THEORY AND APPLICATION OF MAJORITY VOTE – FROM CONDORCET JURY THEOREM TO PATTERN RECOGNITION

Louisa Lam

Department of Mathematics
Hong Kong Institute of Education
10 Lo Ping Road, Tai Po, Hong Kong

Abstract. Majority vote is a much studied topic; in particular, the well-known Condorcet Jury Theorem (CJT) had provided validity to the belief that the majority opinion of a group is superior to those of individuals provided the individuals have reasonable competence. The mathematics of the theorem can be easily explained to teachers with a basic knowledge of statistics, and its interpretation has considerable social implications.

Recently, the voting principle has been applied to the technological area of optical character recognition, with significant results. In this domain, computer programs are designed and implemented to read or identify characters and words. When the data is handwritten, the immense diversity in writing styles makes it extremely difficult for one program operating alone to achieve the high levels of accuracy that are required for practical applications. On the other hand, many different programs have been designed for this purpose. Consequently, researchers began to combine the results of different programs by majority vote, in order to obtain more accurate performance. This has created a new trend in pattern recognition, and in the process this author has also derived new results about majority vote. These theoretical findings are reflected in experimental results, and the process has provided an example of the interaction between basic mathematical ideas and their applications in advanced technology.

1. Introduction

This paper is concerned with extensions of majority vote, and with its application to the domain of Optical Character Recognition (OCR), in which computers are used to identify characters. In this domain, computer algorithms are designed and implemented to read and determine the identity of characters. Such algorithms can be applied to a very wide range and high volume of applications, including the processing of postal mail, credit card slips, income tax and other forms, and bank cheques. Because of the usefulness of these applications, many different algorithms have been designed by various researchers for. However, when the data is handwritten, the wide variety of writing styles makes it extremely difficult for one algorithm operating alone to achieve very high levels of accuracy.

Consequently, researchers began to consider combining the decisions of different algorithms, to see if more reliable performances could be obtained. Among the various means of combining decisions, majority vote was the first combination method to be applied by a number of researchers, including this author. Initially, the process was applied to the computer recognition of handwritten digits; each algorithm would determine the identity (0–9) of each input character, after which the combined decision of a number of algorithms would be obtained by majority vote. In the process, certain patterns of behavior had been observed in the results; and in attempting to understand and explain these patterns, this author has derived new theoretical results on the behavior of majority vote.

These theoretical explorations had been motivated by a desire to account for observed experimental results. It has been very satisfying that the experimental results are actually supported by theoretical findings, and these elements will be presented in this paper. In addition, it is very interesting that the topic of majority vote, which has the binomial theorem for its mathematical foundations and has been much studied by social scientists for generations, is now being applied in advanced technology as a basic procedure. This can certainly serve as a vivid demonstration of the power and applicability of basic mathematical ideas.

2. The Classical Majority Vote Problem

Majority vote has been a much studied topic for many years, especially by social scientists. Mathematically, if we assume that n independent people have the same probability p of being correct, then the probability of the majority opinion being correct, denoted by $P_C(n)$, can be computed using the binomial distribution as

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

where the value of k is determined by

$$k = \begin{cases} \frac{n}{2}+1 & \text{if } n \text{ is even}, \\ \frac{n+1}{2} & \text{if } n \text{ is odd}. \end{cases}$$

The following theorem, known as the Condorcet Jury Theorem (CJT) [2], has provided validity to the belief that the judgement of a group is superior to that of individuals, provided the individuals have reasonable competence in the sense that they would make correct decisions with reasonably high probabilities p.

Theorem (CJT): Suppose n is odd and $n \geq 3$. Then the following are true:

1. If $p > 0.5$, then $P_C(n)$ is monotonically increasing in n and $P_C(n)$ ®1 as n ®∞.
2. If $p < 0.5$, then $P_C(n)$ is monotonically decreasing in n and $P_C(n)$ ®0 as n ®∞.
3. If $p = 0.5$, then $P_C(n) = 0.5$ for all n.

This work had provided the basis for much modern research in voting and decision-making, especially for the cases when n is odd (see, for example, [1], [3], [6]).

