

지능형 교육 시스템의 학습자 분류를 위한 Variational Auto-Encoder 기반 준지도학습 기법

정승원[†], 손민재^{**}, 황인준^{***}

Variational Auto-Encoder Based Semi-supervised Learning Scheme for Learner Classification in Intelligent Tutoring System

Seungwon Jung[†], Minjae Son^{**}, Eenjun Hwang^{***}

ABSTRACT

Intelligent tutoring system enables users to effectively learn by utilizing various artificial intelligence techniques. For instance, it can recommend a proper curriculum or learning method to individual users based on their learning history. To do this effectively, user's characteristics need to be analyzed and classified based on various aspects such as interest, learning ability, and personality. Even though data labeled by the characteristics are required for more accurate classification, it is not easy to acquire enough amount of labeled data due to the labeling cost. On the other hand, unlabeled data should not need labeling process to make a large number of unlabeled data be collected and utilized. In this paper, we propose a semi-supervised learning method based on feedback variational auto-encoder(FVAE), which uses both labeled data and unlabeled data. FVAE is a variation of variational auto-encoder(VAE), where a multi-layer perceptron is added for giving feedback. Using unlabeled data, we train FVAE and fetch the encoder of FVAE. And then, we extract features from labeled data by using the encoder and train classifiers with the extracted features. In the experiments, we proved that FVAE-based semi-supervised learning was superior to VAE-based method in terms with accuracy and F1 score.

Key words: Intelligent Tutoring System, Semi-supervised Learning, Variational Auto-encoder

1. 서 론

최근 ICT(Information and Communication Technology) 기술의 발전으로 온라인에서 다양한 학습 시스템이 등장하고 있다. 지능형 교육 시스템(Intelligent Tutoring System)[1-3]은 이러한 학습 시

템 중 하나로, 인공지능 기법을 적용하여 사용자가 효율적으로 학습할 수 있도록 돕는 시스템이다. 지능형 교육 시스템은 사용자와의 상호작용이 발생했을 때 이를 기록한 로그 데이터를 지속적으로 수집하며, 수집된 데이터를 활용하여 사용자 개인에게 적합한 커리큘럼을 설계해주거나 사용자에게 알맞은 학습

※ Corresponding Author: Eenjun Hwang, Address: (136-713) Anam-ro 145, Seongbuk-gu, Seoul, Korea, TEL: +82-2-3290-3256, FAX: +82-, E-mail: ehwang04@korea.ac.kr

Receipt date: Sep. 10, 2019, Revision date: Oct. 28, 2019
Approval date: Nov. 11, 2019

[†] School of Electrical Engineering, Korea University
(E-mail: jsw161@korea.ac.kr)

^{**} School of Electrical Engineering, Korea University
(E-mail: smj5668@korea.ac.kr)

^{***} School of Electrical Engineering, Korea University

※ This study was partly supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1A09919590) and by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HG19C0682)

방법을 추천해주거나, 또는 사용자의 관심사에 맞는 콘텐츠를 제시하는 등 다양한 사용자 맞춤형 서비스를 제공한다[4, 5]. 그러나 이러한 활용에 있어 가장 중요한 것은 데이터에서 사용자의 특성을 정확하게 구분해내야 한다는 것이다. 구분된 사용자의 특성을 기반으로 사용자 맞춤형 서비스가 제공되므로, 만약 잘못된 특성을 이용한다면, 적절하지 못한 서비스가 사용자에게 제공되며, 사용자는 이에 부정적인 반응을 보이게 된다. 따라서 주어진 데이터에서 사용자 특성을 정확하게 구분하는 것이 지능형 교육 시스템의 핵심 연구 주제이다.

