

공학석사학위 논문

Deep metric을 사용한 개인 신용 예측 모델

Model for Predicting Personal Credit Using Deep Metric

2023년 2월

인하대학교 대학원

전기컴퓨터공학과(컴퓨터공학전공)

오 영 민

목 차

국문 요약	1
영문 요약	2
제1장 서론	3
제2장 관련 연구	5
2.1 신용 예측 모델	5
2.2 신용 예측 데이터 불균형 문제	5
2.3 Deep Metric	6
제3장 연구 동기 및 제안 방법	9
3.1 연구 동기	9
3.2 제안 방법	9
3.2.1 Deep Metric을 신용 예측에 적용	9
3.2.2 Deep Metric 학습 모델 구조 변경	10
제4장 실험 방법 및 결과	12
4.1 데이터	12
4.2 실험 개요	12
4.3 성능 평가 및 비교	13
4.3.1 모델별 실험 결과 비교	13
제5장 결론	16
참고 문헌	17

요 약

개인이 신용 대출, 담보 대출 등의 행위로 부채가 생겼을 때 정해진 기간 내에 정상적으로 대출금을 상환할 수 있을지 또는 갚지 못하고 연체될 것인지를 예측하는 개인 신용 예측 모델을 제안한다. 신용 데이터는 정상적으로 대출을 상환한 데이터에 비해 연체된 데이터의 비율이 낮은 데이터 불균형 문제가 존재하기 때문에 일반적인 기계학습 모델로는 학습의 어려움이 발생한다. 이러한 문제를 개선하기 위해 데이터 불균형 문제에 강점이 있는 Deep Metric 모델을 사용하여 개인 신용 예측 모델을 제안하였다. 또한, Deep Metric 여러 모델 중 Lifted Structure를 사용한 신용 예측 모델이 기존의 다른 신용 예측 모델들보다 향상된 성능을 보였다.

ABSTRACT

We suggest a model for predicting personal credit that determines whether a borrower will be able to repay the loan normally within a predetermined time frame or will be past due without repaying it when a borrower becomes in debt as a result of an action like taking out a credit loan or mortgage loan. It is challenging to learn with a general machine learning model because credit data has a data imbalance problem where the ratio of past-due data is smaller than that of data for which loans are typically repaid. A Deep Metric model that is effective in the data imbalance problem is suggested as a solution to this issue in order to predict personal credit. The credit prediction model using the Lifted Structure also outperformed other current credit prediction models among a variety of Deep Metric models.

제 1 장 서론

개인 신용 예측은 대출 기관의 손실 가능성을 감소시키기 위하여, 신용 대출이나 주택 담보 대출, 주식 담보 대출 등의 대출을 받은 개인 차입자가 정해진 기한 내에 상환할 것인지 연체할 것인지를 예측하는 것이다[1, 2, 3].

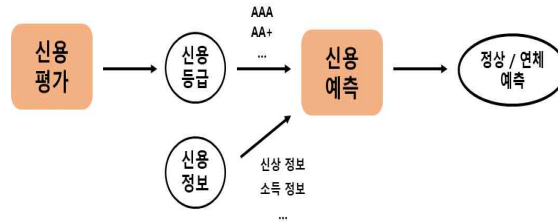


그림 1. 신용 예측과 신용 평가

개인 신용 예측은 그림1과 같이 대출 신청 주체에 대한 신용 평가 기관의 신용 등급과 대출 기관이 보유하거나 수집한 대출자의 신상 정보, 과거 대출 정보, 소득 정보 등의 다양한 종류의 신용 정보를 사용하여 정상(정상 상환) 또는 연체(기한이이익상실) 결과를 예측한다. 개인 신용 예측 모형은 정상 또는 연체의 2개의 레이블(label)을 구분하는 이진 분류 학습 방법인 지도학습(supervised learning) 방식의 기계학습을 사용하여 구현된다[1, 4]. 사용되는 방법으로는 머신러닝 기반의 모델과 통계 기반의 모델들이 사용되고 있고 이 중에서 최근에 XGB(XGBoost) 방법의 성능이 가장 좋은 것으로 알려지고 있다[4, 5, 6, 7, 8, 9].

실제 대출 상황에서는 대출자의 신용 평가로 연체가 예상되는 대출 신청자를 대출에서 많이 차단하기 때문에, 연체된 경우가 정상 상환된 경우보다 매우 적은 현상이 나타난다. 따라서 신용 예측 모델에 사용되는 데이터에는 연체 부류(class)에 속하는 데이터의 비율이 정상 부류에 속하는 데이터의 비율보다 현저히 낮은 데이터 불균형 현상인 데이터 불균형(data imbalance) 문제가 발생한다[4, 5, 7].

이 데이터 불균형 문제는 학습 모델의 학습 성능을 현저히 저하시키기 때문에 SMOTE, ADASYN 등의 resampling 기법을 적용하여 모델의 학습 성능을 개선시킨다[10]. 그러나 resampling 기법을 적용하면 원본 데이터 내에 존재

하는 속성 간 상관관계 등의 특징을 훼손할 수 있기 때문에 학습 성능을 개선하는 측면에는 한계가 있다[11].

