A Theoretical Analysis of the Application of Majority Voting to Pattern Recognition

Louisa Lam and Ching Y. Suen

Centre for Pattern Recognition and Machine Intelligence
Concordia University, Montréal, Canada

Abstract

Recently, it has been demonstrated that combining the decisions of several classifiers can lead to improved recognition results. The combination can be implemented using a variety of strategies, among which majority vote is by far the simplest, yet it has been found to be just as effective as more complicated schemes.

However, all the results reported thus far on combinations of classifiers have been experimental in nature. The intention of this research is to analyze the foundations of the majority vote method in order to gain a deeper understanding and new results about its mode of operation.

1. Introduction

In OCR, experimental results have shown that majority vote is an effective means of combining classifier decisions ([2], [3], [4]). This work is concerned with understanding *how* and *why* the combination of expert opinions by majority vote can produce improved recognition results, and the assumptions under which this can be expected to happen. To achieve this purpose, we will examine the topic starting from its logical foundations of the classical voting problem. In this setting, the binomial distribution has been used to determine the probabilities of consensus, but only for an odd number of voters, and only in the context of binary choice. We will extend these results to even numbers of voters in a multiple choice situation, and compare their results with those for odd numbers so that an ordering of probabilities can be obtained for combining different numbers of experts.

At the same time, our theoretical findings are used to examine results obtained experimentally, and these results are shown to reflect the trends predicted by the theoretical considerations.

2. The Voting Problem

In the rest of this article, we assume that n classifiers or experts are used, and that for each input sample, each expert produces a unique decision regarding the identity of the sample. This identity could be one of the allowable classes, or a rejection when no such identity is considered possible. In combining the decisions of the n experts, the sample is assigned the class for which there is a consensus, or when more than half of the experts are agreed on the identity. Otherwise the sample is rejected.

While each classifier has the possibilities of being correct, wrong, or neutral, the combined (correct) recognition rate is really the probability of the consensus being correct, assuming each vote to have only 2 values – correct or not. Due to the nature of consensus, the combined decision is wrong only when a majority of the votes are wrong *and* they make the same mistake. This is a strength of this combination method – due to the large number of possible mistakes, the majority would not often make the same one. As a result of this need for consensus, we can only calculate the probability of the consensus committing a particular error from the individual probabilities of committing the same error. To assess this particular (mistaken) probability of consensus, we can also consider each vote to have only 2 values – it makes this particular mistake or not.

Consequently, we are led to examine the probabilities of consensus for a group of voters.

3. Results on the Probability of Consensus

If we assume that n independent experts have the same probability p of being correct, then the probability of the consensus being correct, denoted by $P_C(n)$, can be computed using the binomial distribution [1] as

$$P_C(n) = \sum_{m=k}^{n} \binom{n}{m} p^{m} (1-p)^{n-m}$$

where k is the margin of majority.

Since previous studies have always assumed an odd

1051-4651/94 $04.00 © 1994 IEEE

number of voters, we establish the relation between odd and even numbers of voters by proving the following theorem for $n \geq 1$.

Theorem 1: $P_C(2n+1) = P_C(2n) + p^{n+1}(1-p)^n \binom{2n}{n}$, *and*

$$P_C(2n) = P_C(2n-1) - p^n(1-p)^n \binom{2n-1}{n}$$

The next 3 corollaries are direct consequences of *Theorem 1*:

$$C1\text{: } P_C(2n+1) - P_C(2n-1) = p^n(1-p)^n \binom{2n-1}{n}(2p-1)$$

$$C2\text{: } P_C(2n+2) - P_C(2n) = p^{n+1}(1-p)^n \binom{2n}{n}\left[\frac{2np+p-n}{n+1}\right]$$

$$C3\text{: } P_C(2n+2) - P_C(2n-1) = p^n(1-p)^n \binom{2n}{n} * \left[\frac{(4n+2)p^2 - 2np - (n+1)}{2(n+1)}\right]$$

From the preceding results, we can deduce the following remarks when $0 < p < 1$:

(1) As immediate consequences of *Theorem 1*, $P_C(2n) < P_C(2n+1)$ and $P_C(2n) < P_C(2n-1)$ for all n and p.