사용자 특성을 구분하기 위해 일반적으로 사용되는 방법은 수집된 데이터로 분류 모델을 학습시키는 것이다. 이 분류 모델의 정확도를 높이는 데 있어 가장 중요한 요소는 확보한 데이터의 양으로, 만약 방대한 데이터로 모델을 학습시킬 수 있다면 분류 모델은 모든 경우에 대응할 수 있게 되어 정확도 높은 결과를 낼 수 있다. 그러나 이러한 경우는 많지 않다[6]. 지능형 교육 시스템에서 확보할 수 있는 데이터 종류는 많지만, 개인정보 보호 등의 이슈로 인해 정작 우리가 판별하고자 하는 특성이 밝혀진 데이터를 수집하기 어렵다. 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. 이러한 방법은 많은 시간과 비용을 요구하며, 이를 감수하더라도 분류 모델을 학습시키기 위한 양의 데이터를 확보하지 못할 수 있다.

이러한 문제를 해결하기 위해 사용되는 방법이 준지도학습(Semi-supervised Learning)이다. 우리가 분류하고자 하는 특성, 즉 클래스를 아는 데이터를 라벨 데이터(Labeled Data), 모르는 데이터를 언라벨 데이터(Unlabeled Data)라 부르는데, 준지도학습은 이 두 데이터를 혼합하여 활용하는 방법이다. 라벨 데이터의 적은 양을 상대적으로 구하기 쉬운 언라벨 데이터의 많은 양으로 보완함으로써 모델의 정확도를 높일 수 있기 때문에 감염병 발생 예측, 질병 진단과 같은 의료 분야[7, 8], 감정 인지, 기계 번역이 대표적인 자연어 분야[9, 10] 등 여러 분야에 활용되고 있다. 학습 분야에서도 준지도학습을 이용한 연구가 진행되었다[11-14].

Tam et al.[11]는 Bag-of-Word 기반 분류기와 명시 의미 분석(Explicit Semantic Analysis) 기반 분류기를 활용하는 준지도학습 방법을 제안했다. 두 분류기를 라벨 데이터로 학습시키고 언라벨 데이터를 분류한 후, 높은 신뢰도로 분류된 언라벨 데이터를 새로이 라벨 데이터로 활용하였다. Wang et al.[12]는 대규모 공개 온라인 과정(Massive Open Online Course, MOOC)에서 동료 평가를 수행하는데 있어 학생들 사이의 채점 편차를 줄이기 위해 소수의 교사 채점 예시를 활용하는 준지도학습 방법을 제안하였다. 각 학생의 채점안과 교사가 제공한 채점안 사이의 유사도를 구해 이를 학생의 채점 능력으로 간주하고, 각 학년의 가중 집계를 활용하여 최종 점수를 결정함으로써 편차를 교정하였다. Livieris et al.[13]는 Yet Another Two-Stage Idea(YATSI)을 기반으로 발표 점수, 과제 점수 등이 주어졌을 때 학생들의 최종 성적을 예측하는 준지도학습 방법을 제안하였다. 확보된 라벨 데이터로 분류기를 학습시키고, 분류기로 언라벨 데이터의 클래스를 예측한 후, 원래 라벨 데이터와 클래스가 예측된 언라벨 데이터 모두 사용한 K최근접 이웃(K-Nearest Neighbor) 알고리즘으로 최종 예측 결과를 도출하였다. Klingler et al.[14]은 학생이 계산 장애를 가졌는지 분류하기 위해 Variational Auto-encoder(VAE)[15]를 활용한 방법을 제안하였다. 언라벨 데이터로 학습된 VAE의 인코더(Encoder) 부분을 통해 라벨 데이터에서 특징을 추출하고, 추출된 특징으로 분류 모델을 학습시킴으로써 기존의 준지도학습 방법보다 뛰어난 정확도를 보였다. 더불어, 특징 선택(Feature Selection) 과정을 중간에 추가하여 라벨 데이터의 클래스 비율이 불균형한 경우에도 정확한 분류가 가능하도록 하였다.

