ARTICLE IN PRESS

Discrete Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Finding a non-minority ball with majority answers

Dániel Gerbner [a], Balázs Keszegh [a], Dömötör Pálvölgyi [b], Balázs Patkós [a,c], Máté Vizer [a,*], Gábor Wiener [d]

[a] MTA Alfréd Rényi Institute of Mathematics, Reáltanoda u. 13-15 Budapest, 1053, Hungary

[b] Eötvös Loránd University, Department of Computer Science, Pázmány Péter sétány 1/C Budapest, 1117, Hungary

[c] MTA-ELTE Geometric and Algebraic Combinatorics Research Group, Pázmány Péter sétány 1/C Budapest, 1117, Hungary

[d] Department of Computer Science and Information Theory, Budapest University of Technology and Economics, 1117 Budapest Magyar tudósok körútja 2, Hungary

ARTICLE INFO

Article history:
Received 24 September 2015
Received in revised form 6 November 2016
Accepted 9 November 2016
Available online xxxx

Keywords:
Combinatorial search
Majority
Median

ABSTRACT

Suppose we are given a set of n balls $\{b_1, \dots, b_n\}$ each colored either red or blue in some way unknown to us. To find out some information about the colors, we can query any triple of balls $\{b_{i_1}, b_{i_2}, b_{i_3}\}$. As an answer to such a query we obtain (the index of) a *majority ball*, that is, a ball whose color is the same as the color of another ball from the triple. Our goal is to find a *non-minority ball*, that is, a ball whose color occurs at least $\frac{n}{2}$ times among the n balls. We show that the minimum number of queries needed to solve this problem is $\Theta(n)$ in the adaptive case and $\Theta(n^3)$ in the non-adaptive case. We also consider some related problems.

1. Introduction

This paper deals with search problems where the input is a set of n balls, each colored in some way unknown to us and we have to find a ball possessing a certain property (or show that such a ball does not exist) by asking certain queries. Our goal is to determine the minimum number of queries needed in the worst case. It is possible that the queries are all fixed beforehand (in which case we speak of *non-adaptive* search) or each query might depend on the answers to the earlier queries (in which case we speak of *adaptive* search). We say that a ball b is a *majority ball* of a set A if there are more than $\frac{|A|}{2}$ balls in the set that have the same color as b. Similarly, a ball b is a *non-minority ball* of a set A if there are at least $\frac{|A|}{2}$ balls in the set that have the same color as b. Note that these two notions are different if and only if n is even and $\frac{|A|}{2}$ balls are colored of the same color, in which case each of these balls is a non-minority ball and there are no majority balls (moreover, if there are only two colors, then the size of the other color class is also $\frac{|A|}{2}$, thus every ball is a non-minority ball). For more than two colors it is possible that even non-minority balls do not exist. A ball b is said to be a *plurality ball* of a set A if the number of balls in the set with the same color as b is greater than the number of balls with any of the other colors. In this paper we focus on the case of just two colors.

The most natural non-trivial question is the so-called *majority problem* which has attracted the attention of many researchers. In this problem our goal is to find a majority ball (or show that none exists), such that the possible queries are pairs of balls $\{b_1, b_2\}$ and the answer tells us whether b_1 and b_2 have the same color or not. In the adaptive model,

* Corresponding author.
E-mail addresses: gerbner.daniel@renyi.mta.hu (D. Gerbner), keszegh.balazs@renyi.mta.hu (B. Keszegh), dom@cs.elte.hu (D. Pálvölgyi), patkos.balazs@renyi.mta.hu, patkosb@cs.elte.hu (B. Patkós), vizer.mate@renyi.mta.hu (M. Vizer), wiener@cs.bme.hu (G. Wiener).

http://dx.doi.org/10.1016/j.dam.2016.11.012
0166-218X/© 2016 Published by Elsevier B.V.

Fisher and Salzberg [9] proved that $\lfloor 3n/2 \rfloor - 2$ queries are necessary and sufficient for any number of colors, while Saks and Werman [15] showed that if the number of colors is known to be two, then the minimum number of queries is $n - b(n)$, where $b(n)$ is the number of 1's in the binary representation of n (simplified proofs of the latter result were later found, see [2,12,16]). In the non-adaptive model with two colors, it is easy to see that the minimum number of queries needed is $n-1$ if n is even and $n-2$ if n is odd.

There are several variants of the majority problem [1]. The plurality problem, where we have to find a plurality ball (or show that none exists) was considered, among others, in [1,8,10]. Another possible direction is to use sets of size greater than two as queries [6,5].

The main model studied in this paper is the following. In the original comparison model the answer to the query $\{b_1, b_2\}$ can be interpreted as the answer to the question whether there is a majority ball in the subset $\{b_1, b_2\}$. If the answer is yes, then obviously both b_1 and b_2 are majority balls. Therefore we obtain a generalization of the comparison model if for any query that is a subset of the balls the answer is either (the index of) a majority ball, or that there is no majority ball in the given subset (which cannot be the case if the size of the subset is odd and there are only two colors). We study this model in case of two colors, and mostly when only queries of size three are allowed, although we also prove some results for greater query sizes.

