A Fast Algorithm for Finding Ideal Response Pattern Vectors of any Test Q-Matrix

Hsiang-Chuan Liu, Hsun-chih Kuo, Hsien-Chang Tsai

1, First Author
Department of Biomedical Informatics, Asia University, lhch@asia.edu.tw
2
Department of Statistics, National Chengchi University, seankuo@nccu.edu.tw
3, Corresponding Author
Department of Biolog, National Changhua University of Education, tsaihc00@gmail.com

Abstract
In this paper, based on Gauss’ symbol function, a novel fast algorithm for finding the ideal item response pattern vector to correspond to a given attribute state vector of any test Q-matrix is proposed. It is simpler and faster than Tatsuoka’s Boolean description function algorithm. Furthermore, a novel discriminant validity index of any test Q-matrix is also proposed, from now on, we can find the better test Q-matrix to improve the cognitive analyses of educational tests by using this validation index.

Keywords: Q-Matrix, Test Q-Matrix, Attribute State Vector, Ideal Response Pattern, Boolean Description Function

1. Introduction

Knowledge management and diagnosis is an important issue in education. Some well-known theories or models about this issue were proposed such as Ordering theory [1], Item relational structure theory [2-3], Q-matrix theory [4-7], Item Response Theory [8], Cognitive diagnostic assessment Theory [9] Concept map theory [10], Attribute Management Method Theory [11] and Knowledge Sharing Model [12]. In this paper, we will focus on Q-matrix theory.

For cognitive diagnosis, Airasian and Bart proposed the Ordering theory (OT) [1], for more considering the item relationship, Takeya proposed the improve theory, named Item Relational Structure theory (IRS) [2]. Furthermore, our previous work [3] provided an improved IRS theory by using the dynamic threshold limit value based on the empirical distribution critical value of all the values of the relational structure indices between any two items, it is more sensitive and effective than before. However, the OT or IRS may not provide efficient items compatible with the cognitive structure, therefore, Tatsuoka [4-5] proposed her cognition diagnosis method based on Q-matrix theory, called Rule Space Model (RSM), Leighton, Gier and Hunka [6] proposed the Attribute Hierarchy Method (AHM) for cognitive assessment based on Q-matrix theory as well, Liu [7] provided the theoretical approach to reduced Q-matrix theory based on Boolean matrix operations. All of the Q-matrix theories emphasize that exam questions can be described by specific cognitive attributes, and they can include other different attributes that an examinee must possess to solve a test item. The relations of all specific attributes and all of the possible items can be represented by an attributes-items incident matrix, called Q-matrix. Furthermore, we can obtain the reduced Q-matrix which contains all of the efficient items fitted for the requirement of the attributes structure, the item bank corresponded to the reduced Q-matrix is an efficient item bank, and then, any test Q-matrix, which is a sub-matrix of the reduced Q-matrix fitted for the requirement of the attributes structure, can be obtained as well, and the reduced Q-matrix itself is also a test Q-matrix.

In this paper, based on Gauss’ symbol function, we consider to propose a novel fast algorithm for finding the ideal item response pattern vector to correspond to a given attribute state vector of a test Q-matrix. It is simpler and faster than Tatsuoka’s Boolean description function algorithm. Furthermore, we consider to provide a novel validation index of any test Q-matrix for finding the better test Q-matrix to improve the cognitive analyses of educational tests.
2. Attribute-item incidence matrix, Q-matrix

2.1. Attribute structure and its matrix representation

**Definition 1.** Precondition attribute of an attribute. [4, 7]

Let $A = \{a_i\}_{i=1}^m$ be the set of $m$ cognitive skills, called attributes decided by experts before the test. (i) If $a_i, a_j \in A$, any examinee before master the attribute $a_j$, he must master $a_i$, then $a_i$ is called a precondition attribute of $a_j$, denoted as $a_i \rightarrow a_j$, and $a_i \not\rightarrow a_j$ otherwise. (1)

