신경망을 이용한 효율적인 학업 성취도 측정 시스템 모델

김명순* · 정환묵**
*동주대학 컴퓨터 정보통신 계열
**대구효성 가톨릭대학교 전자정보 공학부

A Model of Efficient Estimation System for Learning Accomplishment Using Neural Network

Kim Myeong Soon* , Chung Hwan Mook**
*School of Computer information & Communications, Dongju College
**School of Electronic & Information Engineering, Catholic Univ. of TaeguHyosung

요 약

인간의 잠재적 능력을 측정하기 위하여 많은 검사가 이루어진다. 이러한 검사를 위하여 양질의 문항 제작이 필수적이다. 그러나 문항 제작의 중요성을 인식하지 못하고 단시간에 쉽게 제작된 문항들을 쓰고 폐기하는 경우가 많다. 본 논문에서는 양질의 문항 제작을 위하여 문항 내용 분석, 문항 응답 자료 분석, 문항 난이도, 문항 변별도, 문항 추측도, 오답의 능력도를 이용하여 자료분석에 유용한 효과를 보이는 신경망을 이용한 학업 성취도 측정 시스템 모델을 제안하였다.

1. 서론

인간의 능력에 대한 평가는 일반적으로 검사에 의하여 실시된다. 검사는 문항으로 구성되므로, 인간의 능력을 측정하고 그 결과에 의하여 인간의 능력을 평가하기 위해서 양질의 문항으로 구성된 검사가 요구된다. 수없이 많은 검사에 의하여 인간의 능력이 측정되고 있으나 문항제작에 대한 세심한 배려와 문항의 응답결과에 대한 분석은 검사의 중요성에 비추어 볼 때 못 미치는 실정이다. 문항제작이 단시간에 이루어지고 한번 실시된 문항은 분석 없이 폐기되는 일이 되풀이 되면서 문항제작·분석의 중요성을 인식하지 못하는 경향이 있다. 본 논문에서는 이와 같은 문제점을 해결하기 위해서, 문항 제작 및 분석에 있어서 신경망 이론을 적용함으로서, 문항평가의 학업성취도 측정에 대한 효율을 높일 수 있는 결과를 얻었다. [6]

2. 문항분석 및 평가

문항분석에 대한 평가는 크게 두 가지 방법으로 나뉜다. 하나는 질적 평가이고 다른 하나는 양적 평가이다. 문항에 대한 질적 평가는 문항이 측정의 목적에 부합되게 제작되었는지를 점검하는 방법으로 이는 내용 타당도를 확인하는 과정이며 검사내용 전문가의 주관적 판단에 의존한다.

문항에 대한 양적 평가는 피험자의 응답결과를 검사이론에 입각하여 문항난이도, 문항변별도, 문항추측도, 선다형의 경우 답지의 매력도 등을 분석한다. [2]

2.1 문항내용 분석

문항을 평가하기 위하여 문항의 내용을 우선 분석하여야한다. 첫째, 문항내용이 문항을 제작하기 이전에 작성한 이원분류표와 일치하는가를 검토하여야 한다. 이원 분류 표에 의하여 제작된 문항이라야 내용타당도가 있는 문항이 될 수 있기 때문이다.

둘째, 각 문항유형에 따른 문항제작법에 따라 제작되었는지를 점검하여야 한다. 문항의 내용분석 뿐 아니라, 문항의 형식에 대한 분석 또한 주관적일 수 있으나, 보다 객관적인 방법을 통하여 체계적으로 문항을 분석하기 위하여 문항점검표를 사용할 수 있다. [6]

2.2 문항 응답자료 분석

문항제작 방법에 의하여 제작된 문항들이 양질의 문항인지를 평가하기 위하여, 응답한 자료를 가지고 문항을 분석하는 이론으로 고전검사이론과 문항반응이론이 있다. [2]

(1) 고전검사이론(classical test theory)

고전검사이론은 검사도구의 총점에 의하여 분석되는 이론으로, 검사에 의한 관찰점수는 실제 점수와 오차점수에 의하여 합성됨을 가정한다. 또한 피험자의 실제 점수를 알 수 없기 때문에 이론적으로 동일 검사를 동일 피험자에게 무한히 반복 실시하여 얻은 점수들의 평균점수로 추정한다. 고전검사이론에 의한 문항난이도, 문항변별도, 문항추측도를 이용하여 평가한다.

