



# 로지스틱 회귀 모형과 서포트 벡터 머신을 이용한 개인 맞춤형 학습 가이드 추천 알고리즘 제안



OnE Education

최인재 <sup>1\*</sup>, 장대일 <sup>1</sup>, 윤옥수 <sup>2</sup>, 박연숙 <sup>3</sup>, 문찬호 <sup>3</sup>, 이재화 <sup>3</sup>, 박영선 <sup>1</sup>

<sup>1</sup> 한양대학교 수학과

<sup>2</sup> 목포해양대학교 교양과정부

<sup>3</sup> 오앤이교육 수학교육공학연구소

## 연구 배경

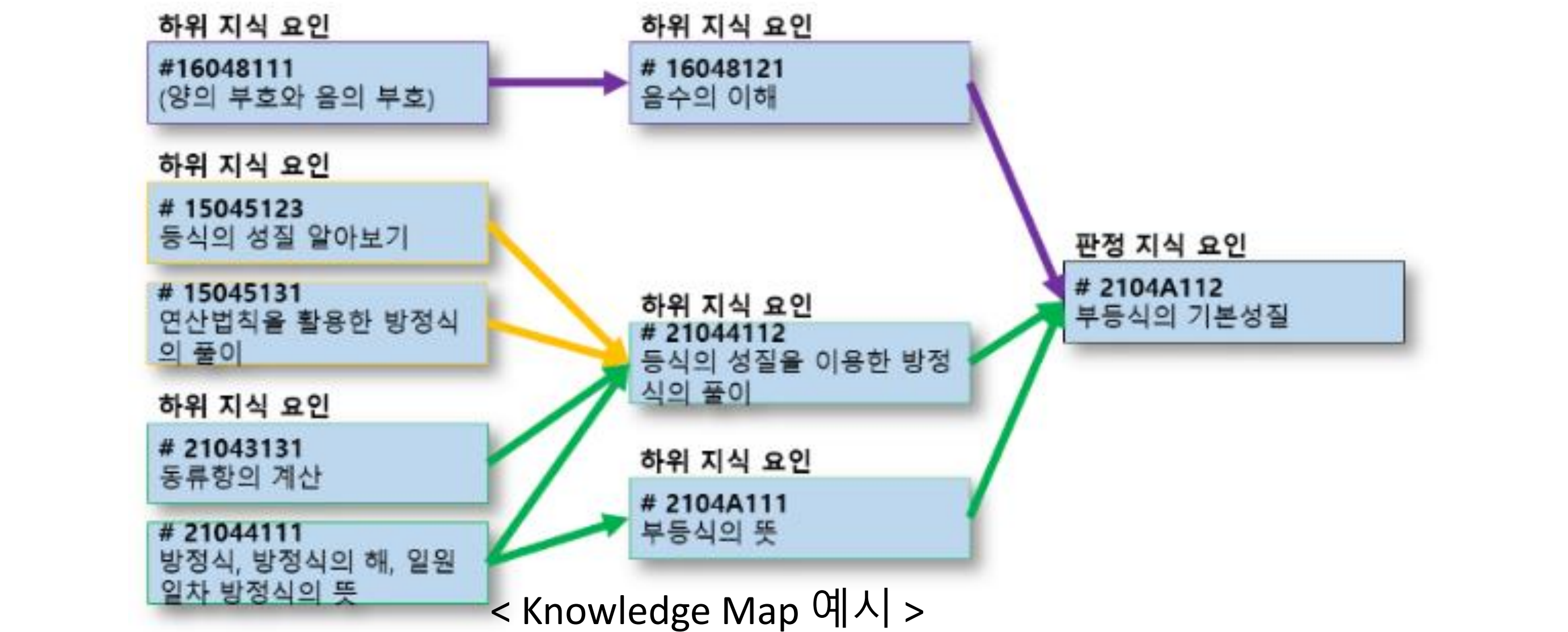
- 온라인 교육 플랫폼에서 지식 오류 상황에 따른 규칙 기반의 처방 학습은 개인화된 맞춤형 가이드 제시를 하지 못하는 한계가 존재
- 오앤이교육의 학습 진행 과정을 통해 학습자의 지식 상황에 따라 인공지능을 기반으로 학습 네비게이션을 제안
- 오앤이교육의 문항반응이론에 의한 추정 솔루션과 Q-matrix를 통해 도출되는 이상반응 변수를 추가한 개인화된 처방 학습 시스템에 인공지능을 결합한 최적의 학습 경로를 제시

