Generalizing the majority voting scheme to conditional voting

Laszlo Kovacs, Henrietta Toman, Agnes Jonas,
Lajos Hajdu, and Andras Hajdu

University of Debrecen
Egyetem ter 1, 4032 Debrecen, Hungary
{kovacs.laszlo.ipgd,toman.henrietta}@inf.unideb.hu
jonasagn@gmail.com,hajdul@math.klte.hu
hajdu.andras@inf.unideb.hu

Abstract. In this paper we propose a new voting scheme for generalizing the classical majority voting system. In contrast to the classical voting method we can make good decision if the number of classifiers assigning the correct class label is less than the half of the overall number of classifiers. This new method can be applied to such problems when the decision is not only logical but is also required to satisfy a pre-defined (e. g. geometrical) condition. In addition, we will show the results corresponding to the concept of pattern of success and failure using the existence theorem for the combinatorial (0,1)-matrices.

Keywords: majority voting, combinatorial (0,1)-matrices, biomedical imaging, diabetic retinopathy

1 Introduction

In this paper we work out a theoretical model which generalizes the classical majority voting scheme, where more than the half of the votes are needed to make a final decision. Our model is based on the idea that in the case of less good votes we still have some chance to make a good decision. This idea was motivated by certain medical imaging problems to detect the optic disc and the macula where bad votes can overcome good ones only if a further geometrical condition is fulfilled. Applying more different optic disc/ macula detectors [1] for voting we can achieve better performance for the automatic detection system than for each individual algorithm. We were interested in the upper and lower bounds of the system accuracy and discussed the concept of the pattern of success and failure in our model. This generalized method can be applied to several problems corresponding to spatial location with additional constraints (e.g. detecting a certain pixel or region).

In the rest of the paper, section 2 presents the classical voting system. In section 3 we introduce a generalization of voting system and show some theoretical results, while in section 4 our results for the pattern of success and failure are discussed. In section 5 we present an illustrative game for this model. The following section contains the medical imaging application. Section 7 gives conclusion and further recommendations.

2 Majority voting

Let $D = (D_1, D_2, \ldots, D_n)$ be a set of classifiers, $D_i : R^k \to \Omega$ $(i = 1, \ldots, n)$ where $\Omega = (\omega_1, \omega_2, \ldots, \omega_c)$ be a set of class labels. In the majority vote method of combining classifier decisions the class label ω_i supported by the majority of the classifiers D_i is assigned to x. Most often ties are broken randomly.

In [2] Kuncheva et al. discuss exhaustively the following special case. Let n be odd, $\Omega = (\omega_1, \omega_2)$ (each classifier output is a binary vector) and all classifiers have the same classification accuracy p. An accurate class label is given by the majority vote if at least $[n/2]$ classifiers give correct answers. The majority vote method with independent classifier decisions gives an overall correct classification accuracy calculated by the binomial formula:

$$P = \sum_{k=0}^{[n/2]} \binom{n}{k} p^{n-k}(1-p)^k$$

Several interesting results can be found in [3] applying the majority voting in pattern recognition. This method is guaranteed to give a higher accuracy than the individual classifiers if the classifiers are independent and $p > 0.5$.

3 The generalization

Let $\eta = (\eta_1, \ldots, \eta_n)$ be an n-dimensional random variable. Assume that the coordinates η_i of η are independent random variables with

$$P(\eta_i = 1) = p, \quad P(\eta_i = 0) = 1 - p \quad (i = 1, \ldots, n),$$

where $p \in [0, 1]$. Execute the experiment η independently t times, and write the outcomes in a table of size $n \times t$. (The j-th column of the table contains the realization of η in the j-th experiment $(j = 1, \ldots, t)$.) Define now the random variables $\chi_1, \ldots, \chi_t$ in the following way. If in the j-th column there are k ones then let

$$P(\chi_j = 1) = p_{nk}, \quad P(\chi_j = 0) = 1 - p_{nk} \quad (j = 1, \ldots, t),$$

where the p_{nk}-s $(k = 0, 1, \ldots, n)$ are given numbers with

$$0 \le p_{n0} \le \ldots \le p_{nn} \le 1.$$

Observe that the χ_j-s are independent. Finally, put

$$\xi = |\{j \ : \ \chi_j = 1\}|,$$

that is, ξ is the number of "good" decisions. Observe that all the individual decisions η_i $(i = 1, \ldots, n)$ are of binomial distribution with parameters (t, p). As we shall see, ξ is also of binomial distribution, with the appropriate parameters. To show this, first we need the following lemma.