3. Extensions of the Classical Problem

When the number of voters n can be even as well as odd, voting can result in a tied vote, and the requirement of a strict majority for a combined decision would result in a lack of majority or no decision in these cases. Under these conditions, the values of $P_C(n)$ would not be monotonic in n as stated in CJT, but would depend on the value of p as well. It has been established [4] that for small values of p ($p < p_l = 0.1208$), the consensus probabilities are ordered as:

$$P_C(2n+2) < P_C(2n) < P_C(2n+1) < P_C(2n-2) < P_C(2n-1)$$

for all n. For example, we would have:

$$P_C(8) < P_C(9) < P_C(6) < P_C(7) < P_C(4) < P_C(5) < P_C(2) < P_C(3).$$

On the other hand, for large values of p ($p \geq p_u = 0.8090$), the ordering would be

$$P_C(2n) < P_C(2n-1) < P_C(2n+2) < P_C(2n+1) < P_C(2n+4) < P_C(2n+3),$$

which means, for example,

$$P_C(2) < P_C(1) < P_C(4) < P_C(3) < P_C(6) < P_C(5) < P_C(8) < P_C(7).$$

These patterns have been proved theoretically in [4], while they can be observed from the table of binomial distributions, a part of which is shown in TABLE I. From this table, it can be observed that the orderings stated above are true for $p = 0.1$ and $p = 0.9$, but not for values of p such that p_1 £ $p < p_u$.

TABLE I
Values of $P_C(n)$ for different values of p and n

P	Values of n								
	2	3	4	5	6	7	8	9	10
0.10	0.0100	0.0280	0.0037	0.0086	0.0013	0.0027	0.0004	0.0009	0.0001
0.20	0. 0400	0.1040	0.0272	0.0579	0.0170	0.0333	0.0104	0.0196	0.0064
0.30	0.0900	0.2160	0.0837	0.1631	0.0705	0.1260	0.0580	0.0988	0.0473
0.70	0.4900	0.7840	0.6517	0.8369	0.7443	0.8740	0.8059	0.9012	0.8497
0.80	0.6400	0.8960	0.8192	0.9421	0.9011	0.9667	0.9437	0.9804	0.9672
0.90	0.8100	0.9720	0.9477	0.9914	0.9842	0.9973	0.9950	0.9991	0.9984

4. Application of Majority Vote to Optical Character Recognition

For this application, different computer programs/algorithms have been developed to read characters automatically, with different levels of performance. This being the case, the problem is more general than the one stipulated in the CJT, where all voters are assumed to have the same level of competence. In addition, it is not clear in such cases whether the decisions are really independent.

In one case, seven such programs [5] had been developed by different researchers to read (or classify) handwritten numerals. These programs (also called *classifiers*) had been used to read handwritten numerals extracted from addresses of mail envelopes handled by the United States Postal Service. Some examples of such data are shown in Fig. 1.

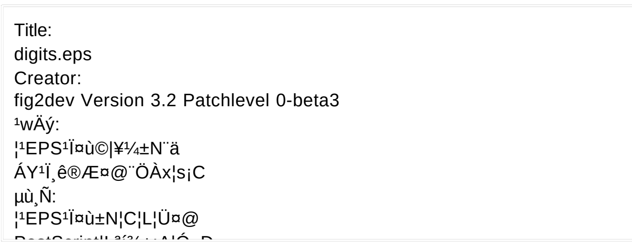

Fig. 1 Examples of handwritten numerals obtained from US mail envelopes

When these classifiers were used to read 2711 handwritten numerals in a standard database, the results shown in TABLE II had been obtained for the individual classifiers.

TABLE II
Performance of individual classifiers

Classifier	C1	C2	C3	C4	C5	C6	C7
Correct rate (%)	93.99	96.38	95.17	96.20	97.05	93.88	95.76

Apart from the correctly classified samples, the rest of the data had been wrongly classified, so the error rate had been 2.95 – 6.01%. In an attempt to reduce the error rate, researchers started to use more than one classifier to determine the identity of a numeral, and then combine the decisions (0 – 9) of the different classifiers by majority vote. The majority decision was adopted if it existed; otherwise the sample would be rejected. In this way, there are $\binom{7}{2} = 21$ combinations of 2 classifiers, $\binom{7}{3} = 35$ combinations of 3 classifiers, and so on. The results obtained from all the 120 possible combinations are shown in Fig. 2, in which the correct rate is plotted against the error rate for each combination. Of course, highly desirable results would be those in the upper left, where the correct rate is high and the error rate is low.