## 연구 목적

- 본 연구를 통해 학습자의 지식이해상태와 각 지식 요인 별 문항 반응 이론에 따른 이상반응을 도출하여, 학습자 개인 맞춤형이 가능한 학습 지도를 추천
- 지식이해상태와 이상반응 변수에 따른 판정 지식 학습을 위한 학습 지도에서 인공지능 기반의 최적의 경로를 제시

## 데이터 설명

- 학습자의 이전 학습 이력과 현재 단계에서 학습한 데이터를 통해서 판정하고자 하는 지식 요인에 대한 학습 성취 여부 판단
- Knowledge Map : 수학 교육과정에 필요한 지식 요인을 학년별/단원별로 구분하여 모든 판본 별 두 지식 요인 간의 직접적인 인과관계(선행되어야 할 지식)를 정의하여 포함 관계로 연결한 지식 지도
- 지식이해상태 : [측] 단계 진단 검사 후 학습자의 각 지식 요인 별 정답률을 1과 0으로 구분하고, 그 사이의 값을 가지는 경우 2로 구분한 것.
- 이상반응 : [측] 단계 진단 검사에 Q-Metrics를 적용하여 학습자의 현재 지식이해상태와 가장 근접되는 정상 반응을 1 또는 0으로 나타낸 것.
- 최종 결정 : 판정 지식에 대한 Knowledge Map을 Rule 기반으로 학습한 뒤, 7문항의 Small Step 진단 검사를 통해 Mastery, Partial Mastery, Non Mastery로 판정한 최종 학습 결과