(ii) If $a_i \rightarrow a_j$, there is no any attribute $a_k \in A$ such that $a_i \rightarrow a_k, a_k \rightarrow a_j$, then $a_j$ is called a direct precondition attribute of $a_j$, denoted as $a_i \rightarrow a_j$, and $a_i \not\rightarrow a_j$ otherwise. (2)

**Definition 2.** The adjacency matrix of the attributes set. [4, 7]

If $A = \{a_i\}_{i=1}^m$ is the attributes set, then the Boolean matrix $A_m = [a_{ij}]_{m \times m}$ is called the adjacency matrix of $A$, where $a_{ij} = 1$, if $a_i \rightarrow a_j$, and $a_{ij} = 0$ otherwise (3)

**Definition 3.** The reachability matrix of the attributes set [4, 7]

If $A_m$ is the adjacency matrix of the attributes set $A = \{a_i\}_{i=1}^m$, and the Boolean matrix $R_m$ satisfying

$$R_m = (A_m + I_m)^{i+1} = (A_m + I_m)^i, i \leq m$$

Where the addition operators are Boolean addition operation, then $R_m$ is called the reachability matrix of $A$.

2.2. Example

**Example 1.** Let $A = \{a_1, a_2, a_3, a_4\}$ be the set of 4 cognitive attributes, the graph of attributes structure of $A$ is shown in the following figure:

![Figure 1](image)

The adjacency matrix, $A_4$, and the reachability matrix, $R_4$, of the attributes set $A = \{a_1, a_2, a_3, a_4\}$ are shown as follows:

$$A_4 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad R_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$ (5)
2.3. Attribute- item incident matrix, Q- matrix

**Definition 4. Prerequisite attribute of an item** [4,7]
Let $\{a_i\}^m_{i=1}$ be the attribute set, $\{I_j\}^n_{j=1}$ be the item set of a test. If the examinee can answer item $I_j$ correctly, he must master attribute $a_i$ first, then $a_i$ is called a precondition attribute of item $I_j$, denoted as $a_i \rightarrow I_j$, and $a_i \not\rightarrow I_j$ otherwise.

**Definition 5. Attribute- item incident matrix, Q- matrix** [4,6,7]
Let Boolean matrix $Q = [q_{ij}]_{m \times n}$ represent the incidence matrix of $\{a_i\}^m_{i=1}$ and $\{I_j\}^n_{j=1}$, where $q_{ij} = 1$, if $a_i \rightarrow I_j$, and $q_{ij} = 0$ otherwise. (6)

**Example 2.** In Example 1., the incidence matrix of $\{a_i\}^4_{i=1}$ and $\{I_j\}^4_{j=1}$ is:

\[
Q_{4 \times 4} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

2.4. Attribute- efficient item incident matrix, reduced Q- matrix

**Definition 6. Efficient item, inefficient item.** [4,7]
Let $Q = [q_{ij}]_{m \times n}$ be the incidence matrix of $\{a_i\}^m_{i=1}$ and $\{I_j\}^n_{j=1}$, If $a_s, a_t, a_j \in A \Rightarrow a_j \Rightarrow q_{ij} \geq q_{ik}$, then item $j$ is efficient, otherwise, it is inefficient.

**Definition 7. Attribute- efficient item incident matrix, Reduced Q- matrix.** [4,7]
(i) After deleting the inefficient items from the Q- matrix, the new matrix is called reduced Q- matrix, denoted by $Q_R$.
(ii) Let the all possible items set corresponded to $Q = [q_{ij}]_{m \times n}$ be denoted as

\[
\{I_j\} = \{I_j\}^{n-1}_{j=1}
\]

Then the all efficient items set corresponded to $Q_R$ can be denoted as

\[
\{I_{j_e}\} = \{[q_{ij}]_{m \times n} \rightarrow a_j \Rightarrow q_{ij} \geq q_{ik} | i, j = 1, 2, ..., 2^m - 1 \}
\]

