Fuzzy Basis on Clustering of Knowledge Structure with Cognition Diagnosis for Algebra Learning

JENG-MING YIH
Department of Mathematics Education, National Taichung University
140 Min-Sheng Rd., Taichung City 403, Taiwan
Taiwan
yih@mail.ntcu.edu.tw

Abstract: The main purpose of this study is to provide an integrated method and algorithm for knowledge structure analysis and cognition diagnosis. Fuzzy clustering and algorithm of graphic structures analysis are combined so that features of knowledge structures of each cluster are clearly displayed. Concept structure analysis (CSA) could provide individualized knowledge structure. CSA algorithm is the major methodology and it is based on fuzzy logic model of perception (FLMP) and interpretive structural modeling (ISM). CSA could display individualized knowledge structure and clearly represent hierarchies and linkage among concepts for each examinee. Furthermore, fuzzy clustering is used to classify examinee based on response pattern of testing data. Therefore, CSA will be more effectively to display features of each cluster. In this study, the author provide the empirical data for concepts of linear algebra from university students. The results show that students of varied cluster own distinct knowledge structures. CSA combined with fuzzy clustering could be very feasible for cognition diagnosis. Based on the findings and results, some suggestions and recommendations for future research are provided.

Key-Words: cognition diagnosis, FLMP, fuzzy clustering, ISM, knowledge structure

1 Background and Motivation
Cognition diagnosis is an important issue related to cognition science and psychometrics. This issue will influence educational measurement so that quite a few psychometric researchers focus on methodological development [12]. Zadeh developed fuzzy theory and it flourishes methodologies in many fields [3] [4]. One of these fields is cognition diagnosis and it help represent knowledge structure [5] [9] [17]. It is a common viewpoint that human knowledge is stored in the form of structural relationship among concepts and their subordinate relationship is fuzzy, not crisp. Moreover, knowledge structures is unique for person according to the viewpoints of cognition science [13] [14]. For the feasibility of remedial instruction based on the cognition diagnosis, clustering method is needed so that students within the same cluster own similar knowledge structures and students among different clusters have the most variance on knowledge structures [16] [19] [22].

Therefore, an integrated approach for cognition diagnosis will be discussed. This integrated method combines concept structure analysis (CSA) and fuzzy clustering. CSA could display individualized knowledge structure and fuzzy clustering will help present traits of each cluster. An empirical testing data on concepts of linear algebra from university students will be analyzed and discussed.

2 Literature Review
CSA and fuzzy clustering are the main algorithms. CSA is based on the foundation of fuzzy logic model of perception (FLMP), interpretive structural modeling (ISM). All the related foundation for CSA and fuzzy clustering will be discussed as follows.

2.1 Fuzzy Logic Model of Perception
Fuzzy logic model of perception FLMP can be reparameterized as a simple 2-category logit model. Suppose there be a combination of two factors and there are levels within each factor. The fuzzy true value is to express the degree that the combination of two levels from distinct factor will support prototype [6]. The probability that this combination could be viewed as prototype is as follows [6].

\[ p(c_i, o_j) = c_i o_j [c_i o_j + (1-c_i)(1-o_j)]^{-1} \]  \hspace{1cm} (1)

2.2 Interpretive Structural Modeling
Interpretive structural modeling (ISM) is developed to deal with the relationship among elements within a
complex system [1]. ISM aims to arrange elements in the form of graphic structure and relationship [7]. Suppose there be \( K \) elements within a complex system and \( A = (a_{ij})_{K \times K} \) is the subordinate relationship matrix among \( K \) elements. \( a_{ij} = 1 \) means element \( i \) is the precondition of element \( j \); otherwise \( a_{ij} = 0 \) means element \( i \) is not the precondition of element \( j \). \( \hat{A} \) is the transitive closure of \( A \) and \( R \) is the reachability matrix of \( A \). With \( A \) and \( R \), the hierarchical graph of elements could be plotted. Figure 1 is an example of matrix \( A \) and its hierarchical graph is depicted in the form of structural relationship. There are three kinds of levels and they are bottom level, medium level and top level. In Figure 1, \( A_2 \) is the precondition of \( A_3 \) and \( A_4 \).