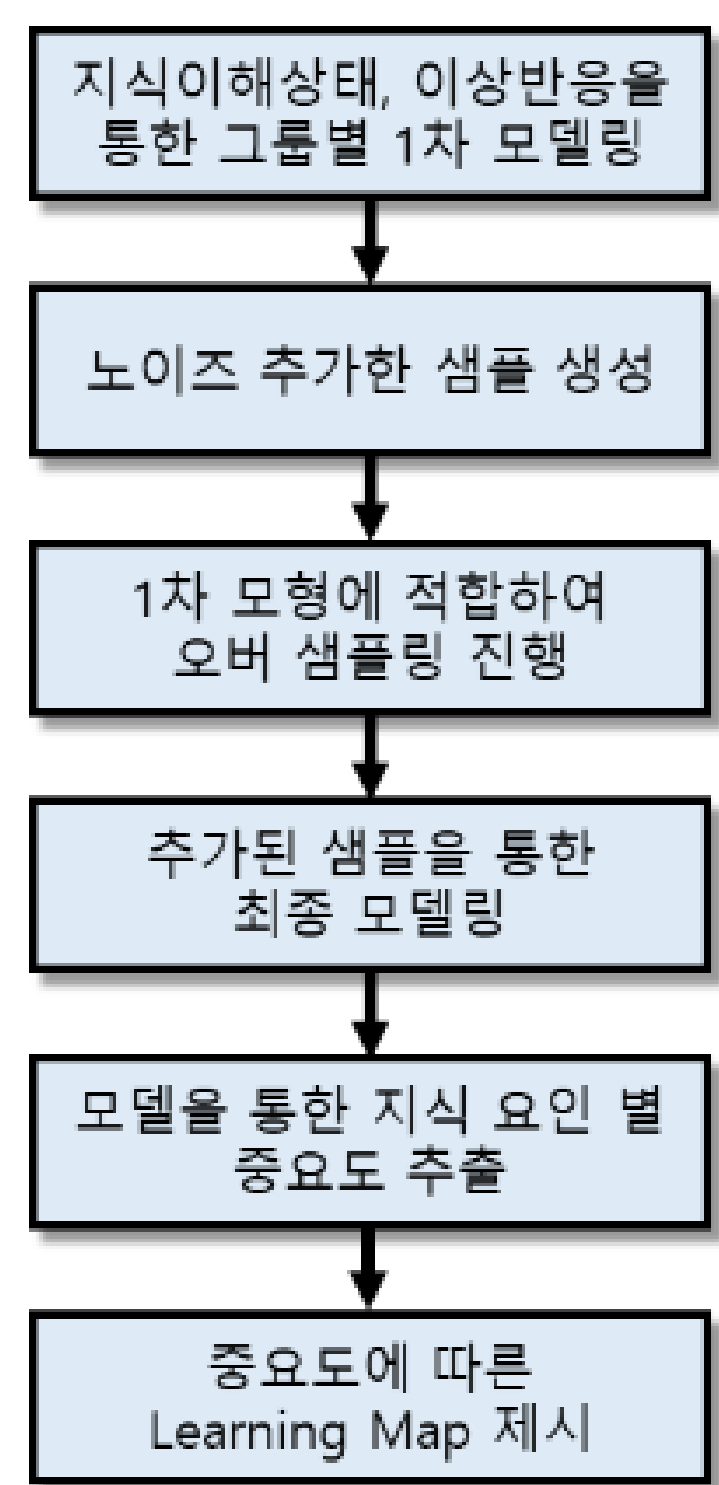
| 문항 번호 I21030902514 |        |                 |      |    |          |           |
|--------------------|--------|-----------------|------|----|----------|-----------|
|                    | 지식이해상태 | 판정지식요인의 [측] 정답률 | 이상반응 | 등급 | θ(THETA) | 표준오차 (SE) |
| 학생1                | 2      | 0.85            | 1    | 1  | 3.5487   | 0.6310    |
| 학생2                | 2      | 0.79            | 1    | 1  | 3.1652   | 0.5897    |
| 학생3                | 2      | 0.65            | 1    | 3  | 1.5901   | 0.4312    |

| I11180300096 |          |          |          |          |          |          |          |          |       |
|--------------|----------|----------|----------|----------|----------|----------|----------|----------|-------|
| Level 1      | Level 2  | Level 1  | Level 2  | Level 2  | Level 2  | Level 2  | Level 1  | Level 2  |       |
| 16048121     | 16048111 | 21044112 | 15045123 | 15045131 | 21043131 | 21044111 | 2104A111 | 2104A111 | 학생    |
| 1            | 1.00     | 1        | 1.00     | 1.00     | 1.00     | 1.00     | 1        | 1.00     | 최종 결정 |
| 1            | 1.00     | 1        | 1.00     | 1.00     | 1.00     | 1.00     | 1        | 1.00     | 학생1   |
| 1            | 1.00     | 1        | 1.00     | 1.00     | 1.00     | 1.00     | 1        | 1.00     | 학생2   |
| 1            | 1.00     | 1        | 1.00     | 1.00     | 1.00     | 1.00     | 1        | 1.00     | 학생3   |

< 데이터 예시 >

## 분석 방법

- 이전 학습 이력을 Q-matrix를 통해서 지식 이해상태와 이상반응 측정하여 2개의 그룹으로 분리하여 그룹 별 모델링을 통한 최적의 결과를 제시
- 기존 샘플에 노이즈를 추가한 샘플을 생성하여 모형에 적합함으로써 종속변수를 얻는 방법으로 오버샘플링한 뒤, 최종 모델링 진행
- 로지스틱 회귀 분석과 서포트 벡터 머신을 이용하여 판정 지식을 Master하기 위한 중요도 높은 변수를 제시, 이를 통해 새로운 Learning Map 추천



## 분석 알고리즘

- 로지스틱 회귀 분석은 종속변수가 참일 확률을 거짓일 확률로 나눈 odds(오즈)를 로그 변환하여 구한 logit(로짓)에 대한 회귀 분석을 진행하여 독립변수와와의 관계를 파악

