COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA 11 (4) (1991) 383–387

ON COMPUTING MAJORITY BY COMPARISONS

MICHAEL E. SAKS[1] and MICHAEL WERMAN

Received January 20, 1989
Revised February 6, 1990

The elements of a finite set X (of odd cardinality n) are divided into two (as yet unknown) classes and a member of the larger class is to be identified. The basic operation is to test whether two objects are in the same class. We show that $n - B(n)$ comparisons are necessary and sufficient in worst case, where $B(n)$ is the number of 1's in the binary expansion of n.

1. Introduction

Let X be a finite set of cardinality n (n odd) whose elements are partitioned in an unknown way into two parts A and B. The problem is to identify at least one element that belongs to the larger part. The basic operation permitted is a comparison of any two elements $x, y \in X$. The comparison is denoted $x : y$ and its outcome is either "$x = y$" or "$x \neq y$". Since comparing every element to a fixed element $a \in X$ completely determines the partition, $n-1$ comparisons suffice to solve the problem. There is a comparison strategy that solves the problem using at most $n - B(n)$ comparisons, where $B(n)$ is the number of ones in the binary expansion of n (Such a strategy is described in section 3). The main result of this note is that this is tight.

Theorem 1.1. *$n - B(n)$ comparisons are necessary and sufficient to identify a member of the majority class.*

The upper bound has apparently been observed by several people. The new result of this paper is the lower bound. It is proved by a generating function argument related to that introduced by Best, van Emde Boas and Lenstra ([1]) and Rivest and Viullemin ([5]) (see [2], Chap. 8, Theorem 2.1) for analyzing the number of variables that must be probed to evaluate a Boolean function.

The problem was presented by Steve Grantham at the AMS-INS-SIAM meeting on "Graphs and Algorithms" June, 1987, but the original source is unclear. In the more general situation in which the variables take on arbitrary values, Fisher and Salzberg ([3]) showed that $\lceil (3n/2) + 1 \rceil$ comparisons are necessary and sufficient to

AMS subject classification (1991): 68 P 10, 68 R 05

[1] Supported in part by NSF grant DMS87 03541 and Air Force Office of Scientific Research grant AFOSR–0271.

identify a majority element if one exists. Also, for general values of the variables and arbitrary k, Misra and Gries ([4]) obtained estimates of the number of comparisons needed to determine whether there are at least k variables with the same value.

2. Reformulation of the problem

Consider the state of knowledge after some set C of comparisons has been performed. The set C of comparisons can be divided into two sets $C_=$ and $C_{\neq}$ according to the results. Let $\chi(C_=, C_{\neq})$ be the set of partitions $\{A', B'\}$ of X that are consistent with this information.

The condition for terminating the algorithm is:

(T) There exists an element x such that for every partition in $\chi(C_=, C_{\neq}), x$ belongs to the larger part.

Let $X_1, \ldots, X_k$ be the partition of X into the connected components of the graph with edge set C. For each $i \in [k](= \{1, \ldots, k\})$ the results of the comparisons within X_i determine the restriction $\{A \cap X_i, B \cap X_i\}$ of the unknown partition $\{A, B\}$ to X_i. Let (L_i, S_i) be the partition of X_i ordered so that $|L_i| \geq |S_i|$, and let $m_i = |L_i| - |S_i|$. For $I \subseteq [k]$, let $L_I = \bigcup_{i \in I} L_i$, $S_I = \bigcup_{i \in I} S_i$ and $m_I = \sum_{i \in I} m_i$. Then it is easily seen that

$$\chi(C_=, C_{\neq}) = \{\{L_I \cup S_{[k]-I}, L_{[k]-I} \cup S_I\} | I \subseteq [k]\}.$$

Furthermore, $|L_I \cup S_{[k]-I}| - |L_{[k]-I} \cup S_I| = m_I - m_{[k]-I}$. Thus an element x satisfies (T) if and only if x belongs to L_j for some j and $m_I \geq m_{[k]-I}$ for all I containing j. From this we obtain the following equivalent termination condition:

(T') There exists j such that $m_j \geq \sum_{i \neq j} m_i$.

Note that further comparisons between two elements in the same X_i provide no new information. A comparison between any $x \in X_i$ and $y \in X_j$ determines $\{A \cap (X_i \cup X_j), B \cap (X_i \cup X_j)\}$ to be either $\{L_i \cup L_j, S_i \cup S_j\}$ or $\{L_i \cup S_j, L_j \cup S_i\}$, and hence causes m_i and m_j to be replaced by either $m_i + m_j$ or $|m_i - m_j|$.