본 논문에서는 Klingler et al.가 제안한 준지도학습 방법보다 분류 정확도를 향상시키기 위해 라벨 데이터를 사용하여 피드백(Feedback)을 줄 수 있는 구조가 추가된 VAE, Feedback Variational Auto-Encoder(FVAE)를 제안한다. FVAE에는 피드백을 주기 위한 다층 퍼셉트론(Multi-Layer Perceptron, MLP)이 존재하며, 인코더 부분에 다층 퍼셉트론이 연결되어 있다. 학습 과정에서 FVAE는 언라벨 데이터, 라벨 데이터 모두 사용하여 특징(Feature)을 추출하기 위한 학습을 수행하는 동시에, 라벨 데이터와 FVAE 내부의 다층 퍼셉트론을 통해 분류를 수행한

다. 이 분류 결과가 인코더 부분에 영향을 주어, 기존 모델보다 분류 모델의 성능을 향상시키는 특징을 추출하고 준지도학습의 분류 정확도를 높일 수 있다. 이를 입증하기 위해 MOOC에서의 중퇴자 예측 데이터를 활용하여 기존 VAE 기반 준지도학습과 비교한다.

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.

2.1 Variational Auto-encoder

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.

Fig. 1은 VAE의 구조를 그린 그림이다. VAE는 입력변수 x 가 주어졌을 때 출력변수인 x' 이 x 와 최대한 유사하도록 학습한다. 이 학습의 핵심은 내부의 은닉층(Hidden Layer)으로, x 의 차원을 축소하여 잠재변수 z 를 생성하는 것이다. z 는 x 보다 축소된 차원을 갖지만 x' 로 복원되었을 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 $p(z|x)$ 를 직접 계산하기 어렵기 때문에 변분 베이저안 방법(Variational Bayesian Methods)을 이용한다. $p(z|x)$ 를 구하는 대신, 확률 분포

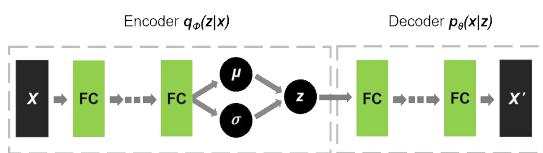


Fig. 1. The structure of VAE.

포 $q(z|x)$ 를 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 ϕ 는 $q_\phi(z|x)$ 의 매개변수(Parameter)를 나타낸다. 디코더(Decoder)는 z 에서 x' 로 복원하는 부분으로, 확률 분포 $p_\theta(x'|z)$ 로 표현한다. $p_\theta(x'|z)$ 에서 θ 는 ϕ 와 마찬가지로 확률 분포의 매개변수를 의미한다. z 는 식 (1)에 의해 결정된다. 이때 μ, σ 는 각각 x 가 주어졌을 때 z 의 평균과 표준편차를 의미하며, ϵ 는 확률 분포 $\mathcal{N}(0,1)$ 를 따르는 상수로, 매 학습 마다 임의로 선택된다.

$$z = \mu + \sigma \cdot \epsilon \tag{1}$$

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

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

2.2 VAE 기반 준지도학습

Fig. 2는 VAE를 이용한 준지도학습 방법을 그린 그림이다. VAE를 이용한 준지도학습은 인코더를 학습시키는 것부터 시작한다. 언라벨 데이터를 이용하

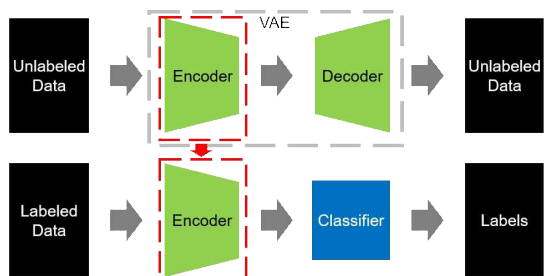


Fig. 2. The process of VAE-based semi-supervised learning.

여 VAE를 학습시키면 VAE의 인코더와 디코더가 동시에 학습된다. 학습된 인코더는 언라벨 데이터에 나타나는 일반적인 특징을 z 를 통해 표현할 수 있게 되며, 디코더는 z 에서 원래의 데이터로 복원할 수 있는 능력을 갖추게 된다. VAE의 학습이 끝나면, 인코더만을 가져와 라벨 데이터에서 특징을 추출하고, 이 특징은 원래의 입력변수 x 대신 분류기의 입력 변수가 된다. 그 뒤 과정은 일반적인 지도학습(Supervised Learning)과 동일하게 진행된다.

