Mathematical Problems of Computer Science 22, 2001, 96–99.

On One Characterizer of Greedy Algorithm of the Solution of the Problem of Finding a Checking Test for 0-1 Matrices

Edward V. Yeghiazaryan, Vahagn R. Tonoyan

Abstract

This article discusses the behavirour of greedy algorithm for the problem of finding a checking test for 0-1 matrix of order $m \times n$. It proves that for any number $\varepsilon > 0$ and in the case $n = o(m)$ the greedy algorithm finds a checking test for almost all matrices with a complexity equal to or less than the number $\log_{2(1-\varepsilon)^{-1}} n + c$, where c is a constant number.

The discussion of some questions in discrete mathematics (finite set-covering, clarification of the condition of a scheme) draws to the following problem.

The matrix T of elements 0-1 is given, with columns each containing at least one 1. Each column of the submatrix, derived from the set of rows of the matrix also contains one 1. The set of the rows of the matrix is called a checking test. We will call number of rows of the submatrix the complexity of the checking test.

The problem is the following: for a given matrix T it is necessary to find a checking test with the minimum complexity, which will be called the optimal or minimal test for the matrix T.

Let's define some concepts.

$B^m = \{\tilde{\alpha} / \tilde{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_m)\}$ is the set of vertices of a unit cube of size m. We will call the number $\|\tilde{\alpha}\| = \sum_{i=1}^{m} \alpha_i$, the norm of the collection $\tilde{\alpha} \in B^m$.

Suppose $M(m)$ is some finite set, and the number of elements of the set is denoted by $|M(m)|$. Let's also suppose that some property Q is given and each element of the set $M(m)$ either satisfies or not to that property. We will denote the subset of elements of the set $M(m)$ satisfying to the property Q by $M(Q,m)$. We will call the relationship $\frac{|M(Q,m)|}{|M(m)|}$ the part of elements of the set M satisfying to the property Q.

We will say that almost all the elements of the set $M(m)$, satisfy to the property Q, if

$$\lim_{m \to +\infty} \frac{|M(Q,m)|}{|M(m)|} = 1.$$

Let's prove some auxiliary statements.

Lemma 1: For almost all collections $\tilde{\alpha} \in B^m$ the following is true: $\|\tilde{\alpha}\| \sim \frac{m}{2}$.

Let's observe B^m as a space of events, where the probability of occurrence of each event $\tilde{\alpha} \in B^m$ is $\frac{1}{2^m}$. Let's define the arbitrary value $u = u(\tilde{\alpha})$ as $|\tilde{\alpha}| = k$. It acquires the values $k = 0,1,2,...,n$ by the probability $P(u = k) = \frac{1}{2^m}\binom{m}{k}$.

Let's calculate Mu mathematical expectation and Du dispersion of the arbitrary value u.

Let's observe the Newton Bynum $(1+x)^m = \sum_{k=0}^{m}\binom{m}{k}x^k$. We will take the derivative and multiple the result by x:

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

After taking the derivative once again, we will get

$$m(1+x)^{m-1} + m(m-1)x(1+x)^{m-2} = \sum_{k=1}^{m} k^2\binom{m}{k}x^{k-1}.$$

By plugging $x = 1$ into the resulted recurrences, it will be easy to see that

$$Mu = \sum_{k=0}^{m} kp(u = k) = \sum_{k=0}^{m} k\frac{1}{2^m}\binom{m}{k} = \frac{m}{2},$$

$$Du = Mu^2 - (Mu)^2 = \sum_{k=0}^{m} k^2 \cdot \frac{1}{2^m}\binom{m}{k} - \frac{m^2}{4} = \frac{m(m-1)}{4} + \frac{m}{2} - \frac{m^2}{4} = \frac{m}{4}:$$

Using the Chebishev inequality for any number $t \geq 0$ we will get

$$P(|u - Mu| \geq t) \leq \frac{Du}{t^2}.$$

Let's apply this inequality to the above-mentioned arbitrary value u. We will have

$$P(\left|u - \frac{m}{2}\right| \geq \frac{m}{\log_2 m}) \leq \frac{\log_2 m}{4m},$$

or as the same

$$P(\left|u - \frac{m}{2}\right| < \frac{m}{\log_2 m}) \geq 1 - \frac{\log_2 m}{4m}.$$

Therefore $P(u \sim \frac{m}{2}) \to 1$, when $m \to \infty$. This fact implies the proof of the lemma 1.

