Q-Matrix Construction

Robert Henson
The University of North Carolina at Greensboro

And

Jonathan Templin
University of Kansas

Introduction

- Several different cognitive diagnosis models incorporate the use of the Q-matrix.
  - Examples include the DINA, NIDA, and RUM
- The Q-matrix is what specifies which skills are required to correctly answer each item.

Introduction

- For example if we have a test intended to measure basic math:

- Possible items to a basic math test may be:
  - 2+3-1
  - 4/2
  - (4 x 2) + 3

- Because not all items measure all skills, we use a Q-matrix to indicate which skills are required for each item.

The Q-Matrix

- An example of a Q-matrix using our math test.

	Add	Sub	Mult	Div
2+3-1	1	1	0	0
4/2	0	0	0	1
(4 x 2)+3	1	0	1	0

The Q-Matrix

- Notice that by specifying the Q-matrix we have defined the skills of interest.
- If this is done carelessly, it is possible that the skills are not well defined and as a result your parameters will be meaningless.
- In many ways, development of the Q-matrix is one of the most important steps of cognitive diagnosis.

Introduction

- In this session we will:
  - Discuss a few different methods of Q-matrix development.
  - Discuss two methods that are being developed based on empirical development of the Q-matrix.

Introduction

- The methods will include:
- Basic Methods:
  - Simple inspection of the items.
  - Multiple Rater Methods.
  - Iterative procedures based on item parameters.

Introduction

- Advanced Methods.
  - Probabilistic Q-matrix Estimation using the DINA.
  - Empirical-Based design of the Q-matrix using the RUM.

Simple Inspection

- In using simple inspection, we are evaluating the item and determining what skills are required to answer each item.

- In doing this two possible situations can occur.

  - The test was constructed with the intent to measure a certain set of skills (skills are known).

  - The set of measured skills is unclear (skills are not known).

Skills are Known

- Here we assume that the test was constructed to measure a specific set of skills.
- In this case, because the skills are already known, all one must do is determine which of the skills are required to correctly answer each item.
- To do this, we recommend working through each question and making note of which skills were used.

Examples

- A basic math test designed to measure (addition, subtraction, multiplication, and division).
  - 2+3-1
- A questionnaire designed to measure the 10 criteria used to define a pathological gambler.
  - *For example, "I find it difficult to stop gambling."*

Examples

- Other examples may include tests that have been designed to measure specific parts of speech or verbal ability.
- The important thing is that the tests were created to measure multiple skills or traits and so determination of the required skills is simpler than in many cases.

Skills are Unknown

- There may be other cases where the test was not originally developed with a cognitive diagnosis model in mind.
- In these situations, the skills or traits measured by the test or questionnaire are unknown.
- This means that first we must determine the basic set of skills measured by the test.

Skills are Unknown

- Before moving on, we give a brief word of caution.
- From our experience, a common situation where skills are unknown is a unidimensional test.
  - One would like additional information about the examinees while also getting the unidimensional ability.

Skills are Unknown

- Some difficulty may arise if a test was initially developed to measure a continuous unidimensional skill and now the purpose is to determine multiple dichotomous skills.
  - The basic result will be categories that can be defined as a discrete ability scale.
- Cognitive diagnosis models are most beneficial for tests that are not truly unidimensional.

Determine the Skills

- Again, we recommend working through the items to determine the required skills.

- In determining the reasonableness of the model and the skills required remember that:

  – There is only one strategy used to answer each item.

  – The nature of the skills may be different than a typical compensatory model.

Develop the Q-matrix

- Once the basic set of skills measured by the test have been determined you can work back through the items and develop the Q-matrix.

- The following example, taken from the 1999 Third International Mathematics and Science Study (TIMSS), demonstrates this process with several Chemistry items.

Example Chemistry Items

- F06: Paint applied an iron surface prevents the iron from rusting. Which ONE of the following provides the best reason?

- L06: Filtration using the equipment shown above can be used to separate which materials?

- N07: Which is an example of a chemical reaction?

