

학습 평가를 위한 최적의 SPRT 판정에 관한 연구

전미란[†], 오암석^{**}

요 약

컴퓨터 기반 교육의 잠재적인 이점은 개별화 교육이다. 결국 수준별 교육과정의 운영과 열린 교육의 실천을 위하여 어떻게 학습자 개개인에 대해서 신속하고 정확하게 평가할 수 있는가에 대한 문제가 제기된다. 본 연구에서는 평가를 위한 문항 수가 두드러지게 감소하면서 반면에 합격과 불합격 판정이 비교적 정확한 기존의 SPRT(Sequential Probability Ratio Testing) 판정방법에 대한 문제점을 보완하여 보다 최적의 판정을 내릴 수 있는 방안을 시뮬레이션 연구를 통하여 제안한다. 이를 위하여 실제 측정에 의해서 얻어진 피험자 자료를 활용하여 SPRT 판정의 정확도 검증에 위한 시뮬레이션과 최적의 SPRT 판정을 위한 초기 설정치와 다단계(Multilevel) SPRT 모형 및 판정기준을 제시하고 시뮬레이션을 통하여 타당성을 알아본다.

A Study on Judgement of Optimal SPRT for Learning Estimation

Mi Ran Jeon[†], Am Seok Oh^{**}

ABSTRACT

A potential advantage of computer-based education is individualized instruction. In this regard, issues regarding how to evaluate individual learners fast and accurately in order to implement customized instruction program and to open education are raised. In this study, we suggest through a simulation study a way to reach an optimal decision by complementing the insufficient point of the way of the decision of the existing SPRT(Sequential Probability Ratio Testing) in which the number of questions in a test decreases while decision on pass or fail is relatively exact. For the purpose, the data of the examinees obtained from actual tests were used in this study and we present that simulation to verify the accuracy of decision and Multilevel SPRT decision standard for initial set point for optimal SPRT decision and evaluate the ground of the same through simulation.

Key words: SPRT(순차적 확률비율 검사), Multilevel SPRT(다단계 순차적 확률비율 검사), CAT(컴퓨터 활용 적응적 평가)

1. 서 론

최근 컴퓨터 기술의 발달과 함께 멀티미디어 응용 기술에 기초한 컴퓨터 교육 매체 개발이 활발히 진행되고 있다. 이는 곧 교육 개혁의 쟁점이 되고 있는 개별화 수업의 가능성을 한층 더 열어 주고 있다. 수준별 교육과정의 운영과 열린 교육의 실천을 위하여 어떻게 개별 학습자의 학습에 대하여 신속하고 정확

하게 평가를 할 수 있을 것인가에 대한 적절한 방안이 요구되고 있다[1-3]. 평가란 교수학습의 질적 개선을 위해 학습자의 학습활동에 관한 정보를 수집하여 의사결정을 하는 과정을 말한다. 여기서 의사결정이라는 말의 의미는 학습자의 학습 능력에 따라 우수한 학생은 해당 학습을 심화하거나 다음 단계로 나아갈 수 있는 기회를 주고, 열등한 학생에게는 적절한 피드백을 제공하여 부족한 부분을 다시 학습하게 하

※ 교신저자(Corresponding Author): 전미란, 주소: 부산시 남구 신신로 179번지(608-711), 전화: 051)629-1211, FAX: 051)610-8746, E-mail: lkj7880@paran.com
접수일: 2009년 4월 27일, 완료일: 2009년 5월 29일

[†] 정회원, 동명대학교 컴퓨터미디어공학과 박사수로
^{**} 중신회원, 동명대학교 컴퓨터미디어공학과 교수
(E-mail: asoh@tu.ac.kr)

는 과정이라 볼 수 있다[4].

수준별 교육을 위하여 CAT(Computer-based Adaptive Testing)를 이용한 개별적 평가에 관한 연구들이 꾸준히 있어 왔다. CAT는 되도록 짧은 시간 내에 적은 수의 평가문항을 사용하면서도 학습자의 능력을 오차가 최소한이 되도록 정확하게 평가하는 방식이다. 이 방식은 필요에 따라 측정오차의 크기 등 다른 변인들을 조절하면서 개별적으로 검사의 길이와 신뢰도를 조절할 수 있게 하는 등의 특성이 있다[5,6].