(2) When even numbers $2n$ of experts are combined, $P_C(2n)$ is monotonically increasing if $p > n/(2n+1)$, which is true for all n if $p \geq 1/2$. Conversely, $P_C(2n)$ is monotonically decreasing with n if $p < n/(2n+1)$, which is true for all n if $p < 1/3$. When $1/3 \leq p < 1/2$, however, the behavior of $P_C(2n)$ would depend on the relative magnitudes of p and $n/(2n+1)$.

(3) When both even and odd numbers of experts are considered together, we know from Remark (1) that $P_C(2n+2) < P_C(2n+3)$ *and* $P_C(2n+2) < P_C(2n+1)$ for all p. The relation between $P_C(2n+2)$ and $P_C(2n-1)$ is given by $C3$, from which it follows that

$$P_C(2n+2) > P_C(2n-1) \textit{ iff } p > f_1(n) = \frac{n + \sqrt{5n^2+6n+2}}{4n+2},$$

and $P_C(2n+2) < P_C(2n-1)$ iff the reverse inequality holds for p. Since $f_1(n)$ is increasing and approaches $p_u = (1 + 5^{1/2})/4$ as $n \rightarrow \infty$, $P_C(2n+2) > P_C(2n-1)$ for all n if $p \geq p_u$ ($\doteq 0.8090$). For example, when $p = 0.75$, $P_C(8) < P_C(5)$ since $p < [3 + (65)^{1/2}]/14$, while the opposite holds for $p = 0.8$.

(4) As a consequence of Remarks (1) and (2), we can conclude that when $p \geq p_u$, $P_C(2n) < P_C(2n-1) < P_C(2n+2) < P_C(2n+1) < P_C(2n+4)$ for all n. For example, we would have the ordering $P_C(2) < P_C(1) < P_C(4) < P_C(3) < P_C(6) < P_C(5) < \ldots$.

(5) Remark (4) defines the ordering of $P_C(n)$ for sufficiently large values of p. We now consider small values of p (to consider the probabilities of consensus errors). Using *Theorem 1* and its corollaries, it can be shown that when $p < p_l \doteq 0.1208$, the consensus probabilities are ordered as $P_C(2n+2) < P_C(2n) < P_C(2n+1) < P_C(2n-2) < P_C(2n-1) < \ldots$ for all n. An example of this ordering would be $P_C(9) < P_C(6) < P_C(7) < P_C(4) < P_C(5) < P_C(2) < P_C(3)$.

4. Application to Pattern Recognition

The discussion in Section 3 presents a mathematical model for comparing the recognition rates obtained from the majority vote of n independent experts when each expert has recognition rate p. We assume the experts to have reasonable performance, so that p is greater than the threshold p_u given above, in which case the ordering of the consensus probabilities is stated in Remark (4). In particular, it follows that a combination of an even number n of experts would yield a recognition rate that is lower than those obtained from both $n+1$ and $n-1$ experts.

In pattern recognition applications, it is also an important consideration that the results should have low error or substitution rates. For these error rates, we can consider the consensus probabilities for small p. If the probabilities of each expert making a particular mistake are approximately equal to p, then we can certainly assume that $p < p_l$, in which case a combination of an even number n of experts would be less likely to commit this error than the consensus of $n+1$ or $n-1$ experts. With this assumption of approximate uniformity, the same conclusion regarding the number of experts can be applied to the overall substitution rates, which are after all the summation of the probabilities of particular errors.

The assumption of equal probabilities has made possible the computation of exact differences in the likelihoods of consensus, whether it is the correct or wrong decision. Admittedly, the assumption of equal probability, while convenient in theory, is impossible to achieve in practice – different experts cannot be expected to operate with equal probabilities in real-life situations. For this reason, we have also examined (in a separate article) the consequences of relaxing this condition. However, it is worth noting that the ordering of the probabilities derived in Remarks (4) and (5) have been demonstrated in experiments where the performances of experts do differ. The results of one such experiment are given below.

In this experiment, the data consists of the recognition results obtained by 7 classification algorithms developed at CEDAR in Buffalo, USA. The test set BS contains 2711 handwritten numerals extracted from USPS mailpieces, and these are contained on CEDAR CDROM together with the recognition results. Fuller descriptions of the classifiers are given in [3]. When only the top choice is considered, the individual algorithms produce the results shown in Table 1 with no rejections, i.e., each input sample is assigned its nearest class.