Lemma 1. *For any $j = 1, \ldots, t$ we have*

$$P(\chi_j = 1) = \sum_{k=0}^{n} p_{nk} \binom{n}{k} p^k (1-p)^{n-k}.$$

Proof. The statement is trivial in view of the definitions of the objects involved.

We introduce the following notation: put

$$q = \sum_{k=0}^{n} p_{nk} \binom{n}{k} p^k (1-p)^{n-k}. \tag{1}$$

Lemma 2. *The random variable ξ is of binomial distribution with parameters (t, q), where q is given by (1)*

Proof. Let $k \in \{0, 1, \ldots, t\}$. Then, since the χ_j-s are independent, we have

$$P(\xi = k) = \binom{n}{k} q^k (1-q)^{n-k},$$

and the statement follows.

In order to have the majority voting be "better" than the individual decisions, we need only to guarantee that $q \geq p$. The next statement yields a guideline along this way.

Proposition 1. *Let $p_{nk} = k/n$ $(k = 0, 1, \ldots, n)$. Then we have $q = p$, and consequently $E\xi = tp$.*

Proof. By Lemma 2 we have

$$E\xi = tq.$$

Thus we need only to show that $q = p$ whenever $p_{nk} = k/n$ $(k = 0, 1, \ldots, n)$. Indeed, using that a random variable of binomial distribution with parameters (n, p) has expected value np, in this case we have

$$q = \sum_{k=0}^{n} p_{nk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} k/n \binom{n}{k} p^k (1-p)^{n-k} = np/n = p.$$

Hence the statement follows.

As a trivial consequence we obtain the following statement.

Corollary 1. *Suppose that for all $k = 0, 1, \ldots, n$ we have $p_{nk} \geq k/n$. Then $q \geq p$, and consequently $E\xi \geq tp$.*

Theorem 1. *Suppose that $p \geq 1/2$ and for any $n \geq k_1 > k_2 \geq 0$ with $k_1+k_2 = n$ we have $p_{nk_1}+p_{nk_2} \geq 1(= (k_1+k_2)/n)$ and $p_{nk_1} \geq k_1/n$; further, that $p_{n\frac{n}{2}} \geq 1/2$ if n is even. Then $q \geq p$, and consequently $E\xi \geq tp$.*

Proof. One can easily check that for any k_1, k_2 as in the statement we have

$$p_{nk_1}\binom{n}{k_1}p^{k_1}(1-p)^{n-k_1} + p_{nk_2}\binom{n}{k_2}p^{k_2}(1-p)^{n-k_2} \geq$$

$$\geq \frac{k_1}{n}\binom{n}{k_1}p^{k_1}(1-p)^{n-k_1} + \frac{k_2}{n}\binom{n}{k_2}p^{k_2}(1-p)^{n-k_2}.$$

This by Proposition 1 clearly implies the statement.

As a trivial consequence we obtain the following result of Kuncheva et al.

Corollary 2. *Suppose that n is odd, $p \geq 1/2$ and for all $k = 0, 1, \ldots, n$ we have $p_{nk} = 1$, if $k > n/2$, and $p_{nk} = 0$ otherwise. Then $q \geq p$, and consequently $E\xi \geq tp$.*

Of particular interest is the value $P(\xi = t)$, since this expresses the probability that we make only "good" decisions. In case of an individual decision, the corresponding probability is p^t. So we need to choose the probabilities p_{nk} so that $P(\xi = t) \geq p^t$. In fact we can characterize a much more general case. For this purpose we need the following lemma, due to Gilat [4].

Lemma 3. *For any integers t and l with $t \geq 1$ and $1 \leq l \leq t$ the function*

$$f(x) = \sum_{k=l}^{t}\binom{t}{k}x^k(1-x)^{t-k}$$

is strictly monotone increasing on $[0, 1]$.