범수균 등[1], 김영환 등[2]의 연구에서는 컴퓨터에 기초한 평가길이를 적절히 조절함으로써 학습자의 학습능력을 평가할 수 있는 CAT 프로그램의 개발의 가능성을 시사하고, 평가를 위한 문항 수를 두드러지게 감소시키면서 합격과 불합격 판정에 대한 오류를 극소화시킬 수 있다고 피력하고 있다. 또한 SPRT 방식을 이용한 적응적 멀티미디어 평가시스템을 설계하고 구현하였다. 적응적 멀티미디어 평가시스템은 적은 문항으로 빠른 시간 내에 많은 학습자의 학습 성취도를 분류하고자 할 때 가장 효과적이며, 학습자 스스로 자신의 기준 점을 정하고 여러 번 반복 평가를 시도해 보고자 할 때, 교습 진행 과정에서 학습자의 사전 지식을 평가할 때, 그리고 학습 이후의 학습 성취도를 평가하기 위한 도구로써 적합하다고 피력하고 있다. 따라서 범수균 등의 연구를 통하여 SPRT를 CAT 진단 시스템에서 효과적으로 수용할 수 있는 가능성을 입증하였다고 할 수 있다. 그러나 실질적인 적용에 앞서 SPRT 방식의 시스템에 대한 문제점들을 보완하여 보다 신뢰성 있고 효과적인 CAT 시스템의 구현과 활용 방안이 요구된다.

따라서 본 논문에서는 SPRT의 실제 적용에 앞서 시뮬레이션을 통하여 그 타당성을 알아보고 다단계 SPRT 판정 방법을 통하여 SPRT의 단점들을 보완하고 최적의 SPRT 초기 설정치에 관련한 근거를 제시한다.

2. 관련 연구

2.1 컴퓨터 활용 적응적 평가

CAT는 컴퓨터를 활용하여 최소한의 문항을 가지고 오류(제 1종 오류 또는 제 2종 오류)를 최소화하면서 피험자들의 학습 성취도를 측정하는 것이다.

CAT의 장점으로서 첫째, 전통적인 지필식 검사에 비해 검사 제작, 시행, 채점, 그리고 결과보고 등에 소요되는 시간이나 경비를 절감할 수 있다. 둘째, 멀티미디어 오브젝트의 특성을 이용하여 다양한 형태의 문항 제작이 가능하며, 문항의 제시 시간의 통제나 검사 실시 시간의 통제 등 활용상의 표준화가 쉬워 활용이 용이하다. 셋째, 평가 문항들은 디지털화되어 있어 보관, 수정, 비밀보장에 용이하다는 등이 있다[3].

이러한 CAT의 다양한 장점에 따라 국내에서도 다수의 학자들이 이 방법의 가치를 인정하고 학문적으로 연구를 하고 있음에도 불구하고, 그 실질적인 활용은 극히 미약한 실정이다. 여기에는 다음과 같은 이유들이 있다. 첫째, CAT의 기본 원리가 되는 이론적 배경이 수학과 통계학에서 시작된 것으로서 난해하기 때문에 일반 교육 실무자들이나 비전문가들이 접하기가 어려워 쉽게 활용하게 되지 않는다. 둘째, 현재 연구된 CAT의 모형들이 아주 다양해서 어떤 경우에 어떤 모형이 좋은지를 현장의 교사와 실제 활용자들이 구별하여 사용하기 힘들다. 셋째, 대부분의 CAT 모형들이 사전에 많은 양의 데이터를 모으고 분석하여 정리할 것을 요구하고 있어서 장시간 체계적인 연구를 할 수 없는 일반 실무자들은 활용의 실제적인 제한을 받게 된다. 넷째, 대부분의 모형들이 구체적으로 어떻게 개발해야 하며, 컴퓨터 프로그램으로 어떻게 구현되는지에 대한 지침이 없으며 설계를 위해 표준화된 자료도 없어 실제 개발이 많은 제한을 받고 있다. 따라서 이러한 문제점을 극복할 수 있는 효율적이며 효과적인 적응적 평가 도구의 개발 연구가 시급하다[1,3].

또한 CAT는 적응적 평가측면에서 분류하면 다음과 같은 세 가지 부류로 나눌 수 있다; 1) 베타 분포의 사용(use of the Beta Distribution), 2) 적응적 숙달 검사(AMT : Adaptive Mastery Testing), 3) 순차적 확률비율 검사(SPRT : Sequential Probability Ration Test). 본 논문에서는 적은 비용에 비해 정확도가 높은 SPRT에 대해서 중점적으로 알아본다.