Classifier	Recognition
Binpoly	93.99
Chaincode	96.38
Gabor	95.17
Gradient	96.20
GSC	97.05
Histogram	93.88
Morphology	95.76

Table 1 Performance of CEDAR classifiers

These classifiers have 120 possible combinations, whose recognition results on the BS database are shown in Fig. 1, where the scatter plot shows the recognition versus the substitution rates. The two disjoint clusters of points (one resulting from combinations of even numbers of experts and the other from odd numbers) serve to illustrate the conclusions of Section 3. Combining the decisions of odd numbers of experts produces higher correct as well as higher error rates, so the corresponding points in Fig. 1 are positioned to the upper right, while the even combinations result in points located to the left and below (representing lower substitution and correct rates).

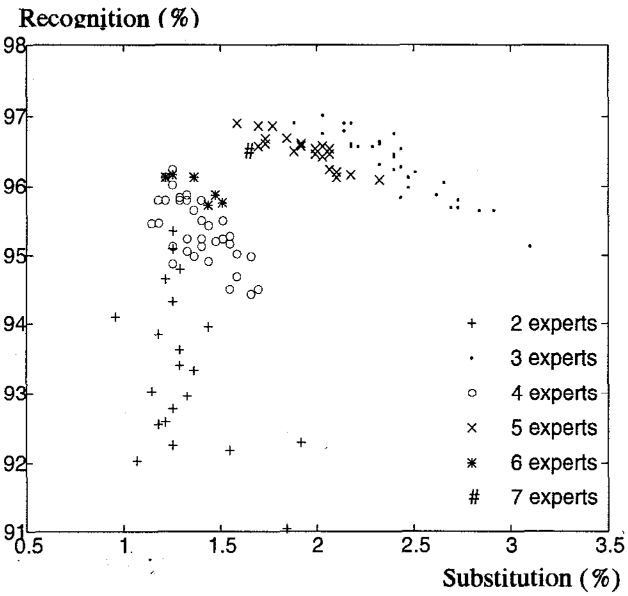

Fig. 1 Combined results of CEDAR classifiers

5. Conclusion

The majority vote method has been used successfully to combine the results of classifiers for character recognition. The intention of this study is to examine its mode of operation and gain a deeper understanding of how this method works, so that we can have confidence in its performance when applied to different data and/or experts. By this theoretical analysis, we have partly achieved our objective of providing a more reliable basis for using this method, and further studies have been made in this direction.

In the course of our research, we have derived many conclusions about the expected behavior of the consensus. Nevertheless, a number of decisions remain with the user. For example, it is clear that the performance of the combined decision is an increasing function of the number of experts, provided each expert can perform at an appropriately high level. The number of experts that can be used would naturally depend on practical limitations. Within these confines, the choice of an odd or even number of experts also depends on the requirements of the specific application. The former produces a higher recognition rate, and the consensus of *2n-1* experts would outperform that of *2n* experts in this respect. However, it is often the case in pattern recognition applications that errors are much more costly than rejections, in which case the performance of an even number of experts would be more reliable.

Acknowledgements

The authors wish to thank the researchers at CEDAR of Buffalo, New York for making their recognition results publicly accessible.

This research was supported by the Natural Sciences and Engineering Research Council of Canada, the National Networks of Centres of Excellence program of Canada, and the FCAR program of the Ministry of Education of the province of Québec.

References

[1] N. C. de Condorcet, *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris, Imprimérie Royale, 1785.

[2] J. Franke, L. Lam, R. Legault, C. Nadal and C. Y. Suen, "Experiments with the CENPARMI data base combining different classification approaches," *Pre-Proc. 3rd Int. Workshop on Frontiers in Handwriting Recognition*, Buffalo, U.S.A., May 1993, pp. 305-311.

[3] D.-S. Lee and S. N. Srihari, "Handprinted digit recognition: a comparison of algorithms," *Pre-Proc. 3rd Int. Workshop on Frontiers in Handwriting Recognition*, Buffalo, U.S.A., May 1993, pp. 153-162.

[4] C. Y. Suen, C. Nadal, T. A. Mai, R. Legault and L. Lam, "Computer recognition of unconstrained handwritten numerals," *Proc. IEEE*, vol. 80, no. 7, pp. 1162-1180, July 1992.