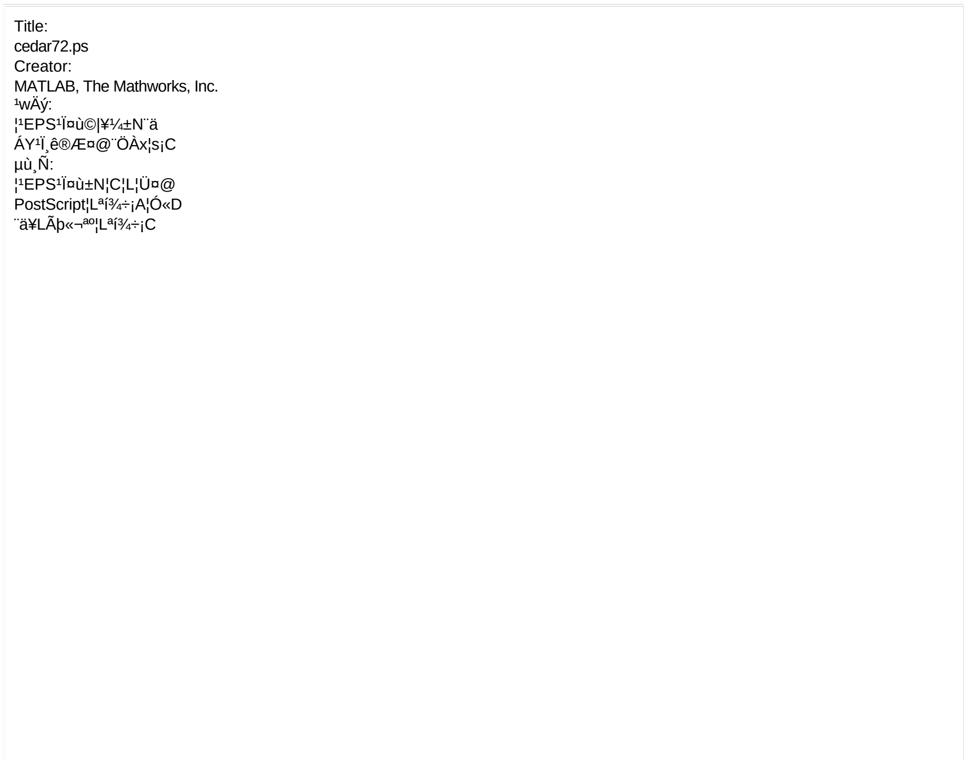

Fig. 2 Results of combining classifier decisions

From Fig. 2, it can be observed that combinations of even numbers of classifiers tend to produce both lower correct and lower error rates (and higher rejection rates) than those of odd numbers. This causes results of even numbers of classifiers to be placed to the lower left of results of odd numbers of classifiers. Adding one classifier to an even number would increase both the correct and error rates, while adding this to an odd number would decrease both rates. Naturally, increases in both the correct and error rates are accompanied by a decrease in the rejection rate, and vice versa.

These observations had given rise to theoretical investigations by this author as to why these phenomena should occur, and had resulted in new findings on the theory of majoity vote. The effects of adding one vote can be easily explained, since increasing the number of classifiers (or voters) from an odd number $2n-1$ to an even number $2n$ can only change some decisions to tied votes, resulting in indecisions or rejections. The votes which can be changed this way are among the ones in which the initial voting had a majority of only one vote. Similarly, when we add one vote to an even number of votes, the added vote can have the effect of breaking some ties, thus decreasing rejections by changing them to correct decisions or errors. These changes would occur whether the classifier decisions are independent or not, even though the magnitude of the changes would depend on the performances of the particular classifiers. They cause the graph to move upwards and to the right when one vote is added to an even number, and in the reverse direction when one vote is added to an odd number.

When we add two votes to an even (or odd) number of votes as repeated additions of one vote, it is not clear what the net effect of the two additions would be, since the second addition appears to reverse the trend of the first. For this reason, we have to examine the results when the two votes are added together to an existing group of votes. In a comparison of probabilities (before and after the addition of two votes), the assumption of independence of the votes was very useful as it allowed the joint probability to be calculated as a product of probabilities. Under this assumption, it can be established theoretically [4] that the results depend on the classical entity of the odds ratio, where the odds ratio r_i of vote i is defined as

$$r_i = \frac{probability(correct)}{1 - probability(correct)}.$$

If the original n votes have odds ratios r_i for $i = 1, \ldots, n$, and the new votes have odds ratios s_1 and s_2, then adding the new votes would increase the combined correct rate if $s_1 s_2 > r_i$ for all i. This is a sufficient but not necessary condition, and the proof makes use of the result established in the "marriage" problem.

Moreover, comparisons of the probabilities indicated that adding two votes to an even number would be more effective in increasing the correct rate than adding the votes to an odd number, given similar levels of performance. However, if reducing the error rate is the more important objective, then adding two votes to an odd number would be more effective.

These theoretical findings can be observed in the experimental results shown in Fig. 2, from which it is clear that increases from two to four, then to six classifiers result in mainly an upward trend (increase in correct rate). This is in marked contrast to the leftward movement (decrease in error rate) produced by increasing the number of classifiers from three to five and seven. These results are particularly noteworthy given that independence of classifier decisions (as stipulated in the theoretical results) cannot be assumed in the experiments.