2.2 SPRT 기본개념

2.2.1 SPRT의 특징

SPRT형 모델은 여러 가지 적응적 평가 시스템 중에서 가장 단순하면서도 비교적 그 판정의 오차가

다른 평가 방식에 비해 크게 되지 않는다. 쉽게 활용 할 수 있는 SPRT 방식은 수학자 Bayes의 이론에 바탕을 두고 있는데, 초기에는 전기회사에서 한 상자 당 전구의 불량률을 측정하기 위해서 활용되기도 하였다. 전통적인 방식에 비하여 소모되는 전구의 폐기율이 아주 적어짐에도 불구하고 불량률에 대한 검사의 결과는 만족스러웠다. 이 방법은 처음에 한 개의 전구를 상자에서 꺼내어 검사를 하는데 이때 전체 상자 안에 있는 모든 전구들이 불량률의 기준을 통과할 초기의 확률과 통과하지 못할 확률은 각각 0.5이다. 이때 한 개의 전구가 불량률이 아니었다면, 전체 전구가 불량률이 아닐 확률은 0.5에서 어느 정도 증가하고 반면에 불량일 확률은 다시 어느 정도 감소한다. 또 하나를 무작위 추출하여 검사한 결과가 불량률이 아니라면 다시 전체 상자가 불량일 확률과 그렇지 않을 확률은 조정된다. 이런 과정을 반복하다가 이 확률이 통계적 오차의 허용치 안에서 검증이 되면, 이 과정은 끝나게 되고 결정이 내려지는 방식이다. 이 방법은 위와 같은 제조업 분야에서 상품의 불량률을 검사하는데 널리 사용되어졌으며, SPRT의 가장 큰 이점 중의 하나는 고전적 통계방법에 비하여 결론에 도달하기 위해 소요되는 표본(sample)의 수가 현격히 줄어든다는 것이다[3].

2.2.2 SPRT에 의한 의사결정

측정이 순서적으로 진행될 때, 합격과 불합격을 판정하는 방법으로 Abraham Wald에 의해 개발된 SPRT는 두 이산가설간에 선택을 위한 의사결정 방법론으로서 품질관리(quality control)에서 주로 활용되어 왔다[11]. SPRT는 다음의 규칙에 의거하여 의사결정을 수행하는데, 본 연구에서는 문항이론에 근거해서 다음의 규칙을 적용한다.

규칙 S1 : 만약 두 판단들의 사후확률비(PR)가 $(1-\beta)/\alpha$ 보다 크거나 같으면 첫 번째 대안이 선택된다. 즉, 합격 판정을 한다.

규칙 S2 : 만약 두 판단들의 사후확률비(PR)가 $\beta/(1-\alpha)$ 보다 작거나 같으면 두 번째 대안이 선택된다. 즉, 불합격 판정을 한다.

규칙 S3 : 만약 규칙 S1과 S2를 만족하지 않으면, 또 다른 관찰이 수행되고 사후확률이 다시 계산되어 지고 S1부터 S3까지의 단계를 반복해서 적용한다. 여기서 α 는 주어진 첫 번째 판단이 참인데 거짓으

로 잘못 판단할 확률(제 1종 오류)이며, β 는 주어진 두 번째 판단이 참인데 거짓으로 잘못 판단할 확률(제 2종 오류)이다. 컴퓨터를 통하여 학습종료 후 학습평가(합격, 불합격)를 판정하는 상황에서 상술한 법칙들을 적용하여 설명한다.

먼저, 한 문항을 무작위로 선택해서 학생에게 제시했을 때, 학생의 반응을 관찰한 후에 다음과 같은 확률 비율(PR : Probability Ratio)을 계산한다.

$$PR = \frac{P_{om}P_m^r(1-P_m)^w}{P_{on}P_n^r(1-P_n)^w} \quad \text{수식(1)}$$

P_{om} 과 P_{on} 각각은 합격과 불합격에 대한 사전확률(Prior Probability)에 대한 초기값이다. 만약 여기서 P_{om} 과 P_{on} 이 동일하다면 위의 공식에서 서로 약분된다. P_m 은 합격자가 제시한 문항을 맞출 확률을 의미하고, P_n 은 불합격자가 해당 문제를 맞출 확률을 의미한다. r 과 w 는 어떤 시점까지 제시된 문항 중에서 피험자(불합격자, 합격자)가 각각 응답한 정답 수와 오답 수를 의미한다[7-11].

3. SPRT 시뮬레이션

3.1 자료

SPRT 방식의 효과를 검증하고 효과적으로 적용하기 위하여 237명의 피험자를 표본으로 시뮬레이션을 실시했다. 그리고 SPRT에 있어서 충족되어야 할 가정과 처방에 따라 만들어진 “교육방법 및 교육공학” 교재 내용에서 엄선한 100문항을 K-SPRT[3]를 통해 부산대학교에 재학 중인 사범계 교직과목 수강자들을 대상으로 시험을 실시하였다. 이들은 모두 교직과목인 “교육방법 및 교육공학”을 수강하고 있는 학생들이었다. 시험은 SPRT 방식의 기본 가정에 위배되지 않도록 충실히 시행되었다. 학생들에게는 같은 문제를 다시 풀 수 있는 기회가 주어지지 않았으며 모든 문제는 무작위로 추출되었다. 그리고 피험자들에게는 판정결과가 중간에 제시되지 않았으며, 모든 문제를 다 풀 때까지 자신의 시험결과를 알 수 없었다. 무작위로 추출된 문항에 따른 피험자의 정답/오답 반응의 결과와 문항의 번호, 시험 소요시간 등의 정보를 일정한 형식으로 기록하였다. 기록된 결과는 본 연구 실험을 위한 시뮬레이션 자료로서 활용되었다.

