

# 사용자 특성 분류를 위한 Variational Auto-Encoder 기반 준지도학습 기법

정승원, 손민재, 황인준  
고려대학교 전기전자공학과  
e-mail : {jsw161, smj5668, ehwang04}@korea.ac.kr

## Semi-supervised Learning Scheme based on Variational Auto-Encoder for Classification of User Characteristics

Seungwon Jung, Minjae Son, Eenjun Hwang  
School of Electrical Engineering, Korea University

### 요 약

지능형 교육 시스템은 인공지능 기법을 적용해 사용자가 효율적으로 학습할 수 있도록 돕는 시스템으로, 수집된 데이터를 활용하여 사용자 개인에게 적합한 커리큘럼을 설계해주거나 학습 방법을 추천해준다. 이러한 서비스의 기반은 사용자의 특성을 정확히 분류하는 것으로, 이를 위해선 분류하고자 하는 특성을 아는 데이터가 필요하다. 그러나 데이터 수집 방법의 한계로, 충분한 데이터를 확보하는 데 많은 시간과 비용이 요구된다. 이 문제를 해결하기 위하여 본 논문에서는 Variational Auto-Encoder(VAE)를 언라벨 데이터로 학습시킨 후 VAE의 인코더 부분을 특징 추출 방법으로 사용하는 기존 기법에서, VAE 대신 Feedback Variational Auto-Encoder(FVAE)을 활용한 기법을 제안한다. FVAE는 VAE에서 다층 퍼셉트론이 더해진 구조를 가지고 있어, 라벨 데이터에서 주어진 특성 정보를 반영하여 특징을 추출할 수 있다. 실험 결과, 기존의 VAE 기반 준지도학습 기법보다 정확도, F1 점수 측면에서 뛰어남을 보였다.

### 1. 서론

현재 IT 기술의 발전으로 온라인 환경에서 다양한 학습 시스템이 운영되고 있다. 대표적인 예로 인공지능 기법을 적용해 사용자가 효율적으로 학습할 수 있도록 돕는 지능형 교육 시스템(Intelligent Tutoring Systems, ITS)을 들 수 있다. 이러한 시스템은 사용자가 시스템과 상호작용했을 때 정보를 기록한 로그 데이터를 지속적으로 수집하고 있다. 수집된 데이터는 사용자 개인에게 적합한 커리큘럼을 설계해주거나 사용자에게 맞춘 학습 방법을 추천해주거나, 또는 사용자의 행동을 예측하는 등 다양하게 활용되고 있다 [1, 2]. 그러나 이러한 활용에 있어, 데이터에서 사용자의 특성을 정확히 뽑아내지 못한다면 시스템은 사용자에게 적절한 대응을 할 수 없다. 따라서 주어진 데이터에서 원하는 정보를 정확히 분류해내는 것이 학습 시스템 연구에서의 핵심 과업이 되고 있다.

분류 모델의 성능을 높이는 데 가장 중요한 요소는 데이터의 양이다. 방대한 데이터를 가지고 모델을 학습시킬 수 있다면 모델은 모든 경우에 대응할 수 있게 되어 정확도 높은 결과를 낼 수 있다. 그러나 이러한 상황은 많지 않다. 예를 들어, ITS에서 확보한 데이터는 많지만, 개인

정보 보호 등의 이슈로 인해 정작 우리가 판별하고자 하는 특성을 아는 데이터를 수집하기 힘들다. 이를 해결하기 위해선 직접 사용자에게 설문을 요청하거나 테스트를 진행해야 한다. 특히, 사용자에게 학습 장애가 있는지, 어떤 성격적 특징이 있는지 등과 같이 전문지식을 요구하는 특성일 경우 전문가의 도움이 필수적이다. 이 방법은 많은 시간과 비용을 요구하며, 이를 감수하더라도 분류 모델을 충분한 양의 데이터를 확보하지 못할 수 있다.