3. 제안한 방법

FVAE는 본 논문에 제안하는 모델로, 기존 VAE로 추출한 특징보다 분류 모델의 정확도를 높일 수 있는 특징을 뽑아낼 수 있도록 VAE에 피드백 구조가 추가된 모델이다. 이 피드백 구조는 다층 퍼셉트론으로 구성되어 있으며, 구조는 Fig. 3에서 확인할 수 있다. 피드백 부분의 다층 퍼셉트론은 완전연결계층(Fully - Connected layer, FC)로 이루어져 있으며, 입력변수는 인코더의 결과물인 잠재변수 z 가, 출력변수는 라벨 데이터의 클래스가 된다. 이러한 구조 변화로 인해 FVAE의 손실함수, $l_{FVAE}(\theta, \Phi)$ 는 식 (3)과 같다.

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

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

간단히 말해, FVAE의 손실함수는 VAE의 손실함수에 식 (4)의 교차 엔트로피(Cross Entropy) 손실에 계수 α 을 곱한 값을 더한 결과다. 여기서 y_i 는 i 번째 데이터의 라벨 값을 의미하고, y'_i 는 i 번째 데이터 입력 시에 다층 퍼셉트론이 분류한 결과다. 그리고 α 는 VAE의 손실이 교차 엔트로피 손실 사이의 크기를

조정해주기 위한 계수다. 추가된 교차 엔트로피 손실은 다층 퍼셉트론을 이용하여 지금까지 VAE 부분인 인코더와 디코더를 통해 학습된 특징 z 가 과연 분류 학습이 적합한지 확인할 수 있는 지표가 된다. 교차 엔트로피 손실은 다층 퍼셉트론이 정확하게 분류할 수록 작아지기 때문에, VAE 부분, 특히 인코더에 영향을 주어 분류 모델의 정확도를 향상시킬 수 있는 특징을 추출하도록 유도한다. 다시 말해, FVAE의 인코더를 통해 생성된 z 가 기존 VAE의 인코더로 추출한 z 보다 분류 모델의 학습에 좀 더 적합한 형태를 가지게 된다는 것이다.

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 $l(\theta, \Phi)$ 항을 계산하는 동시에 라벨 데이터를 통해 $loss_{cross-entropy}$ 항을 계산한다. 이 두 손실의 합이 최종적으로 $l_{FVAE}(\theta, \Phi)$ 값이 되며, FVAE는 이를 최소화하도록 학습한다.

FVAE은 VAE보다 분류 모델에 적합한 특징을 뽑아낼 수 있다는 장점 말고도 다른 장점을 하나 더 갖고 있다. 바로 특징 추출 학습과 분류기의 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 결과 및 고찰

본 논문에서는 실험을 위해 2015년 Knowledge Discovery in Databases cup(KDDcup)라는 데이터 마이닝 경진대회[16]에서 사용한 데이터를 사용하였다. 이 데이터는 MOOC에서의 중퇴자 예측을 위해 준비된 데이터로, 총 96409개 데이터 샘플이 존재한다. 강의 비디오나 과제, 토론 페이지에 접근한 횟수, 일일 MOOC 활동 횟수, MOOC에서 마우스, 키보드를 통한 상호작용 횟수 등으로 구성된 90개 특징을 가지고 있으며, 이러한 특징을 통해 해당 사용자가 학기 동안 강의를 끝까지 청강하지 않고 도중에 강의를 포기할 것인지 그 여부를 분류해낸다. 제안한 모

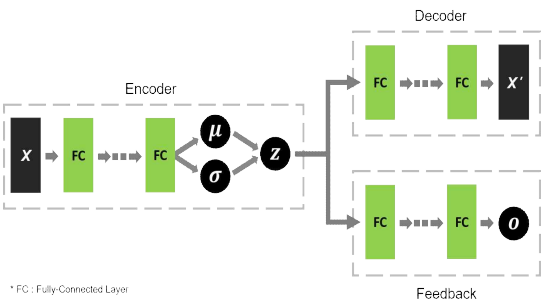


Fig. 3. The structure of FVAE.

델이 준지도학습 모델이므로 전체 데이터 중 언라벨 데이터로 96209개, 라벨 데이터로 라벨 간 비율을 맞춘 데이터 200개로 나누어 실험에 사용하였다. 모든 실험은 파이썬(Python) 환경에서 수행되었으며, Scikit-learn[17]과 Tensorflow[18] 라이브러리가 활용되었다.

4.1 분류 정확도 비교 실험

비교 실험에 사용한 특징 추출 모델은 VAE, VAE에 특징 선택 과정을 추가한 모델(VAE-FS)[14] 그리고 본 논문에서 제시하는 FVAE이다. 모든 모델의 인코더와 디코더는 각각 5개 층으로 이루어져 있으며, 인코더 각 층의 노드 수는 90개, 60개, 40개, 30개, 8개로 설정하였다. 디코더의 각 층 노드는 인코더와 대칭이 되게끔 설정하였다. 다만, VAE-FS 모델은 중간에 특징 선택이 존재하여 특징의 수를 절반 줄이기 때문에, 세 모델에서 최종적으로 추출된 특징의 차원을 모두 8차원으로 통일하기 위해 인코더의 마지막 층과 디코더의 첫 번째 층은 16개 노드로 이루어져 있다. FVAE 내부의 다층 퍼셉트론은 8개 노드로 이루어진 은닉층 3개로 구성되었다. 과적합(Overfitting)을 피하기 위해 모든 모델은 학습률이 0.001로 고정된 상태에서 50 Epoch만큼 Adam[19] 방법으로 학습하였으며, 배치 크기는 200으로 설정하였다. FVAE 학습 시에 α 는 1로 고정하였으며, 배치 크기 200에서 40을 라벨 데이터가, 나머지 160을 라벨 데이터가 차지하였다.

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(Ada-boost, AD), 다층 퍼셉트론(MLP), 총 7가지를 사용하였다. 단, FVAE 실험에 사용된 다층 퍼셉트론은 FVAE 내부의 다층 퍼셉트론이다. 대부분의 분류기 설정은 Scikit-learn의 기본 설정으로 통일하였으나, 다층 퍼셉트론의 경우 FVAE에 내장된 다층 퍼셉트론과 모두 같은 구조를 갖도록 설정하였다. 모든 분류기는 5겹 교차검증(5-fold Cross Validation)으로 실험했으며, 평가하기 위한 수치로는 정확도(Accuracy)와 F1 점수(F1 Score)를 이용하였다.

분류기별 결과는 다음 Fig. 4, 5와 같다. Fig. 4를 보면, 로지스틱 회귀 결과를 제외한 나머지 모든 분류기보다 FVAE의 정확도가 높은 것을 확인할 수 있다. 특히 결정 나무에서 VAE보다 약 0.06만큼 향상된 것을 확인할 수 있다. 로지스틱 회귀에서 정확도가 낮아지긴 했으나 차이는 0.005 정도로, 다른 분류기에서 향상된 정도에 비해 작았다. 다층 퍼셉트론에서 큰 향상을 보이진 못했지만 기존 기법은 특징 추출 단계, 분류기 학습 단계의 두 단계를 요구하는 반면, 제안된 기법은 한 단계만 거친다는 점을 고려한다면 제안된 기법이 우수함을 확인할 수 있다. VAE-FS는 해당 논문에서 제시한 것만큼 좋은 성능을 보이지 못했다. Fig. 5의 F1 점수 역시 분류 정확도와 마찬가지로 로지스틱 회귀에서만 FVAE가 낮은 뿐, 모든 분류기에서 우수한 성능을 보이고 있다.

