2007 IEEE International Conference on Control and Automation
Guangzhou, CHINA - May 30 to June 1, 2007

FrC2-5

Accuracy of Classifier Combining Based on Majority Voting

Peng Hong*, Lin Chengde, Luo Linkai, Zhou Qifeng
Department of Automation
Xiamen University
Xiamen, P.R.China, 361005
*Corresponding Author: penghong@xmu.edu.cn

***Abstract*—In this paper, we formulate the accuracy of classifier combining which is based on majority voting, there are only two parameter involved, one is the average accuracy of individual classifiers, the other we call it Lapsed Accuracy (*LA*) is related with the efficiency of classifier combining, and we discuss the theoretical bounds of majority voting via the formula.**

Keywords*- *classifier combination, majority voting, combination accuracy, Lapsed Accuracy, theoretical bounds, classifier design and evaluation

I. INTRODUCTION

Majority voting has attracted much attention in the literature on classifier combination scheme for its simplicity and its good performance on real data processing. The impressive performance of majority voting has been demonstrated in various applications such as handwriting recognition [1], [2], and person authentication [3], et al. A simple analytical justification for majority voting is given by the well-known Condorcet's Theorem [4]. Under the assumption of independent classifiers, if the individual classifier error rate $e < 0.5$ for odd number of classifiers (voters) M, the correct decision rate increases as M increases. Here, for simplicity, it is assumed that all classifiers have the same error rate. Moreover, the problem of theoretical analysis of majority voting has been investigated by a number of papers, see, for example, [3], [5], [6], [7], [8]. Specifically, Lam and Suen [5] give an analysis of majority voting under the assumption that the classifiers are independent. In [3], Kittler et al develop a theoretical framework for combining classifiers which use distinct pattern representations. They show that majority voting is a special case of the sum rule. The sum rule is developed under two key assumptions: (1) statistical independence; and (2) a posteriori probabilities computed by the respective classifiers do not deviate significantly from the prior probabilities. They also show that the sum rule is more resilient to estimation errors compared to the other combination strategies discussed in their paper. More recently, Kuncheva et al [8] address the performance of majority voting empirically. They provide some insights based on pair wise dependence statistics since majority vote with dependent classifiers can offer improvement over independent classifiers and over the individual accuracies.

In the research of classifier combination, it's well known that base classifiers to be combined should be diverse (negative dependent, independent, or complementary) [8-10]. We can imagine it meaningless to combine several identical classifiers. Sometimes, altering diversity among classifiers can be the key to the success (or failure). For example, classifier ensemble methods such as bagging, boosting, and arcing [11] are all based on the idea of promoting diversity [12]. Diversity measures aim to describe the diversity among classifiers in multiple classifier systems. However, up to now, the researches on diversity measures are still very limited. The existing measures are always not able to show strong correlation between diversity and classifier combining accuracy [6]. In this paper, we propose a combination efficiency measure based on majority voting, and educe the relationship of combination accuracy with average individual accuracy and the measure. In [13], Narasimhamurthy formulates the problem of determining the theoretical bounds of majority voting performance for a binary classification problem as a Linear Program (*LP*). We will describe this problem more clearly by the combination efficiency measure proposed in this paper.

II. PROBLEM FORMULATION

In this section, we propose a combination efficiency measure named Lapsed Accuracy (*LA*) and derive a formula of combination accuracy with average individual accuracy and the Lapsed Accuracy.

A. *Notation and Representation*

We use the following notation for the rest of the paper. Let $\mathcal{C} = \{c_1, c_2, \cdots, c_M\}$ denote a set of classifiers, $\mathcal{X} = \{x_1, x_2, \cdots, x_L\}$ is a set of training samples, and $\Omega = \{\omega_1, \omega_2, \cdots, \omega_K\}$ represents a set of class labels. Let $x \in \mathbb{R}^n$ be a vector with n features to be labeled in Ω. We take the oracle output of classifier combining as follow

$$c_m(x) = \begin{cases} 1 & \text{if } x \text{ is recognized correctly by } c_m \\ 0 & \text{otherwise} \end{cases}$$

Denote a random variable x as an input drawn randomly from the distribution of the problem.