**Example 3.** In Example 2., the reduced Q- matrix of $A$ and $\{I_{j_e}\}$ is:

\[
Q_R = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
\end{bmatrix}
\]
2.5. Test Q- matrix, $Q_T$

**Definition 8. Test Q- matrix, efficient test**

(i) For a given attributes set, $A = \{a_i\}_{i=1}^m$, with the reduced Q- matrix, $Q_R$,

A matrix, $Q_T$, is called a test Q- matrix of $A$ or a test Q- matrix, if each of its column is a column of $Q_R$, and each of its row is not a zero vector.

(ii) A test corresponded to a test Q- matrix is called an efficient test of $A$ or an efficient test.

**Theorem 1.** $Q_R$ is a test Q- matrix.

**Example 4.** In Example 3,

\[
Q_1 = \begin{bmatrix}
1 & 1 & 1 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix},
\]

\[
Q_2 = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 \\
\end{bmatrix},
\]

\[
Q_3 = \begin{bmatrix}
1 & 1 & 1 & 1 \\
0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 \\
\end{bmatrix},
\]

Then $Q_1$ and $Q_2$ are two test Q- matrices, and $Q_3$ is not a test Q- matrix.

$I_{Q_1}$ and $I_{Q_2}$ are two efficient tests, and $I_{Q_3}$ is not an efficient test.

Note that each item of $I_{Q_3}$ is efficient, but $I_{Q_3}$ is not an efficient test.

3. Ideal response pattern vector [4-7]

3.1. Definitions about ideal response pattern vector

**Definition 9. Attribute state vector, possible attribute state vector**

Let $A = \{a_i\}_{i=1}^m$ be a given attribute set $A = \{a_i\}_{i=1}^m$.

(i) The Boolean vector $\alpha_k = (\alpha_{ik})_{i=1}^m$ is called the attribute state vector of an examinee $k$ for $A$,

where $\alpha_{ik} = 1$, if the examinee $k$ masters $a_i$, and $\alpha_{ik} = 0$ otherwise, $k = 1, 2, ..., N$.

(ii) The Boolean vector $\alpha^* = (\alpha^*_i)_{i=1}^m$ is called a possible attribute state vector of a possible examinee for $A$, where $\alpha^*_i = 1$, if the possible examinee masters $a_i$, and $\alpha^*_i = 0$ otherwise.

(iii) $\alpha^*_A = \left\{ \alpha^*_i \mid \alpha^*_i = (\alpha^*_i)_{i=1}^m, \alpha^*_i \in \{0, 1\} \right\}$ is called the possible different attribute state vector set for $A$, satisfying $|\alpha^*_A| = 2^m$.

**Definition 10. Response pattern vector, ideal response pattern vector**
Let $Q_T = \left[ q_{ij} \right]_{m \times n}$ be a test Q-matrix of attribute set $A = \{ a_i \}_{i=1}^m$ and an efficient test $I = \{ i_j \}_{j=1}^n$

(i) The vector $r_k = \left( r_{ik} \right)_{i=1}^n$ is called a response pattern vector of an examinee $k$ for $Q_T$, where $r_{ik} = 1$ if the examinee $k$ answers item $i_j$ correctly, and $r_{ik} = 0$ otherwise, $i = 1, 2, \ldots, n$, $k = 1, 2, \ldots, N$

(ii) $R^*_k (Q_T) = \{ \mathbf{L}^* | \mathbf{L}^* = (r_{ij}^*)_{j=1}^n, r_{ij}^* \in \{0,1\} \}$ is called the possible different response pattern vector set for $Q_T$, satisfying $\left| R^*_n \right| = 2^n$.