Note that obviously, for any $x \in [0, 1]$ we have

$$\sum_{k=0}^{t}\binom{t}{k}x^k(1-x)^{t-k} = 1.$$

As a simple consequence of Lemma 3 we obtain the following result. Recall that the η_i-s $(i = 1, \ldots, n)$ are just "individual" random variables, of binomial distribution with parameters (t, p).

Theorem 2. *Let t and l be integers with $t \geq 1$ and $1 \leq l \leq t$. Then $P(\xi \geq l) \geq P(\eta_1 \geq l)$ if and only if $q \geq p$, i.e. $E\xi \geq tp$.*

Proof. Let t and l be as given in the statement. Then we have

$$P(\xi \geq l) = \sum_{k=l}^{t}\binom{t}{k}q^k(1-q)^{t-k}$$

and

$$P(\eta_1 \geq l) = \sum_{k=l}^{t} \binom{t}{k} p^k (1-p)^{t-k}.$$

Thus by Lemma 3 we obtain that

$$P(\xi \geq l) \geq P(\eta_1 \geq l)$$

if and only if $q \geq p$, and the theorem follows.

4 The deterministic case

Suppose that in each row of the above mentioned $n \times t$ table we have exactly r ones. We keep the previous notation for the random variables χ_j $(j = 1, \dots, t)$ and ξ. Now we should choose the probabilities in a way that $E\xi \geq r$.

Proposition 2. *If $p_{nk} = k/n$ for all $k = 0, 1, \dots, n$ then $E\xi = r$.*

Proof. Denote by u_j the number of ones in the j-th column. Then we have $E\chi_j = u_j/n$. Thus

$$E\xi = \sum_{j=1}^{t} E\chi_j = \sum_{j=1}^{t} u_j/n = rn/n = r.$$

As a simple consequence we get that both Theorem 1 and Theorem 2 are valid also in this case, with the obvious modifications.

4.1 Pattern of success

The concept of pattern of success and failure was introduced in [2]. We analyze these notions in the generalized model. First we consider the pattern of success in additive case where we would like to maximalize $\sum_{j=1}^{t} p_{nu_j}$. If $p_{nk} = k/n$, then the problem is solved concerning the Proposition 2. In general case the pattern of success can be characterized by the following theorem.

Theorem 3. *Let the probabilities p_{nk} be arbitrary, up to $p_{n0} = 0$. Let $k_0 \neq 0$ be an index such that $p_{nk_0}/k_0 \geq p_{nk}/k$ for all $k = 1, \dots n$, and suppose that $k_0 \mid tr$ and $k_0 \leq r$. Then $E\xi \leq rp_{nk_0}/k_0$, and the maximum can be attained.*

Proof. The statement follows by observing that in this way the ones in the table gets the largest possible weights.

The multiplicative case for the pattern of success means to make only good decisions. In other words, we consider the case when $P(\xi = t)$ and $\prod_{j=1}^{t} p_{nu_j}$ is maximal. For $p_{nk} = k/n$ we have the following theorem.

Theorem 4. *Let $p_{nk} = k/n$ for all $k = 0, 1, \ldots, n$ and suppose that $nr \geq t$. Then $P(\xi = t)$ is maximal for the tables in which $|u_i - u_j| \leq 1$ for all $1 \leq i, j \leq t$. Here u_i denotes the number of ones in the i-th column.*

Proof. We have $u_1 + \ldots + u_t = nr$, whence

$$t(E\chi_1 + \ldots + E\chi_t) = nr.$$

Thus if $|u_i - u_j| \geq 2$, and say $u_i > u_j$, then $(u_i - 1)(u_j + 1) > u_i u_j$, and the statement follows.

In general case we get the following result:

Theorem 5. *Let the probabilities p_{nk} be arbitrary, up to $p_{n0} = 0$. Let $k_0 \neq 0$ be an index such that $-log(p_{nk_0})/k_0 \geq -log p_{nk}/k$ for all $k = 1, \ldots n$, and suppose that $k_0 \leq r$. Then $E\xi \leq r p_{nk_0}/k_0$, and the maximum can be attained if $t \to \infty$ while r/t remains constant.*

Proof. The statement can be proved by applying the result of Theorem 3 for the logarithm of $\prod_{j=1}^{t} p_{nu_j}$.