Skills Used In Chemistry

- The TIMSS defines several types of processes at work for each item:
  - Understanding simple information.
  - Understanding complex information.
  - Theorizing, analyzing, and solving problems.
  - Using tools, routine procedures, and science processes.
  - Investigating the natural world.
- For the items listed previously:
  - F06 – Understanding simple information.
  - L06 – Using tools, routine procedures, and science processes.
  - N07 – Understanding simple information.

Possible Problems to Avoid

- After the Q-matrix has been developed there are certain considerations that must be made.
  - Have I tried to measure too many skills?
  - Are there skills that are very similar?
  - Are some skills required by most or all items?
  - Have I specified too many skills on a single item?

Too Many Skills?

	Skill 1	Skill 2	…	Skill 20
1				
.				
30				

- Must consider reducing the number of skills.
- You will not have enough information to estimate all of these skills.
- Skills are too finely defined.

Similar Skills?

	Skill 1	Skill 2	Skill 3
1	1	1	0
2	1	1	1
3	1	1	0
4	0	0	1
…	…	…	…
20	1	1	0

- In this example Skill 1 and Skill 2 are measured by most of the same items.
- It will be difficult to determine whether items are being missed because of lacking Skill 1, Skill 2, or both ("blocking").
- Consider combining the two skills or selecting one of the two skills for each item.

Skills Required by Many Items?

- In this case, a single attribute is measured by every item.
- The item alone will determine whether you will have a high or low score.
- Also, if you lack this skill it may be difficult to determine mastery of the other skills.
- Consider breaking the skill into two skills (difficult level of skill and easy level of skill).

	Skill 1	Skill 2	Skill 3
1	1	0	0
2	1	0	1
3	1	1	0
4	1	1	1
20	1	0	1

Too Many Skills for an Item

- In some cases, it may be tempting to specify several (more than 4 or 5) skills for an item.

- This can begin to cause problems if it is frequent.

- Re-evaluate your skills.

  - Are they too fine grade?

  - Can the meaning of each skill be broadened so that fewer defined skills are required on each item?

Simple Inspection Summary

- In general, Simple Inspection relies on intuition and knowledge of the topic area.

- Once the skills have been defined and the Q-matrix determined, we must consider the expectations that are placed on the model.

  – By eliminating specific situations your initial Q-matrix results will be more informative.

Multiple-Raters

- A more likely situation is where a set of experts/researchers are working on the same project.
- In that case, each of the researchers may follow the same procedures as previously outlined.
  - Determine the skills.
  - Specify required skills for each item.
  - Refine Q-matrix.

Multiple-Raters

- However, it is unlikely that they will all provide the same answer.

- Therefore, as a second possibility, we consider the procedures of Q-matrix development for multiple raters.

Determine Skills

- To begin, we recommend that all experts (or a sub-committee) be selected to determine the required skills.
  - This procedure is the same as before only now they must agree on the set of skills
  - Given that the basic set of skills have been determined, a thorough definition should be written out for each.
  - These definitions should be given to all experts.

Development of the Q-matrix

- Each expert is now asked to create the Q-matrix.
- Here we have two possible options:
  - Use 0/1 for the Q-matrix.
  - Rate each skill based on his or her impression of its relevance to each item (e.g. on a scale of 1 to 5).

Development of the Q-matrix

- When they have finished, they should consider the same set of questions as specified earlier for possible refinement of the Q-matrix.
- The experts ratings are collected and aggregated.
- Next, we consider how this information is used.

Multiple Rater Results

- Use the results to determine the most likely Q-matrix.

- Use an iterative procedure asking raters for justifications if they deviate from the most common conclusions.

- Use rater scores to determine probabilities each skill is required for each item.
  - I will discuss this later.

Multiple Rater Summary

- In general, Multiple Raters is no different than a single rater, only now more information is obtained.
- This allows for more options of how one determines the final Q-matrix to be used.
- Summary of raters conclusions can range from very simple (e.g., the most common Q) to more complicated statistical procedures in aggregating the ratings.

Refinement based on Item Parameters

- Finally, we get to the last of the basic methods for Q-matrix construction.

- Even if a lot of care has been placed in determining an initial Q-matrix, it is possible that the Q-matrix is incorrect.

