

Hidden Markov Model을 이용한 학습자 성향 파악에 관한 연구

A study for classification of students' learning-styles
with HMM

정영모¹, 이지형², 차현진³, 박선희⁴, 윤태복⁵, 김용세⁶,

^{1 3 4 5 6} 성균관대학교 창의적 설계추론 지적 교육 시스템 연구단

E-mail: (ewmhb, lois6934, seonhp2, tbyoon, yskim)@skku.edu

² 성균관대학교 정보통신공학부

E-mail: jhlee@ece.skku.ac.kr

요 약

지능형 학습 시스템(ITS, Intelligent Tutoring System)은 학습자의 학습 스타일을 인지하여 학습자에 맞는 학습전략을 세우고 적절한 학습 서비스를 제공하는 시스템이다. 기존의 학습시스템은 학습자의 학습 스타일 보다는 학습 콘텐츠에 중심을 두어 학습자에게 맞는 학습 전략을 적절히 세우는 과정이 부족했다. 이에 본 논문에서는 학습자의 학습과정에서 발생한 데이터를 기반으로 학습자의 학습 스타일을 파악하는 방법을 제안한다. 이를 위해 서양 건축양식 학습을 위한 교육 콘텐츠를 이용하였으며, 수집된 데이터를 분석하여 Felder & Silverman 이 제시한 학습 스타일에 근거한 학습자의 학습 스타일을 추출하였다. 실험에서는 70명의 데이터를 수집하였고, 학습자가 교육 콘텐츠를 학습한 순서에 대한 시계열 데이터를 기반으로 학습자 성향을 알아보기 위하여 은닉 마코프 모델(Hidden Markov Model)을 사용하였다. 은닉 마코프 모델을 적용하여 얻은 분석 결과를 가지고 각 학습자에게 맞는 학습 스타일을 진단하였다. 은닉 마코프 모델에서 얻은 학습 스타일 진단 모델은 향후에 학습자 학습 스타일을 파악하는데 사용할 수 있으며, ITS에 있어 학습자 성향 분석 모듈로 고려해볼 수 있다.

Key Words : HMM, ITS

1. 서 론

과거에서부터 지금까지 교육과 관련된 다양한 연구가 이루어지고 있다. 교육 지도 방법, 학습전략 세우기, 학습 콘텐츠 구성 등 여러 가지가 있는 데 이중 교육의 지도 방법에 있어 최근엔 인터넷의 보급으로 인해 사이버교육, 원격교육, 온라인 교육, 인터넷 교육 등 다양한 교육 시스템이 구축되어 학교 및 가정, 회사 등에서 사용되고 있다. 이러한 시스템에 최근에는 지능형 학습 시스템(ITS, Intelligent Tutoring System)이라는 개념을 도입하여 기존의 교수 시스템의 단점인 콘텐츠 기반의 단순한 지식 전달이 아닌 학습자의 학습 성향 및 학습 수준 등을 파악하여 학습자에게 맞는 차별화된 교육 시스템을 제공하고자 하는 연구가 활발히 진행되고 있다.

이러한 서비스를 위해서는 교육 시스템에서

자동으로 학습자의 학습 스타일을 파악하여 이에 적절한 교육방식을 선택할 수 있는 기술이 필요하다. 본 논문에서는 학습자의 학습스타일을 파악하는 시스템을 개발하였다.

사용한 개발 방법은, 먼저 Felder & Silverman[1]이 제시한 학습 스타일 분류 방법을 이용하여 피실험자의 학습 스타일을 파악한다. Felder & Silverman은 이미 학습자 성향 분류, 교육 전략 개발 및 학습 전략의 평가와 관련된 연구를 하였고 이는 본 논문에서 파악하고자하는 학습자의 학습스타일에 대한 근거를 제시하였다. 이후 제작된 서양건축양식 학습용 교육 콘텐츠[2]를 피 실험자에게 사용하게 한 다음 그들의 교육 콘텐츠 사용패턴을 추출하고, 추출된 피 실험자들의 사용패턴과 학습 스타일 간의 상관관계는 은닉 마코프 모델을(Hidden Markov Model) 이용하여 모델링하였다.

서양건축양식 학습용 교육 콘텐츠는 크게 서양 건축양식 학습, 퀴즈, 생각 나눔, 테스트로 구성되었다. 각 콘텐츠를 통해 실제 피 실험자는 학습을 하게 된다. 이때 학습자의 학습기록 (교육내용 버튼클릭, 콘텐츠 이동을 위한 버튼클릭, 학습에 걸린 시간등..)을 XML 데이터로 수집한다. 이후 수집한 XML 데이터를 적절히 가공처리한 후 은닉 마코프 모델에 적용하여 각 학습 스타일별 학습자 모델을 생성한다. 이렇게 제작된 은닉 마코프 모델을 통하여 학습 성향을 모르는 학습자의 학습 스타일을 진단하게 된다. 자세한 내용은 2장과 3장을 통해 기술할 것이다.

본 논문의 구성은 다음과 같다. 2장에선 학습자 성향 파악을 위한 관련연구를 기술하며, 3장에선 은닉 마코프 모델에 대한 소개 및 실험방법, 실험 결과를 서술하며, 4장에선 결론 및 향후 연구를 맺는다.

2. 학습자 성향을 파악하기 위한 관련 연구

2.1 학습자 성향 파악을 위한 관련 연구

학습자들은 정보를 받아들이고 처리하는 방식에서 다른 선호도와 스타일을 가지고 있다. Felder & Silverman은 정보를 이해하는 차원에서 Global vs. Sequential (G vs. Q), 정보 습득하는 차원에서 Visual vs. Auditory (V vs. A), 정보를 인지하는 차원에서 Sensing vs. Intuitive (S vs. N), 그리고 정보를 활용하는 차원에서 Active vs. Reflective (C vs. R)로 네 가지 영역에서 학습 스타일을 분류하였다. 다음 표1은 학습 스타일에 관한 설명이다.

학습스타일	설명
Global	Global 학습자는 부분적으로 보고 이해 하지 못하며 전체 학습의 큰 그림이나 개요들을 통해 더 잘 이해하는 경향이 있다.
Sequential	Sequential 학습자는 세부내용을 참을성 있게 학습하며, 표조적인 정해진 방법을 통해 더 잘 이해하는 경향이 있다.
Visual	Visual 학습자는 그림, 차트, 영화, 데모 등으로 본것을 잘 기억하며, 글이나 말로 하는 설명보다는 실습 동영상 등을 통해 더 잘 이해하는 경향이 있다.
Auditory	Auditory 학습자는 학습을 통해 들은 것, 말한것을 좀 더 잘 기억하고 토론을 통해 많은것을 얻는 경향이 있다.
Sensing	Sensing 학습자는 학습의 세부 내용을 주의깊게 공부하며, 구체적인 정보로 구성된 학습자료를 활용할때 효과적인 경향이 있다.
Intuitive	Intuitive 학습자는 어떤 상징화 된것을 다루는데 익숙하며 추상화된 개념을 해석하는 것을 잘하는 경향이 있다.
Active	Active 학습자는 토론하고 설명하고 테스트 하는등의 실험적인 성향을 가지고 있는 경향이 있다.
Reflective	Reflective 학습자는 습득한 지식과 정보를 시험해보고, 처리하는 성향을 가지며, 혼자 또는 한사람과 적을 이루어 공부할 때 효과적인 경향이 있다.

표 1. 학습 스타일 및 설명

2.2 학습 스타일별 인터페이스 구성

G vs. Q : 정보를 이해할 때 학습자의 선호에 의해 교육 콘텐츠를 선택 하는지, 교육 전문가가 의도 하는 학습 순서에 따라 학습 하는지를 알아보기 위하여 개요보기 버튼, 목차를 통한 이동버튼과, 항목이동 화살표 버튼 등으로 인터페이스를 설계하였다.

V vs. A : 정보를 습득할 때 그림위주의 설명을 선호하는지 혹은 텍스트 위주의 설명을 더 선호하는지를 알아보기 위하여 그림위주의 학습설명버튼과 텍스트위주의 학습 설명버튼 등으로 인터페이스를 설계하였다.

S vs. N : 정보를 인지할 때 주위 깊게 문제를 풀어나가는 경향이 있는지 직관적으로 문제를 풀어나가는 경향이 있는지를 알아보기 위하여, 퀴즈 풀기와 부가 학습을 구성 하였다.

C vs. R : 정보를 활용할 때 적극적인인지 수동적인지를 알아보기 위하여 선생님께 질문하기버튼, 의견달기버튼, 의견 보기버튼을 구성 하였다.

다음 그림 1은 학습 스타일 관점에 따른 서양 건축양식 학습용 교육 콘텐츠 인터페이스의 일부분 이다.



그림 1. 서양 건축양식 학습용 교육 콘텐츠 인터페이스

3. 은닉 마코프 모델 (Hidden Markov Model) 소개 및 실험

3.1 은닉 마코프 모델 (Hidden Markov Model)

데이터 마이닝 기법에는 신경망(Neural Network), 결정트리(Decision Tree), 은닉 마코

프 모델(Hidden Markov Model)등 여러 가지 기법이 있다. 이 중 본 연구에선 은닉 마코프 모델을 적용하고자 한다. 은닉 마코프 모델은 통계학적 방법으로 시스템의 출력 값에 해당하는 시퀀스(sequence)정보를 가지고 시스템을 모델링 하는데 사용하는 방법이며 시간적, 순차적인 데이터를 다루는데 적합하다.[3,4] 본 연구에서는 학습자가 서양 건축양식 학습용 교육 콘텐츠를 학습함으로써, 즉 시간에 따라 순차적으로 데이터가 기록됨으로써 어떠한 시퀀스 패턴이 나올 수 있으며, 이러한 시퀀스는 분석할 필요가 있다. 다음 단원은 시퀀스 패턴을 얻고 분석하는 실험과정을 다룬다.

3.2 실험 과정 및 가공처리(Preprocessing)

학습 스타일을 분류하기 위한 전체 프로세스는 Felder & Silverman이 제시한 학습 스타일 일을 기반으로 학습 성향을 미리 알아보기 위한 ILS Questionnaire 프로세스와 학습 성향 데이터를 수집하기 위한 서양 건축양식 학습용 교육 콘텐츠 인터페이스 학습 프로세스로 구성 되어 있다.

실험과정에 앞서 ILS(Index of Learning Style)에 대한 간략한 소개를 다루겠다.

ILS는 Felder & Silverman이 제시한 학습자의 학습 스타일 경향지수이며 이는 온라인 설문지를 통해 알아볼 수 있다. ILS는 표1에서와 같이 학습 스타일을 정수 값으로 표현하는데, 이때 이 학습 경향 지수를 본 연구에선 학습 선호도(LoP : Level of Preference)라고 명칭 한다. 선호도는 1, 3, 5, 7, 9, 11인 6개의 정수 값으로 학습자의 학습 스타일경향을 분류하는데, 예를 들어 Global의 학습 성향을 갖는 학습자가 두 명이고 각각 선호도 값이 9와 1이라고 한다면 선호도 값이 9인 학습자는 선호도 값이 1인 학습자에 비해 Global의 성향을 강하게 가지고 있는 학습자라고 볼 수 있다. 이렇게 선호도 값이 비교적 높은 값을 가지고 있는 학습자는 해당 학습 성향이 강한 학습자로 분류할 수 있으며, 반대로 선호도 값이 낮은 학습자는 학습 성향을 분류하기에 불분명하다고 볼 수 있다. 이는 반대의 학습 성향도 어느 정도 갖고 있다고 볼 수 있기 때문이다. 실험 과정은 다음과 같다.

우선 학습자는 ILS 온라인 설문에 참여하여, 학습 스타일별로 선호도 값을 알아본다. 이후 학습자는 학습자 성향 수집을 위한 서양 건축양식 학습용 교육 콘텐츠를 학습하며, 이 인터페이스에서 제공하는 건축양식 학습, 퀴즈, 생각 나눔, 테스트를 수행한다. 이 때 학습자의 데이터(시간, 이동 및 학습을 위한 버튼 클릭

등)는 XML DATA로 기록된다. 이후 전체 피실험자의 XML 데이터가 수집이 되면 학습자 성향을 분석하기 위해 은닉 마코프 모델에 적용하게 되는데 이를 적용하기 전에 적절한 가공처리(Preprocessing)를 해야 한다. 왜냐하면 학습자에게서 수집한 학습 데이터는 학습 스타일별로 무작위로 섞여있기 때문이다. 가공처리는 다음과 같이 수행한다.

첫 번째 ILS 온라인 설문지를 통해 학습 스타일별로 파악한 선호도 값이 3이하라면 해당 데이터는 분석에서 제외해야한다. 이 데이터는 학습 성향을 분류하기에 앞서 언급한 바와 같이 상반된 학습 성향도 가질 수 있는 모호한 데이터이기 때문이다.

두 번째 은닉 마코프 모델은 입력 값으로 시퀀스 정보를 고려하므로, 시퀀스 정보가 필요한 속성(attribute)을 정의해야한다. 정의한 속성에 따라 학습과정에서 발생한 시퀀스를 만들어야한다. 다음은 속성에 관한 설명과 시퀀스를 만드는 방법에 대해 기술한다.

속성(Attribute)은 학습자의 학습 스타일을 알아내기 위한 것으로 교육 전문가가 학습 스타일 영역별로 정의하였다. 학습자가 서양 건축양식 학습용 교육 콘텐츠를 사용하고 수집한 데이터는 정의한 속성에 따라 추상화 시켰다. 예를 들어 A와 B라는 학습 콘텐츠가 있고 A와 B에는 각각 A₁, B₁이란 콘텐츠가 있고 다시 A₁, B₁에는 A₁₁, B₁₁ 이란 콘텐츠가 계층적으로 있다고 가정한다. 여기서 학습자는 A, A₁, A₁₁, B₁₁인 순으로 학습 하였다. 이때 속성을 상위 레벨에 있는 콘텐츠의 이동을 고려해서 정의하게 되면 앞선 학습 순서는 A, A, A, B인 시퀀스 정보로 만들 수 있다.

G vs. Q : 큰 목차를 가지고 속성을 정의하였다. 예를 들어 그림 1에서 보듯이 서양 건축양식 학습에는 그리스 양식, 로마양식, 고딕 양식, 로마네스크 양식이 있다. 이때 교육전문가가 의도한 학습 순서인 그리스, 로마, 고딕, 로마네스크 양식 순서로 학습을 하였는지를 알아봄으로써 학습자의 학습스타일을 분류할 수 있으며, 이러한 9개의 속성을 가지고 시퀀스를 만들었다.

V vs. A : 그림위주의 학습버튼, 텍스트위주의 학습버튼 등 18개를 속성으로 정의하여 이를 기반으로 학습자를 관찰 하였다.

S vs. N : 퀴즈를 풀기위한 시도를 얼마나 했는지 등 13개를 속성으로 정의하였다.

C vs. R : 선생님께 질문하기버튼, 의견 보기버튼 등 10개를 속성으로 정의 하였다.

이와 같이 학습 스타일 영역별로 정의한 속성을 가지고 시퀀스를 만들며 이 시퀀스는 은닉 마코프 모델의 입력 값으로 적용한다. 은닉

마코프 모델은 이 시퀀스를 입력 값으로 하여 학습 스타일별로 학습 모델을 구성 한다. 학습 모델은 학습 스타일을 알지 못하는 어떤 학습자의 시퀀스 정보가 입력 값으로 들어온다면 해당 학습자의 학습 스타일을 진단해주는 역할을 한다.

3.3 은닉 마코프 모델 구성 및 실험 결과

3.2 실험 과정에서 학습 스타일별로 속성의 정의에 따른 시퀀스 데이터를 가지고 은닉 마코프 모델을 구성하고 테스트 하였다. 은닉 마코프 모델의 구성과 테스트 비율은 5 : 2 로 하였다. 다음 표 2는 학습 스타일 영역별 전체 데이터 개수와 학습, 테스트 데이터 개수이다.