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

Fig. 1. The Linkage of Elements in Hierarchies

3 Integrated model and Procedure

The procedure of the integrated model is depicted in Figure 2. Firstly, FCM is to classify examinee based on their response patterns. Secondly, concept structure analysis (CSA) will analyze individualized knowledge structures. CSA includes three algorithms, which are AMC (algorithm for mastery of concept), ASC (algorithm for subordination of concepts) and AFISM (algorithm for fuzzy ISM). By the integrated procedure, examinee within the same cluster represent similar knowledge structures and remedial instruction could be feasible based on the information of cognition diagnosis within the same cluster.

![Fig. 2. Procedure of the Integrated Model](image)

Three algorithms, AMC, ASC and AFISM, are combined in order to analyze individualized knowledge structure. Some basic definitions are as follows [21].

1. There are \( M \) \((m = 1, 2, \ldots, M)\) items in a test which measures \( A \) \((a = 1, 2, \ldots, A)\) concepts. There are \( N \) \((n = 1, 2, \ldots, N)\) examinees who take the test.
2. Response data matrix is \( X = (x_{nm})_{N \times M} \) is the response matrix. \( x_{nm} = 1 \) means student \( n \) gives correct answer on item \( m \); otherwise \( x_{nm} = 0 \) means student \( n \) gives wrong answer on item \( m \).
3. It exists \( 2^A \) ideal concept vectors based on \( A \) concepts and the ideal concept matrix is \( Y = (y_{ma})_{M \times A} \) is item-concept matrix. It is \( y_{ma} = 1 \) if item \( m \) exactly measures concept \( a \); otherwise \( y_{ma} = 0 \) means item \( m \) does not measure concept \( a \).
4. It exists \( 2^A \) ideal concept vectors based on \( A \) concepts and the ideal concept matrix is \( Z = (z_{ia})_{1 \times A} \) with \( z_i = (z_{i1}, z_{i2}, \ldots, z_{iA}) \), \( i = 1, 2, \ldots, I \), \( I = 2^A \). \( z_{ia} = 1 \) means the ideal concept vector \( z_i \) contains concept \( a \);
Proceedings of the 9th WSEAS Int. Conference on APPLIED COMPUTER and APPLIED COMPUTATIONAL SCIENCE

A-cut is applied to it, it is fuzzy recognition and

\[ \sum_{i=1}^{l} c_{ni} = c_{ni} \]

The standardized closeness satisfies \( 0 \leq s_{crni} \leq 1 \) and \( \sum_{i=1}^{l} s_{crni} = 1 \).

3.2 ASC

(1) Let \( D = (d_{na})_{N \times A} = (SC)(Z) \) be the matrix of student and with formula of FLMP, the probability of concept \( a \) to be the precondition of concept \( a' \) (probability of subordination) is

\[
P_{aa} = \begin{cases} 
1 & , d_{na} = d_{na'} = 1 \\
0 & , d_{na} = d_{na'} = 0
\end{cases}
\]

\[
(\frac{(d_{na})(1-d_{na'})}{(d_{na})(1-d_{na'}) + (1-d_{na})(d_{na'})}, \text{ else})
\]

3.3 AFISM

(1) For student \( n \), the fuzzy relation matrix from ASC algorithm is \( F_{n}^{\alpha}(p_{aa'})_{A \times A} \). \( \alpha \)-cut is applied to get binary relation matrix [7]. \( \alpha \) value \( (0 \leq \alpha \leq 1) \) is selected and the binary relation matrix \( F_{n}^{\alpha} \) is

\[
F_{n}^{\alpha} = (p_{aa'})_{A \times A} \quad \text{and} \quad p_{aa'} = \begin{cases} 
1 & , p_{aa'} \geq \alpha \\
0 & , p_{aa'} < \alpha
\end{cases}
\]

(2) Adjacent value between concept \( a \) and concept \( a' \) is

\[
p_{aa'}^{\alpha} = \begin{cases} 
1 & , p_{aa'} \geq \alpha \\
0 & , p_{aa'} < \alpha
\end{cases}
\]

(3) For matrix \( F_{n}^{\alpha} = (p_{aa'})_{A \times A} \), ISM is used to get the individualized and hierarchical knowledge structures.