(iii) The vector $L_k^+ = \left( L_{ik}^+ \right)_{i=1}^m$ fitted for the requirement of the attributes structure is called the ideal response pattern vector of an examinee $k$ for $Q_T$, where $L_{ik}^+ = 1$ if the examinee $k$ answers item $i_j$ correctly without guessing, and $L_{ik}^+ = 0$ if the examinee $k$ does not answer item $i_j$ correctly without slipping, $j = 1, 2, \ldots, n$

(iv) The vector $P_k^+ = \left( P_{ij}^+ \right)_{i=1}^m$ fitted for the requirement of the attributes structure is called the ideal response pattern vector of a possible examinee, where $P_{ij}^+ = 1$ if the possible examinee answers item $i_j$ correctly without guessing, and $P_{ij}^+ = 0$ if the possible examinee does not answer item $i_j$ correctly without slipping, $j = 1, 2, \ldots, n$

(v) $R_{kn}^+ (Q_T) = \{ \mathbf{L}^+ | \mathbf{L}^+ = (r_{ij}^+)_{j=1}^n, r_{ij}^+ \in \{0,1\} \}$ is called the ideal response pattern vector set of $N$ well-known examinees for $Q_T$.

$R_{kp}^+ (Q_T) = \{ \mathbf{L}^+ | \mathbf{L}^+ = (r_{ij}^+)_{j=1}^n, r_{ij}^+ \in \{0,1\} \}$ is called the possible ideal response pattern vector set of possible examinees for $Q_T$.

3.2. Example of ideal response pattern vector

**Example 5.** In Example 4.

The possible different attribute state vector set is $\alpha_A^* = \left\{ \alpha_p^* | \alpha_p^* = \left( \alpha_{ip}^* \right)_{i=1}^4, \alpha_{ip}^* \in \{0,1\} \right\} = \left\{ \alpha_p^* | p = 1, 2, \ldots, 2^m \right\}$

Where

$\alpha_1^* = (0,0,0,0), \alpha_2^* = (0,0,0,1), \alpha_3^* = (0,0,1,0), \alpha_4^* = (0,1,0,0), \alpha_5^* = (1,0,0,0), \alpha_6^* = (0,0,1,1), \alpha_7^* = (0,1,0,0), \alpha_8^* = (0,1,0,1), \alpha_9^* = (1,1,0,1), \alpha_{10}^* = (1,1,0,0), \alpha_{11}^* = (0,1,1,1), \alpha_{12}^* = (1,0,1,1), \alpha_{13}^* = (1,0,1,0), \alpha_{14}^* = (1,1,1,1), \alpha_{15}^* = (1,1,1,0), \alpha_{16}^* = (0,1,1,1), \alpha_{17}^* = (1,1,1,0), \alpha_{18}^* = (1,1,0,0), \alpha_{19}^* = (1,1,0,1), \alpha_{20}^* = (0,1,1,0), \alpha_{21}^* = (0,1,1,1), \alpha_{22}^* = (1,1,1,1)$.

(15)

The possible different response vector set is $R_p^+ (Q_T) = \{ \mathbf{L}^+ | \mathbf{L}^+ = (r_{ij}^+)_{j=1}^n, r_{ij}^+ \in \{0,1\} \} = \{ \mathbf{L}^+ | p = 1, 2, \ldots, 2^m \}$

A Fast Algorithm for Finding Ideal Response Pattern Vectors of any Test Q- Matrix

Hsiang-Chuan Liu, Hsun-chih Kuo, Hsien-Chang Tsai
4. A fast algorithm for finding ideal response pattern vectors

How to find the ideal response pattern vectors in cognitive diagnosis analyses of educational tests is the most important issue. Based on the theory of lattice and Boolean algebra, Tatsuoka proposed her Boolean Description Function algorithm for finding ideal response pattern vectors [5]. However, her algorithm is too theoretical and complex to understanding for average person. In this paper, based on Gauss’ symbol function, a simple and fast algorithm for finding the ideal item response vectors to correspond to given attribute state vectors of a test Q-matrix, called Liu’s algorithm, is proposed as follows,

4.1. Liu’s algorithm for finding ideal response pattern vectors

**Theorem 2. Liu’s algorithm**

Let $Q_T = [q_{ij}]_{m \times n}$ be a test Q-matrix of attribute set $A = \{a_i\}_{i=1}^m$ and an efficient test $I = \{I_j\}_{j=1}^n$.