3.2 SPRT 판정의 정확도 검증을 위한 시뮬레이션

SPRT 판정의 정확도를 검사하기 위하여 SPRT 방식에 의한 판정과 전통적인 판정을 비교하였다. 합격 판정점은 SPRT 공식에 따라 합격 추정선(합격자가 정답을 맞출 확률)과 불합격 추정선(불합격자가 정답을 맞출 확률)의 중간값을 선택하였다[1,3]. 첫 번째 실험은 SPRT의 효과를 검증하고 SPRT의 특성을 파악하기 위하여 기존의 실험 연구와 같이 SPRT 통제요인(parameter)들은 사전에 다음과 같이 설정하였다. 합격 추정선 = 0.85, 불합격 추정선 = 0.60, 그리고 α , β 의 값을 모두 같은 값으로 하고 0.025와 0.05의 두 경우로 구분하여 실험하였다. 기존의 연구 실험에서는 합격자 또는 불합격자로 판정할 수 없는 수험자는 설정치의 중간점수와 총점을 기준으로 판정하였다. 이 시뮬레이션의 첫 번째 결과는 표 1과 표 2에 주어져 있다.

표 1과 표 2를 살펴보면, 오류율 α , β 를 작게 잡을 수록 수험자의 판정에 있어서 합격과 불합격을 판단할 수 없는 미정판정(0.5인 경우는 2, 0.025인 경우에는 7)이 많이 나왔고, 이것은 조건을 엄격히 한 결과가 되기 때문에 당연한 결과라고 할 수 있다. 이러한 미정판정을 오류로 처리하였을 경우에도 SPRT 방식

의 정확도 면에서는 비교적 높은 값(0.8396 \rightarrow 0.8354)을 유지하고 있음을 알 수 있다. 표 1과 표 2에서 SPRT는 본 연구에서 고려한 방식의 SPRT를 의미하고, Conventional은 일반적인 평가방식으로 100 문항 전체를 풀었을 경우의 결과를 보여 주고 있다.

표 1의 SPRT에 의한 결과에서 합격자로 91명, 불합격자로 144명, 그리고 미정 판정이 2명의 결과가 얻어졌음을 의미하고, 이러한 결과가 얻어졌을 때의 평균 점수와 평균 문항수가 제시되어 있다. 즉, 합격한 피험자의 평균점수는 79.82점이고, 이 때, 평균적으로 24문항만으로 합격자를 판정할 수 있었음을 의미한다. 'NoDecision'은 미정판정을 의미하는데, 100 문항을 다 풀어도 2명은 합격과 불합격으로 판정할 수 없었음을 나타내고 있다. 그리고 Total 부분은 100 문항에 대한 237명 모든 학생들의 평균점수를 의미한다.

본 실험 연구는 피험자들의 수험 과정과 결과를 토대로 소급 적용하였으므로 보다 정확도를 높이면서 출제 문항을 줄일 수 있는 SPRT 초기값 설정구간(합격 추정선/불합격 추정선)을 찾아 다시 SPRT 방식의 실험을 하였다.

SPRT 검증 실험 결과 표 1, 표 2에서 알 수 있듯이 SPRT 판정 방식의 정확도는 약 84%로 비교적 높게 나타났으나 본 연구에 참여한 피험자 평균은 70.58점으로 합격 판정점, 즉 72.5점에 근사해 있다는 점을 착안하여 초기 설정치를 바꾸어 실험하였다. 즉 해당 피험자 평균보다 낮은 점수를 합격 판정점으로 초기치를 설정하여 재 실험하였다. 실험 초기치는 합격 판정선을 75점, 불합격 판정선을 45점으로 설정하여 그 중간값 60점을 합격 판정점으로 설정하여 실험을 실시하였다. 실험결과는 표 3과 표 4에 주어져 있다.