Define a stochastic variable as follow

$$\xi = \frac{\sum_{m=1}^{M} c_m(x)}{M} \in \{0, \frac{1}{M}, \frac{2}{M}, \cdots, \frac{M-1}{M}, 1\}$$

Denote the probability distribution of ξ with σ_i

$$\sigma_i = Prob\{\xi = \frac{i}{M}\}$$

Denote the probability with $\boldsymbol{p_m}$ that the classifier $\boldsymbol{c_m}$ correctly classify a random input $\boldsymbol{x}$

$$p_m = Prob\{x, c_m(x) = 1\}$$

Denote the average individual accuracy as $\boldsymbol{P}$, we have

$$P = \frac{1}{M}\sum_{m=1}^{M} p_m$$

Finally, we denote the probability with P_{maj} that the ensemble classify correctly via majority votes .

B. Lapsed Accuracy (LA)

Proposition 1.

1.1 $$\sum_{i=0}^{M}\sigma_i = \sum_{i=0}^{M} Prob\{\xi = \frac{i}{M}\} = 1$$

1.2 $$P = E\xi = \frac{1}{M}\sum_{i=0}^{M} i \bullet \sigma_i$$

1.3 $$P_{maj} = \sum_{i=N+1}^{M}\sigma_i = \sum_{i=N+1}^{M} Prob\{\xi = \frac{i}{M}\}$$

Proof.

Result 1.1 and 1.3 is obvious, result 1.2 is proved as follow

$$P = \frac{1}{M}\sum_{m=1}^{M} p_m = \frac{1}{M}\sum_{m=1}^{M} Prob\{x, c_m(x) = 1\}$$

$$= \frac{1}{M}\sum_{m=1}^{M} Ec_m(x) = E\frac{\sum_{m=1}^{M} c_m(x)}{M}$$

$$= E\xi = \sum_{i=0}^{M}\frac{i}{M} Prob\{\xi = \frac{i}{M}\} = \frac{1}{M}\sum_{i=0}^{M} i \bullet \sigma_i$$

For discussion convenience, we assume odd classifier number $\boldsymbol{M = 2N + 1}$, now, we introduce the Lapsed Accuracy denoted as $\boldsymbol{LA}$.

Definition 1.

$$LA \triangleq \frac{1}{N}\sum_{i=1}^{N} i(\sigma_i + \sigma_{i+N+1})$$

In the definition, $\boldsymbol{LA}$ figures the percentage of lapsed correct output in the combination, by majority voting, for a sample $\boldsymbol{x}$, if number of individual classifier which judge it correct is less than half, then those take no effect, in another hand, if number of individual classifier judging correctly is more than half, then we also consider the exceed part is redundant.

Proposition 2.

$$0 \leqq \boldsymbol{LA} \leqq 1$$

Proof.

$$LA = \frac{1}{N}\sum_{i=1}^{N} i(\sigma_i + \sigma_{i+N+1}) = \frac{1}{N}\sum_{i=0}^{N} i(\sigma_i + \sigma_{i+N+1})$$

$$\leq \frac{1}{N}\sum_{i=0}^{N} N(\sigma_i + \sigma_{i+N+1}) = \sum_{i=0}^{N}(\sigma_i + \sigma_{i+N+1}) = 1$$

$$LA = \frac{1}{N}\sum_{i=1}^{N} i(\sigma_i + \sigma_{i+N+1}) \geq \frac{1}{N}\sum_{i=1}^{N} 0(\sigma_i + \sigma_{i+N+1}) = 0$$

Proposition 3.

$$LA = \frac{2N+1}{N}P - \frac{N+1}{N}P_{maj}$$

Proof.

$$LA = \frac{1}{N}\sum_{i=1}^{N} i(\sigma_i + \sigma_{i+N+1}) = \frac{1}{N}(\sum_{i=0}^{N} i \bullet \sigma_i + \sum_{i=0}^{N} i \bullet \sigma_{i+N+1})$$

$$= \frac{1}{N}(\sum_{i=0}^{N} i \bullet \sigma_i + \sum_{i=N+1}^{2N+1}(i - N - 1)\sigma_i)$$

$$= \frac{1}{N}(\sum_{i=0}^{2N+1} i \bullet \sigma_i - (N+1)\sum_{i=N+1}^{2N+1}\sigma_i) = \frac{2N+1}{N}P - \frac{N+1}{N}P_{maj}$$

