successfully been used for such applications. We compared the performance and performed the survey of these methods on a collection of machine-learning benchmarks. This paper reports the general survey and the results of applying Information Fusion (Multiple Classifier System) methods to a system that learns from various models, problems that may arise in implementing these models (algorithms) are explored and the research issues where the further work is expected to be done.**

Index Terms—**Information Fusion, Multiple Classifier, Machine Learning, Ensemble, Diversity, Data Mining, bagging, boosting.**

1 * Associate Professor, CE-IT, Faculty of Engineering & Technology
** Dean, Faculty of Engineering & Technology, Charotar University of Science & Technology, Changa-388421, Anand, Gujarat (INDIA)

I. INTRODUCTION

Data Mining is a continuous and an ongoing process used to find useful and understandable patterns and to predict, associate and group the patterns to make the correct and timely decisions. Old View was to learn one good model. New View is to learn a good set of models. Probably this is the best example of interplay between "theory & practice" in Machine Learning with Data Mining. Fusion often produces accuracy gains. It can combine "classifiers" of various types. E.g., decision trees, rule sets, neural networks, etc. Information Fusion (Ensemble Classification) is an aggregation of predictions of multiple classifiers with the goal of improving accuracy. It is an improvement of basic multi classifier algorithm to avoid over fitting and noise [5].

Information fusion is becoming a major need in Data Mining and Knowledge Discovery in database. Typical applications of these techniques include the preprocessing step or data (Information) modeling (e.g. multiple classifier system, ensemble methods). [1-2] Data Mining uses information fusion techniques for improving both the predictive accuracy and quality of the extracted knowledge. In predictive model, a set of fields as an input are used to predict the values of the field as output. Nevertheless the gap between both data mining & information fusion areas is large. The purpose of the study is to improve understanding of why and when these algorithms can be used, combination techniques, how to reweight, and, how it affects classification error and the future scope of these methods [25]. We compared the performance and performed the survey of these methods on a collection of machine-learning benchmarks.

Due to this need the interest on Information Fusion techniques is increasing in the Data Mining community. Fusion could be defined as a merging of diverse, distinct, or separate elements into a unified whole. An Information Fusion method (system) is thus seen as a function that transforms the input into an output given some conditions with improved predictive accuracy [11]. These conditions might include controls, input parameters, background knowledge, and goals. Information Fusion is the process of combining data to refine state estimates and predications. Information fusion is the process of acquisition, filtering, correlation and integration of relevant information from various sources into one representational format that is appropriate for deriving decisions regarding the interpretation of the information, system goals like classification, prediction, tracking etc[1-2].

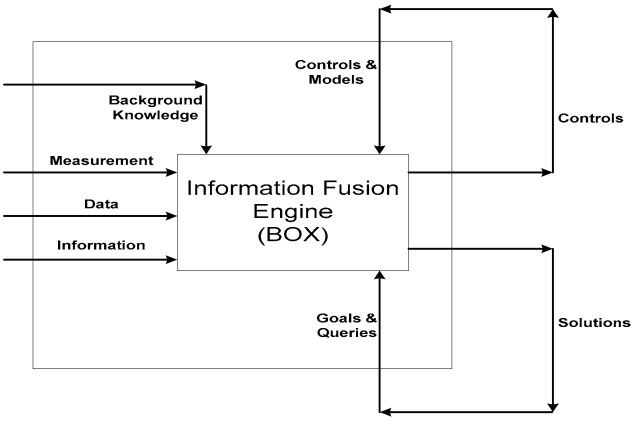

Figure 1: Information Fusion System

The purpose of information fusion is to produce information from different sources in order to support the decision-making process. In theory, the fusion of redundant information from different sources can reduce overall uncertainty and thus increase the accuracy of the system. The final information generated to be of superior quality, in some tangible way, than the information available from the primary sources. Fusion, by definition, requires a qualitative difference between the final output and the output of the original sources [2].

