

# 다차원 문항반응이론에 기반한 문항 응답 데이터 생성 알고리즘

김병욱\*, 이원규\*\*

\*고려대학교 컴퓨터교육학과

\*\*고려대학교 컴퓨터교육과

e-mail:byoungwook.kim@inc.korea.ac.kr, lee@inc.korea.ac.kr

## Algorithm Generating Item Response Data Based on Multidimensional Item Response Theory

ByoungWook Kim\*, WonGyu Lee\*\*

\*Dept. of Computer Science Education, Graduate School, Korea University

\*\*Dept. of Computer Science Education, College of Education, Korea University

### 요 약

본 논문은 다차원 문항반응이론 모델에 기반하여 시뮬레이션을 위한 피험자들의 문항 응답 데이터를 생성하는 알고리즘을 개발하는 것이 목적으로 하였다. 본 알고리즘은 시험지를 구성하고 있는 문항들의 모수를 읽고, 각각의 차원에 대해 피험자들의 능력 수준을 나타내는 정규 분포 확률 변수를 생성한다. 본 알고리즘은 다차원 문항반응이론 모델에 기반하여 피험자들이 각 문항에 대해 정답으로 응답할 확률을 계산한다. 피험자들의 문항 응답을 결정하는 균일 분포 난수와 비교한다. 만약 확률이 난수보다 크면 피험자는 올바른 답을 한 것으로 보고 그렇지 않을 경우 틀리게 답할 것으로 한다. 본 프로그램은 피험자 수, 문항 수를 조절할 수 있다. 본 알고리즘을 통해 교육 측정 분야에서 다차원 문항반응이론을 이용하여 학습자들의 문항 응답 데이터를 이용한 시뮬레이션 연구에 기여할 수 있을 것으로 기대한다.

### 1. 서론

교육 평가 분야에서는 문항 모수 및 피험자 능력을 추정하기 위해 시뮬레이션 데이터를 기반으로 연구가 진행된다. 시뮬레이션 데이터가 주로 사용되는 이유는 검사지를 구성하고 있는 문항 및 학습자들의 능력 모수를 추정하는 연구에 필요한 데이터는 쉽게 구할 수 없기 때문이다. 특히, 평가 결과를 통해 상위학급으로의 진학과 같이 평가 결과가 민감한 데이터에서 연구 목적으로 통제를 가한다는 것은 사실상 불가능한 일이다. 따라서 일반적으로 교육 평가에서 문항 모수에 대한 통계적 결과를 얻기 위해 통제 실험을 해야 하는 상황에서는 시뮬레이션을 통해 생성된 문항 응답 데이터를 기반으로 연구가 진행된다 [1][2]. 시뮬레이션 연구의 주된 장점은 연구자들이 피험자들의 능력 및 문항 특성을 정의할 수 있다는 점이다. 피험자와 문항에 대한 특성을 알고 있는 상태에서 연구자들은 데이터를 통제할 수 있고 특정한 요인을 조작하여 그 효과를 평가할 수 있다[3].

지금까지 문항응답데이터를 생성은 문항반응이론(Item Response Theory: IRT)을 중심으로 진행되었다[4]. 그러나 IRT는 일차원 독립성이라는 가정을 전제로 되어야 하거나 현실적으로 일차원 독립성을 완벽히 충족시키는 검사는 존재하지 않기 때문에 교육 평가 분야에서는 정확한

분석을 위해 다차원 문항반응이론(Multidimensional Item Response Theory: MIRT)을 활용하는 연구가 많아지고 있다[5]. 그러나 아직까지 MIRT를 기반으로 한 문항응답 데이터 생성 프로그램이 개발되지 않아 시뮬레이션을 통한 연구에 대한 제약 사항으로 남아 있다. 이제 본 논문에서는 기존의 IRT를 기반으로 한 문항응답데이터 생성 알고리즘을 개선하여 MIRT를 기반으로 한 문항응답데이터 생성 알고리즘을 개발하였다.

### 2. 이론적 배경

#### 2.1 문항반응이론

문항응답데이터는 IRT를 기반으로 한 시뮬레이션 연구에서 사용되기 위해 생성된다. IRT는 평가 문항들에 대한 응답에 근거하여, 피험자의 특성( $\theta$ )이나 평가문항의 난이도, 변별도를 측정하기 위한 검사이론이다[6]. IRT의 특징은 개인의 능력, 문항의 난이도라고 하는 모수를 구할 때, 개별 평가 문항에 대해 정답 여부와 같은 이산적인 결과를 가지고 확률론적으로 접근한다는 점이다. 즉 검사 자료에 수학적 모형들을 적용하는 것이다[4].