Proposition 3 illustrate the relationship of combination accuracy P_{maj} with average individual accuracy $\boldsymbol{P}$ and Lapsed Accuracy $\boldsymbol{LA}$. we also can write it as follow

$$P_{maj} = \frac{2N+1}{N+1}P - \frac{N}{N+1}LA \qquad (2.1)$$

Generally, $\boldsymbol{P}$ represent the quality of individual classifier, and $\boldsymbol{LA}$ reflect the efficiency of classifier combining, the improvement of accuracy of classifier combining, under the restricted of accuracy of individual classifier, we should fixate on the combination efficiency, decrease the lapse of individual accuracy. utilizing the formula, we can design and evaluate a classifier combining.

C. Empirical Approach

For design and evaluation of a classifier combining, we must get a estimation of Lapsed Accuracy $\boldsymbol{LA}$. we also need estimate p_m, σ_i when the $\boldsymbol{M}$ classifiers in ensemble $\boldsymbol{C}$ run on the sample set $\mathcal{X}$. assume $x_l \in \mathcal{X}$, $\forall l$ are ***i.i.d*** random variables.

Take

$$\hat{p}_m = \frac{1}{L}\sum_{l=1}^{L} c_m(x_l) \quad x_l \in \mathcal{X}$$

$$c^i(l) = \begin{cases} 1 & if \ \sum_{m=1}^{M} c_m(x_l) \ equals \ i \quad x_l \in \mathcal{X} \\ 0 & otherwise \end{cases}$$

$$\hat{\sigma}_i = \frac{\sum_{l=1}^{L} c^i(l)}{L}$$

In the empirical approach, $\hat{p}_m$ represent the empirical accuracy of classifier m, and $\hat{\sigma}_i$ show the percentage of sample on which total i classifiers output 1(correct). Obviously, $E\hat{p}_m = p_m$ and $E\hat{\sigma}_i = \sigma_i$.

Then we can estimation of $\hat{P}$, $\hat{P}_{maj}$ and $\hat{L}$ as follow: $\hat{P} = \frac{1}{M}\sum_{m=1}^{M}\hat{p}_m$, $\hat{P}_{maj} = \frac{1}{L}\sum_{l=1}^{L}\sum_{i=N+1}^{M} c^i(l)$ and $\hat{L} = \frac{1}{N}\sum_{i=1}^{N} i(\hat{\sigma}_i + \hat{\sigma}_{i+N+1})$, the $\hat{P}$, $\hat{P}_{maj}$ and $\hat{L}$ represent average accuracy, combination accuracy and Lapsed Accuracy of classifier combining on the sample set $\mathcal{X}$. Similar to proposition 1, proposition 2 and 3, there is proposition 4 and the proof is also similar.

Proposition 4.

4.1 $\sum_{i=0}^{M}\hat{\sigma}_i = 1$

4.2 $\hat{P} = \frac{1}{M}\sum_{i=0}^{M} i\bullet\hat{\sigma}_i$

4.3 $\hat{P}_{maj} = \sum_{i=N+1}^{M}\hat{\sigma}_i$

4.4 $0 \le \hat{L} \le 1$

4.5 $\hat{L} = \frac{2N+1}{N}\hat{P} - \frac{N+1}{N}\hat{P}_{maj}$

Based on proposition 4.5, we can get the empirical accuracy formula of classifier combining, it's consistent to the formula (2.1)

$$\hat{P}_{maj} = \frac{2N+1}{N+1}\hat{P} - \frac{N}{N+1}\hat{L} \qquad (2.2)$$

Estimation of LA has a more convenient approach, denote the number of classifiers with $\sigma(l)$ in ensemble C that classify correctly the sample $x_l \in \mathcal{X}$, where

$$\sigma(l) = \sum_{m=1}^{M} c_m(x_l) \qquad x_l \in \mathcal{X}$$

Denote

$$\bar{\sigma}(l) \triangleq \begin{cases} \sigma(l) & \sigma(l) \le N \\ \sigma(l) - (N+1) & \sigma(l) > N \end{cases}$$

Proposition 5.

$$\hat{L} = \frac{1}{N}\sum_{i=1}^{N} i(\hat{\sigma}_i + \hat{\sigma}_{i+N+1}) = \frac{1}{N\times L}\sum_{l=1}^{L}\bar{\sigma}(l)$$