A. Rationale and significance of the study

The degree and method by which information fusion (multiple classifiers) systems share training resources among their components can be a measure of co-operation. Training resources that are sharable in a multiple classifier system are training patterns, algorithms or information. Co-operative training as a means of sharing training information amongst an ensemble during training. To achieve the improvements in classification accuracy by sharing, or co-operation, amongst classifiers during their training is useful in a multiple classifier system [11].

Information Fusion is becoming a major requirement in data mining and knowledge discovery in database for the following uses. In this paper we mainly focused towards model building.

B. Three main uses

1) Information Fusion in preprocessing
   Fusion is used to increase the quality of raw data prior to the application of DM methods.
2) Information Fusion for building models
   The model built from data uses some kind of Information Fusion techniques (particular aggregation operator).
3) Information Fusion to extract information
   The knowledge extracted from the data is the result of a particular fusion technique e.g. aggregated value computed from the data.

C. Design of Information Fusion (Multi Classifier System) can be accomplished at 4 levels

1) Aggregation Level
2) Classifier Level
3) Feature level
4) Data Level

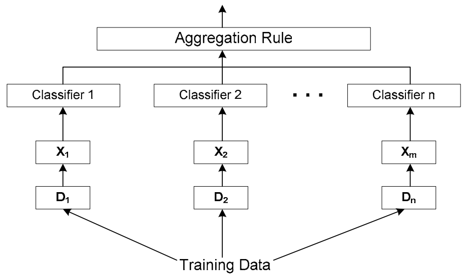

Figure 2: Levels of Information Fusion System

D. Combining Schemes for Information Fusion Methods

1) Static vs. Adaptive, Fixed vs. Trainable
2) Voting methods: Max, average, majority
3) Weighted average, fuzzy integrals, belief theory.
4) Decision Template, Behavior Knowledge space
5) Data Level combining: partitioning technique for training multiple classifiers that generates nearly optimal training partitions.

E. Three ideas

1) Simple (outweighted) votes
2) Weighted votes
3) Train a combining function

F. Architecture

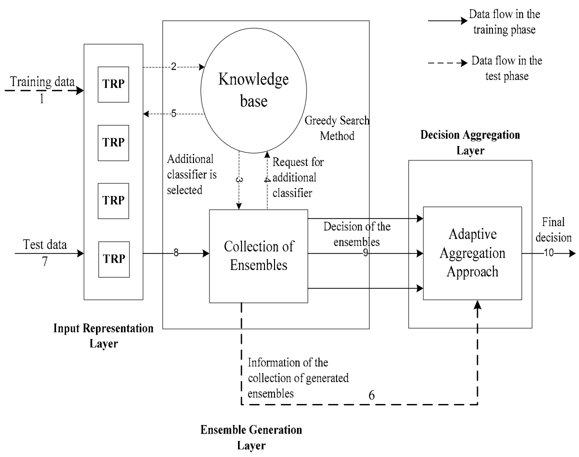

Figure 3: Architecture of Information Fusion System

G. Steps of Information Fusion

1) Learning the application domain
2) Creating a target data set: data selection
3) Data cleaning: (may take 60% of effort!)
4) Data reduction and transformation
5) Choosing methods of data mining
6) Choosing the mining algorithm(s)
7) Data mining: search for patterns of interest
8) Pattern evaluation and knowledge presentation
9) Use of discovered knowledge
10) Multiple classifier systems consist of a set of classifiers and a combination strategy.

H. Motivations

1) Existence of many alternative classifiers each has its own feature and representation space
2) Existence of different training sets collected at different times and may even have different features.
3) Each classifier may have good performance in its own region of the feature space.

II. Multiple Classifier System

Multiple Classifier (Ensemble) learning is a type of machine learning that studies algorithms and architectures that build collections, or *ensembles*, of statistical classifiers that are more accurate than a single classifier. Multiple classifier systems (MCSs) based on the combination of outputs of a set of different classifiers have been proposed in the field of machine learning as a method for the development of high performance classification systems[11]. Previous work clearly showed that multiple classifier systems are effective only if the classifiers forming them are accurate and make different errors. Therefore, the fundamental need for methods aimed to design "accurate and diverse" classifiers. We cannot normally expect to obtain base models that misclassify examples in completely separate parts of the input space and ensembles that classify all the examples correctly. However, there are many algorithms that attempt to generate a set of base models that make errors that are as uncorrelated as possible [25-26].