Lemma 2: If $\varphi(m) \to \infty$, when $m \to \infty$ and $\varphi^2(m) = o(m)$, then $\|\tilde{\alpha}\| > \frac{m}{2}(1 - \frac{1}{\varphi(m)})$ for almost all collections $\tilde{\alpha} \in B^m$. Moreover, the part of collection for which the above-mentioned inequality holds is at least $1 - \frac{\varphi^2(m)}{m}$.

By plugging $t = \frac{m}{2\varphi(m)}$ into the above-mentioned Chebishev inequality, we will get

$$P(\left|u - \frac{m}{2}\right| \geq \frac{m}{2\varphi(m)}) \leq \frac{\varphi^2(m)}{m}.$$

Therefore,

$$P(\left|u-\frac{m}{2}\right|<\frac{m}{2\varphi(m)})\geq 1-\frac{\varphi^2(m)}{m}.$$

We can also imply from here that the inequality

$$P(u>\frac{m}{2}(1-\frac{1}{\varphi(m)}))\geq 1-\frac{\varphi^2(m)}{m}$$

holds and the lemma 2 is proven.

We will denote the set of all matrices of $m\times n$ order containing the elements 0 or 1 by T_{mn}.

Lemma 3: If $n=o(m)$, then it is possible to mention $\varphi(m)\to\infty$, when $m\to\infty$, and such a function, satisfying to the conditions $\varphi^2(m)=o(m)$, that for almost all matrices $T\in T_{mn}$ each column contains at least $\frac{m}{2}(1-\frac{1}{\varphi(m)})$ many 1s.

We have $\frac{m}{n}=\psi(m)\to\infty$, when $n\to\infty$. Let's select some function φ, which satisfies the conditions $\varphi^2(m)=o(\psi(m))$, $\varphi^2(m)=o(m)$ and $\varphi(m)\to\infty$, when $m\to\infty$.

We will denote by z_{mn} the part of such matrices $T\in T_{mn}$, for which each column contains at least $\frac{m}{2}(1-\frac{1}{\varphi(m)})$ many 1s.

It's easy to see that $z_{mn}>(1-\frac{\varphi^2(m)}{m})^n$.

As the inequality $(1-\frac{1}{x})^x>e^{-1-\frac{1}{x}}$ holds for the case $x\geq 2$, then $z_{mn}>e^{-\frac{n}{m}\varphi^2(m)(1+\frac{\varphi^2(m)}{m})}$.

It is easy to check that in the case the above mentioned conditions are true $\frac{n}{m}\cdot\varphi^2(m)(1+\frac{\varphi^2(m)}{m})\to 0$, when $m\to\infty$, and, therefore, $z_{mn}\to 1$, when $m\to\infty$.

Now let's define the greedy algorithm Θ of finding a checking test.

The algorithm Θ includes the row with the least number and containing the most number of 1s of the set T in the checking test composed step by step. Afterwards, it deletes that row from the matrix T, as well as all the columns passing through the elements 1 of that row. The mentioned action is repeated with the resulted matrix unless all the columns of the matrix are deleted.

Theorem 1. For any $\varepsilon>0$ number in the case of $n=o(m)$ the greedy algorithm finds a checking test for almost all matrices $T\in T_{mn}$, which has a complexity not exceeding the number $\log_{2(1-\varepsilon)^{-1}} n+c$, where c is a constant number.