Proof.

$$\begin{aligned}\sum_{l=1}^{L}\bar{\sigma}(l) &= \sum_{l=1}^{L}\left(\sum_{i=0}^{N} i\bullet c^i(l) + \sum_{i=N+1}^{2N+1}(i-(N+1))c^i(l)\right) \\ &= \sum_{i=0}^{N} i\sum_{l=1}^{L} c^i(l) + \sum_{i=N+1}^{2N+1}((i-(N+1))\sum_{l=1}^{L} c^i(l)) \\ &= \sum_{i=0}^{N} i\bullet L\bullet\hat{\sigma}_i + \sum_{i=0}^{N} i\bullet L\bullet\hat{\sigma}_{i+N+1} = L\sum_{i=0}^{N} i(\hat{\sigma}_i + \hat{\sigma}_{i+N+1}) \\ \Rightarrow \hat{L} &= \frac{1}{N}\sum_{i=0}^{N} i(\hat{\sigma}_i + \hat{\sigma}_{i+N+1}) = \frac{1}{N\times L}\sum_{l=1}^{L}\bar{\sigma}(l)\end{aligned}$$

Now we describe the validity of the empirical method, the purpose of ensemble is improve P_{maj} via majority voting, our method propose a way to increase the estimation of P_{maj}. Denote ζ as the output of ensemble

$$\zeta = \begin{cases} 1 & \textit{if ensemble classify x correctly} \\ 0 & \textit{otherwise} \end{cases}$$

We can educe the equation of expectation and variance of ζ with P_{maj}

$$E\zeta = P_{maj}\ ,\ D\zeta = (1 - P_{maj})P_{maj}$$

Then the efficient estimation of P_{maj} is illustrated as follow:

$$E\hat{\sigma}_i = \frac{1}{L}\sum_{l=1}^{L} Ec^i(l) = \frac{1}{L}\sum_{l=1}^{L} P\{\xi(x_l) = \frac{i}{M}\} = \sigma_i$$

$$D\hat{\sigma}_i = D(\frac{1}{L}\sum_{l=1}^{L} c^i(l)) = \frac{1}{L}Dc^i(l) = \frac{1}{L}(1-\sigma_i)\sigma_i$$

$$\Rightarrow E\hat{P}_{maj} = E\sum_{i=N+1}^{M}\hat{\sigma}_i = \sum_{i=N+1}^{M}\sigma_i = P_{maj} = E\zeta$$

$$D\hat{P}_{maj} = D(\sum_{i=N+1}^{M}\hat{\sigma}_i) = D(\frac{1}{L}\sum_{l=1}^{L}\sum_{i=N+1}^{M} c^i(l)) = \frac{1}{L}D\zeta$$

III. Theoretical Bounds

In this section, we derive the theoretical upper and lower bounds of majority voting performance for a classification problem given a set of classifiers whose accuracies are known. Based on formula (2.2)

$$\hat{P}_{maj} = \frac{2N+1}{N+1}\hat{P} - \frac{N}{N+1}\hat{L}$$

Take the average accuracy $\hat{P}$ as a invariable , we fixate on empirical Lapsed Accuracy $\hat{L}$, as we know, $\hat{L}$ vary from 0 to 1, consider the upper bound of $\hat{P}_{maj}$, we should take the possible minimum of $\hat{L}$, or the possible maximum of $\hat{L}$ when we consider the lower bound. Based on $\hat{L} = \frac{1}{N}\sum_{i=1}^{N} i(\hat{\sigma}_i + \hat{\sigma}_{i+N+1})$, $\sum_{i=0}^{M}\hat{\sigma}_i = 1$ and $\hat{\sigma}_i \ge 0$, to minimum $\hat{L}$, we should let $\hat{\sigma}_i$ take greater value on $i = 0$ and

N +1 as possible, and to maximum $\hat{L}$, we should let $\hat{\sigma}_i$ take greater value on ***i = N*** and ***2N +1*** as possible.