실험 결과 표 3과 표 4에서 볼 수 있는 바와 같이

표 1. SPRT 시뮬레이션 결과 ($\alpha=\beta=0.05$)

	SPRT			Conventional		
	Mastery	Nonmastery	NoDecision	Mastery	Nonmastery	Total
수험자의 수	91	144	2	111	126	237
평균 점수 (표준편차)	79.82 (6.06)	64.70 (7.68)	74 (1.00)	79.71 (5.49)	62.54 (8.53)	70.58 (10.1)
평균 문항수	24.07	21.07	100	100	100	100

※ 합격 추정점수=85, 불합격 추정점수=60, 합격 판정점 =72.5, 정확도(유사도)=0.8396

표 2. SPRT 시뮬레이션 결과 ($\alpha=\beta=0.025$)

	SPRT			Conventional		
	Mastery	Nonmastery	NoDecision	Mastery	Nonmastery	Total
수험자의 수	88	142	7	111	126	237
평균 점수 (표준편차)	80.55 (5.15)	64.23 (7.38)	74 (0.75)	79.71 (4.64)	62.54 (8.03)	70.58 (10.1)
평균 문항수	30.72	26.61	100	100	100	100

※ 합격 추정점수=85, 불합격 추정점수=60, 합격 판정점 =72.5, 정확도(유사도)=0.8354

표 3. SPRT 시뮬레이션 결과 ($\alpha=\beta=0.05$)

	SPRT			Conventional		
	Mastery	Nonmastery	NoDecision	Mastery	Nonmastery	Total
수험자의 수	188	49	0	199	38	237
평균 점수 (표준편차)	73.70 (8.48)	58.63 (6.79)	-	73.54 (8.24)	55.07 (8.70)	70.58 (10.1)
평균 문항수	19.01	25.51	-	100	100	100

※ 합격 추정점수=75, 불합격 추정점수=45, 합격 판정점 =60, 정확도(유사도)=0.9071

표 4. SPRT 시뮬레이션 결과 ($\alpha = \beta = 0.025$)

	SPRT			Conventional		
	Mastery	Nonmastery	NoDecision	Mastery	Nonmastery	Total
수험자의 수	189	42	6	199	38	237
평균 점수 (표준편차)	74.02 (8.09)	56.54 (5.11)	60.5 (0.76)	73.54 (7.90)	55.07 (5.59)	70.58 (10.1)
평균 문항수	24.99	31.73	100	100	100	100

※ 합격 추정점수=75, 불합격 추정점수=45, 합격 판정점=60, 정확도(유사도)=0.9409

수험자의 평균 점수를 고려한 합격 추정선과 불합격 추정선으로 초기화하여 SPRT 방식으로 실험하였을 때, 출제 문항 수는 합격 판정 부분에서는 줄어 들고, 불합격 판정을 위해서는 늘어나고 있음을 알 수 있다. 그리고 판정을 하지 못한 미정의 피험자는 오류 처리 하였음에도 불구하고, 정확도에서는 각각 90%, 94%로 높은 결과를 보였다. 기존의 연구 방식과 같이 미정 판정인 수험자의 맞음/틀림 반응의 총점을 합격 판정점을 기준으로 합격과 불합격을 판정하였을 경우에는 정확도가 0.9662(96%)로 향상 되었다. 그러나 수험자의 학습 성취도 평가를 위하여 교사가 합격 추정선과 불합격 추정선의 중간값인 합격 판정점을 정확히 설정하여도 합격 추정선과 불합격 추정선의 구간 거리를 지나치게 많이 잡거나 작게 잡으면 정확한 결과를 얻지 못하는 경우가 발생하게 된다. 다음 예는 합격 추정선과 불합격 추정선을 각각 85와 35점으로 초기 설정 값의 범위 구간을 넓게 잡은 경우로써 중간값으로 앞의 예에서와 같이 60(합격 판정점)을 설정하여 실험하였다. 그 결과는 표 5와 표 6에 주어져 있다.

표 5와 표 6의 결과를 보면, 추정 구간 간격이 넓을수록 합격인 학생이 불합격 처리가 되는 경향이 나타나고 있음을 볼 수 있다. 즉, 합격 판정점을 넘어선 수험자라 하더라도 합격 추정선에 충분히 도달하지 못한 학생은 불합격으로 판정하여 정확도는 크게 떨어졌다(0.7469, 0.7552). 또한 표 5와 표 6의 결과에서 오류율은 높음에 비하여 필요한 평균 문항 수는 현저하게 떨어지는 상반된 경향을 나타내고 있다. SPRT 방식을 이용한 판정에 앞서 초기 설정값인 합격 추정선과 불합격 추정선의 구간 간격에 따라 판이하게 다른 결과가 초래되었다. 이와 같이 일반적인 SPRT 방식은 적은 문항수로 정확한 판정을 위한 최적의

표 5. SPRT 시뮬레이션 결과 ($\alpha = \beta = 0.05$)

