

시각화 방법을 이용한 학습자 학습 성향 진단기의 개선

윤태복^{○*}, 최미애^{**}, 이지형*, 김용세^{**}

성균관대학교 컴퓨터공학과*, 성균관대학교 창의연구단**

{tbyoon, machoi, yskim}@skku.edu, jhlee@ece.skku.ac.kr

Improvement of Learner's Learning Style Diagnosis used Visualization Method

Tae Bok Yoon*, Mi Ae Choi**, Jee-Hyong Lee*, Yong Se Kim**

Computer Engineering Dept. SungKyunKwan Univ. *

CREDITS Research Center, SungKyunKwan Univ. **

1. 서론

지능형 학습 시스템은 학습자의 학습 과정에서 수집된 데이터를 분석하여 학습자에게 맞는 전략을 세우고 적합한 서비스를 제공하는 시스템이다. 학습자에게 적합한 서비스를 위해서는 학습자 모델링 작업이 우선시 되며, 이 모델 생성을 위해서 학습자의 학습 과정에서 발생한 데이터를 수집하고 분석하게 된다. 하지만, 수집된 데이터가 학습자의 일관되지 못한 행위나 비예측 학습 성향을 포함하고 있다면, 생성된 모델을 신뢰하기 어렵다. 본 논문에서는 학습자의 학습 과정에서 얻은 학습 행위 데이터를 이용하여 학습 성향을 진단하는 방법과 시각화 방법을 이용하여 비정상적인 데이터를 감소(Reduction)하여 진단기의 성능을 향상시키는 방법을 제안한다. 시각화로 표현된 학습자의 데이터는 전문가의 지식(Expert Knowledge)을 이용하여 분류/선별하고, 학습자 모델 생성에 이용한다. 비정상 학습 데이터를 분류하여 생성된 진단기와 비정상 데이터 분류하지 않는 데이터를 이용하여 만든 진단기의 성능을 비교/분석하였다. 실험에서는 홈 인터리어 콘텐츠 기반에 학습자의 학습 행위에 대한 학습 성향을 진단하기 위한 DOLLS-HI를 이용하여, 수집된 학습자의 데이터에서 비정상 데이터를 분류하고 학습 성향 진단을 위한 모델을 생성하였다.

2. 학습자 성향 파악을 위한 연구 및 환경 구현

학습자는 학습 과정에서 정보를 받아들이고 이해하는 방식에서 다양한 모습을 보여주고 있다. 예를 들면 문자로 설명된 내용 보다는 그림으로 설명된 학습 콘텐츠를 더 선호하거나, 학습 과정에 있어서 순서대로 학습하는 것보다 순서에 상관없이 자신이 원하는 정보를 자유롭게 찾아보면서 학습하는 것을 더 선호하는 학습자가 있을 것이다. Felder & Silverman[1]은 학습 정보를 이해하는 차원에서 Global과 Sequential, 정보를 습득하는 차원에서 Visual과 Auditory, 정보를 인지하는 차원에서 Sensing과 Intuitive, 그리고 정보를 활용하는 차원에서 Active와 Reflective로 네 가지 영역에서 학습 성향을 분류하였다. 앞서 설명한 학습자 성향의 분류를 이용하여 홈 인터리어 학습 콘텐츠를 이용한 학습 성향 진단을 위한 DOLLS-HI(Diagnosis of Learner's Learning Styles Housing Interior)[2]를 개발하였다. 홈 인터리어 학습을 위한 DOLLS-HI는 학습과정에서 수집된 버튼 클릭 정보, 시간 정보, 퀴즈 풀이과정에서의 정답률 등을 수집하고 분석하여 학습자의 성향을 알아내는데 이용된다. 그림1은 학습자 성향 진단을 위한 DOLLS-HI의 전체 흐름도이다.

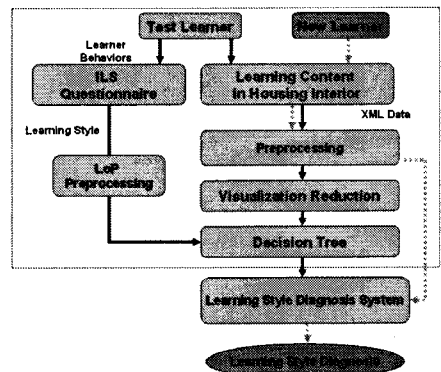


그림 1. 학습자 성향 진단을 위한 흐름도

3. 시각화 표현을 이용한 학습자 데이터의 분류

앞서 설명한 바와 같이 학습자의 데이터, 즉 인간으로부터 수집된 데이터는 매우 다양하고 복잡한 모습을 보인다. 이 같은 데이터를 효율적으로 분석하여 유용한 정보 또는 지식을 생성할 수 있는 분석방법으로 본 논문은 시각화(Visualization) 방법을 사용하였다. 시각화는 시각적으로 표현이 불가능한 텍스트 기반의 데이터를 여러 시각적 요소(색, 위치, 선, 크기, 모양 등)를 사용하여 사용자가 이해하기 쉽도록 표현하여, 궁극적으로 사용자의 분석업무를 지원하는 것이 정보시각화라고 할 수 있다. 사용자는 정보시각화 기술을 사용하여 매우 빠르고 정확하게 시각적으로 표현된 대량의 데이터와 상호작용을 하게 되고, 그들로부터 의미 있는 지식을 발견하게 된다[4]. 인지과학 분야에서의 다양한 연구들은 정보시각화가 데이터 탐색을 통한 지식생성 또는 지식획득에 효율적인 도구임을 뒷받침하

고 있다. 이러한 시각화의 장점을 살펴보면 다음과 같다. 첫째, 시각화는 복잡한 구조를 가지는 데이터의 이해를 쉽게 하는 능력이 있다. 둘째, 시각화는 예측되지 못한 급작스런 속성의 인지를 가능하게 해준다. 셋째, 시각화는 데이터가 가지고 있는 결함을 빠르고 분명하게 보여준다(데이터 수집의 오류, 시스템 오류 등). 넷째, 시각화는 대용량 데이터의 특징을 이해 할 수 있도록 한다. 다섯째, 시각화는 가설 형성을 도와준다.