Unfortunately, even asking all triples cannot guarantee that we can solve the majority problem for two colors. Suppose we have an even number of balls that are partitioned into two sets of the same size, X and Y, and suppose that the answer for any triple T is a ball from $T \cap X$ if and only if $|T \cap X| \geq 2$. In this case we cannot decide whether all balls have the same color or all balls in X are red, but all balls in Y are blue. In the former case all balls are majority balls, while in the latter there exists no majority ball.

Because of this, our aim will be to show a non-minority ball (which always exists if there are only two colors). Let us assume the balls are all red or blue and all queries are of size q. We will denote the minimum number of queries needed (in the worst case) to determine a non-minority ball by $A_q(n)$ in the adaptive model and by $N_q(n)$ in the non-adaptive model.

At first sight the model we have just introduced seems to be rather artificial. Let us, however, state a more natural problem that is equivalent to this model.

Suppose that our input is a binary sequence of length n, i.e., n numbers such that each is either 0 or 1. Our task is to find a median element, such that the queries are *odd* subsets of the input elements and the answer is one of the median elements of the subset. Let us assume queries are of size q. Denote the minimum number of queries needed in the worst case to determine a median element by $A_q^{med}(n)$ in the adaptive model and by $N_q^{med}(n)$ in the non-adaptive model.

Proposition 1. *$A_{2l+1}(n) = A_{2l+1}^{med}(n)$ and $N_{2l+1}(n) = N_{2l+1}^{med}(n)$.*

Proof. If we replace 0 and 1 by red and blue, then the median elements of any set are exactly the non-minority balls of the set. □

We obtain a natural generalization that also works for even sized subsets if the answer is the tth element for some fixed t. More precisely, for a query Q, the answer may be a if and only if there exist $t-1$ elements $e \in Q \setminus \{a\}$, such that $e \geq a$ and $|Q| - t$ elements $e' \in Q \setminus \{a\}$, such that $e' \leq a$. Note that in this model there might be more than one valid answer to a given query. This can be outruled by assuming that all elements are different (in which case we do not deal with just the numbers 0 and 1, obviously). This approach was proposed by G.O.H. Katona and studied by Johnson and Mészáros [13]. They have shown that if all elements are different, then they can be almost completely sorted[1] using $O(n \log n)$ queries in the adaptive model and $O(n^{q-t+1})$ queries in the non-adaptive model and both results are sharp. However, their algorithms fail if not all elements are different. Our results imply that the same bound holds in the adaptive model with no restriction. However, the bound in the non-adaptive model cannot be extended to the general case. We discuss our related results in Section 4.

To state our results concerning $A_3(n)$ and $N_3(n)$ we introduce the following notations. We write $[n] = \{1, 2, \ldots, n\}$ for the set of the first n positive integers and the set of balls is denoted by $B = [n]$. For a set S, the set of its k-subsets will be denoted by $\binom{S}{k}$. Let $\mathcal{Q} \subseteq \binom{B}{3}$ be a query set. Then for any ball $b \in B$ let $d_{\mathcal{Q}}(b) = |\{Q \mid b \in Q \in \mathcal{Q}\}|$ denote the *degree* of b in $\mathcal{Q}$ and for any two balls $b_i, b_j \in B$ let $d_{\mathcal{Q}}(b_i, b_j) = |\{Q \mid \{b_i, b_j\} \subset Q \in \mathcal{Q}\}|$ denote the *co-degree* of b_i and b_j in $\mathcal{Q}$. Furthermore, let us write $\delta(\mathcal{Q}) = \min\{d_{\mathcal{Q}}(b) \mid b \in B\}$ and $\delta_2(\mathcal{Q}) = \min\{d_{\mathcal{Q}}(b_i, b_j) \mid b_i, b_j \in B\}$.

Throughout the paper we use the following standard notation to compare the asymptotic behavior of two functions $f(n)$ and $g(n)$. We write $f(n) = o(g(n))$ if $\lim_n \frac{f(n)}{g(n)} = 0$ holds. We write $f(n) = O(g(n))$ ($f(n) = \Omega(g(n))$) if there exists a positive constant C such that $f(n) \leq Cg(n)$ ($f(n) \geq Cg(n)$) holds for all values of n and we write $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$ hold. Sometimes, the function f might have two variables k and n. Then $f(k, n) = O_k(g(n))$ means that for every k there exists a constant C_k such that $f(k, n) \leq C_k g(n)$ holds for all values of n. Finally, we write $f(n) = \tilde{O}(g(n))$ if there exist positive constants C and k such that $f(n) \leq Cg(n) \log^k n$ holds.

Theorem 2. *$A_3(n) = O(n)$.*

[1] Note that the $t-1$ largest and the $q-t$ smallest elements cannot ever be differentiated with such questions, so we only want to determine these and sort the rest.

Please cite this article in press as: D. Gerbner, et al., Finding a non-minority ball with majority answers, Discrete Applied Mathematics (2016), http://dx.doi.org/10.1016/j.dam.2016.11.012