(i) If $\alpha_k = (\alpha_{ik})_{i=1}^m$ is a given attribute state vector of examinee $k$, then the corresponded ideal response pattern vector of the examinee $k$, $r^+_{jk} = \left(\sum_{i=1}^m q_{ij} \alpha_{ik}\right) / \sum_{i=1}^m q_{ij}$, $j = 1, 2, \ldots, n$, $k = 1, 2, \ldots, N$.

(ii) If $\alpha_p^+ = (\alpha_{jp}^+)_{i=1}^m$ is any given possible attribute state vector, then the corresponding possible ideal response pattern vector, $r^+_{jp} = \left(\sum_{i=1}^m q_{ij} \alpha_{jp}^+\right) / \sum_{i=1}^m q_{ij}$, $j = 1, 2, \ldots, n$.

Where $[x]$ is the Gauss’ symbol function, $[x]$ is the greatest integer that is less than or equal to $x$.

**Proof:**

(i) In a given test Q-matrix, $Q_T = [q_{ij}]_{m \times n}$, not lose the generality, let $q_{ij} = q_{2j} = 1$ and $q_{ij} = 0$, $i = 3, 4, \ldots, m$, then $a_i \mapsto I_j$, $a_2 \mapsto I_j$, $a$, $I_j$, $i = 3, 4, \ldots, m$. 

Where

\[
\begin{align*}
\text{If } q_{ij} = q_{2j} = 1, \text{ then } a_i \mapsto I_j, \text{ and } a_2 \mapsto I_j.
\end{align*}
\]
If attribute states of \( a_1 \) and \( a_2 \) are \( a_1 = a_2 = 1 \), then the examinee \( k \) can answer item \( I_j \) correctly without guessing, in other words, the value of his ideal response pattern of item \( I_j \) must be \( r_{jk}^* = 1 \), and \( r_{jk}^* = 0 \) otherwise, and the same results can be obtained by using the following function

\[
r_{jk}^* = \left[ \frac{\sum_{i=1}^{m} q_{ij} \alpha_{ik}^*}{\sum_{i=1}^{m} q_{ij}} \right] = \left[ \frac{1 \times \alpha_{4k} + 1 \times \alpha_{2k} + 0 \times \alpha_{5k} + \ldots + 0 \times \alpha_{mk}}{1 + 0 + 0 + \ldots + 0} \right]
\]

(19)

(a) if \( \alpha_{1j} = \alpha_{2j} = 1 \), then the ideal response pattern of item \( I_j \) for examinee \( k \) is \( r_{jk}^* = [1] = 1 \)

(b) if \( \alpha_1 = 1, \alpha_2 = 0 \) or, \( \alpha_1 = 0, \alpha_2 = 1 \) then \( r_{jk}^* = \left[ \begin{array}{c} 1 \\ 2 \end{array} \right] = 0 \)

(c) if \( \alpha_1 = \alpha_2 = 0 \), then \( r_{jk}^* = [0] = 0 \)

In generally, it can be completed by using the mathematical induction. (ii) the proof is the same as (i).