Suppose $n=o(m)$. Let's select such function $\varphi(m)$, that the conditions of the lemma 3 are satisfied. Suppose $T\in T_{mn}$ is a matrix, each column of which contains at least $m\gamma=\frac{m}{2}(1-\frac{1}{\varphi(m)})$ many 1s.

During the work of the algorithm Θ rows and columns are sequentially deleted from the matrix T. Let's notice that after such deletion each column of the resulted matrices will also contain at least $m\gamma$ many 1s.

We will denote the number of columns of matrix T_r, resulted after r steps of work of the algorithm Θ, by $n\delta_r$. As the matrix T_r contains at least $n\delta_r m\gamma$ many 1s, then the selected row in the $r+1$ step of the algorithm Θ will contain at least $n\delta_r\gamma$ many 1s.

Therefore $n\delta_{r+1} \le n\delta_r - n\delta_r\gamma$, or as the same $\delta_{r+1} \le \delta_r(1-\gamma)$.

Thus $\delta_r \le (1-\gamma)^r$, $r = 1,2,3,...$ and $\delta_0 = 1$.

It's not difficult to see that for any value $r = 1,2,3,...$ the number of rows selected by the algorithm - $\Theta(T)$, does not exceed the number $r + n\delta_r$.

Let's observe the function $f(r) = r + n(1+\gamma)^r$.

We have $f'(r) = 1 + n(1+\gamma)^r \ln(1-\gamma)$ and $f''(r) > 0$.

We will find the value of $(1-\gamma)^{r_0} = -\dfrac{1}{n\ln(1-\gamma)}$ from the equation $f'(r) = 0$.

It's clear that

$$\Theta(T) \le r_0 - \frac{1}{\ln(1-\gamma)},$$

and, moreover, $r_0 = \log_{(1-\gamma)^{-1}} n\ln(1-\gamma)^{-1} = \log_{(1-\gamma)^{-1}} n + \log_{(1-\gamma)^{-1}} \ln(1-\gamma)^{-1}$.

For a given number $\varepsilon > 0$ we will select such value of m, that the following condition holds $\dfrac{1}{\varphi(m)} < \varepsilon$. In this case $\gamma = \dfrac{1}{2}(1 - \dfrac{1}{\varphi(m)}) > \dfrac{1}{2}(1-\varepsilon)$. It's easy to check that it is possible to select such constant number c, that the following inequality also holds $\Theta(T) \le \log_{2(1+\varepsilon)^{-1}} n + c$.

References

[1] Э.В.Егиазарян, "Длина минимального проверяющего теста для некоторого класса таблиц", Межвуз. сб. "Прикладная математика", вып. 2. Ереван, с. 34-40, 1981.

[2] Ю.Ш.Коспанов, "Об одном алгоритме построения достаточно простых тестов", Дискретный анализ, Сборник трудов, Институт математики СО АН СССР, выпуск 8, с. 43-47, 1966.

[3] Э.И.Нечипорук, "О сложности вентильных схем, реализующих булевские матрицы с неопределенными элементами", ДАН СССР, т. 163, 1, с. 40-43, 1965.

0-1 մատրիցների համար ստուգող կամ ախտորոշող թեստ գտնելու խնդրի լուծման ագահ ալգորիթմի մի բնութագրիչի մասին

Է.Վ. Եղիազարյան, Վ.Ռ. Տոնոյան

Ամփոփում

Աշխատանքում հետազոտվում է ագահ ալգորիթմի վարքը $m \times n$ կարգի 0-1 մատրիցի ստուգող թեստ գտնելու խնդրի համար: Ապացուցվում է, որ ցանկացած $\varepsilon > 0$ թվի համար $n = o(m)$ դեպքում ագահ ալգորիթմը համարյա բոլոր մատրիցների համար գտնում է ստուգող թեստ, որի բարդությունը չի գերազանցում $\log_{2(1-\varepsilon)^{-1}} n + c$ թվին, որտեղ c-ն հաստատուն է: