

조건부 적대적 생성 신경망을 이용한 학생들의 수업 추천 모델

서은성, 김인수, 석준희*

고려대학교, 고려대학교, *고려대학교

seoun4612@korea.ac.kr, dlstn0910@korea.ac.kr, *jseok14@korea.ac.kr

Student Course Recommendation Model Using Conditional Generative Adversarial Networks

Seo Eun Seong, Kim In Soo, Seok Jun Hee*

Korea Univ., Korea Univ., *Korea Univ.

요 약

본 연구에서는 현재 수강 신청 시스템의 한계를 극복하기 위해 학생들이 커리큘럼을 설계할 수 있도록 인공지능 모델을 개발하였다. 학생들의 수강 희망/최종 수강 과목을 전처리를 위해 비음수 행렬분해를 도입했고, 학생들의 수강 희망 과목 벡터로부터 수강 추천 벡터를 생성하기 위해 조건부 적대적 생성 신경망(cGAN)을 구현했다. cGAN 모델은 다른 학생들의 수강 희망 과목을 조건부 데이터로, 최종 수강 신청 과목을 정답으로 사용하여 훈련되었으며, cGAN이 생성한 수업 벡터와 과목 특성 매트릭스를 곱하여 추천과목을 생성했다. 추천과목은 코사인 유사도 함수로 평가하여 학생이 실제 최종 수강했던 과목과 유사성을 판단했다. 그 결과, 본 논문에서 제시한 cGAN 기반 추천 모델이 학생들이 원하는 수업에 대한 요구 사항을 충족하는 추천과목을 생성할 수 있음을 보여주었다.

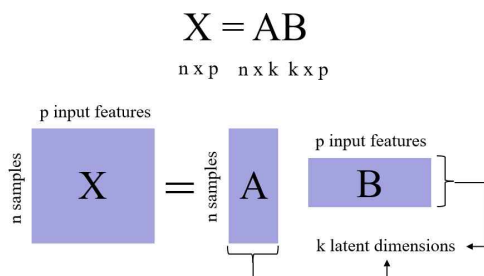
I. 서 론

현재 수강 신청 시스템은 학생이 수강할 수 있는 과목 수에 대한 제한과 학업 및 개인 목표를 충족하는 과정을 찾는 것이 어렵다는 문제로 인해 학생들의 커리큘럼 설계에 문제가 존재한다. 이러한 문제를 해결하기 위해 본 논문에서는 학생들이 원하는 수업과 등록할 수 있는 과목 사이의 격차를 극복하는 데 도움이 되는 인공지능 모델을 개발했다.

II. 비음수 행렬분해와 조건부 적대적 생성 신경망

학생이 원하는 과목과 수강 신청 시스템의 한계에 따라 학생에게 수강 과목을 추천하기 위해 수학적 기법인 비음수 행렬분해(Non-negative Matrix Decomposition)[1]와 수강 추천과목을 생성하기 위한 인공지능 모델 조건부 적대적 생성 신경망(Conditional Generative Adversarial Networks, cGAN)[2]을 조합하여 사용하였다.

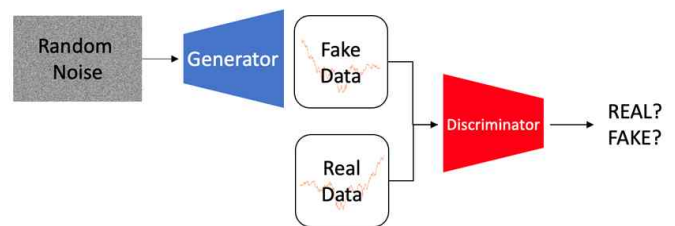
비음수 행렬분해는 학생 정보 행과 과목정보 열로 이루어진 수강행렬 X 를 학생 및 과목 특성 행렬로 분해하는 데 사용되었다. 학생 특성 행렬 A 는 학생의 특성을 나타내고, 과목 특성 행렬 B 는 교과목의 특성을 나타낸다. 본 연구에서는 잠재 차원인 k 를 17로 지정하였다. 이러한 매트릭스를 분석함으로써 학생들과 학생들이 수강 신청 사이의 상관관계를 확인할 수 있었다.



<그림 1> 비음수 행렬분해의 모식도

적대적 생성 신경망(GAN)[3] 알고리즘은 서로 경쟁하는 두 모델인 생성기(Generator)와 판별기(Discriminator)로 이루어진 인공지능 모델이다. 이러한 적대적 인공 생성 신경망 모델은 두 인공지능 모델의 적대적 학습을 통해 최적의 학습을 이루었을 때, 생성기를 이용해 가우시안 랜덤 노이즈로부터 진짜와 유사한 그럴듯한 가짜 샘플을 생성할 수 있게 된다.

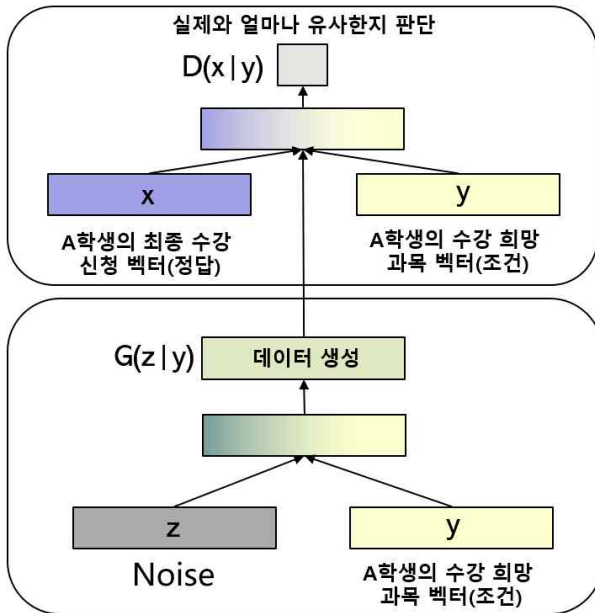
그러나 학생-희망 벡터로부터 학생-수강 벡터를 생성해내는 인공지능 모델은 단순 노이즈로부터 샘플을 생성하는 GAN 모델만으로는 구현해내기가 어렵다. 이러한 목표를 이루기 위해 일정한 조건을 충족하는 샘플을 생성해내는 GAN 모델인 cGAN을 본 연구에 도입하였다. cGAN 모델은 기본적인 GAN과 구조는 똑같으나, 생성기와 판별기에 조건을 입력해주어 특정 조건을 충족하는 샘플을 생성할 수 있다.



<그림 2> GAN 구조의 모식도

III. cGAN을 이용한 수강 과목 추천 모델

본 연구에서는 비음수 행렬분해를 통해 취득한 각 학생의 수강 희망 과목 벡터를 조건으로 다양한 학생의 수강 추천과목 벡터를 생성하기 위해 cGAN 모델을 사용하였다. 다른 학생들의 수강 희망 과목 벡터를 조건부 데이터로, 최종 수강 신청 과목 벡터를 정답으로 사용하여 cGAN 모델을 훈련 시켰고, cGAN이 생성한 수강 추천과목 벡터에 과목 특성 행렬을 곱하여 나온 결과에 행렬을 오름차순으로 정렬하여 추천과목을 생성했다. 이를 통해 학생이 수강할 가능성이 가장 높은 과목을 식별할 수 있었다.



<그림 3> 본 연구에서 사용된 cGAN 모델의 모식도

IV. 성능 평가

생성된 추천과목 벡터와 최종 수강 신청과목 벡터의 유사도를 평가하기 위해 코사인 유사도 함수를 사용하였다. 코사인 유사도 함수는 벡터 간 각도의 코사인을 계산하는 0이 아닌 두 벡터 간의 유사성 측정이다. 벡터 크기의 곱으로 나눈 벡터의 내적으로 정의된다. 두 벡터 A와 B가 주어지면 코사인 유사도는 다음과 같이 계산된다.

$$\text{코사인 유사도} = (A \cdot B) / (\|A\| \cdot \|B\|)$$

여기서 A와 B는 벡터이고 *는 내적이며 $\|A\|$ 그리고 $\|B\|$ 는 각각 벡터 A와 B의 크기이다. 코사인 유사도 함수는 -1에서 1까지의 값 범위를 가지며, 값이 1이면 벡터가 완벽하게 유사함을 나타내고, 값이 0이면 벡터가 직교(즉, 유사하지 않음)함을 나타내며, 값이 -1은 벡터가 완전히 다르다는 것을 나타낸다. 이 방법은 특정 사용자가 주어진 유사 사용자를 찾기 위해 협업 필터링과 같은 추천 알고리즘에서 일반적으로 사용된다.