  – Think in terms of a confirmatory factor analysis.

- For this reason, we consider typical signs of an incorrect Q-matrix based on the item parameters.

Refinement based on Item Parameters

- We consider two common models.
  - DINA
  - RUM
- In doing this, we revisit the definition of each item parameter and discuss signs of a mis-specified Q-matrix.

DINA

- Recall that the DINA model has two parameters:
- The slip parameter (s_j)
  - 1-s_j indicates the probability of a correct response for someone classified as having all required skills.
  - A high s_j indicates many individuals classified as mastering all required attributes are still missing the item.
  - May indicate that a required skill has not been specified.

DINA

- The guess parameter (g_j)
  - This quantity is defined as the probability of a correct response for someone classified as lacking at least one skill.
  - High values imply many of the individuals classified as not having all required attributes are still correctly responding to the item.
  - May indicate that too many required skills have been specified for that item.

RUM

- Recall that the RUM has three parameters:
- The π^* parameters
  - The probability of a correct response given that all required attributes have been mastered and has a high ability score η.
  - A low value indicates that many individuals classified as mastering all required attributes are still missing the item.
  - May indicate that a required skill has not been specified.

RUM

- The r^* parameters
  - Are defined as the factor for which the probability of a correct response is reduced if that skill has not been mastered.
  - A high value means that nonmastery of that skill has little influence on the probability of a correct response.
  - May indicate that the skill should be removed from the Q-matrix.

RUM

- The c parameters
  - Is a measure of the extent that abilities not specified in the Q-matrix can impact the probability of a correct response (the opposite of a 1-PL IRT difficulty parameter).
  - Low values imply a stronger influence of abilities not specified in the Q-matrix.
  - May indicate that a required skill has not been specified.

Additional Indicators of Q-matrix Misspecification

- Slow convergence/lack of convergence if using an MCMC.
- If many of the class probabilities are very low.
  - In many models this can be detected using skill associations.
- Poorly fit test score distribution.

Refinement based on Item Parameters

- In any event, these are simply indicators of possible problems.
  - There are other reasons that these item parameters may be estimated as previously described.
- Given these results one should:
  - Revisit any “trouble” items .
  - Consider if the entries of the Q-matrix should be changed.
    - Look for theoretically supported reasons.

Basic Approaches

- Generally speaking, whether you have a set of experts or it is only you.
  - You should determine the skills.
  - Determine which items require which skills.
  - Consider possible refinements of the Q-matrix.
  - Fit a preliminary model and evaluate item parameters.
  - Consider refinements and fit model (repeat).

Advanced Methods

- The previous methods were based on basic methods of developing and refining the Q-matrix.

- Next, we move to two methods that can be used in estimation of the model to empirically determine a possible Q-matrix.

  – Essentially, we also estimate parameters for the Q-matrix.

Advanced Methods

- The two different methods are:
  - Probabilistic Q-matrix estimation using the DINA.
  - Empirical-Based design of the Q-matrix using the RUM.

Probabilistic Q-matrix DINA

- In the Probabilistic Q-matrix algorithm:
  - Uses a Bayesian estimation procedure that estimates selected entries in the Q-matrix.
  - Users are allowed to specify Q-matrix entries in terms of the (subjective) probability an item requires a given attribute.
  - Posterior probabilities of Q-matrix entries are obtained, indicating the likelihood an skill is required for a successful response to an item.

Probablistic Q-matrix Example

- Fraction subtraction test (Tatsuoka, 1990).
  - A 20 item math test given to 2,144 middle school students.
- Fraction subtraction Q-matrix (de la Torre and Douglas, 2004).
  - Eight skills (average 2.75 attributes per item).