For analysis convenience, we let $\hat{\sigma}_i$ take nonzero value on ***i*** only two point which maybe 0, ***N***, ***N +1*** and ***2N +1***, make sure that $\sum_{i=0}^{M}\hat{\sigma}_i = 1$, $\hat{\sigma}_i \geq 0$ and $\hat{P} = \frac{1}{M}\sum_{i=0}^{M} i\bullet\hat{\sigma}_i$. The discussion is separated into three parts which is correspond to the three cases of $\hat{P}$. In each case, we can choose the proper $\hat{\sigma}_i$ that $\hat{P}_{maj}$ is taken to two kinds of extreme situation: the upper bound $\hat{P}_{maj}^{\max}$ and the low bound $\hat{P}_{maj}^{\min}$.

Case 1: $\hat{P} \leq \frac{N}{2N+1}$

Upper Bound

$$\hat{\sigma}_i = \begin{cases} \frac{2N+1}{N+1}\hat{P} & i = N+1 \\ 1 - \frac{2N+1}{N+1}\hat{P} & i = 0 \\ 0 & otherwise \end{cases}$$

$\hat{L}$ gets the minimum value 0, so the corresponding $\hat{P}_{maj}$ reaches the maximum

$$\hat{L} = 0 \Rightarrow \hat{P}_{maj}^{\max} = \frac{2N+1}{N+1}\hat{P}$$

Lower Bound

$$\hat{\sigma}_i = \begin{cases} \frac{2N+1}{N}\hat{P} & i = N \\ 1 - \frac{2N+1}{N}\hat{P} & i = 0 \\ 0 & otherwise \end{cases}$$

In this situation, $\hat{P}_{maj}$ reaches its minimum value 0, it's the certain lower bound

$$\hat{L} = \frac{2N+1}{N}\hat{P} \Rightarrow \hat{P}_{maj}^{\min} = 0$$

Case 2: $\hat{P} \geq \frac{N+1}{2N+1}$

Same as case 1, $\hat{\sigma}_i$ can be chosen carefully as follow:

Upper Bound

$$\hat{\sigma}_i = \begin{cases} \frac{2N+1}{N}(1-\hat{P}) & i = N+1 \\ \frac{2N+1}{N}\hat{P} - \frac{N+1}{N} & i = 2N+1 \\ 0 & otherwise \end{cases}$$

$\hat{P}_{maj}$ reach its maximum value 1, it's the certain upper bound

$$\hat{L} = \frac{2N+1}{N}\hat{P} - \frac{N+1}{N} \Rightarrow \hat{P}_{maj}^{\max} = 100\%$$

Lower Bound

$$\hat{\sigma}_i = \begin{cases} \frac{2N+1}{N+1}(1-\hat{P}) & i = N \\ \frac{2N+1}{N+1}\hat{P} - \frac{N}{N+1} & i = 2N+1 \\ 0 & otherwise \end{cases}$$

In this situation, $\hat{L}$ take the maximum value 1, so corresponding $\hat{P}_{maj}$ is the lower bound

$$\hat{L} = 1 \Rightarrow \hat{P}_{maj}^{\min} = \frac{2N+1}{N+1}\hat{P} - \frac{N}{N+1}$$

Case 3: $\frac{N}{2N+1} < \hat{P} < \frac{N+1}{2N+1}$

Identical, in this case we can choose the proper $\hat{\sigma}_i$:

Upper Bound

$$\hat{\sigma}_i = \begin{cases} \frac{2N+1}{N+1}\hat{P} & i = N+1 \\ 1 - \frac{2N+1}{N+1}\hat{P} & i = 0 \\ 0 & otherwise \end{cases}$$

$\hat{L}$ take the minimum value 0, so corresponding $\hat{P}_{maj}$ is maximum, it's the upper bound

$$\hat{L} = 0 \Rightarrow \hat{P}_{maj}^{\max} = \frac{2N+1}{N+1}\hat{P}$$

Lower Bound

$$\hat{\sigma}_i = \begin{cases} \frac{2N+1}{N+1}(1-\hat{P}) & i = N \\ \frac{2N+1}{N+1}\hat{P} - \frac{N}{N+1} & i = 2N+1 \\ 0 & otherwise \end{cases}$$

$\hat{L}$ take the maximum value 1, so corresponding $\hat{P}_{maj}$ reaches the minimum, we can compute the lower bound as follow

$$\hat{L} = 1 \Rightarrow \hat{P}_{maj}^{\min} = \frac{2N+1}{N+1}\hat{P} - \frac{N}{N+1}$$