IRT 모델에서  $\hat{\theta}$ 는 검사문항에 의해 측정된 피험자들의 추정된 잠재적 특성 능력치를 나타낼 때 사용된다. IRT는

다양한 모델이 개발되었다[7]. 본 논문은 3모수를 기반으로 맞추어져 있다. IRT에 따르면 피험자의 능력( $\theta$ )이 높을수록 문항에 올바르게 반응할 확률이 높다. 3모수 모델 로지스틱 IRT 모델에서 피험자가 문항에 올바르게 반응할 확률은 다음과 같이 정의된다.

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-D a_i(\theta_j - b_i)}} \quad (\text{식 1})$$

$e$ : 자연상수,  $i$ : 검사 문항( $i=1,2,3,\dots,n$ ),  $j$ : 피험자 ( $j=1,2,3,\dots,N$ ),  $a_i$ : 문항 $i$  변별도로  $\theta_j = b_i$  지점에서 문항 반응 함수의 기울기에 비례한다.  $b_i$ : 문항 $i$  난이도로 문항 $i$ 에  $(1-c)/2$ 의 확률로 정답으로 응답할 확률을 가지고 있는 능력 지점이다.  $c_i$ : 문항 $i$ 에 대한 문항응답 함수의 최저 근접 파라미터로 문항의 정답을 맞출 능력이 없음에도 추측으로 정답에 반응할 확률,  $\theta_j$ : 피험자 $j$ 의 능력,  $P_i(\theta_j)$ :  $\theta_j$ 의 능력을 가지고 있는 피험자  $j$ 가 문항 $i$ 에 정답으로 응답할 확률,  $D$ : 계수인자로 1.702.

IRT 모델은 두 가지 가정을 전제로 하는데[7], 그 중 하나는 일차원성 가정이다. 일차원성 가정이란 하나의 검사는 단일한 특성을 측정하여야 한다는 가정이다. 인간의 능력 혹은 특성은 다양한 잠재적 특성으로 합성된 다차원의 잠재적 공간에 의해 구성될 수 있다. 지적 능력만 보더라도 여러 종류의 잠재적 특성이 있음을 알 수 있다. 그러나 이론적, 실제적 제한점 때문에, 특히 수학적 어려움 때문에 하나의 검사로 인간의 다차원 특성을 측정하기란 쉽지 않다. 문항 분석을 통해 인간의 능력 추정에 IRT를 적용하기 위하여 검사는 하나의 잠재적 특성을 측정하여야 한다. 즉 한 검사는 피험자가 지니고 있는 하나의 특성을 측정하여야 IRT를 적용하여 문항 분석뿐 아니라, 피험자의 능력까지 추정할 수 있다는 것이다. 만약 수리검사의 문항들이 어려운 단어들로 구성되어 있다면 이 검사는 수리력 뿐만 아니라 어휘력까지 측정하게 된다. 이와 같은 검사는 일차원성 가정을 위배하고 있다고 볼 수 있다. 한 검사가 피험자의 여러 능력을 측정하는 문항들로 구성되어 있다면, 검사도구의 일차원성 가정을 충족시키지 못하므로, 문항반응이론이 적용되지 않는다.

결국, 이러한 문항에 대하여 전통적인 일차원 문항 반응 모형을 적용하는 경우 측정되는 잠재능력 공간을 과소하게 추정하여 문항반응 간 지역 독립성의 가정은 만족되지 못하며, 능력 및 문항 모수 추정에 오류가 발생한다. 즉 실제 문항과 피험자의 상호작용을 정밀하게 분석하고 설명하기 위해서 활용되는 측정모형이 수리적 간명화 보다는 현실적 복잡성을 보다 적절히 반영하여야 한다는 점이 다차원 문항반응모형의 이론적 근거라 할 수 있다. 따

라서 일차원성이 아닌 다차원성 문항반응이론은 최근 몇몇 교육·심리 측정이론가에 의해 연구가 계속되고 있다.