A. Ensemble Data Mining Methods

Ensemble Methods began about fifteen years ago as a separate area within machine learning and were motivated by the idea of wanting to leverage the power of multiple models and not just trust one model built on a small training set. Significant theoretical and experimental developments have occurred over the years and have led to several methods, especially bagging and boosting, being used to solve many real problems. However, ensemble methods also appear to be applicable to current and upcoming problems of distributed data mining and online applications. Therefore, practitioners in data mining should stay tuned for further developments in the vibrant area of ensemble methods. This paper balance between theory, algorithms, and applications of ensemble methods gives a comprehensive idea of the work being done in the field [22-26].

A supervised machine learning task involves constructing a mapping from input data (normally described by several features) to the appropriate outputs. In a classification learning task, each output is one or more classes to which the input belongs. The goal of classification learning is to develop a model that separates the data into the different classes, with the aim of classifying new examples in the future. In a regression learning task, each output is a continuous value to be predicted. Many traditional machine learning algorithms generate a single model (e.g., a decision tree or neural network). Ensemble learning methods instead generate multiple models [22-26].

There are many algorithms that attempt to generate a set of base models that make errors that are as uncorrelated as possible. Methods such as Bagging and Boosting promote diversity by presenting each base model with a different subset of training examples or different weight distributions over the examples [12]. Running the same learning algorithm on different subsets of training example can yield very different classifiers which can be combined to yield an effective ensemble [6].

B. Background: Ensemble learning

Combines multiple learned models to construct better generalizations. Classifiers that always agree won't give new information. Combining predictions of an ensemble will often be more accurate than any of single prediction. Ideal learning ensemble: individually accurate classifiers with high level of disagree. Applications: bagging, boosting. Reasons for using small training data, large hypothesis space. Many possible classifiers remain with equal accurate. Compensate for non-optimal search (several local solutions) [24-25].

C. Reasons for using ensemble based system

1) Statistical Reasons:
   A set of classifiers with similar training performances may have different generalization performances. In fact, even classifiers with similar generalization performances may perform differently in the field
2) Large volume of data
3) Too little data
4) Divide and conquer
5) Data fusion

D. Information Fusion (Ensemble) Techniques comparison

TABLE 1: Comparison of Ensemble Methods

	Bagging	Boosting	Stacking
Sampling Method	Random, with replacement	Least squares (proportionate)	Round-robin (cross-validation)
Splitting of Data	Length-wise	Width-wise	Length-wise
Guaranteed improvement of weak classifiers?	No	Yes	No
Hierarchical?	No, but can be extended	No	Yes
Training	N/A	Multiple passes	Single bottom-up pass
Wrapper or mixture?	Wrapper	Wrapper	Both

Sampling Method	Length-wise	Least squares (proportionate)	Round-robin (cross-validation)
Weak learner	Same	Same	Different
Training data become available	All at once	All at once	All at once
Can learn new classes?	No	No	No
Run weak learner on	Bootstrap replicates of the training set	Same, but differently weighted set	Same, but differently weighted set
Emphasizing "difficult" examples?	No	Yes	Yes
Is ensemble of ensemble?	No	No	Yes
Combination rule	Simple majority voting	Weighted majority voting	Meta learner

A. *Two Key Components of an Ensemble System*

All ensemble systems consist of two key components.

1) A strategy is needed to build an ensemble that is as diverse as possible some of the more popular ones, such as bagging, boosting, AdaBoost, stacked generalization, and mixture of experts [25].
2) A strategy is needed to combine the outputs of individual classifiers that make up the ensemble in such a way that the correct decisions are amplified, and incorrect ones are cancelled out. Several choices are available for this purpose as well [25].