(2) 문항반응이론(item response thoery)

문항반응이론은 검사 총점에 의하여 문항을 분석하는 것이 아니라, 문항은 문항 하나 하나의 불변하는 고유한 속성을 지니고 있으므로, 그 속성을 나타내는 문항특성 곡선에 의하여 문항을 분석하는 이론이다.

2.3 문항 난이도(item difficulty)

문항 난이도란 한 문항의 어려운 정도를 나타내는 것으로서 피험자의 집단에 따라 변하며, 한 개인에 대한 문항 난이도란 존재하지 않는다. 문항 난이도는 집단의 표준편차와 평균에 의해 결정되는 통계치 이다. (식 1)은 문항난이도 지수를 나타낸 것이다.

$$P=\left(\frac{R-\frac{W}{K_i-1}}{N-NR}\right)\times 100 \qquad \text{--- (식 1)}$$

여기서, P : 문항난이도 지수

R : 정답 수 W : 오답 수

N : 답을 한 전체 학생 수(=사례 수)

NR : 각 문항의 미달한 학생 수

K_i : 답지의 수

2.4 문항변별도(item discrimination)

문항타당도(item validity)는 "문항이 무엇을 측정하고 있느냐", "측정해야 할 것을 측정하고 있느냐", "학생의 능력을 변별하는 힘이 있느냐"는 개념으로 표현할 수 있다. 이와 같은 문항 타당도의 개념 중에 한 문항이 피험자의 능력 차이를 얼마나 예리하게 변별하는 능력이 있느냐를 보는 것이 문항변별도 이다. (식 2)는 문항 변별도 지수를 나타낸 것이다.

$$D.I.=\frac{R_U-R_L}{f} \qquad \text{--- (식 2)}$$

여기서, R_u : 상위 집단의 정답자 수

R_L : 하위 집단의 정답자 수

f : 각 집단의 학생 수

2.5 문항의 추측도(item inference)

진위형 문항이나 선다형 문항에서 문항의 답을 맞힌 피험자 중 추측에 의하여 문항의 답을 맞힌 피험자도 있게 된다.

문항추측도를 추정하기 위하여 우선 전체 피험자 중에서 문항의 답을 알지 못하여 추측한 피험자가 몇 명이고 그 중 문항의 답을 맞힌 피험자 수가 몇 명인지 파악하여야 한다. (식 3)과 같이 구할 수 있다.

$$G_R = G * \frac{1}{Q} \quad \text{--- (식 3)}$$

여기서, G_R : 추측하여 문항의 답을 맞힌 피험자 수

G : 추측한 피험자 수

Q : 답지 수

반대로 추측을 하여 문항의 답을 맞히지 못한 피험자는 정답이 아닌 오답을 선택하였으므로, 추측에 의하여 문항의 오답을 택한 피험자 수는 (식 4)와 같이 표현한다.

$$G_W = G * \frac{Q-1}{Q} \quad \text{--- (식 4)}$$

여기서, W : 문항의 답을 맞히지 못한 피험자 수

Q : 답지 수

2.6 오답의 능률도(effectiveness of distractors)

오답의 능률도란 선다형 문항에서 오답이 정답처럼 보여 응답자가 오답을 정답으로 택할 수 있는 가능성을 의미한다. 문항 오답의 매력도가 어느 한곳에 편중되지 않는가를 알아볼 때에 오답의 이론적 빈도를 참고한다. (식 5)는 오답의 능률도를 구하는 식이다.

$$ED = \frac{\sum(100-P)}{k-1} \quad \text{--- (식 5)}$$

여기서, P : 문항 난이도

K : 보기 수

오답의 매력도를 계산하는 방법은 100에서 문항 난이도를 빼고 $k-1$로 나누어 답지별 오답의 이론빈도를 구한다. 그런 다음 오답의 이론빈도에서 답지별로 실제 오답빈도를 뺀 값을 모두 더한 것이 오답의 능률도 (ED)가 된다.

오답의 매력도가 낮은 문항으로 판정하는 일반적인 기준은 답지별로 실제 오답반응과 오답의 이론 빈도의 차이가 오답의 이론빈도보다 20~30%를 초과하면 오답의 매력도가 낮은 문항으로 인정한다.