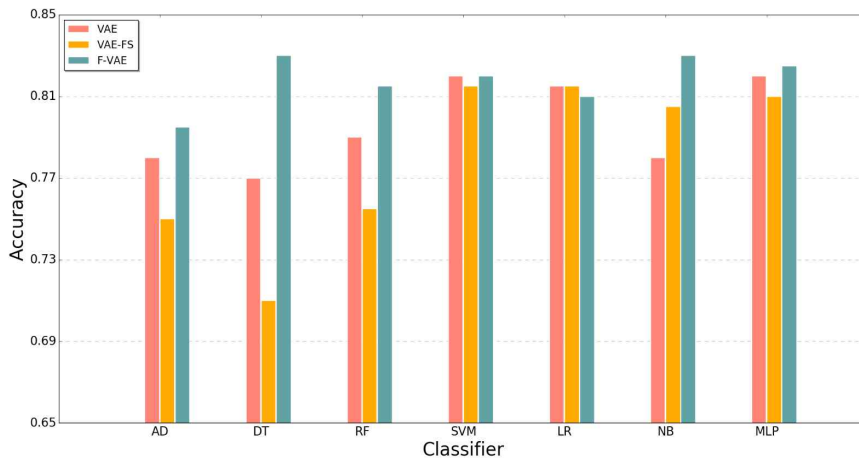


Fig. 4. The accuracy comparison results.

4.2 α 의 영향 비교 실험

이 실험은 α 가 추출된 특징에 어떤 영향을 미치는지 확인하기 위한 실험으로, α 가 1일 때부터 1,000,000이 될 때까지 10배씩 값을 늘리면서 FVAE를 학습시켰다. α 를 10배씩 증가시킨 이유는 $l(\theta, \phi)$ 와 $loss_{cross-entropy}$ 사이의 크기 차이가 학습 데이터, 모델 구조 등 여러 요인에 따라 작게는 몇 배에서 크게는 몇 만배 이상까지 다양하게 발생하기 때문에 좀더 확연한 차이를 관찰하기 위함이다. α 를 제외한 각 분류 모델의 나머지의 경우 위의 4.1 실험과 동일한 환경으로 실험을 수행하였으며, 각 분류기마다의 정확도를 측정하고 7가지 분류기의 평균 정확도를 계산하였다. Fig. 6이 해당 실험의 결과를 나타낸 그

림이다. α 가 1일 때부터 100이 될 때까지는 평균 정확도가 증가하고 있음을 확인할 수 있다. 그러나 α 가 더 커질 경우, 평균 정확도가 감소하기 시작하는 것을 확인할 수 있다. 이러한 결과가 발생한 이유는 α 의 크기가 점점 커지면 $l_{FVAE}(\theta, \phi)$ 내에서 $loss_{cross-entropy}$ 가 차지하는 비중이 점점 늘어나 피드백으로 인한 이득이 커지지만 어느 순간이 지나면 $loss_{cross-entropy}$ 가 $l(\theta, \phi)$ 보다 우선시되면서 피드백 구조가 내부의 인코더 학습을 돕지 않고 오히려 방해하기 때문으로 추측된다.

5. 결 론

본 논문에서는 준지도학습에서 VAE가 잘 분류하

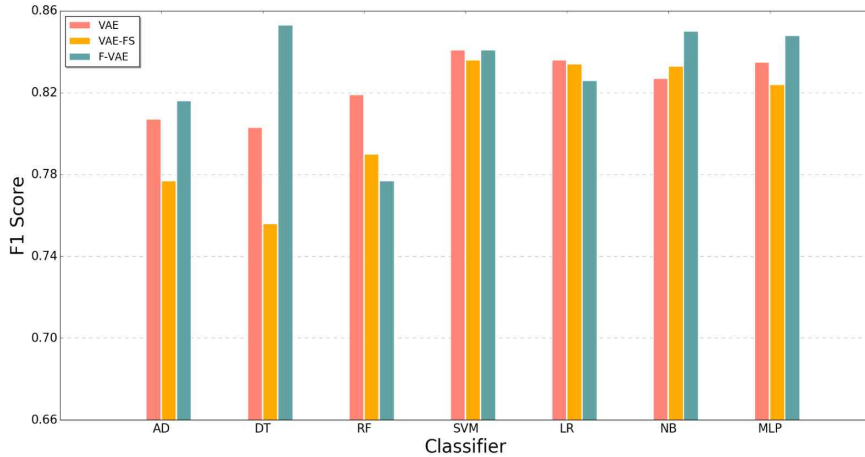


Fig. 5. The F1-score comparison results.