이 문제를 해결하기 위해 사용되는 방법이 준지도학습(Semi-supervised Learning)이다. 준지도학습은 우리가 분류하고자 하는 특성, 즉 라벨을 아는 데이터(라벨 데이터)와 라벨을 모르는 데이터(언라벨 데이터)를 혼합하여 분류 모델을 학습시키는 방법이다. 준지도학습은 적은 라벨 데이터 양을 많은 양의 언라벨 데이터로 보완하여 모델의 정확도를 높일 수 있기 때문에 의학 분야 [3], 자연어 이해 [4] 등 다양한 분야에 활용되고 있다. 학습 분야에서도 준지도학습을 이용한 연구가 진행되었다 [5, 6].

Klingler et al. [6]은 학생들이 계산 장애를 가졌는지 분류하기 위하여 심층학습(Deep Learning) 모델 중 하나인 Variational Auto-Encoder(VAE) [7]를 언라벨 데이터로

학습시킨 후 VAE의 인코더(Encoder) 부분만 떼어내어 특징 추출 부분으로 사용하는 방법을 제시하였다. 이 방법은 기존의 준지도학습 방법보다 뛰어난 정확성을 보였으며, 특징 선택(Feature Selection) 과정을 중간에 추가하여 균형이 맞지 않는 데이터로 학습하더라도 모델이 특성을 잘 분류해냈음을 증명하였다.

본 논문에서는 위 연구보다 분류 정확도를 향상시키기 위해 VAE에 라벨 데이터를 사용하여 피드백(Feedback)을 줄 수 있는 구조가 추가된 VAE, Feedback Variational Auto-Encoder(FVAE)를 제안한다. FVAE의 구성은 VAE와 유사하지만 VAE의 인코더 부분에 피드백을 주기 위한 다층 퍼셉트론(Multi-Layer Perceptron, MLP)이 있다는 차이가 존재한다. 구체적으로 말하면, 학습 과정에서 FVAE는 언라벨 데이터, 라벨 데이터 모두 사용하여 특징(Feature)을 추출하기 위한 학습을 수행하는 한편, 라벨 데이터와 FVAE 내부의 다층 퍼셉트론을 통해 분류를 수행한다. 이 분류 결과가 VAE에 영향을 주어, VAE가 좀 더 분류가 잘 될 수 있는 특징을 추출할 수 있도록 만든다. Kingmar et al. [8]의 연구에서 유사한 방법을 제안했으나 본 논문에서는 해당 연구와 달리 특징 추출 기법으로써 여러 분류기와의 결합에 초점을 맞춘다.

논문의 구성은 다음과 같다. 2장에서는 VAE와 VAE를 이용한 준지도학습 방법에 대해, 3장에서 FVAE에 대해 설명한다. 4장에서 FVAE의 성능 실험 결과를 보인 후, 마지막으로 5장에서 결론을 기술한다.

## 2. Variational Auto-Encoder 기반 준지도학습

이 장에서는 제안한 모델 설명에 앞서 기존의 VAE를 이용하여 수행한 준지도학습 방법에 관해 서술한다. 먼저 VAE의 구조를 설명한 후 VAE를 이용하여 어떻게 준지도학습을 수행하는지 그 방법에 대해 기술한다.

VAE는 입력값을 넣어 차원을 축소하고 다시 복원하는 과정을 반복 수행하여 입력값과 출력값이 최대한 같아지도록 학습하는 오토인코더(Auto-Encoder, AE)를 확장한 모델이다. 오토인코더와의 차이는 확률 분포에 기반하고 있다는 것이다.

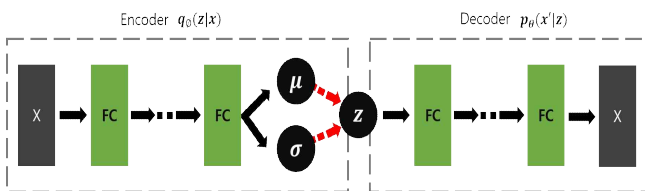


그림 1. Variational Auto-Encoder의 구조도

그림 1은 VAE의 구조를 그린 그림이다. VAE는 입력변수  $x$ 를 축소하였다가 복원한 출력변수  $x'$ 를  $x$ 와 최대한 유사하도록 학습한다. 이 학습의 핵심은 축소하는 부분의 은닉층(Hidden Layer)으로,  $x$ 의 차원을 축소하여 잠재변수  $z$ 를 생성하는 것이다.  $z$ 는  $x$ 보다 축소된 차원을 갖지만  $x'$ 로 복원