3. 신경망에 의한 문항분석 및 평가

신경망에서 가장 중요한 것은 학습자료의 준비과정으로서 학습자료가 목표로 하고 있는 문제를 잘 반영하지 못하는 경우에는 실패할 가능성이 매우 높다. 하지만 학습자료만 잘 준비된다면 매우 빠른 시간 안에 원하는 시스템을 개발 할 수 있다는 장점을 가진다.

오류 역전파 학습 알고리즘이라고도 부르는 BP (Back Propagation) 알고리즘은 순방향 다층 신경망의 학습에 효과적으로 적용할 수 있어서 여러 가지 분야에 보편적으로 활용되는 학습 알고리즘이다.

출력층 오차 신호를 이용하여 은닉층과 출력층간의 연결강도를 변경하고, 또한 출력층 오차 신호를 은닉층에 역전파하여 입력층과 은닉층간의 연결강도를 변경하는 학습방법이다. [7]

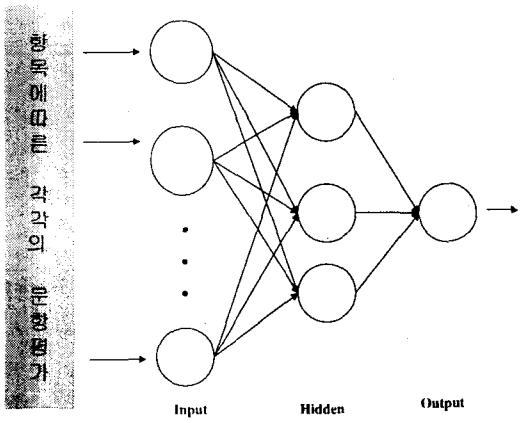

그림 1. 신경망을 이용한 문항평가

4. 문항 제작 모델

4.1 기존의 문항 평가 제작 모델

(그림 2)는 기존의 문항 평가 제작 모델을 나타낸 것이다.

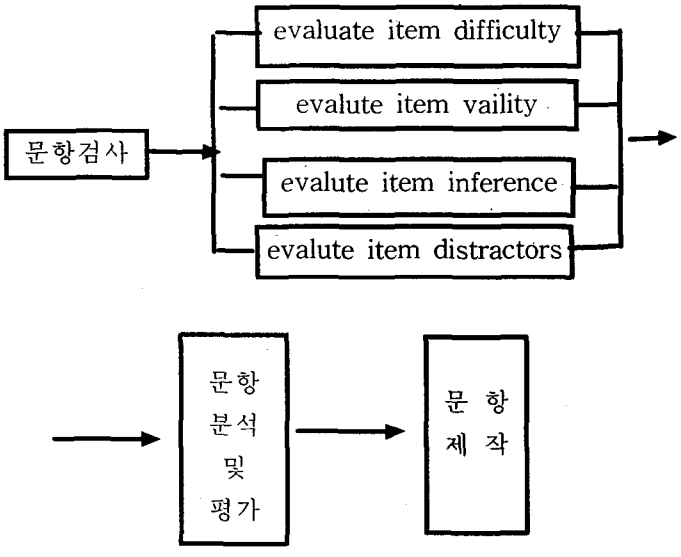

그림 2. 문항 평가 제작 모델

4.2 제안된 신경망 학업 성취도 측정 시스템 모델

(그림 2)와 같은 기존의 모델에서는 문항제작 및 분석의 중요성을 알지 못하는 문제점이 있었다.

이와 같은 문제점을 해결하기 위하여, 본 논문에서는 신경망을 이용한 효과적인 학업 성취도 시스템 모델(EESM; Efficient Estimation System Model for Learning Accomplishment Using Neural Network)을 제안하였고, (그림 3)과 같이 표현한다.

EESM은 신경망을 이용하여 양질의 문항 제작을 위한 효과적인 모델이다.

ESSM은 먼저 문항 응답 자료 분석을 하고, P평가, DI평가, G평가, ED평가 등을 한 결과 값을 신경망에 적용하여 양질의 문항 제작을 완료하는 모델이다.