Same results can be proved for the concept of pattern of failure because it is the dual version of pattern of success. For example, in additive general case $\sum_{j=1}^{t} p_{nu_j}$ is minimal if we consider the index $k_0 \neq 0$ such that $(1 - p_{nk_0})/k_0 \geq (1 - p_{nk})/k$ for all $k = 1, \ldots n$.

4.2 Combinatorial (0,1)-matrices

If we consider the individual decisions $\eta = (\eta_1, \eta_2, \ldots, \eta_n)$ as an n-dimensional random binary vector and execute η independently t times, we can collect the outcomes into a matrix A of zeros and ones of size $n \times t$. In this way we can get the accuracy of individual algorithms from the sum of ones contained in each row. We are interested in the sum of column vector of A in the case of known row sum. We can check whether a given layout of this (0,1)-matrix A beside given row sum exists or not by using Gale-Ryser theorem in [7],[6],[8]. Let the row sum vector $R = (r_1, r_2, \ldots, r_n)$ and the column sum vector $S = (s_1, s_2, \ldots, s_t)$ be non-negative integral vectors. Denote by $\mathfrak{U}(R, S)$ the set of all $n \times t$ matrices $A = [a_{ij}]$ satisfying $a_{ij} = 0$ or 1 and

$$\sum_{j=1}^{t} a_{ij} = r_i \text{ and } \sum_{i=1}^{n} a_{ij} = s_j$$

for $i = 1, .., n$ and $j = 1, .., t$. The vectors R and S satisfy the fundamental equation

$$r_1 + r_2 + \cdots + r_n = s_1 + s_2 + \cdots + s_t.$$

Without loss of generality we can choose the ordering of the members of vectors R and S so that

$$r_1 \geq r_2 \geq \cdots \geq r_n \text{ and } s_1 \geq s_2 \geq \cdots \geq s_t.$$

Then vectors R and S are said to be monotone. The existence of a (0,1)-matrix in $\mathfrak{U}(R,S)$ can be formulated in terms of conjugation and majorization of vectors R and S. The conjugate of $R = (r_1, r_2, \ldots, r_n)$ (where $r_k \leqslant t, (k = 1, 2, \ldots, n)$) is the non-negative integral vector $R^* = (r_1^*, r_2^*, \ldots, r_t^*)$ where

$$r_k^* = \mid \{i : r_i \geq k, i = 1, 2, \ldots, n\} \mid .$$

We can show a geometric way to interpret the conjugate vector R^*. Consider an array of n rows and t columns which has a layout that there are exactly r_i ones in the first position of row i, $(i = 1, 2, \ldots, n)$. Then $R^* = (r_1^*, r_2^*, \ldots, r_t^*)$ is the vector of column sums of the array.
Now let $E = (e_1, e_2, \ldots, e_n)$ and $F = (f_1, f_2, \ldots, f_n)$ be two monotone, non negative integral vectors. Then E is majorized by F $(E \preceq F)$ if all the partial sums of E and F satisfy

$$\sum_{i=1}^{k} e_i \leqslant \sum_{i=1}^{k} f_i,$$

$k = 1, 2, \ldots, n$, with equality for $k = n$.

Theorem 6. (Gale [7] and Ryser [8]) *Let $R = (r_1, r_2, \ldots, r_n)$ be non-negative integral vector where $r_i \leqslant t$ $(i = 1, 2, \ldots, n)$ and $S = (s_1, s_2, \ldots, s_t)$ be monotone non-negative integral vector. Then there exists an $n \times t$ (0,1)-matrix in $\mathfrak{U}(R,S)$ if and only if $S \preceq R^*$.*

For equal accuracy p of each algorithm we got the following result.