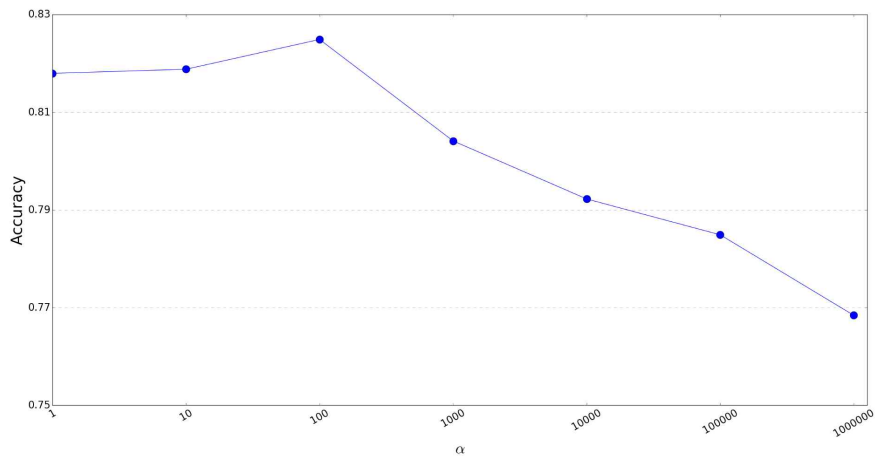


Fig. 6. The accuracy variations according to α .

는 특징을 추출할 수 있도록 만들기 위해 VAE의 구조를 변화시키는 방법에 대해 연구를 진행하여, 그 결과로 FVAE를 제안하였다. FVAE는 내부에 다층 퍼셉트론 구조가 포함되어 있으며, 언라벨 데이터와 라벨 데이터 모두 활용하여 분류 성능을 높일 수 있는 특징을 추출한다. 실험을 통해 기존에 있는 VAE를 활용한 준지도학습보다 정확도와 F1 점수가 향상됨을 보였고 특히 인코더와 분류기를 동시에 학습시킬 수 있어 두 단계로 학습할 필요 없이 한 단계 과정만으로도 충분한 성능을 보일 수 있음을 증명하였다. 향후 연구로 FVAE 손실 함수의 최적화에 대한 연구를 진행할 것이며, 감염병 발생 예측과 같이 다른 분야의 연구에 활용하거나 지능형 교육 시스템의 새로운 데이터에 적용해보는 등 다양한 상황에서 FVAE 성능을 테스트할 예정이다.

REFERENCE

[1] SmartTutor, <http://www.smarttutor.com/> (accessed October 25, 2019).
 [2] AutoTutor, www.autotutor.org/ (accessed October 25, 2019).
 [3] Tutor.com, www.tutor.com/ (accessed October 25, 2019).
 [4] S. Adjei, K. Ostrow, E. Erickson, and N.T. Heffernan, "Clustering Students in ASSISTments: Exploring System- and School-Level Traits to Advance Personalization," *Proceedings of the 10th International Conference on Educational Data Mining*, pp. 340-341, 2017.
 [5] R.S. Baker, A.T. Corbett, and K.R. Koedinger, "Detecting Student Misuse of Intelligent Tutoring Systems," *Proceedings of the International Conference on Intelligent Tutoring Systems*, pp. 531-540, 2004.
 [6] S. Lee and S. Lee, "Data Augmentation for DNN- based Speech Enhancement," *Journal of Korea Multimedia Society*, Vol. 22, No. 7, pp. 749-758, 2019.
 [7] L. Zhao, J. Chen, F. Chen, W. Wang, C.T. Lu, and N. Ramakrishnan, "Simnest: Social Media Nested Epidemic Simulation Via Online Semi-supervised Deep Learning," *Proceeding of 2015*