Fraction Subtraction Skills

1. Convert a whole number to a fraction.

3. Separate a whole number from fraction.

5. Simplify before subtracting.

7. Find a common denominator.

1. Borrow from whole number part.

3. Column borrow to subtract the second numerator from the first.

5. Subtract numerators.

7. Reduce answers to simplest form.

Example Items

2. $\frac{3}{4} - \frac{3}{8}$ (Skills 4 and 7)

10. $4\frac{4}{12} - 2\frac{7}{12}$ (Skills 2, 5, 7, and 8)

19. $7 - 1\frac{4}{3}$ (Skills 1, 2, 3, 5, and 7)

Imagine you had no clue what the Q-matrix entries for these items might be…

Probabilistic Q-matrix Entries

- For each of the three items from before, the Q-matrix entries would look like:

Skill 1	Skill 2	Skill 3	Skill 4	Skill 5	Skill 6	Skill 7	Skill 8
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Results of Probabilistic Q-matrix Procedure

- For each uncertain entry in the Q-matrix, a posterior probability is obtained:

	Skill							
Item	1	2	3	4	5	6	7	8
2.	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00
10.	0.00	0.99	0.74	0.00	1.00	0.00	1.00	1.00
19.	1.00	0.98	1.00	0.00	1.00	0.05	1.00	0.00

- Here, green entries agree with original Q-matrix.
- Red indicates an entry that was in addition.
  - The red referred to having the item $4\frac{4}{12} - 2\frac{7}{12}$ need the skill “simplify before subtracting.”

Empirical Model

- Next we consider a method of Q-matrix refinement using the RUM.
- As with most models the RUM assumes that the Q-matrix is fixed and is known.
- The goal is to develop a method that relaxes this assumption.
  - Similar to what was done with the probabilistic Q-matrix using the DINA.

Complications

- While we would like to simply generalize the procedure to the RUM there are two complications.

1. The number of estimated item parameters depends on the Q-matrix.

3. We cannot estimate all r^* values in a simple algorithm because the model is not identified.

Estimation

- To estimate the model, we will use a two stage Markov chain Monte Carlo simulation.

- Stage 1:

  – We estimate the reduced RUM using an initially defined Q-matrix.

  – This procedure fixes r^* to 1 where q_{ij}=0.

  – The goal of this step is to get reasonable “starting values”.

Estimation

- Stage 2:
  - Change to using a Q-matrix with all 1s in estimation.
  - Continue the chain (previous step in the chain is used as starting values).
    - Estimated r* values are near where they would have been estimated.
    - Originally fixed r* start at 1.

Estimation

- Even using good starting values, our model is unidentified and so given a long enough chain we will still have problems.

- To correct this we "slow down" the chains of the r* values that were estimated in step 1.

  – Propose new values less frequently.

- This procedure helps keep us in the orientation we were in with the initially specified **Q**.

Simulation Study

- To test the effectiveness of the 2-stage method we used a simulation study.
- The goal was to simulate realistic data where the true Q-matrix was known
- Then, we systematically misspecify the initial Q that is used in the 2-stage procedure.
- Compute a new Q-matrix based on the estimated r^* values and compare back to the true generating Q-matrix.

Conclusion

- Simulation studies are encouraging, showing that in many cases item parameters and the correct Q-matrix are recovered.
  - Even when 20% of the Q-matrix has been misspecified nearly complete recovery of the true Q-matrix can occur.
- Using the 2-Stage MCMC is a method that allows the researcher to provide a basic orientation for the estimation of all r^* values.
- In doing this, the attributes definition is based on what is provided by the researcher.

Advanced Methods Conclusions

- By defining a Q-matrix one also determines the value of his or her results.

- However, no one person will always define the correct Q-matrix in much the same way that confirmatory factor analysis does not always work as one had intended.

- In these cases, it is important that we develop methods that allow the data to suggest Q-matrix entries that we may have over looked.

Conclusions

- There is no substitution for a well defined theory and well defined skills.
- Given these skills multiple raters can provide their opinion of a possible Q-matrix and refine it to a Q-matrix that will be used in a basic analysis.
- In many cases, simple inspection of the results from the estimation algorithm may provide additional insight as to a reasonable Q-matrix.

Conclusions

- However, there cases were simple inspection if the estimated item parameters will not provide the necessary information to determine a reasonable Q-matrix.
- Therefore, we discussed two methods that allow for an additional aid to Q-matrix construction.
- These methods are used to provide more information that a researcher may have missed.
- Not to provide an alternative to Q-matrix development.