III. Parameters to be Considered

A. *The role of models in predictive validation*

Models are described by equations based on statistics. They only examine linear relationships. They contain many assumptions about the data. So, Data mining models (algorithms) must be smart data wise and dumb business wise. Business in general is a combination of people, processes and products. There are various models available to find patterns in the data like predict (category membership or numerical variables) the values. We have to combine technology with the process(s), but obvious question is how? The steps for constructing and using the models are: Build, Train, Validate (check on fresh data) and predict.

B. *Evaluating Classifier Accuracy*

1) Holdout
2) Random subsampling
3) Cross validation (K-fold Cross Validation, Leave-one-out Cross Validation)
4) Bootstrap

C. *What can be done?*

Observe: the Individual accuracies of the ensemble members, the ensemble accuracy, the ensemble diversity, advance classification of multi-class imbalanced data. To develop cost-sensitive algorithm. To improve the identification performance on the important classes.To balance classification performance among several classes. What is the reasoning behind the prediction? How is the model working out the predictions? Narrow down the number of fields, can it work? Use rules before and after the model construction to interpret the results.

Model choice and validation have central role in analysis and predictive modeling. It is natural to use predictive accuracy as the decisive criterion in the final choice of the predictive model. Any assessment of predictive accuracy, assumes a data mechanism (like sampling) that relates model predictions to the population that is the target of the predictions.

IV. Conclusion

This survey has explored published work by the research community and future work to be done in the area. After identifying the limitations in methods and techniques of information fusion, this paper shows that this field can benefit from other related fields and still there is a lot of room to work for the future. This paper discussed a formal introduction of a fusion system to better understand the fundamentals of information fusion.

Ensemble learning needs individual learning. Individual learning has to interact in a dynamic environment in order to contribute to ensemble learning. Ensemble learning is goal-directed. Ensemble learning is driven by expectation failure. One has to find out the elements of dependence, independence and inter-dependence and ability to apply the skills in the context that requires them. Perspectives would differ according to particular analysis and experience. It perceives the situation in the framework of acquaintance of theoretical concepts; the past experience, current preoccupation and also ensuring that all significant facts are covered and those irrelevant points are not pursued. It serves to inform and canalize the discussion from one step in the analysis to another. The distinct elements in the situation are identified and the salient ones segregated. This paper will serve like a spiral of knowledge creation.

We explore the relationship between diversity measures and ensemble performance. We further investigate the issue of diversity within the generated ensembles of classifiers which emerges as an important concept in classifier fusion. Rather, some type of trade off seems to be necessary between participant classifiers' accuracy and ensemble diversity in order to achieve maximum recognition gains [25-26].

V. Research Issues & Future Work

Theoretical concepts and research findings could be examined for their application to the practical issues on hand. The inter-relationships between the elements and their implications are thought out. Diverse alternatives are evaluated to deal with the elements in the situation. Decisions if any are evaluated. Brings out the actual learning by collaboration, team-work, presentation, discussion and analysis.

As new data arrives, the ensembles can be updated without reviewing any past data. There are several traditional ensemble learning issues that remain in online ensemble learning framework such as the number and types of base models to use, the combining method to use, and how to maintain diversity among the base models [14]. Other issues are:

1) Ensemble learning framework such as the number and types of base models to use and which is combining method to use? And how to maintain diversity among the base models? [17].
2) How Statistical methods can be better tied with data mining. e.g. How to apply ensemble methods to other types of data such as time series data?
3) Various methods for observing the behavior of the algorithms like using error estimation, the bias and variance decomposition, average (tree) sizes, and graphs showing progress over trials, robustness to noise, attempted to drop instances with very high weight [14-15].
4) Boosting is not consistence in performance with noisy data. Are there methods to make boosting algorithms more robust when the dataset is noisy?
5) Boosting stops if one of the classifiers achieves zero training set error. That classifier gets infinite voting power and is effectively the single classifier. Are there better methods for handling this extreme situation?
6) Are there ways to make boosting comprehensible for general models?
7) Are there situations where one of the classifiers that were learned from a sample with a skewed distribution performs well by itself on the unskewed test set?
8) Error in training (learning) can lead to bias. Can the bias be made more explicit?
9) Can some better method to combine classifiers be devised?
10) How can boosting be applied to other algorithms, such as k-nearest-neighbors?
11) Boosting is sequential. A cascade of classifiers can be used. Can some efficient parallel implementations be devised?
12) How aggregating classifiers can be generated? How to aggregate?
13) Diversity can be achieved with input parameter setting. How to reweight to improve diversity?
14) How boosting can be modified for regression?
15) How AdaBoost.M1 can be modified to handle categorical output?
16) Resampling is done foe diversity. Which resampling method is used for classifier design?
17) Does feature selection resulting in relatively simple classifier? How to apply it for feature (attribute) subset selection as a preprocessing? How to filter out irrelevant features? How an AdaBoost can be used for improving the classification and efficient feature selection?
18) Training error drops exponentially fast as we add more weak classifiers, how many such weak classifiers should be there?
19) How the boosting algorithm can be extended for the prediction of continuous values?
20) How to deal with multi-class problems? Can it be possible explicitly to change the class labels by altering the attribute(s) for classification?
21) How do we qualitatively understand relationship between predictors and output?
22) Classifier is based on likelihood ratio test. The likelihoods are obtained from probability tables constructed while training. How to construct probability tables?
23) GAs are good at optimization. How to perform Genetic search for classifiers that are accurate yet different? How to perform Genetic search for good diverse classifiers?
24) How to create diverse classifiers by using different parameters and using different training sets?
25) How to decompose a problem into subtasks?
26) How to create Mixture of experts? How each expert takes care of a certain input space?
27) How to construct Hierarchical neural networks? How to route the cases to pre-defined expert networks?
28) How significance tests and ROC curves are useful for model selection?
29) Issues such as accuracy, training time, robustness, interpretability, and scalability must be considered and can involve trade-offs, further complicating the quest for an overall superior method.
30) How to find a relationship between initial performance, number of cases, and number of classifiers within an ensemble?
31) Removing outliers is an important preprocessing step to remove random error. How to analyze effect of prior outlier removal on boosting?
32) If finally we reach a simpler classifier, is there anyway to compress it? Or can we bypass boosting but reach a simple classifier?
33) The performance is dependant on the data. How to find the relationship between these two? How to train the specific classifier with a fixed training set?
34) Diversity can also be achieved with changing data sets at different timeb(iterations).How to train the specific classifier using a different training set at each epoch?
35) How to train the specific classifier by directly weighting the cost function of the t-th neural network?
36) How do we optimally decide the number of classifiers?# (number) of classifiers in ensemble?
37) How to calculate cost and time efficient “Accuracy” as evaluation measure?
38) Weight adapts as iteration progress. A sample with misclassification has higher weight. Can it be possible to drop instances with very high weight? But the experiments did not show this to be a successful approach.
39) Attempt to stack a Naive-Bayes on top of the base classifiers built by AdaBoost and Bagging. Can some better method to combine classifiers be devised?
40) Changes are inevitable. If input data changes, if

parameter(s) changes then how to find classifier ensembles for changing environments? How to find classifier ensembles for change detection?

ACKNOWLEDGMENT

We are thankful to The Omnipotent GOD for making us able to do something. We express our gratitude to the management of CHARUSAT; Shri Charotar Moti Sattavis Kelavani Mandal, for providing us research opportunities and their wholehearted support for such activities. Finally, our acknowledgement can not end without thanking to the authors whose research papers helped us in making this research.

REFERENCES

[1] E. Bosse, A. Guitouni and P. Valin, Defence Research & Development Canada Valcartier, Decision Support Systems section An Essay to Characterize Information Fusion Systems.

[2] Tuve Lofstrom, Rikard Konig, Ulf Johansson, Tom Ziemke School of Humanities and Informatics, University of Skovde, Sweden. JDL: Joint Directors of Laboratories, a US DoD government committee overseeing US defence technology R&D; the Data Fusion Group of the JDL created the original JDL Data Fusion Model. Benefits of relating the Retail Domain and Information Fusion

[3] Merz, C.J. (1999). A Principal Component Approach to Combining Regression Estimates. Machine Learning, 36, 9-32.