IEEE International Conference on Data Mining, pp. 639-648, 2015.
 [8] K.M. Adal, D. Sidibé, S. Ali, E. Chaum, T.P. Karnowski, and F. Mériaudeau, "Automated Detection of Microaneurysms Using Scale-adapted Blob Analysis and Semi-supervised Learning," *Computer Methods and Programs in Biomedicine*, Vol. 114, No. 1, pp. 1-10, 2014.
 [9] F.H.H. Mahyoub, M.A. Siddiqui, and M.Y. Dahab, "Building an Arabic Sentiment Lexicon Using Semi-supervised Learning," *Journal of King Saud University-Computer and Information Sciences*, Vol. 26, No. 4, pp. 417-424, 2014.
 [10] Y. Cheng, W. Xu, Z. He, W. He, H. Wu, M. Sun, and Y. Liu, "Semi-supervised Learning for Neural Machine Translation," *Proceeding of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1965-1974, 2016.
 [11] V. Tam, E.Y. Lam, S.T. Fung, W.W.T. Fok, and A.H. Yuen, "Enhancing Educational Data Mining Techniques on Online Educational Resources with a Semi-supervised Learning Approach," *Proceeding of 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering*, pp. 203-206, 2015.
 [12] Y. Wang, H. Fang, Q. Jin, and J. Ma, "SSPA: An Effective Semi-supervised Peer Assessment Method for Large Scale MOOCs," *Interactive Learning Environments*, pp. 1-19, 2019.
 [13] I.E. Livieris, K. Drakopoulou, V.T. Tampakas, T. Mikropoulos, and P. Pintelas, "Predicting Secondary School Students' Performance Utilizing a Semi-supervised Learning Approach," *Journal of Educational Computing Research*, Vol. 57, No. 2, pp. 448-470, 2019.
 [14] S. Klingler, R. Wampfler, T. Käser, B. Solenthaler, and M.H. Gross, "Efficient Feature Embeddings for Student Classification with Variational Auto-encoders," *Proceedings of*

the 10th International Conference on Educational Data Mining, pp. 72-79, 2017.

- [15] D.P. Kingma and M. Welling, "Auto-encoding Variational Bayes," *arXiv Preprint arXiv:1312.6114*, 2013.
- [16] KDD Cup 2015, <https://biendata.com/competition/kddcup2015/> (accessed October 25, 2019).
- [17] Scikit-learn, <https://scikit-learn.org/stable/> (accessed October 25, 2019).
- [18] Tensorflow, <https://www.tensorflow.org/> (accessed October 25, 2019).
- [19] D.P. Kingma and J. Ba, "Adam: A method for Stochastic Optimization," *arXiv Preprint arXiv:1412.6980*, 2014.



정 승 원

2016년 고려대학교 전기전자전파공학부(학사)
 2016년~현재 고려대학교 전기전자공학(석·박사통합과정)
 관심분야: 기계 학습, 딥러닝, 데이터 마이닝, 데이터베이스, 추천 시스템 등



손 민 재

2017년 동서대학교 컴퓨터공학과(학사)
 2019년 고려대학교 전기전자공학과(석사)
 관심분야: 불균형 데이터 처리, 수요 예측, 기계 학습, 데이터 마이닝 등



황 인 준

1988년 서울대학교 컴퓨터공학과(학사)
 1990년 서울대학교 컴퓨터공학과(석사)
 1998년 Univ. Maryland at College Park 전산학과(박사)
 1998년~1999년 Bowie State Univ. 조교수
 1999년 Hughes Research Lab. 연구교수
 1999년~2004년 아주대학교 정보통신전문대학원 조/부교수
 2004년~현재 고려대학교 전기전자공학과 교수
 관심분야: 빅데이터 처리, 멀티미디어 처리 및 검색, 기계 학습, 딥러닝 등