Lemma 4. *Let $r_i = r = pt$ for all $i = 1, 2, \ldots, n$, then $S \preceq R^*$ is always satisfied.*

Proof. In the case $r_i = r = pt$ we get $r_1 + r_2 + \cdots + r_n = s_1 + s_2 + \cdots + s_t = nr$. Then we have $r_k^* = n$ for all $k = 1, 2, \ldots, n$ where $r_k \leqslant t$, since there are exactly r ones in each row so each r_k^* $(k = 1, 2, \ldots, n)$ has the same cardinality for $k \leq r$. On the other hand $s_i \leqslant n$ for all $i = 1, 2, \ldots, t$ because ones can be only in the first n rows of the matrix. Then we have

$$s_1 + s_2 + \cdots + s_k \leq r_1^* + r_2^* + \cdots + r_k^* = kn$$

for $k = 1, 2, \ldots, n$ so the condition $S \preceq R^*$ is completed.
In a special case where $R = (r_1, r_2, \ldots, r_n)$ and $S = (s_1, s_2, \ldots, s_t)$ are monotone we can get the rearrangement of the matrix A where $r_i = r$ $(i = 1, 2, \ldots, n)$ i.e. each row contains exactly r ones and the first r column sum is $s_j = n$ $(j = 1, 2, \ldots, r)$, so equality holds in the assumption $S \preceq R^*$.

Assume that we have different accuracies $p_1, p_2, \dots, p_n$ for the individual algorithms. Then we have the following restrictions for these accuracies. Let $A \in \mathfrak{U}(R,S)$, where $R = (r_1, r_2, \dots, r_n) = (tp_1, tp_2, \dots, tp_n)$ is the row sum vector that gives the individual accuracies. If we transpose the matrix $A \in \mathfrak{U}(R,S)$ then $A^T \in \mathfrak{U}(S,R)$ i.e. each entry of $A^T = [b_{ij}]$ satisfies that $b_{ij} = 0$ or 1 and $r_1 + r_2 + \dots + r_n = s_1 + s_2 + \dots + s_t$ also holds. Consequently the Gale-Ryser theorem holds for $A^T \in \mathfrak{U}(S,R)$ as well.

5 Special illustrative game

We illustrate the results mentioned before with a game. Suppose that N players play a game. Players can tell the truth with probability p or lie. Each player say a number, if one says $1/N$ that means telling the truth, if one says a number x_i independently from the interval $[-1/N, -1/(N-k)N]$ that means telling a lie. Let k mean the number of true answers, in this way $N-k$ people tell lie. We get the final decision by adding the numbers told by players. So we obtain the final decision by evaluating the expression below:

$$\sum_{i=1}^{N-k} x_i + \frac{k}{N}$$

If the value of this expression is positive then we make a good decision, otherwise we make bad decision. We can calculate the distribution function for the probability of good decision applying the result in [5] with some modifications. We have

$$F\left(\frac{k-1}{N}\right) = \frac{G(N,k)}{(N-k)!\left(\frac{N-k-1}{(N-k)N}\right)^{N-k}},$$

where $G(N,k) = \sum_{j=0}^{N-k} (-1)^j \binom{N-k}{j} \left[\left(\frac{k-1}{N} - \frac{j(N-k-1)}{(N-k)N}\right)^+\right]^{N-k}$ and $(x)^+$ denotes $max\,\{x, 0\}$. Then we get that if $k = \frac{N+1}{3}$ i.e. approximately one third of the players give the right answer then the probability of good final decision is 0.5 given by

$$F\left(\frac{k-1}{N}\right) = F\left(\frac{\frac{N+1}{3} - 1}{N}\right) = F\left(\frac{N-2}{3N}\right) = \frac{1}{2}.$$

The majority vote accuracy of n independent classifiers with individual accuracy p calculated in [2] is shown below in Table 1.

The accuracy of our decision system for the game illustrated in Table 2 is above the individual accuracy in all cases.