4 Data Resource

Linear algebra test for university students is designed by author. The instrument consists of 19 dichotomous items which measure 6 concepts. There
are 933 university students from Taiwan in this test. Attributes of concepts are depicted in Table 1.

### Table 1. Concept Attributes of Test

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Concept Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>operation of matrix</td>
</tr>
<tr>
<td>2</td>
<td>system of linear equations</td>
</tr>
<tr>
<td>3</td>
<td>determinants</td>
</tr>
<tr>
<td>4</td>
<td>vector space and the property of $\mathbb{R}^n$</td>
</tr>
<tr>
<td>5</td>
<td>eigen value and eigen vector</td>
</tr>
<tr>
<td>6</td>
<td>geometry of linear algebra</td>
</tr>
</tbody>
</table>

Item-concept matrix $Y = (y_{na})_{M \times A}$ and correct ratio of each item are depicted in Table 2. $\alpha = 0.65$ is selected in the AFISM step.

### Table 2. Item-Concept Matrix of Test

<table>
<thead>
<tr>
<th>Item</th>
<th>Concept</th>
<th>Correct Ratio</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.8328</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.8071</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.4544</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.4009</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.6806</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.3323</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.6356</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.0139</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>0.1061</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.5520</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>0.4566</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>0.3901</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>0.4019</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>0.2444</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>0.0171</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>0.2444</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>0.1490</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>0.1179</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>0.1854</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### 5.1 Number of Cluster

Table 2 is the information of partition coefficient and partition entropy. One is conclude that there exist the largest partition and the smallest partition entropy when the cluster number is 3. Therefore, the proper number of cluster is 3. Table 3 depicts number of students within each cluster.

### Table 2. Partition Coefficient and Partition Entropy

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Partition Coefficient</th>
<th>Partition Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.560761</td>
<td>0.378933</td>
</tr>
<tr>
<td>3</td>
<td>0.566667</td>
<td>0.32840</td>
</tr>
<tr>
<td>4</td>
<td>0.474313</td>
<td>0.948845</td>
</tr>
<tr>
<td>5</td>
<td>0.413319</td>
<td>0.935323</td>
</tr>
<tr>
<td>6</td>
<td>0.376137</td>
<td>0.965708</td>
</tr>
</tbody>
</table>

### Table 3. Number of Students within Each Cluster

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>319</td>
</tr>
<tr>
<td>II</td>
<td>295</td>
</tr>
<tr>
<td>III</td>
<td>319</td>
</tr>
<tr>
<td>Total</td>
<td>933</td>
</tr>
</tbody>
</table>

### 5.2 Knowledge Structures

As shown from Figure 3 to Figure 5, one student is randomly selected from each cluster respectively.

As to student 62, mastery of concept 1 is 0.55. Concept 1 is also the basis for concept 6, 2, 3, 4, 5. For student 340, mastery of concept 3 is 0.58. Concept 3 is also the basis of the other concepts. For student 147, mastery of concept 1, 2, 3, 4 is 0.55, 0.55, 0.53, 0.53. Concepts 1, 2, 3, 4 are also the basis of concept 5 and concept 6. One is concluded that master of concepts for these three students are different. It is also clearly understood that knowledge structures of these three students vary a lot.

![Fig. 3. Knowledge Structure of Student 62 (Cluster I)](image-url)
6 Conclusions

This study investigates an integrated methodology to display knowledge structures based on fuzzy clustering. In addition, empirical test data of linear algebra for university students are discussed. It shows that knowledge structures will be feasible for remedial instruction [20]. This procedure will also useful for cognition diagnosis. Future study could apply this method to other fields and focus on the development internet system [15] [25].

References:
[18] U. Bodenhofer, A Similarity-Based Generalization of Fuzzy Orderings Preserving