V. 실험 결과 및 분석

본 연구에서는 각 학생에 대한 수강 추천과목 벡터와 실제 최종 수강 신청 벡터 간의 코사인 유사도를 계산했다. <그림 4>, <그림5>, <그림6>은 각각 학생의 수강 희망 과목 벡터, 실제 최종 수강 신청과목 벡터 그리고 학생의 수강 희망 과목 벡터를 cGAN의 조건으로 생성된 추천과목 벡터를 보여준다. A학생의 경우 실제 수강 희망했던 과목은 '응용 경제 수학 (영어)', '미시경제분석(영어)', '식품 마케팅', '동양 철학 개론'이다. 수강 희망 벡터는 파란색에서 노란색으로 갈수록 중요도가 높아지며 7번과 11번 특성의 중요도가 높은 것을 알 수 있다. A학생의 경우 실제 최종 수강 신청 벡터와 원하는 수강 신청 벡터가 다르며, 최종 수강 신청에서는 10번째 특성이 가장 중요도가 높았다.

학생이 수강 희망 과목 벡터를 통해 생성된 추천과목 벡터와 실제 최종 수강 신청 벡터의 코사인 유사도는 0.97이었다. 즉, cGAN모델이 최종 수강 신청 벡터와 유사하게 추천과목을 생성해낸 것을 보여준다. 산출된 추천과목 벡터에 과목 특성 행렬을 곱하여 원래의 학생-과목 행렬로 복원하였을 때 A학생에게 추천한 과목은 '심리학의 이해', '회계 중급(영어)',

수강희망 데이터



<그림 4> A학생의 수강 희망 과목 벡터

최종 수강신청 데이터



<그림 5> A학생의 최종 수강 신청 과목 벡터

생성된 추천 과목 데이터



<그림6> A학생의 수강희망 벡터를 조건으로 생성된 추천 과목 벡터

재무관리(영어)', '오퍼레이션 관리', '마케팅원리(영어)'이었다.

전반적으로 제시한 cGAN 모델은 다양한 학생들이 원하는 과정을 기반으로 수강 신청 벡터를 생성하고 원하는 수강 희망 과목과 유사한 과목들을 추천할 수 있었다. 코사인 유사도 함수를 통해 추천과목과 학생이 희망하는 과목의 유사도를 평가할 수 있었고, 수강 희망 과목의 요건을 충족하는 과목을 추천한다는 점에서 유망한 결과를 보였다.

VI. 결론

본 연구에서는 현재 수강신청 시스템의 한계 내에서 학생들이 커리큘럼을 설계할 수 있도록 도와주는 인공지능 모델을 개발하였다. 본 논문에서 제안된 cGAN 모델은 원하는 수강 희망과목에 대한 요구 사항을 충족하는 추천과목을 생성할 수 있었고, 코사인 유사도 함수를 사용하여 학생이 최종 수강 신청했던 과목과 생성된 과목이 유사한 것으로 평가되었다. 이 모델을 통해 학생들의 수강권을 확대하고 학업적, 개인적 목표에 맞는 수강권을 보장할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 교육부 재원으로 한국연구재단의 대학혁신지원사업 예산을 지원받아 수행된 연구입니다.

이 논문은 한국연구재단의 지원을 받아 수행된 연구입니다.

(NRF-2022R1A2C2004003).

참 고 문 헌

- [1] Lee, Daniel, and H. Sebastian Seung. "Algorithms for non-negative matrix factorization." *Advances in neural information processing systems* 13 (2000).
- [2] Mirza, Mehdi, and Simon Osindero. "Conditional generative adversarial nets." *arXiv preprint arXiv:1411.1784* (2014).
- [3] Goodfellow, Ian, et al. "Generative adversarial networks." *Communications of the ACM* 63.11 (2020): 139-144.