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}$$

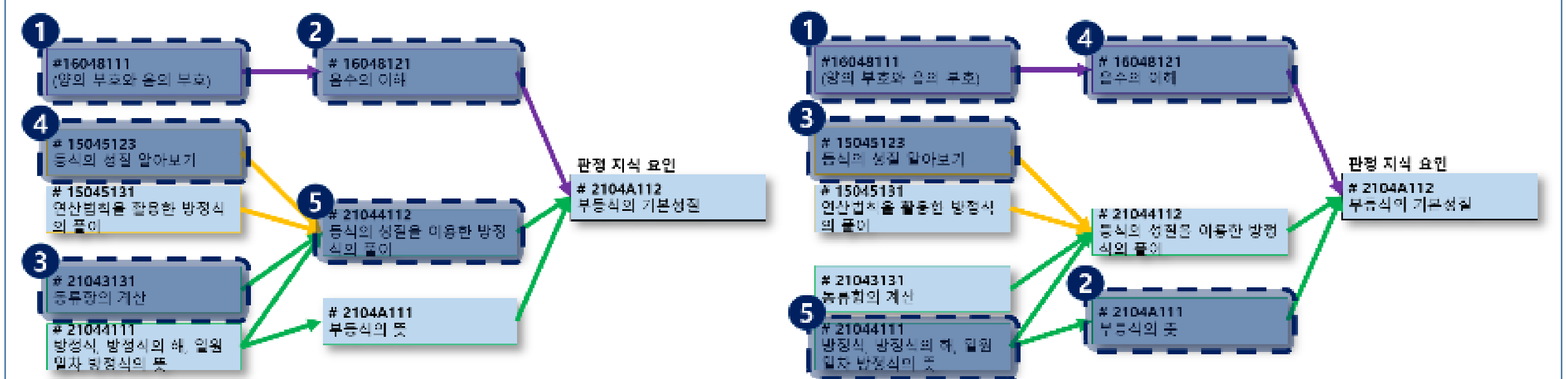
- 위 식에서 회귀 계수에 밑이 e인 지수를 취하면 오즈비를 얻고 이는 변수의 중요도를 나타냄

- 서포트 벡터 머신은 두 클래스를 분류하는 초평면(hyperplane)을 찾아서 분류하는 머신러닝 기법
- 두 클래스와 초평면 간의 거리를 최대화하는 하는 데이터를 서포트 벡터(support vector)라고 하고, 일반적으로 고차원의 kernel function을 통해 비선형의 형태로 초평면을 추정하기 위해 Lagrangian multiplier를 사용
- 여유 변수(slack variable;  $\xi$ )를 통해 일정한 오분류를 허용하면서 두 클래스 사이의 거리인  $\|w\|$ 를 최소화,  $C$ 는 비용(cost)

$$\min_w \frac{\|w\|^2}{2} + C \sum_{i=1}^n \xi_i$$

## 분석 결과

- 그룹 별 로지스틱 회귀 모형과 서포트 벡터 머신을 통해 문항과 지식 요인 변수들의 중요도 추출 후 Learning Map 생성
- 그룹 별 두 모형이 제시하는 Learning Map에서 공통으로 추천하는 지식 요인을 처방 학습으로 제시



< 그룹 별 SVM을 통한 Learning Map >  
 왼편 : 지식이해상태: 0, 이상반응: 1  
 오른편 : 지식이해상태: 2, 이상반응: 1

- 두 모형의 Map을 종합하여 최종 Map을 제시한다.



## 요약 & 추후 연구

- 그룹 별, 모형 별 중요 변수로 학습자의 능력 변수  $\theta$ (THETA)가 항상 등장하는 것으로 보아 학습자에 대한 정보 변수들을 이용한 추가적인 그룹 세분화를 통한 개인 맞춤형 Learning Map 제시
- 시뮬레이션 데이터가 아닌 실제 학습자 데이터를 이용한 모형 사후 검증을 통해 모형 안정성 확보

- 본 논문은 중소기업기술정보진흥원의 지원을 받아 수행된 연구임(S2796484)