Thus the testing problem can be reformulated as a game between two players, the *selector* and the *assigner*. A position of the game is a multiset $M = \{m_1, m_2, \ldots, m_k\}$ of nonnegative integers. In each round of the game, the *selector* selects two members of the multiset and the *assigner* replaces these two numbers by either their sum or by their absolute difference, reducing the size of the multiset by 1. The game ends when the largest number in the multiset is at least half the total. The selector seeks to maximize (and the assigner to minimize) the size of the multiset when the game ends. The value of the game $V(m_1, \ldots, m_k)$ is the size of the set remaining if both players follow their optimal strategy. For n odd, let $V_n = V(\underbrace{1, 1, \ldots, 1}_{n \text{ times}})$. Theorem 1.1. is then equivalent to:

Theorem 2.1. *For n odd, $V_n = B(n)$.*

In the next section we present a strategy for the selector which will prove $V_n \geq B(n)$. The main result, that $V_n \leq B(n)$ is proved in section 4.

3. The lower bound on V_n

Consider a strategy for the selector that obeys the following rule: at each round if there are two nonzero numbers that are the same then select them. We claim that starting from the multiset of n 1's, the first time a position is reached in which all nonzero elements are different the following two conditions hold:

(i) the game is over, i.e. the largest element of the multiset is at least the sum of the others.

(ii) the multiset has at least $B(n)$ elements.

It can be seen by induction that every nonzero element produced during the game is a power of 2, which immediately implies (i). To prove (ii), let $m_1, \dots, m_j$ be the nonzero elements of the terminal multiset and suppose that zero appears t times. Each zero was produced when the assigner replaced two equal powers of two by their difference. Let $a_1, \dots, a_t$ be the numbers obtained if the assigner had chosen to add each such pair. Then $m_1, \dots, m_j, a_1, \dots, a_t$ are powers of two summing to n, implying $j + t \geq B(n)$, and hence $V_n \geq B(n)$.

4. Proof that $V_n \leq B(n)$

Let $M = \{m_1, \dots, m_k\}$ be a game position. For $i, j \in [k] = \{1, \dots, k\}$, let M_{ij}^+ (resp. M_{ij}^-) be the multisets obtained by replacing m_i, m_j by their sum (resp. their absolute difference). Analyzing the situation after one round of the game yields:

Lemma 4.1. *Let $M = \{m_1, \dots, m_k\}$ be a game position that is not terminal. Then*

$$V(M) = \max_{i,j} \min\{V(M_{ij}^+), V(M_{ij}^-)\}.$$

Now the proof of the theorem proceeds as follows. We will construct a function Φ which assigns a positive integer to any game position and show that Φ satisfies

(4.1) $$\Phi(M) \geq V(M) \text{ if } M \text{ is a final position.}$$

(4.2) $$\Phi(M) \geq \min\{\Phi(M_{ij}^+), \Phi(M_{ij}^-)\} \text{ for all } i, j \text{ if } M \text{ is not a final position.}$$

By lemma 4.1 and induction, (4.1) and (4.2) imply

(4.3) $$V(M) \leq \Phi(M) \text{ for any position } M.$$

Finally, we show

(4.4) $$\Phi(\underbrace{1, \dots, 1}_{n}) = B(n) \text{ for } n \text{ odd,}$$

which will complete the proof.

Φ is constructed as follows. Let $M = \{m_1, \dots, m_k\}$ be a game position. For $I \subseteq [k]$, let M_I be the multiset $\{m_i | i \in I\}$ and $m_I = \sum\{m_i | i \in I\}$. Say that I is *heavy* if $m_I > m_{[k]-I}$. Define the generating function

$$f_M(x) = \sum_{I \text{ heavy}} x^{m_I}.$$

For an integer k let $P(k)$ be the largest power of 2 dividing k ($P(0) = \infty$). Finally, define

$$\Phi(M) = 1 + P(f_M(-1)).$$

It suffices to show that Φ satisfies (4.1), (4.2) and (4.4). First note that $M = (m_1, \ldots, m_k)$ is a terminal position if and only if one element, say m_1, is greater than the sum of the rest. Then I is heavy if and only if $1 \in I$ and so

$$f_M(x) = x^{m_1}(1 + x^{m_2})(1 + x^{m_3}) \ldots (1 + x^{m_k}).$$

Thus $|f_M(-1)|$ equals either 0 or 2^{k-1} and so $1 + P(f_M(-1)) \geq k = V(M)$, proving (4.1). Now by symmetry, it suffices to prove (4.2) for the non-terminal position $M = \{m_1, \ldots, m_k\}$ when $i = k-1$ and $j = k$. Assume $m_{k-1} \geq m_k$ and let $M^+ = M^+_{k-1,k}$ and $M^- = M^-_{k-1,k}$.