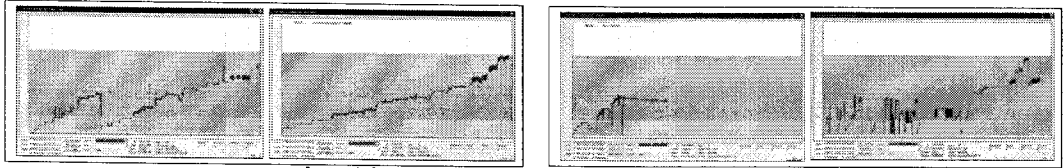


그림 2. (좌) 정상 학습 데이터 (우) 비정상 학습 데이터

그림 2는 DOLLS-HI에서 얻은 데이터를 시각화 표현으로 나타낸 결과이다. 학습자의 학습 시간 (1200초)에 따라 학습에 사용한 콘텐츠(학습하기, 퀴즈풀기, 실습하기 등) 클릭 정보를 이용하여 학습자의 학습 경로(Learning Path)를 나타내었다. 이 시각화를 이용하여 학습자의 학습 행위를 정상학습 행위 4가지와 비정상 학습 행위 4가지로 분류하였다.

4. 학습자 학습 행위 데이터의 진단을 위한 실험

실험을 위하여 DOLLS-HI 학습 콘텐츠를 이용하여 성균관대학교 신입생중 372명을 대상으로 실험을 실시하였다. 먼저 실험 대상학생들에게 ILS 온라인 설문지를 통하여 학습 성향(Global & Sequential, Visual & Auditory, Sensing & Intuitive, Active & Reflective)을 파악하였다. 그리고, 학생들에게 DOLLS-HI를 이용하여 20분 동안 하게 하였다. 학습 후에 수집된 학생들의 데이터는 XML파일로 저장된다. 그 파일에는 시간에 따라 학습한 학습 콘텐츠의 정보가 순서대로 저장되어 있다. 앞서 설명한 시각화 분류 방법을 이용하여 비정상적인 학습자를 모두 분류하였다. 전체 수집데이터 372명중에서 의미없는 학습, 부분 학습, 학습하지 않음, 데이터 수집 오류에 해당하는 학생을 제외하고 남은 데이터의 수는 259명이었다. 선별된 학생들의 데이터는 전처리 과정을 거쳐 의사 결정 트리를 이용하여 분석하였다. 표 1은 전체 372명의 데이터와 가지적 선별법을 통해 선별된 259명의 데이터를 의사결정 트리를 이용하여 각각 분석한 결과이다. 교차검증방법(Cross-Validation)을 이용하여 10회 실시하였으며, 트레이닝 데이터의 비율은 65%이고, 나머지 35%의 데이터는 테스트로 이용하였다.

표 1. 의사결정트리를 이용한 에러율(%)

	Active & Reflective		Global & Sequential	
	Normal	Visualization Reduction	Normal	Visualization Reduction
1	50	44	44	44
2	45	39	45	42
3	42	44	50	47
4	56	49	52	42
5	50	48	47	43
6	56	45	53	40
7	36	47	48	39
8	44	37	42	44
9	53	44	50	35
10	52	48	49	47
avg.	48.4	44.5	48	42.3
style rate	186/186	134/125	193/179	146/113
training/test	248/124	172/87	248/124	172/87

5. 결론

학습자의 데이터를 분석하는 과정에서 시각화(Visualization) 방법을 이용하여 비정상 학습 데이터를 선별하는 방법을 소개하였다. 비정상 데이터 제거후, 실험결과에서 Active & Reflective 와 Global & Sequential의 에러율이 각각 3.9%, 5.7% 향상되었고, Sensing & Intuitive와 Visual & Auditory는 각각 0.3%, 0.5% 향상되었다. 비정상 학습자 데이터를 제거하지 않은 결과를 진단 모델로 사용한다면, 새로운 학습자에 대한 진단 결과의 신뢰성을 의심하지 않을 수 없을 것이다. 비정상 학습자 데이터 선별 방법을 이용하여 생성된 진단 모델의 신뢰성을 높이고 서비스의 질적 향상을 예상할 수 있다. 향후 연구로는 시각화 방법으로 표현된 학습자의 데이터를 분류하는 방법을 전문가의 지식을 활용하여 자동으로 처리 할 수 있는 기법이 요구되며, 시각화로 표현된 학습 경로(Learning Path)간의 유사도 측정을 이용한 데이터 분석 방법이 유용할 것으로 예상된다.

참고문헌

[1] Felder, R., Silverman, L. (1988). Learning and Teaching Styles in Engineering Education, Engineering Education, 78(7), pp674-681.
 [2] Cha, H. J., Kim, Y. S., Park, S. H., Yoon, T. B., Jung, Y. M., and Lee, J. H. (2006). Learning Styles Diagnosis based on User Interface Behaviors for the Customization of Learning Interfaces in an Intelligent Tutoring System, Proc. 8th Int'l. Conf. on Intelligent Tutoring Systems (ITS).
 [3] Card, S. K., Mackinlay, J. D., and Schneiderman, B.(1999). Information Visualization: Using vision to think, San Francisco, Morgan-Kaufmann.