Table 1. System accuracy for classical majority voting scheme

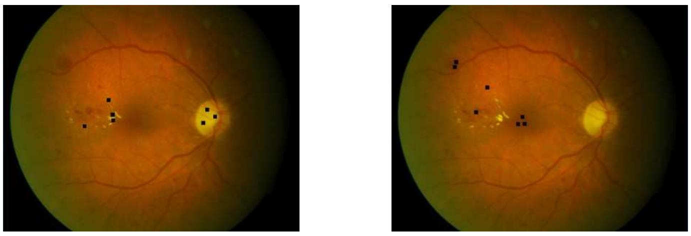
	n=3	n=5	n=7	n=9
p = 0.6	0.6480	0.6826	0.7102	0.7334
p = 0.7	0.7840	0.8369	0.8740	0.9012
p = 0.8	0.8960	0.9421	0.9667	0.9804
p = 0.9	0.9720	0.9914	0.9973	0.9991

Table 2. System accuracy of the generalized voting scheme

	n=3	n=5	n=7	n=9
p = 0.6	0.8208	0.8390	0.8895	0.9247
p = 0.7	0.9163	0.9373	0.9658	0.9823
p = 0.8	0.9728	0.9850	0.9942	0.9980
p = 0.9	0.9963	0.9988	0.9997	0.9999

6 Medical imaging application

Detecting the optic disc/macula on the retinal images plays important role in making diagnosis in the clinical protocol. We have organized more individual OD/ macula detector algorithms into a voting system. In our approach, all of the algorithms return with the center as a single pixel. We have combined the output for each detector and considered the minimal bounding circles for all subgroups of the candidates. The radius must be less than a clinically predetermined constant. The circle with maximal number of candidates is chosen.



(a) outputs of OD detectors

(b) outputs of macula detectors

Fig. 1. Results of the different detecting algorithms

In this combined system with the above mentioned voting scheme we can make a good decision even in the case when the bad candidates have majority such as in the case illustrated in Fig.1. Bad decision is made only when the bad candidates can be bounded by a circle with an appropriate radius. This observation motivated to work out a theoretical model generalizing the majority voting scheme.

7 Conclusion

We worked out a new theoretical model that enables the investigation of majority voting systems being more general than the simple majority voting scheme. In our specific application better overall system accuracy is achieved than in case of individual algorithm. Same results are expected for all image processing problems where the algorithms vote with a single pixel or range as output. The full characterization of the participating algorithms to achieve the best system performance is still an open issue. The essential criterion for the selection of the algorithms to be combined is that $p > 1/2$ for its accuracy.

Our plan is to discuss the dependent case in general, as well. For example, it will be interesting to know how the accuracy of the individual algorithms and the dependency influence the system accuracy. The pattern of success and failure is a useful information in clinical systems since they characterize the expected value of the system error and the boundary of the system accuracy.

Acknowledgments. This work was supported in part by the Janos Bolyai grant of the Hungarian Academy of Sciences, and by the TECH08- 2 project DRSCREEN- Developing a computer based image processing system for diabetic retinopathy screening of the National Office for Research and Technology of Hungary (contract no.: OM-00194/2008, OM-00195/2008, OM-00196/ 2008). Research supported in part by the OTKA grants K67580 and K75566, and by the TÁMOP 4.2.1./B-09/1/KONV-2010-0007 project. The project is implemented through the New Hungary Development Plan, cofinanced by the European Social Fund and the European Regional Development Fund.

References

1. Harangi, B., Qureshi, R.J., Csutak, A., Peto, T., Hajdu, A.: Automatic Detection of the Optic Disc Using Majority Voting in a Collection of Optic Disc Detectors. In: 7th IEEE International Symposium on Biomedical Imaging, pp. 1329–1332. IEEE Press, Rotterdam (2010)
2. Kuncheva, L.I., Whitaker, C.J., Shipp, C.A.: Limits on the Majority Vote Accuracy in Classifier Fusion, Pattern Analysis and Applications 6, 22–31 (2003)
3. Lam, L., Suen, C.Y.: Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 27, 553–568 (1997)
4. Gilat, D.: Monotonicity of a Power Function: An Elementary Probabilistic Proof. The American Statistician 31, 91–93 (1977)
5. Buonocorea, A., Pirozzi, E., Caputo, L.: A Note on the Sum of Uniform Random Variables, Statistics and Probability Letters 79, 2092–2097 (2009)
6. Krause, M.: A Simple Proof of the Gale-Ryser Theorem. The American Mathematical Monthly 103, 335–337 (1996)
7. Gale, D.: A Theorem on Flows in Networks. Pacific Journal of Mathematics 7, 1073–1082 (1957)
8. Ryser, H.J.: Combinatorial Mathematics. The Mathematical Association of America, Washington (1963)