Lemma 4.2. $f_M(x) = f_{M^+}(x) + x^{m_k} f_{M^-}(x)$.

Proof. Write $M^+ = \{m_1^+, \ldots, m_{k-1}^+\}$ and $M^- = \{m_1^-, \ldots, m_{k-1}^-\}$ where $m_i^+ = m_i^- = m_i$ if $i \leq k-2$ and $m_{k-1}^+ = m_{k-1} + m_k$ and $m_{k-1}^- = m_{k-1} - m_k$. Let A_1 be the set of subsets of $[k]$ containing exactly one element from $\{k-1, k\}$ and A_0 be the remaining subsets of $[k]$. Then for $I \in A_0$, $m_I = m^+_{I-\{k\}}$ and since M and M^+ have the same sum, I is heavy for M if and only if $I - \{k\}$ is heavy for M^+. Similarly for $I \in A_1$, $m_I = m^-_{I-\{k\}} + m_k$. Since the sum of all elements of M^- is $2m_k$ less than the sum of all elements of M, we conclude $I \in A_1$ is heavy for M if and only if $I - \{k\}$ is heavy for M^-. The lemma follows. ∎

Now, for any integers a and b, $P(a+b) \geq \min\{P(a), P(b)\}$. Therefore by the above lemma,

$$\begin{aligned}\Phi(M) &= 1 + P(f_M(-1)) = 1 + P(f_{M^+}(-1)) + (-1)^{m_k} f_{M^-}(-1) \\ &\geq 1 + \min(P(f_{M^+}(-1)), P(f_{M^-}(-1))) \\ &= \min\{\Phi(M^+), \Phi(M^-)\},\end{aligned}$$

proving (4.2).

Finally, let $n = 2h+1$ and suppose M is the game position consisting of n 1's. Then I is heavy exactly when $|I| \geq h+1$, so

$$\begin{aligned}f_M(-1) &= \sum_{j=h+1}^{2h+1} (-1)^j \binom{2h+1}{j} = \sum_{j=h+1}^{2h+1} (-1)^j \left(\binom{2h}{j-1} + \binom{2h}{j}\right) = \\ &= \binom{2h}{h}(-1)^{h-1}.\end{aligned}$$

Since $B(2h+1) = B(h) + 1$, (4.4) follows from

Proposition 4.3. $P\left(\binom{2h}{h}\right) = B(h)$.

Proof. By induction on h; for $h = 1$ the result is trivial. For $h > 1$, we apply the induction hypothesis to get

$$P\left(\binom{2h}{h}\right) = P\left(\frac{2h(2h-1)}{h^2}\binom{2h-2}{h-1}\right) = P\left(\binom{2h-2}{h-1}\right) + 1 - P(h)$$
$$= B(h-1) + 1 - P(h) = B(h),$$

where the last equality follows from the fact that the binary expansion of $h - 1$ has a 1 in the last $\varrho(h)$ places and a 0 in position $\varrho(h) + 1$, and the binary expansion for h is obtained by complementing these bits. ∎

References

[1] M. R. Best, P. van Emde Boas, and H. W. Lenstra, Jr.: A sharpened version of the Aanderaa–Rosenberg conjecture, *Math. Centrum, Amsterdam*, 1974.

[2] B. Bollobás: *Extremal Graph Theory*, Academic Press, 1978.

[3] M. J. Fisher, and S. L. Salzberg: Finding a majority among n votes, *Journal of Algorithms,* **3** (1982), 375–379.

[4] J. Misra, and D. Gries: Finding repeated elements, *Science of Computer Programming,* **2** (1982), 143–152.

[5] R. Rivest, and J. Viullemin: On recognizing graph properties from adjacency matrices. *Theor. Comp. Sci.* **3** (1976/77), 371–384.

Michael E. Saks

Department of Mathematics and RUTCOR,
Rutgers University, New Brunswick, NJ 08903

and

Bell Communications Research,
Morristown, NJ 07960.
U. S. A.

Current Address:

Department of Computer Science
and Engineering, Mail Code 0114,
University of California, San Diego,
La Jolla, CA 92093-0114.
U. S. A.
saks@ucsd.edu

Michael Werman

Institute for Mathematics and its Applications,
University of Minnesota,
Minneapolis, Minn. 55455.
U. S. A.

Current Address:

Department of Computer Science,
Hebrew University,
Givat Ram Jerusalem.
Israel
werman@humus.bitnet