## 2.2 다차원 문항반응이론

많은 연구자들은 일차원 독립성이라는 가정을 충족시키지 못하는 문항들에 대해서는 MIRT를 사용하는 것을 권장하고 있다 (Reckase, 1985; Ackerman, 1994a).  $k$ 개의 능력 차원을 상정한 로지스틱 MIRT 모델에서 피험자가 문항  $i$ 에 대해 올바르게 반응할 확률은 다음과 같이 정의된다.

$$P(u_i = 1 | \theta_j) = c_i + (1 - c_i) \frac{\exp(1.702 a_i^T \theta_j + d_i)}{1 + \exp(1.702 a_i^T \theta_j + d_i)} \quad (\text{식 2})$$

$i$ : 검사 문항( $i=1,2,3,\dots,n$ ),  $u_i$ : 문항 $i$ 에 대한 피험자 점수(0, 1),  $a_i$ :  $k$ 차원( $k=1,2,3,\dots,m$ )에서 문항 $i$ 에 대한 문항 변별도 벡터 ( $a_{ik} = a_{i1}, a_{i2}, a_{i3}, \dots, a_{im}$ ),  $d_i$ : 문항 $i$ 에 대한 문항 난이도 스칼라 값으로 (-)는 난이도가 높고 (+)는 난이도가 낮은 것을 의미,  $c_i$ : 문항 $i$ 에 대한 최저 근접 파라미터,  $\theta_j$ : 피험자 $j$  ( $j=1, 2, 3, \dots, N$ )에 대한 능력( $\hat{\theta}$ ) 벡터,  $P(u_i = 1 | \theta_j)$ : 피험자 $j$ 가 문항 $i$ 에 대해 올바르게 응답할 확률.

MIRT 모델에서는 전체적인 문항 난이도 모수 뿐만 아니라, 각 차원마다 문항 변별도 모수가 존재한다. 함수의 각 요소는 가산적이어서 하나의 잠재 능력이 낮다면 다른 잠재 능력이 높은 것으로 보정된다. 선행연구를 보면 MIRT 모델을 기반으로 한 시뮬레이션 데이터가 다른 방법으로 접근하여 생성된 자료보다 좀 더 현실 데이터와 유사하다고 보고하고 있다. 최근의 연구자들은 시뮬레이션 데이터의 기준으로 MIRT를 사용하고 있다.

## 3 문항 응답 데이터 생성 알고리즘

시뮬레이션은 평가 데이터의 통계적 처리의 효율성에 대해 연구하기 위해 많은 평가 기관에서 사용하고 있다. 일반적으로 피험자들의 문항 응답 데이터는 사전에 결정된 문항 모수(난이도, 변별도, 추측도)와 피험자 능력을 기반으로 시뮬레이션 된다. 본 논문에서 개발한 알고리즘은 (그림 1)과 같다.

```

Input: 문항 개수( $n$ ), 피험자 수( $m$ ),
      문항 모수 벡터( $a$ : 난이도,  $b$ : 변별도,  $c$ : 추측도)
Output: 문항응답데이터( $u$ ), 추측여부( $g$ )

for(  $i$  to  $m$  ) {
  for(  $i$  to  $n$  ) {
    theta = Rand( $\theta$ ) (-3 ~ 3) // 피험자의 능력을 설정
    p = Rand(0~1) // 피험자가 문항을 맞출 확률 랜덤
    if(  $p > P(\theta)$  ) { // 맞을 확률이 크면 알고 있음
       $u_{ij} = 1$ ;
       $g_{ij} = 0$ ;
    }
    else { // 모르고 있을 경우 추측
       $g_{ij} = 1$ ;
       $g = \text{Rand}(0\sim 1)$ 
      if(  $1 / \text{답지 수} < g$  ) { // 찍어서 맞은 것
         $u_{ij} = 1$ ;
      }
      else { // 찍어서 틀린 것
         $u_{ij} = 0$ ;
      }
    }
  }
}

```

(그림 1) MIRT 기반 문항응답데이터 생성 알고리즘

MIRT 기반 문항응답데이터 생성 알고리즘은 다음과 같은 입력 데이터를 받는다. 문항 개수( $n$ ), 피험자 수( $m$ ), 문항 모수 벡터( $a$ : 난이도,  $b$ : 변별도,  $c$ : 추측도). 제안된 알고리즘에서 피험자들의 능력은  $-3$ 과  $3$  사이에서 랜덤으로 생성되도록 되어 있다. 만약 실제 피험자들의 문항응답데이터와 시뮬레이션으로 생성된 문항응답데이터를 비교할 경우 피험자의 능력 벡터도 입력값으로 받도록 수정할 수 있다.