출력된 결과 값을 분석하면, 아주 높은 학습율을 얻을 수 있기 때문에, 문항 분석 시스템에 효과적으로 적용할 수 있는 장점을 가지고 있다. [3]

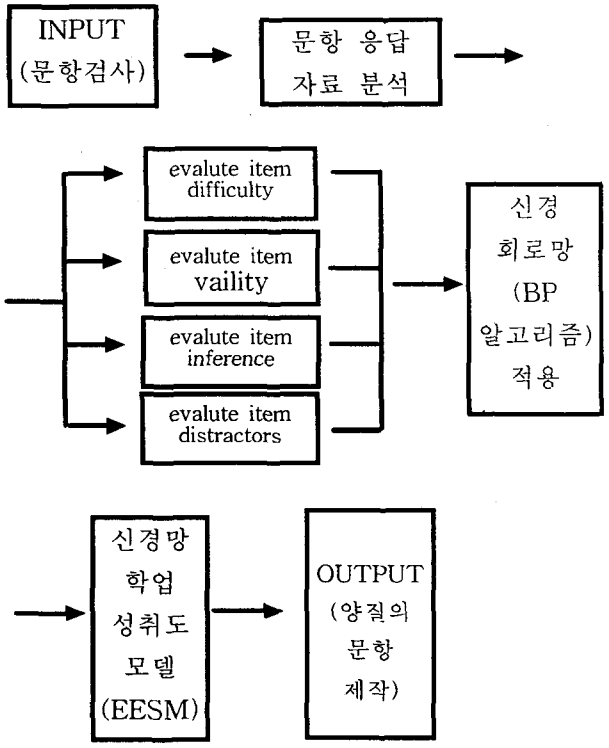

그림 3. 신경망 학업 성취도 모델(EESM)

5. 실행 절차

ESSM을 위한 실행 절차는 다음과 같다.

(단계 1) 문항 검사(test) 실시

(단계 2) 문항 응답 자료 분석

(단계 3) (단계2)에 따라서 평가를 실시

(단계 3-1) evaluate item difficulty(P_v)

$$\begin{cases} 40 \le P_v \le 60 & (very\ good) \\ 20 \le P_v \le 39 \ \ or \ \ 61 \le P_v \le 80 & (good) \\ 0 \le P_v \le 19 \ \ or \ \ 81 \le p_v \le 10 & (bad) \end{cases}$$

(단계 3-2) evaluate item validity(DI_v)

$$\begin{cases} DI_v \ge 0.4 & (\text{매우 좋음}) \\ 0.30 \le DI_v \le 0.39 & (\text{상당히 좋으나 개선 여지}) \\ 0.20 \le DI_v \le 0.29 & (\text{약간 좋으며, 개선 여지}) \\ DI_v \le 0.19 & (\text{좋지 않고, 수정 필요}) \end{cases}$$

(단계 3-3) evaluate item inference (G_v)

(단계 3-4) evaluate effectiveness of distractors(ED_v)

$$F = \frac{(T_{nc} - A_n)}{N_{dn}}$$

여기서, F는 오답의 빈도

T_{nc}는 전체 사례 수

A_n는 정답자 수

N_{dn} 는 오답자 수

$$\begin{cases} F*0.7 \le ED_v \le F*1.3 & (\text{매우 좋은 문항}) \\ F*0.4 \le ED_v \le F*1.6 & (\text{상당히 좋으나, 개선 여지}) \\ F*0.1 \le ED_v \le F*1.9 & (\text{약간 좋은 문항, 개선 필요}) \\ \text{그외의 경우} & (\text{별로 좋지 않은 문항, 버리거나 수정 필요}) \end{cases}$$

(단계 4) (단계 3)의 값에 따라 Neural Network의 B.P algorithm 적용

(단계 5) EESM 최종 결과 값(L_t) 출력

$$\begin{cases} 71 \le L_t \le 100 & (\text{사용 가능}) \\ 41 \le L_t \le 70 & (\text{수정 후 사용 가능}) \\ L_t \le 40 & (\text{사용 불가}) \end{cases}$$

6. 실행 결과

본 논문은 초등학교 4학년 학생을 대상으로 실시한 학업 성취도 검사의 검사 문항중 선다형 문항을 발췌하였다. 각각의 문항에 대한 문항 평가를 실시한 후 10개 데이터를 학습 데이터로 사용하고 10개 데이터를 테스트 데이터로 사용하였을 때의 학습율(Ⅰ)과 20개의 학습데이터와 10개의 테스트 데이터를 사용한 학습 결과(Ⅱ)가 아래의 [표 1]과 같다.