되었을 때  $x$ 와 유사해지므로 입력 변수인  $x$ 의 특징들을 잘 표현한다고 할 수 있다.  $x$ 를  $z$ 로 변환하는 부분을 인코더라고 하며,  $x$ 가 주어졌을 때  $z$ 를 생성하므로 확률 분포  $p(z|x)$ 로 나타낼 수 있다. 하지만 차원이 다른 값들을 직접 계산하기 어려워, Variational Bayesian Methods를 사용한다.  $p(z|x)$ 를 직접 구하는 대신, 확률 분포  $q_\phi(z|x)$ 라는 변수를 선언하고 이를  $p(z|x)$ 에 근사하도록 만든다. 따라서  $x$ 가 주어졌을 때  $z$ 의 확률 분포는  $p(z|x)$  대신  $q_\phi(z|x)$ 로 표현한다. 이때  $\phi$ 는  $q_\phi(z|x)$ 의 매개변수(Parameter)를 나타낸다. 디코더(Decoder)는  $z$ 에서  $x'$ 로 복원하는 부분으로, 인코더와 마찬가지로 확률 분포  $p_\theta(x'|z)$ 로 표현한다.  $\theta$ 는  $\phi$ 와 마찬가지로 확률 분포의 매개변수를 의미한다.  $z$ 는 식 (1)에 의해 결정된다. 이때  $\epsilon$ 는 확률 분포  $N(0,1)$ 를 따르는 상수로, 매 학습 마다 임의로 선택된다.

$$z = \mu + \sigma \cdot \epsilon \quad (1)$$

$$l(\theta, \phi) = -E_{z \sim q_\phi(z|x)} [\log p_\theta(x'|z)] + KL(q_\phi(z|x) \| p(z)) \quad (2)$$

식 (2)는 VAE를 학습시킬 때 사용하는 손실함수(Loss Function)이다. 첫 번째 항은 디코더에서  $z$ 를  $x'$ 로 복원하였을 때  $x$ 와 얼마나 비슷한지를 나타내는 복원오차 식으로, 유사하지 않을 경우 증가하게 된다. 두 번째 항은 쿨백-라이블러 발산(Kullback-Leibler Divergence)를 사용해 입력 변수의 분포와 잠재 변수의 분포 사이의 유사도를 계산하는 정규화 항으로, 분포 간의 차이가 크게 나지 않도록 방지하는 역할을 한다. 따라서 이 손실함수로 학습시키면 VAE가 복원 가능한 특징들을 추출하면서 비슷한 성질을 가진 입력을 잠재 변수 분포상에서 서로 가까운 위치에 놓이도록 만들 수 있다.

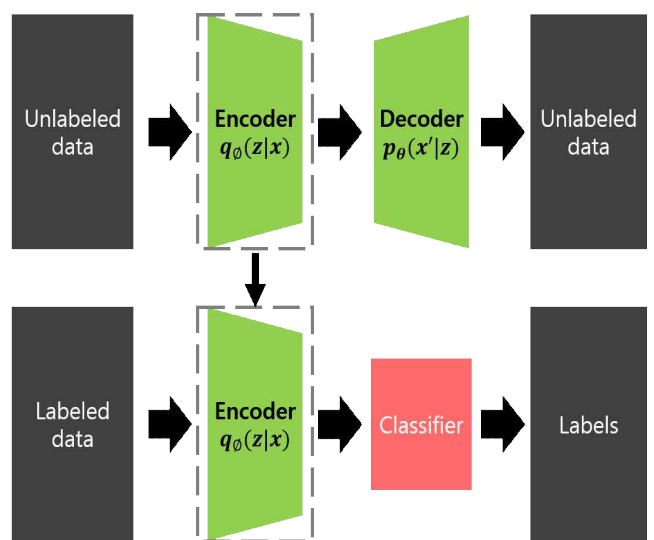


그림 2. 준지도학습 방법 구성도

그림 2는 VAE를 이용한 준지도학습 방법을 그린 그림이다.