1. 피험자 능력치 생성: 알고리즘은 전체적으로 두 개의 반복문으로 구성되어 있다. 첫 번째 반복문으로 모든 피험자들을 선택하며, 두 번째 반복문에서 모든 문항에 대한 문항응답 데이터를 생성한다. 예를 들어,  $k$ 개의 차원에 대한 피험자들의 능력을 시뮬레이션 한다.  $NID(0,1)$ 분포로부터  $k$ 개의 랜덤 수를 생성한다. 이 수는  $k$ 개의 차원에 대한 피험자의 능력치로 사용된다.

2. 정답 확률 판단: 문항 모수와 피험자의 능력치가 설정되면 해당 피험자가 각 문항에 정답으로 반응할 확률이 정해진다. 피험자가 각 문항에 대해 정답으로 응답할지 오답으로 응답할지 설정하기 위해  $U(0, 1)$  분포에서 임의의 수를 생성한다. 균등분포에서 임의로 생성된 수는 (식 2)의 결과값과 비교한다. 균등분포에서 임의로 생성된 수의 값이 (식 2)의 결과값보다 크면 피험자가 해당 문항에 정답으로 반응한 것을 의미하여,  $u_{ij}$ 에 1을 입력하고,  $g_{ij}$ 에 0을 입력한다. 만약에 작을 경우 피험자가 문항의 정답을 모를 경우기 때문에 피험자는 정답을 추측하여 문항 선택지 중에서 임의로 선택하게 된다. 이 때  $g_{ij}$ 에 1을 입력한다. 피험자가 임의로 선택하여 정답을 맞출 확률은 (1/답

지수)이다. 따라서  $U(0, 1)$ 에서 임의로 수를 생성한 후에 (1/답지수)보다 작을 경우 임의로 선택하여 문항의 정답을 맞추게 된 것으로 판단하여,  $u_{ij}$ 에 1을 대입하고, 그렇지 않을 경우  $u_{ij}$ 에 0을 대입한다. 이와 같은 방법으로 각각의 피험자들의 문항 점수 벡터를 계산한다.

#### 4. 결론 및 향후 연구과제

본 논문에서는 다차원 문항반응이론에 기반한 문항응답 데이터 생성 알고리즘을 개발하였다. 본 연구의 목적은 교육 평가 및 측정 분야에서 시뮬레이션을 통해 학습자들의 능력이나 문항 모수의 추정 등에 이용되는 학습자들의 문항응답데이터를 생성하는 것이다. 본 논문에서는 다차원 문항반응이론에 사용되고 있는 확률 모델을 이용하여 알고리즘을 개발하였다. 따라서 본 연구의 결과로 다차원 문항반응이론을 이용하여 시뮬레이션 연구를 하는 연구자들에게 도움을 줄 수 있을 것으로 기대한다. 본 논문에서는 생성된 문항응답데이터에 대한 검증은 이루어지지 않았다. 아직까지 시뮬레이션 문항응답데이터를 검증하는 방법에 대한 합의는 없다. 이에 향후 연구로 MIRT 기반의 문항응답데이터 생성 알고리즘에 의해 생성된 시뮬레이션 문항응답데이터와 대규모 피험자들의 실제 문항응답데이터의 확률분포의 차이를 계산하는 방법 등을 이용하여 시뮬레이션 데이터를 타당성 있게 검증하는 방안에 대한 실험을 지속적으로 수행할 계획이다.

#### 참고문헌

[1] Harwell, M. Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte carlo studies in item response theory. *Applied Psychological Measurement*, 20, 101-125.

[2] Spence, I. (1993). Monte carlo simulation studies. *Applied Psychological Measurement*, 7, 405-425.

[3] Harwell, M. (1997). Analyzing the Results of Monte Carlo Studies in Item Response Theory. *Educational and Psychological Measurement*, 57(2), 266-279.

[4] Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph*, 7.

[5] Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7(4), 255-278.

[6] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage.

[7] van der Linden, W. J. & Hambleton, R. K. (1996, Eds.). Handbook of modern item response theory. New York, NY: Spinger-Verlag.