	SPRT			Conventional		
	Mastery	Nonmastery	NoDecision	Mastery	Nonmastery	Total
수험자의 수	169	68	0	199	38	237
평균 점수 (표준편차)	73.90 (8.97)	62.23 (8.14)	-	73.54 (8.28)	55.07 (14.5)	70.58 (10.1)
평균 문항수	8.10	8.25	-	100	100	100

※ 합격 추정점수=85, 불합격 추정점수=35, 합격 판정점=60, 정확도(유사도)=0.7468

표 6. SPRT 시뮬레이션 결과 ($\alpha = \beta = 0.025$)

	SPRT			Conventional		
	Mastery	Nonmastery	NoDecision	Mastery	Nonmastery	Total
수험자의 수	170	67	0	199	38	237
평균 점수 (표준편차)	73.98 (8.87)	61.97 (8.01)	-	73.54 (8.21)	55.07 (14.0)	70.58 (10.1)
평균 문항수	12.04	11.59	-	100	100	100

※ 합격 추정점수=85, 불합격 추정점수=35, 합격 판정점=60, 정확도(유사도)=0.7552

합격 추정선과 불합격 추정선간의 구간 간격과 중간값을 설정하기 위한 근거가 마련되지 못한 실정이다. 다만 20점에서 30점의 추정구간 간격과 적절한 중간값을 설정하라는 경험적인 예와 결과만을 제시하고 있을 뿐이다. 따라서 3.3절에서 SPRT의 초기 설정치에 따른 시뮬레이션에 대해서 연구한다.

3.3 SPRT 초기 설정치에 따른 정확도 시뮬레이션

SPRT 방식을 적은 문항으로 정확하게 판정하기 위한 각 추정 구간을 잘 대표하는 최적의 추정 구간을 설정하기 위한 근거와 활용 방안을 제시하고자 다음과 같은 연구실험을 행하였다. 본 실험은 합격 및 불합격 판정에 있어서 유의미한 합격 판정점수를 나타내는 구간을 설정하여 모든 경우의 초기치 인자를 고려하여 시뮬레이션을 실시하였다. SPRT 방식으로 합격 추정선을 95점부터 65점까지 그리고 불합격 추정선은 90점부터 25점까지 각각 2.5점 간격으로 설정하여 237명의 수험자 표본으로 시뮬레이션 하였다. 100문항 기준 점수에 합격 판정점수는 각 합격 추정선과 불합격 추정선의 중간값으로 하였으며, SPRT의 판정 결과와 100문항의 기준 결과에 따라 정확도를 계산하고 합격과 불합격 판정까지 필요한

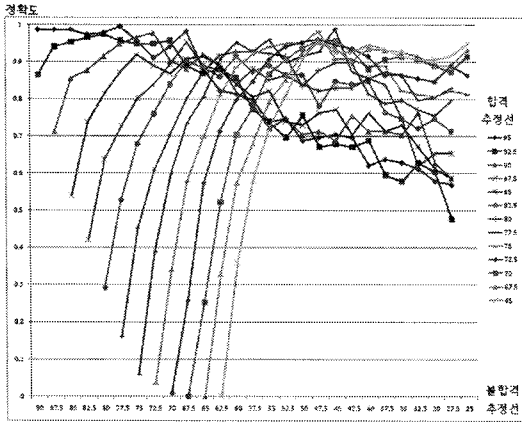


그림 1. 각 추정 구간에 따른 정확도 ($\alpha=\beta=0.05$)

각각의 평균 문항수를 얻었다. 오류를 알파와 베타를 각각 0.025로 설정한 것과 0.05로 설정한 두 실험 결과 테이블을 작성하였다. 정확도는 100문항 모든 문제를 풀어도 SPRT에서 판정할 수 없는 경우, 오류 처리하여 엄격하게 계산하였다. 그림 1은 각 추정 구간에 따른 정확도를 보여 주고 있다.

위 실험 결과에서 볼 수 있듯이 합격 및 불합격 판정에 있어서 정확도가 높은 구간을 탐색할 수 있다. 즉, 각 합격 추정선에서 최고의 정확도를 나타내는 불합격 추정선을 찾아볼 때 판정의 정확도는 가장 높으면서 출제 문항수는 비교적 적은 구간으로써 평균 20, 30개의 문항수를 필요로 한다. 합격 및 불합격 추정선 구간이 좁아질수록 많은 문항을 필요로 하는 경향이 나타났다. 그리고 불합격 추정선은 해당 합격 추정선을 기준으로 볼 때 가장 정확하게 판정하는 불합격 추정선보다 작게 잡을 경우 정확도는 조금씩 감소하나 평균 문항수가 현저하게 감소하는 경향이 나타난다. 즉, SPRT 방식의 판정에 있어서 감소되는 정확도는 합격 판정점을 넘어선 수험자의 점수가 충분한 합격 추정선에 미치지 못할 경우 불합격으로 처리하는 엄격한 판정을 하는 경향을 나타낸다. 따라서 정확도를 조금 낮추어 주는 추정 구간을 선택할 때 최적의 추정 구간의 평균 문항 수 보다 적은 수의 문항만으로 보다 엄격하게 판정할 수 있다.