VAE를 이용한 준지도학습은 인코더를 학습시키는 것부터 시작한다. 언라벨 데이터를 이용하여 VAE를 학습시키면 VAE의 인코더와 디코더가 동시에 학습된다. 학습된 인코더는 언라벨 데이터에 나타나는 일반적인 특징을 표현할 수 있게 되며, 디코더는 이 변수에서 원래의 데이터로 복원할 수 있는 능력을 갖추게 된다. VAE의 학습이 끝나면, 인코더만을 가져와 라벨 데이터에서 특징을 추출한다. 다시 말해, 인코더에서 추출된 특징이 분류기의 입력 변수가 된다. 그 뒤 과정은 일반적인 지도학습과 동일하게 진행된다.

### 3. Feedback Variational Auto-Encoder(FVAE)

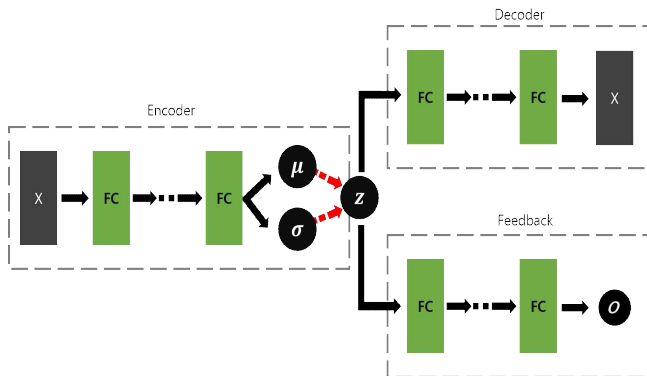


그림 3. Feedback Variational Auto-Encoder 구성도

FVAE는 본 논문에 제안하는 모델로, 더욱 분류를 잘하도록 만드는 특징을 뽑아낼 수 있도록 VAE에 피드백 기능을 추가한 모델이다. FVAE는 기존의 VAE의 구조와 유사하나, 다층 퍼셉트론을 적용하였다. 구성도는 그림 3과 같다. 피드백 부분의 다층 퍼셉트론은 완전연결 계층(Fully-Connected layer, FC)로 이루어져 있으며, 인코더의 결과물인 잠재변수  $z$ 를 입력으로 넣어 학습시킨다. 피드백 기능을 추가하면서 변경된 손실함수는 식 (3)과 같다.

$$l_{FVAE}(\theta, \Phi) = l(\theta, \Phi) + \alpha \cdot loss_{cross-entropy} \quad (3)$$

$$loss_{cross-entropy} = -\sum_i y_i \log(y'_i) \quad (4)$$

FVAE의 손실함수는 VAE의 손실함수에서 식 (4)의 교차 엔트로피(Cross Entropy) 손실에 계수  $\alpha$ 를 곱해 더한 값이다. 여기서  $y_i$ 는 라벨 값을 의미하고,  $y'_i$ 는 다층 퍼셉트론을 통해 나온 결과다.  $\alpha$ 는 VAE의 손실이 교차 엔트로피 손실 사이의 크기를 조정해주기 위한 계수다. 추가된 교차 엔트로피 손실은 다층 퍼셉트론을 이용하여 지금까지 VAE로 학습된 특징들이 과연 잘 분류가 잘 되는 특징인지 VAE에 확인시켜준다. 이는 VAE에, 구체적으로 인코더에 영향을 준다.

FVAE의 학습 과정에 대해 설명하면, 미리 정해진 배치(Batch) 크기에 따라 언라벨 데이터와 라벨 데이터가 같이 배치에 들어가게 된다. 언라벨 데이터와 라벨 데이터 모두 사용하여 VAE의 손실을 계산하는 동시에 라벨 데이터를 통해 교

차 엔트로피 손실을 계산한다. 이 두 손실의 합이 최종적으로 FVAE의 손실이 된다.