입력층 유니트의 개수 4개, 은닉층 유니트의 개수 2개, 출력층 유니트의 갯수 1개로 했을 때의 학습 결과이다.

[표 1] 신경망 학습결과

학습횟수	학습율(Ⅰ)	학습율(Ⅱ)
1000	97.75537	98.78872
3000	97.7665	99.09236
5000	98.01625	99.17195
7000	97.81524	99.16862

위 결과에서 볼 수 있듯이 신경망을 이용한 문항평가 시스템은 97%이상의 높은 학습율을 보이고 있으며 또한 학습횟수 5000번에서 각각 가장 높은 학습율을 보이고 있다.

7. 결론

검사의 좋고 나쁘거나 그 적절성을 검증하기 위해 그 속에 포함되어 있는 문항을 검증하는 필수적이다. 한번 힘들여 만든 문항은 쓰고 버릴 것이 아니라, 다음에도 유효적절하게 이용할 수 있도록 해야 한다. 새롭고 좋은 문항을 만드는데는 시간적 소비와 정신적 노력이 많이 들게 된다. 이러한 문항들을 한번 쓰고 버리기 보다는 저장했다가 유효하게 다시 쓰기 위한 문항은행(item pool) 혹은 문제은행제도의 설치를 위해서도 문항 분석은 필수의 선행조건이 된다. 문항 분석 및 평가를 위하여 문항 내용 분석, 문항 응답 자료 분석, 문항 난이도, 문항 변별도, 문항 추측도, 오답의 능률도를 자료 분석에 유용한 신경망을 이용한 문항 분석 학업 성취도 시스템 모델을 제안하여 보다 신뢰도 높은 양질의 문항을 작성할 수 있었다.

참고문헌

[1] L.A.Zadeh, "A Fuzzy-Approach to the Definition of Complex or Imprecise Concepts", Int · Journal of Man-Machine Studies, Vol.89, pp.249-291, 1996

[2] Baker,F.B.(1992a). Item Response Theory: Parameter Estimation Techniques. New York: Marcel Dekker, Inc.

[3] Baker,F.B.(1992b). Equating tests under the grade response model. Applied psychological Measurement, *16,* 87-96.

[4] Baker,F.B.(1993). Equating tests under the nominal response model. Applied psychological Measurement, *17,* 239-251.

[5] De Ayala,R.J.(1992). The nominal response model in computerized adaptive testing. Applied Psychological Measurement, *16,* 327-343.

[6] Thissen,D.M.(1991). MULTILOG: Multiple, categorical item analysis and test scoring using item response theory [Computer Program]. Mooresville: Scientific Software Inc.

[7] N. DeClaris, M.C Su, "A Neural Network Based Approach to Knowledge Acquisition and Expert System", July 8, 1993.

[8] 부산광역시교육청, "교과 수준별 프로그램 적용을 통한 효율적인 방과후 교육활동 방안", "97 교과교육 연구 보고서, 방과후 교육활동 연구회 동백중학교편, pp.46-48, 1997.12.

[9] 박용출, 정환묵, "수준별 학습을 위한 학습자 평가 시스템의 설계 및 구현", 대구효성가톨릭대학교 대학교육원 교육학석사학위 청구논문,1999.2.

[10] 신동희, 원성현, 최성혜, 정환묵, "지능형 교수 시스템에서 학습자의 인지 상태와 평가 방법에 대한 퍼지 적용", '94 한국퍼지시스템학회 추계학술대회 논문집, Vol.4, No.2, ppl152-156, 1994.

[11] 성태제(1991a). 문항반응이론 입문. 서울:양서원

[12] 성태제(1991b). 목표지향검사를 위한 준거설정방법:오답추출능력에 의한 준거설정방법의 타당성, 교육학연구, 제 29권 제 2호, 147-164.

[13] 한국심리학회(1997). 고전검사이론의 극복: 일반화가능도이론, 문항반응이론, 1997학년도 동계연구세미나.