[4] Zheng, Z. & Webb, G. (1998). Stochastic Attribute Selection Committees. In Proceedings of the 11th Australian Joint Conference on Artificial Intelligence (AI'98), pp. 321-332.

[5] David Opitz and Richard Maclin, Journal of Artificial Intelligence Research 11 (1999) 169-198. Popular Ensemble Methods: An Empirical Study.

[6] Yoav Freund Robert E. Schapire , AT&T Labs _ Research, Shannon Laboratory. Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999. A Short Introduction to Boosting.

[7] Dietterich, T. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization, Machine Learning 40, 139-158.

[8] Oza, N.C. (2001). Online Ensemble Learning, Ph.D. thesis, University of California, Berkeley.

[9] Oza, N.C. & Russell, S. (2001). Experimental Comparisons of Online and Batch Versions

[10] Bagging and Boosting, The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California.

[11] Oza, N.C. & Tumer, K. (2001). Input Decimation Ensembles: Decorrelation through Dimensionality Reduction. In Second International Workshop on Multiple Classifier Systems. Springer-Verlag. Berlin.

[12] Ratsch, G., Onoda, T., & Muller, K.R. (2001). Soft Margins for AdaBoost. Machine Learning, 42, 287-320.

[13] Tumer, K. & Oza, N.C. (2003). Input Decimated Ensembles, Pattern Analysis and Applications, 6(1):65-77.

[14] Oza, N.C. (2003). Boosting with Averaged Weight Vectors. In T. Windeatt and F. Roli (Eds.), Proceedings of the Fourth International Workshop on Multiple Classifier Systems, 15-24, Berlin.

[15] Oza, N.C. (2004). AveBoost2: Boosting with Noisy Data. In F. Roli , J. Kittler, and T. Windeatt (Eds.), Proceedings of the Fifth International Workshop on Multiple Classifier Systems, 31-40, Berlin.

[16] S. B. Kotsiantis, P. E. Pintelas, International Jornal Of Computational Intelligence Volume 1 Number 2004 ISSN:1304-2386. Combining Bagging and Boosting.

[17] Nikunj C. Oza, Ph.D., NASA Ames Research Center, USA, 2005. Ensemble Data Mining Methods.

[18] Robert E. Schapire, Marie Rochery, Mazin Rahim, Senior Member, IEEE, and Narendra Gupta. IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 13, NO. 2, MARCH 2005. Boosting with Prior Knowledge for Call Classification.

[19] Saharon Rosset, IBM T.J. Watson Research Center KDD'05, August 21–24, 2005, Chicago, Illinois, USA. Robust Boosting and its Relation to Bagging.

[20] Amitava Karmaker, Stephen Kwek, Department of Computer Science, University of Texas at San Antonio, Proceedings of the Fifth International Conference on Hybrid Intelligent Systems (HIS'05) 2005 IEEE. A Boosting Approach to remove Class Label Noise.

[21] Abdul Aziz Gill, George D. Smith.IEEE-ICET 2006 2nd International Conference on Emerging Technologies. Peshawar, Pakistan, 13-14 November 2006. The Compound Effect of Boosting and Stratified Sampling on Decision Tree Accuracy.

[22] Nikunj C. Oza, NASA Ames Research Center.Elsevier 15 July 2007. Classifier Ensembles: Select Real-World Applications.

[23] Nima Hatami and Reza Ebrahimpour. IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.1, January 2007. Combining Multiple Classifiers: Diversify with Boosting and Combining by Stacking.

[24] Robi Polikar, IEEE CIRCUITS AND SYSTEMS MAGAZINE. Ensemble based systems in decision making.

[25] P´adraig Cunningham ,Technical Report UCD-CSI-2007-5. Ensemble Techniques.

[26] ABM Shawkat Ali and Tony Dobele, School of Computing Sciences, Central Queensland University. 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007). A Novel Classifier Selection Approach for Adaptive Boosting Algorithms.