	Global vs. Sequential	Visual vs. Auditory	Sensing vs. Intuitive	Active vs. Reflective
전체 데이터 개수	28	49	36	23
학습 데이터 개수	21	35	27	17
테스트 데이터 개수	7	14	9	6

표 2. 영역별 전체, 학습, 테스트 데이터 개수

표 2와 같이 학습 데이터를 가지고 학습 시키면 학습 스타일의 특성을 가진 은닉 마코프 모델이 만들어지고 이 모델에 학습자의 학습 시퀀스 정보를 입력 값으로 넣게 되면 학습 시퀀스에 대한 확률 값이 도출 된다. 예를 들어 G vs. Q의 경우 각각의 특성을 가진 은닉 마코프 모델 두 개를 구성한다. 이 때 학습 성향을 모르는 시퀀스 데이터를 이 두 개의 은닉 마코프 모델에 넣게 되면 두 개의 은닉 마코프 모델은 각각 확률 값을 도출 하게 되는데, Global모델에서 0.084 확률 값, Sequential모델에서 0.053 확률 값을 도출한다면 이 확률 값 중 Global 모델이 큰 확률 값을 취하므로 학습자는 Global의 학습 성향을 좀 더 반영한다고 볼 수 있다. 이렇게 각 학습 스타일별로 8 개의 은닉 마코프 모델을 구성하고 ILS에서 파악한 학습 성향을 알고 있는 데이터를 테스트 데이터로 하여 학습자의 학습 성향을 올바르게 진단했는지에 대한 실험을 하였다. 다음 표 3은 이에 대한 결과다.

DTs	Global vs. Sequential	Visual vs. Auditory	Sensing vs. Intuitive	Active vs. Reflective
테스트 데이터 개수	7	14	9	6
맞춘 데이터 개수	5	14	7	4
틀린 데이터 개수	2	0	2	2
에러율	28.57%	0.0%	22.22%	33.33%

표 3. 학습 스타일 진단 실험에 대한 결과

표 3에서 보듯이 네 가지의 학습 스타일(G vs. Q, V vs. A, S vs. N, C vs. R)에서 각각

28.57%, 0.0%, 22.22%, 33.33%의 에러율을 얻었다. 에러율은 전체 테스트 데이터 개수를 올바르게 분류하지 못한 테스트 데이터 개수로 나누어 백분율을 구한 값이다. 여기서 V vs A는 0%의 에러율을 보였다. 이는 교육 콘텐츠가 그림과 텍스트를 명확하게 구분하여 설계되었기 때문에 분석에도 용이했다. 반면 C vs. R은 33.33 %의 에러율을 보였는데 이는 학습자가 실험시간의 부족으로 C vs. R에 해당하는 콘텐츠를 많이 이용하지 않았기 때문에 데이터 수집에 있어 어려움이 있었다.

4. 결론 및 향후 연구

ITS에서 학습자 성향을 분류하기 위해 수집한 데이터는 다양한 방법으로 분석해야 한다. 여러 분석 방법 중 은닉 마코프 모델은 시간에 따른 행동의 변화를 통해 특정 패턴을 얻고 이를 통해 데이터를 분류해 주는 방법이다. 그러므로 이 논문에선 은닉 마코프 모델을 적용하여 학습자 성향 진단하고자 시도 하였다. 표 3에서와 같은 비교적 적은 에러율을 갖는 은닉 마코프 모델을 구성하였으며, 이 은닉 마코프 모델은 학습자의 학습 성향을 분류할 수 있는 진단기로 사용할 수 있다. 본 연구에선 70명의 비교적 적은 데이터를 가지고 실험하였으나 향후 연구엔 이보다 많은 데이터를 가지고 좀 더 다양한 시도를 통해 은닉 마코프 모델을 구성하고 테스트를 시도할 것이다. 아울러 학습자의 스타일을 분석을 토대로 학습자에 맞는 학습자만의 차별화된 인터페이스를 제공하는 실험도 시도할 예정이다.

참고문헌

- [1] Felder, R., Silverman, L.: Learning and Teaching Styles in Engineering Education, Engineering Education (1988)78(7) 67468
- [2] Cha, H. J., Kim, Y. S., Park, S. H., Cho, Y. J. Pashkin, M.: Adaptive Learning Interface Customization based on Learning Styles and Behaviors, Proc. International Conference on Computers in Education, Singapore (2005)
- [3] Lawrence, R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Morgan Kaufmann Publishers Inc., San Francisco (1990) 267-296
- [4] Fok, A. W. P., Wong, R. H. S., Ip, H. H. S.: Adaptive User Interface for Personalized Education based on Hidden Markov Model, Proc. of the 9th Global Chinese Conference on Computers in Education, Hawaii (2005)