Based the discussion above, we can find the twofold function of theoretical upper and lower bounds of majority voting performance under $\hat{P}$, it can be illustrated in following two formulae

theoretical upper bound

$$\hat{P}_{maj}^{\max} = \begin{cases} \frac{2N+1}{N+1}\hat{P} & \hat{P} \le \frac{N+1}{2N+1} \\ 100\% & \hat{P} > \frac{N+1}{2N+1} \end{cases} \tag{3.1}$$

theoretical lower bound

$$\hat{P}_{maj}^{\min} = \begin{cases} 0 & \hat{P} \le \frac{N}{2N+1} \\ \frac{2N+1}{N+1}\hat{P} - \frac{N}{N+1} & \hat{P} > \frac{N}{2N+1} \end{cases} \tag{3.2}$$

Denote $\hat{P}_{maj}^{\min}(\hat{P})$ and $\hat{P}_{maj}^{\max}(\hat{P})$ as $\hat{P}_{maj}^{\min}$ and $\hat{P}_{maj}^{\max}$ under the empirical average accuracy $\hat{P}$, proposition 6 can be drawn by formula (3.1) and (3.2).

Proposition 6.

$$\hat{P}_{maj}^{\min}(\hat{P}) + \hat{P}_{maj}^{\max}(1-\hat{P}) = 1$$
$$\hat{P}_{maj}^{\min}(1-\hat{P}) + \hat{P}_{maj}^{\max}(\hat{P}) = 1$$

IV. Discussion

In our paper we propose a combination efficiency measure based on majority voting, comparing with diversity measure, Lapsed Accuracy is a direct route to scale the combination accuracy, and it is not only including the inter-individual diversity but also integrating the combination scheme, so we can educe the formula of combination accuracy based on Majority voting.

Lapsed Accuracy contains potential diversity, the theoretical bound which is discussed in section 3 show that we can get higher combination accuracy if the proportion of N+1 machines output correctly (that means N machines output wrong) is greater.

In recent literatures on diversity measure, there almost all assume that the individual classifier have the identical accuracy, in our discussion, this assumption is not necessary, that show us a broad way to design classifier combining.

References

[1] L. Xu, A. Krzyzak, and C. Suen, "Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 22, pp. 418-435, 1992.

[2] T.K. Ho, J.J. Hull, and S.N. Srihari, "Decision Combination in Multiple Classifier Systems," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66-75, Jan. 1994.

[3] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On Combining Classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 3, pp. 226-239, Mar. 1998.

[4] P.J. Boland, "Majority Systems and the Condorcet Jury Theorem," *Statistician*, vol. 38, pp. 181-189, 1989.

[5] L. Lam and C. Suen, "Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 27, no. 5, pp. 553-568, 1997.

[6] L.I. Kuncheva and C.J. Whitaker, "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy," *Machine Learning,* vol. 51, pp. 181-207, 2003.

[7] L.I. Kuncheva, "A Theoretical Study on Six Classifier Fusion Strategies," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 2, pp. 281-286, Feb. 2002.

[8] L.I. Kuncheva, C.J. Whitaker, C. Shipp, and R. Duin, "Limits on the Majority Vote Accuracy in Classifier Fusion," *Pattern Analysis and Applications,* vol. 6, no. 1, pp. 22-31, 2003.

[9] L. Lam, "Classifier Combinations: Implementations and Theoretical Issues", *The 1st Int'l Workshop on Multiple Classifier Systems (MCS2000), Cagliari, Italy,* pp. 78-86, Springer-Verlag, 2000.

[10] L.I. Kuncheva and C.J. Whitaker, C.A. Shipp, "Is Independence Good for Combining Classifiers?" *15th Int'l Conf. Pattern Recognition,* pp. 168-171, 2000.

[11] T. Dietterich, "An Experimental Comparison of Three Methods for Construction Ensembles of Decision Trees: Bagging, boosting, and randomization", *Machine Learning,* 40(2), pp.139-158, 2000.

[12] P. Melvill, R.J Mooney, "Creating Diversity in Ensembles Using Artificial Data", *Information Fusion (6), Special Issue on Diversity in Multi-classifier Systems,* pp.99-111, 2005.

[13] A. Narasimhamurthy, "Theoretical Bounds of Majority Voting Performance for a Binary Classification Problem", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 27, no. 12, pp. 1988-1995, Dec. 2005.