4.2. Example of Liu's algorithm for finding ideal response pattern vectors

Example 6. In Example 5.
Let \( \alpha_p^* = \{ \alpha_p^* \mid p = 1, 2, \ldots, 2^m \} \) be the possible different attribute state vector set,

\[
Q_i = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow (r_{jp}^*)^i = (\alpha_p^*)^i \circ Q_i = \left[ \begin{array}{c} \alpha_{ip}^* \alpha_{ip}^* + \alpha_{ip}^* \\ \alpha_{ip}^* + \alpha_{ip}^* + \alpha_{ip}^* \\ \alpha_{ip}^* + \alpha_{ip}^* \end{array} \right]
\]

(20)

(i)

\[
Q_i = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow (r_{jp}^*)^i = (\alpha_p^*)^i \circ Q_i = \left[ \begin{array}{c} \alpha_{ip}^* \alpha_{ip}^* + \alpha_{ip}^* \\ \alpha_{ip}^* + \alpha_{ip}^* + \alpha_{ip}^* \\ \alpha_{ip}^* + \alpha_{ip}^* \end{array} \right]
\]

(21)

\( \alpha_1 \Rightarrow L_1^* = (1, 0, 0, 0) \), \( \alpha_5 \Rightarrow L_5^* = (0, 0, 0, 0) \), \( \alpha_2 \Rightarrow L_2^* = (1, 0, 1, 0) \), \( \alpha_9 \Rightarrow L_9^* = (1, 1, 0, 0) \), \( \alpha_3 \Rightarrow L_3^* = (1, 0, 1, 1) \), \( \alpha_4 \Rightarrow L_4^* = (1, 1, 0, 1) \), \( \alpha_5 \Rightarrow L_5^* = (1, 1, 1, 0) \), \( \alpha_6 \Rightarrow L_6^* = (1, 1, 1, 1) \)

(ii)

\[
Q_i = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \Rightarrow (r_{jp}^*)^i = (\alpha_p^*)^i \circ Q_i = \left[ \begin{array}{c} \alpha_{ip}^* \alpha_{ip}^* + \alpha_{ip}^* + \alpha_{ip}^* \\ \alpha_{ip}^* + \alpha_{ip}^* + \alpha_{ip}^* \\ \alpha_{ip}^* + \alpha_{ip}^* + \alpha_{ip}^* \\ \alpha_{ip}^* + \alpha_{ip}^* + \alpha_{ip}^* \end{array} \right]
\]

(22)
5. A novel discriminant validity index of any test Q-matrix

How to find a better test Q-matrix from two different ones is the most important issue for testing. In this paper, a novel discriminant validity index of a test Q-matrix is proposed as follows:

**Definition 11. Liu’s discriminant validity index of a test Q-matrix**

Let $Q_T = [q_{ij}]_{mn}$ be a test Q-matrix, if $R^*_n \left( Q_T \right)$ is the possible different response pattern vector set for $Q_T$, and $R^*_{n,p} \left( Q_T \right)$ is the possible ideal response pattern vector set of possible examinees for $Q_T$. Then Liu’s discriminant validity index of $Q_T$, $d_{Liu} (Q_T)$, is defined as

$$d_{Liu} (Q_T) = \frac{\left| R^*_{n,p} \left( Q_T \right) \right|}{\left| R^*_n \left( Q_T \right) \right|}$$

(23)

Where $0 < d_{Liu} (Q_T) \leq 1$, the larger the better.

**Example 6.** In Example 5.

$$d_{Liu} (Q_{T_1}) = \frac{9}{16} \quad d_{Liu} (Q_{T_2}) = \frac{5}{16}$$

(24)

Since $d_{Liu} (Q_{T_1}) > d_{Liu} (Q_{T_2})$, therefore $Q_{T_1}$ is better than $Q_{T_2}$.

6. Conclusion

In this paper, based on Gauss’ symbol function, the Liu’s algorithm for finding the ideal item response pattern vector of any test Q-matrix is proposed. This new algorithm is simpler and faster than Tatsuoka’s Boolean description function algorithm. Furthermore, before then, there is no any discriminant validity index can be used for comparing two different test Q-matrices, a Liu’s discriminant validity index of any test Q-matrix is also proposed, from now on, we can find the better test Q-matrix to improve the cognitive analyses of educational tests by using this discriminant validity index.

7. Acknowledgment

This paper is partially supported by the National Science Council grant of Taiwan Government (NSC 100-2511-S-468-001).

8. References