표 7은 그림 1의 결과를 토대로 작성된 SPRT 판정을 위한 최적의 추정 구간을 평가하기 위한 결과표이다. 각각의 합격 판정점을 기준으로 가장 정확도가 높으면서 비교적 출제 문항의 평균이 적은 합격 및 불합격 추정구간을 나타내고 있다. 표 7의 결과를 고

표 7. SPRT 판정을 위한 최적의 초기 설정치

등급 (i)	합격 판정점 (X_i)	합격추정선-불합격추정선 (구간)	구간 거리
0	90	95 - 85	$[X_i - 5, X_i + 5]$
1	82.5	90 - 75	$[X_i - 7.5, X_i + 7.5]$
2	75	85 - 65	$[X_i - 10, X_i + 10]$
3	67.5	80 - 55	$[X_i - 12.5, X_i + 12.5]$
4	60	75 - 45	$[X_i - 15, X_i + 15]$
5	52.5	70 - 35	$[X_i - 17.5, X_i + 17.5]$
6	45	65 - 25	$[X_i - 20, X_i + 20]$

려해 보면 다음과 패턴이 있음을 알 수 있다.

- (1) 최적의 합격 판정점 간격: 7.5
- (2) 최적의 합격 및 불합격 추정선 구간:

$$[X_i - (5 + 2.5 \times i), X_i + (5 + 2.5 \times i)], \quad i = 0, \dots, 6.$$

위와 같은 패턴으로 합격 추정선과 불합격 추정선을 이동시켜 설정할 때 전체적인 정확도와 평균 출제 문항수는 일관성 있게 바뀌는 것을 알 수 있다. 즉, 가장 정확도가 높은 구간에서 전체적으로 합격 추정선과 불합격 추정선을 2.5점씩 낮추어 줄 때 정확도는 감소하면서 평균 출제 문항의 수도 줄어드는 일관성 있는 경향을 보이고 있다.

각 추정 구간을 가장 잘 대표하는 최적의 합격 추정선과 불합격 추정선을 찾아 각 추정 구간에 대하여 다단계로 SPRT를 실행하여 판정하도록 하였다. 다단계로 SPRT를 실행하는 과정의 예가 그림 2에 주어져 있다.

등급	A (95-85)	B (90-75)	C (85-65)	D (80-55)	E (75-45)	F (70-35)	G (65-25)
초기 판정 상태	D	D	D	D	D	D	D
(판정 미정 상태)							
1차 판정	N	N	D	D	D	M	M
↑ (1차 판정 기준)							
2차 판정	N	N	D	D	M	M	M
↑ (2차 판정 기준 : 구간제 채택)							
최종 판정	N	N	N	D	M	M	M
↑ (최종 판정 기준)							

그림 2. M-SPRT 판정 과정(M: 합격, N: 불합격, D: 미정)

일반적으로 한 구간에서 SPRT를 실행하였을 경우 합격 판정점에 분포한 수험자 성적에 대하여 많은 문항으로도 판정을 못하는 경우가 빈번하게 발생하게 된다. 즉 100문항의 문제를 모두 다 풀어도 판정을 못하게 되는 경우가 많다. 따라서 가장 잘 대표하는 각 추정 구간을 설정하여 SPRT를 다단계로 실행하였을 경우 여러 추정 구간의 판정을 고려할 수 있게 된다. 합격 판정점에 분포한 수험자의 성적은 일반적으로 판정을 하지 못하는 'D' 판정이 내려지게 되는데, 다른 추정 구간의 판정을 고려할 때 마지막 판정 미정 구간은 곧 수험자의 성적으로 유추할 수 있다. 따라서 합격, 불합격 판정만이 아닌 수험자의 학습 성취도를 여러 등급으로 구분할 수 있게 된다.

4. 결 론

본 논문에서는 문항 수가 두드러지게 감소하면서 판정이 비교적 정확한 SPRT 판정방식을 도입하였다. 그리고 실제 237명의 피험자를 대상으로 실험한 데이터를 토대로 SPRT 전 초기 설정치에 대한 시뮬레이션 연구를 실시하였으며, SPRT의 정확도와 특성을 파악하였다.