FVAE는 VAE보다 분류에 적합한 특징을 뽑아낼 수 있다는 장점 말고도 다른 장점을 하나 더 갖고 있다. 바로 특징 추출 학습과 분류기 학습이 동시에 일어난다는 것이다. 기존의 VAE를 사용한 준지도학습은 언라벨 데이터로 인코더를 학습시키고 라벨 데이터에서 인코더로 뽑아낸 특징들로 분류기를 학습시키는 두 단계의 학습 과정이 필요하다. 반면, FVAE는 학습이 끝났을 때 인코더 부분만 떼어내어 기존의 준지도학습처럼 사용할 수도 있고, 학습된 다층 퍼셉트론을 분류기로 사용할 수도 있다.

### 4. 실험 및 결과

본 논문에서는 실험을 위해 2015년 Knowledge Discovery in Databasescup (KDDcup)라는 데이터 마이닝 경진대회 [9]에서 사용한 데이터를 사용하였다. 이 데이터는 MOOC에서의 중퇴자 예측을 위해 준비된 데이터로, 90개 특징을 가지고 있으며, 총 96409개 데이터가 존재한다. 이 중 언라벨 데이터로 96209개, 라벨 데이터로 라벨 간 비율을 맞춘 데이터 200개로 나누어 사용하였다.

비교 실험에 사용한 특징 추출 모델은 VAE, VAE에 특징 선택 과정을 추가한 모델(VAE-FS) [6] 그리고 본 논문에서 제시하는 FVAE이다. 세 모델 모두 과적합(Overfitting)을 피하기 위해 Epoch을 50만쯤 학습했으며, FVAE에서  $\alpha$ 는 1로 설정하였다, 정확한 비교를 위해 추출한 특징은 8차원으로 모두 같게 설정했다. 성능을 평가하기 위해 사용한 분류기는 로지스틱 회귀(Logistic Regression, LR), 나이브 베이즈(Naive Bayes, NB), 지지 벡터 머신(Support Vector Machine, SVM), 결정 나무(Decision Tree, DT), 랜덤 포레스트(Random Forest, RF), 에이다부스트(Adaboost, AD), 다층 퍼셉트론(MLP), 총 7가지를 사용하였다. 단, FVAE 실험에 사용된 다층 퍼셉트론은 FVAE 내부의 다층 퍼셉트론이다. 모든 분류기는 5겹 교차검증(5-fold Cross Validation)으로 실험했으며, 평가하기 위한 수치로는 정확도(Accuracy)와 F1 점수(F1 Score)를 이용하였다.