Another illustration of the results of combining multiple classifier decisions is given in Fig. 3, in which the decisions of nine classifiers on the same set of data are combined in ascending and descending orders of performance. In this figure, the results obtained from combining classifiers in ascending order of performance are denoted by dotted lines, and the number of classifiers is represented by the symbol # following the number. Understandably, combining classifiers in descending order (which means the best classifiers are combined first) give better initial results, as indicated by the solid lines in this figure. However, the same zigzag pattern is obtained from the addition of one vote. In addition, this figure also shows that the addition of two classifiers to an even number is generally more effective in increasing the correct rate, while the addition of two classifiers to an odd number is more effective in reducing the error rate. Eventually, the addition of two votes in descending order to six and seven votes fail to produce better results, showing that the successively weaker votes can no longer satisfy the condition required for improvement in the combined result.

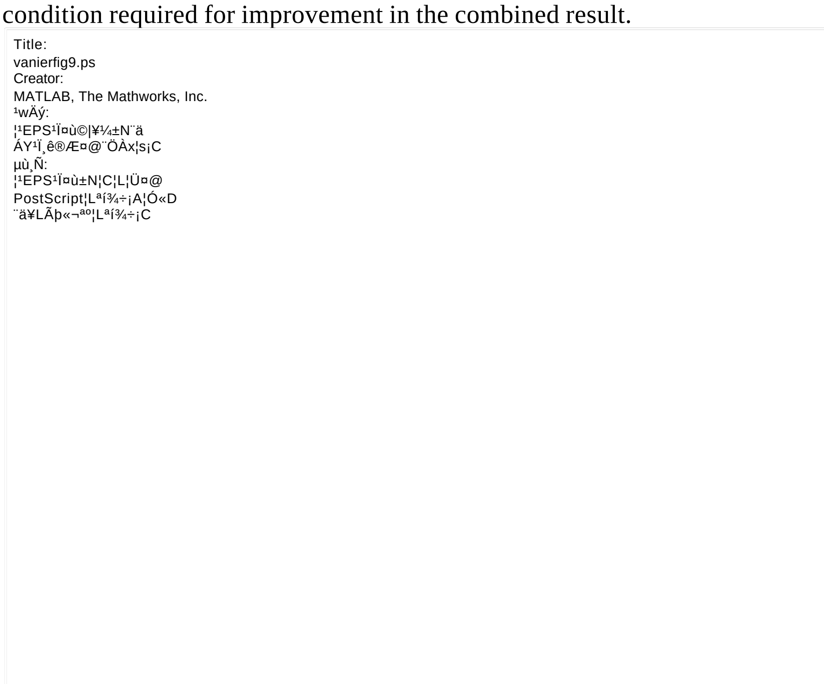

Fig. 3 Combinations of up to 9 classifiers

5. Concluding Remarks

In this paper, we have examined some aspects of majority vote and some extensions of the classical results to include even numbers of voters. These results can be further extended to voters with different levels of competence, and the effects of adding new voters are studied theoretically. As majority vote has been applied to the technological area of pattern recognition (especially OCR), significant quantities of experimental results have been obtained, and it has been found that the theoretical findings are reflected in the empirical results, even though the assumption of independence may or may not hold.

The results have demonstrated that basic mathematical principles can be applied to advanced technological applications, and that experimental results can motivate and lead to further theoretical investigations and developments which can provide theoretical underpinnings for the application.

Acknowledgements

The author is grateful for resources provided at the Centre for Pattern Recognition and Machine Intelligence of Concordia University, Montreal, Canada, and for a conference grant from the Hong Kong Institute of Education.

References

[1] N. C. Condorcet (1785). *Essai sur l''application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris: Imprimerie Royale.

[2] D. Black (1958). *The Theory of Committees and Elections*, London: University Press.

[3] B. Grofman and G. Owen (1986). Condorcet models, avenues for future research. In B. Grofman and G. Owen (Eds.), *Information Pooling and Group Decision-making: Proc. 2nd Univ. of Calif., Irvine, Conf. on Political Economy*, pp. 93-102. Westport, Conn.: JAI Press,

[4] L. Lam and C. Y. Suen (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Trans. Systems, Man, and Cybernetics* 27, 553-568.

[5] D. S. Lee and S. N. Srihari (1993). Handprinted digit recognition: a comparison of algorithms. In *Pre-proc. 3rd Int. Workshop on Frontiers in Handwriting Recognition*, Buffalo, USA, 153-162.

[6] H. P. Young (1988). Condorcet's theory of voting. *American Political Science Review* 82: 1231-1244.