효과적인 SPRT 판정을 위하여 SPRT 판정의 정확도, 판정 미정의 수, 평균 출제 문항 수 그리고 실제 교육 현장에서 이용될 판정점수 등 모두를 고려한 최적의 SPRT 초기 설정치 즉, 최적의 합격 추정선과 불합격 추정선의 패턴을 찾고 근거를 제시하였다.

본 시뮬레이션의 연구결과를 바탕으로 다단계 SPRT 판정 방법 및 기준을 제시하였고 일반적인 SPRT 방식의 문제점을 개선하였다.

시뮬레이션은, 허용오차 α , β (제1종 오류, 제2종 오류) 값을 모두 같은 값으로 설정하되 0.05와 0.025의 두 경우로 구분하여 SPRT 판정의 정확도 검증에 위한 시뮬레이션과 SPRT 초기 설정치에 따른 정확도 시뮬레이션의 측면에서 연구하였다.

본 논문에서 제안한 다단계 SPRT 판정 방법을 통하여 학습자의 학습수준을 세분화하여 여러 판정 구간에서 동시에 평가하고 분류할 수 있었다. 그리고 현실적으로는 많은 문항 수의 문제가 제시되고도 종종 발생하는 판정 미정 상태는 인접한 추정구간에서의 판정 결과를 토대로 판정 미정구간을 기준으로 학습자의 학습 수준을 신속하고 정확하게 판정할 수

있었으며, 결국 최종 판정 미정 구간은 학습자의 학습 수준으로 추정할 수 있었다. 그리고 SPRT를 효율적이며 효과적으로 활용할 수 있도록 최적의 SPRT 초기 설정치에 대한 패턴과 합리적인 판정 방법을 제시하였고 시뮬레이션을 통하여 기준의 타당성을 검토하였다.

참 고 문 헌

- [1] 범수균과 오암석, "멀티미디어가 지원되는 SPRT 시스템 설계 및 구현," 멀티미디어학회 논문지, 제5권, 제6호, pp. 638-644, 2002.
- [2] 김영환, 손미, 정희태, 김영진, "SPRT형 CAT의 판정에 대한 학습자의 풀이시간 자신감의 영향," Korean Association for Educational Information and Broadcasting, 제8권, 제4호, pp. 93-117, 1997.
- [3] 김영환, 손미, "컴퓨터를 활용한 적응적 개별학습 성취도 검사의 제작과 활용을 위한 표준화 지침 개발 및 효과 연구," 한국방송연구회, 제3권, pp. 109-130, 1997.
- [4] 박성익, 최정임, "CAI 코스웨어, 교수자료 개발을 위한 교수설계의 원리와 적용," 교육과학사, 1995.
- [5] 김영환. "컴퓨터에 의한 적응적 평가 프로그램의 교육적 활용," 전국학교컴퓨터교육연구회 소식지, 제14호. 1994.
- [6] 성태제, "컴퓨터 이용검사와 컴퓨터 능력 적응 검사," 교육평가연구, 제5권, 제1호, pp. 73-97. 1992.
- [7] Frick, T., "The Indiana Testing System (ITS, Version 1.0)," Department of Instructional Systems Technology, School of Education, Indiana University, Bloomington, 1986.
- [8] Kingsbury G. & Weiss D., "A Comparison of IRT-Based Adaptive Mastery Testing and Sequential Mastery Testing Procedure," in New Horizons in Testing, Weiss D.(ed.), Academic Press, New Yourk, pp. 257-283, 1983.
- [9] Spary J. & Reckase M., "Comparison of SPRT and Sequential Bayes Procedures for

Classifying Examinees into two Categories Using a Computerized Test," Journal of Education and Behavioral Statistics. Vol.21, No.4, pp. 405-414, 1996.

[10] Theodore W., "Bayesian Adaption During Compter-based Tests and Computer-guide Practice Exercises," Educational Computing Research, Vol.5, No.1, 1989.

[11] Wald A. "Sequential Analysis", Wiley, New York, NY: Wiley., 1947.



오 암 석

1984년 부산대학교 전자계산학과 이학사

1986년 중앙대학교 컴퓨터공학과 공학석사

1991년 부산대학교 컴퓨터공학과 공학박사

1987년~1990년 LG연구소 연구원

1990년~1998년 울산과학기술대학교 전자계산과 부교수

1998년~현재 동명대학교 컴퓨터미디어공학과 교수

관심분야 : 멀티미디어 데이터베이스 및 응용, 상황 인식, 홈네트워크 시스템



전 미 란

2006년 동아대학교 경영대학원 경영학석사

2009년 동명대학교 일반대학원 컴퓨터미디어 공학박사
수료

관심분야 : 컴퓨터 교육, 전자거래, 고객관리시스템