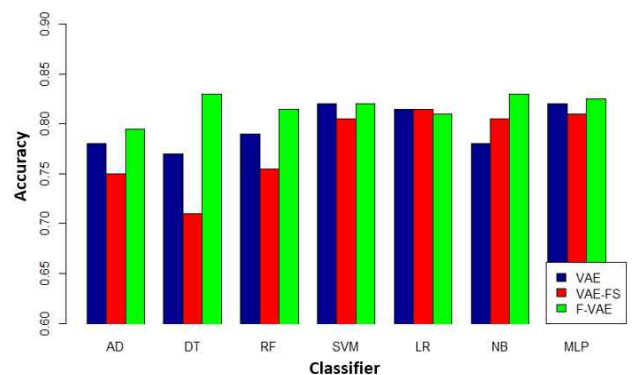


그림 4. 정확도 비교

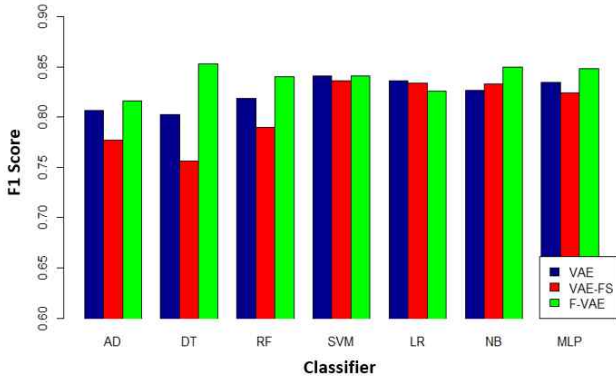


그림 5. F1 점수 비교

분류기별 결과는 다음 그림 4, 5와 같다. 그림 4를 보면, 로지스틱 회귀 결과를 제외한 나머지 모든 분류기보다 FVAE의 정확도가 높은 것을 확인할 수 있다. 특히 결정 나무에서 VAE보다 6% 포인트만큼 정확해진 것을 확인할 수 있다. 로지스틱 회귀에서 정확도가 낮아지긴 했으나 차이는 0.5% 포인트로, 다른 분류기에서 향상된 정확도에 비해 크기가 작았다. 다층 퍼셉트론에서 큰 향상을 보이진 못했지만 기존 기법이 특징 추출 단계, 분류기 학습 단계의 두 단계를 거친 반면, 제안된 기법이 한 단계만 거치면 된다는 점을 고려한다면 제안된 기법이 우수함을 확인할 수 있다. VAE-FS는 해당 논문에서 제시한 것만큼 좋은 성능을 보이지 못했다. 그림 5의 F1 점수 역시 분류 정확도와 마찬가지로 로지스틱 회귀에서만 FVAE가 낮을 뿐, 모든 분류기에서 우수한 성능을 보이고 있다.

## 5. 결론

본 논문에서는 준지도학습에서 VAE가 잘 분류하는 특징을 추출할 수 있도록 만들기 위해 VAE의 구조를 변화시키는 방법에 대해 연구를 진행하여, 그 결과로 FVAE를 제안하였다. 실험을 통해 기존에 있는 VAE를 활용한 준지도학습보다 정확도와 F1 점수가 향상됨을 보였고 특히 다층 퍼셉트론에서 인코더와 분류기를 동시에 학습시켰음에도 뛰어난 성과를 보였다. 향후 연구로 FVAE 손실 함수의  $\alpha$  최적화에 대한 연구를 진행할 것이며, 좀 더 다양한 상황에서 FVAE 성능을 테스트할 예정이다.

## 6. Acknowledgement

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원(No.R0190-16-2012, 빅데이터 처리 고도화 핵심기술개발 사업 총괄 및 고성능컴퓨팅 기술을 활용한 성능 가속화 기술 개발)과 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2016R1D1A1A09919590)

## 참고문헌

[1] S. Adjei, K. Ostrow, E. Erickson and N.T.

Heffernan, "Clustering Students in ASSISTments: Exploring System- and School-Level Traits to Advance Personalization," in Proceedings of the 10th International Conference on Educational Data Mining(EDM 2017), Wuhan, Hubei, China, 25-28 June, 2017, pp. 340-341.

[2] R.S. Baker, A.T. Corbett and K.R. Koedinger, "Detecting student misuse of intelligent tutoring systems," *International conference on intelligent tutoring systems*, pp. 531-540, Springer, Berlin, Heidelberg, 2004.

[3] B.K. Beaulieu-Jones and C.S. Greene, "Semi-supervised learning of the electronic health record for phenotype stratification," *J. Biomed. Inform.*, Vol. 64, pp. 168-178, 2016.

[4] Y. Cheng, W. Xu, Z. He, W. He, H. Wu, M. Sun, Y. Liu, Semi-Supervised Learning for Neural Machine Translation." In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7-12 August, 2016, pp. 1965-1974.

[5] V. Tam, E.Y. Lam and S.T. Fung, "Enhancing educational data mining techniques on online educational resources with a semi-supervised learning approach." in 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE). IEEE, pp. 203-206, 2015.

[6] S. Klingler, R. Wampfler, T. Käser and B. Solenthaler, "Efficient Feature Embeddings for Student Classification with Variational Auto-encoders," in Proceedings of the 10th International Conference on Educational Data Mining(EDM 2017), Wuhan, Hubei, China, 25-28 June, 2017, pp. 72-79.

[7] D.P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[8] D.P. Kingma, S. Mohamed and D.J. Rezende, "Semi-supervised learning with deep generative models." in Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 8-13 December, 2014, pp. 3581-3589.

[9] KDD Cup 2015, <http://kddcup2015.com/>