따라서 본 논문은 Deep Metric 기법을 사용한 신용 예측 모델을 제시함으로써 신용 예측 학습 데이터의 불균형 문제로 인한 학습 성능 저하 문제를 개선하였다.

제 2 장 관련 연구

2.1 신용 예측 모델

신용 예측 모델은 통계에 기반의 방법과 머신러닝에 기반한 방법 등이 사용되었다. 통계 기반의 방법으로는 Logistic Regression, Naive Bayes, Linear Discriminant Analysis 등의 방법이 적용되었으나 이러한 방법들은 비선형 데이터를 분류하기 어렵기 때문에, 커널(Kernel) 기법이 적용된다[5, 6, 7]. 그러나 커널 기법은 적용할 데이터가 달라질 때마다 새로운 커널을 개발하여 적용해야 하므로, 같은 커널을 여러 데이터에 같이 적용하기 어려운 확장성 문제(scalability problem)에 직면한다[6]. 머신러닝 기법으로는 SVM(Support Vector Machine), RF(Random Forest), XGB, DNN(Deep Neural Network) 등의 방법이 사용되었다[5, 6, 7, 8, 9, 10]. 이중에서 SVM도 새로운 커널이 개발되어 적용해야 한다는 문제가 있다[6]. RF나 XGB 등의 방법은 기존의 신용예측에 사용된 모델 중에서 대체로 좋은 성능을 보여주고 있으나 신용 데이터 불균형 문제를 개선하기 위해 데이터 resampling의 추가적인 방법이 필요하다[5, 10].

2.2 신용 예측 데이터 불균형 문제

데이터 불균형 문제란 각 부류에 속하는 데이터의 비율이 같지 않기 때문에 모델이 학습 후에 데이터의 부류를 예측할 때 적은 부류에 속하는 데이터들을 데이터가 많은 부류로 예측하는 등의 잘못된 결과를 나타내는 문제를 뜻한다. 대체로 신용 예측에서는 모델이 연체된 부류에 속하는 데이터들을 정상적으로 상환한 부류로 예측하는 현상이 발생한다[4, 5, 9].

이를 개선하기 위해 데이터의 수를 변형시키는 데이터 resampling 기법이 적용되었다. 데이터 resampling은 데이터가 적은 부류의 데이터를 증가시키는 방법인 oversampling 방법과 데이터가 많은 부류의 데이터를 감소시키는 방법인 undersampling 방법으로 분류된다. 신용 예측에 사용된 데이터 oversampling 방법으로 데이터를 무작위로 증가시키는 Random Over

Sampling과 K-Nearest Neighborhood를 사용한 방법이 존재한다[10]. 대체로 K-Nearest Neighborhood 기반의 방법 중 SMOTE, ADASYN 등의 방법이 적용되었으며 SMOTE가 대체로 좋은 성능을 나타내고 있다[10]. 이와 다르게 신용 예측에 사용된 undersampling 방법으로는 데이터를 무작위로 감소시키는 Random Under Sampling 등의 기법이 사용되었다[10].

2.3 Deep Metric

Deep Metric은 데이터의 label을 구분하기 위한 방법의 일종으로서, 원 데이터 공간에서는 데이터의 구분이 어려운 경우에 Deep Learning을 사용하여 비선형적으로 다른 데이터 공간으로 데이터들을 변환하고 변환된 공간에서는 데이터의 구분이 명확해지도록 학습하는 방법이다. 이를 위해서 Deep Metric은 변환된 공간에서 같은 부류에 속하는 데이터들은 서로 가깝게 위치하고 다른 부류의 데이터들과는 멀리 떨어져 위치하게 하는 데이터 변환 함수를 DNN으로 학습한다[12, 13, 14].

Deep Metric은 대체로 face verification이나 face identification 등의 이미지 데이터를 사용하는 Computer Vision에서 많이 주로 사용되었는데[14], 분류가 잘되지 않는 데이터 공간을 분류가 잘되는 다른 종류의 데이터 공간으로 변형시키는 변환 함수를 DNN으로 학습하기 때문에, 특정 부류에 속하는 데이터의 개수나 비율 등의 영향을 작게 받으므로 데이터 불균형 문제가 있다고 하더라도 분류의 정확성이 향상된다는 강점을 가진다[11, 13, 15, 16].

Deep Metric에는 Siamese Network, Triplet Network, Lifted Structure 등의 여러 가지 종류의 구현 기법들이 있는데, 대체로 부류가 서로 다른 데이터 간의 최소 거리를 유지하기 위해 τ 와 같은 margin을 사용한다.

Siamese Network는 레이블이 같거나 다른 2개의 데이터를 쌍으로 데이터 공간 변환 함수를 학습하는 기법으로 face image classification, face identification 등의 얼굴 이미지와 관련된 분야에서 많이 사용되었다[17, 18, 19].

$$\begin{aligned} \text{contrastive loss} = & \sum_{(i,j) \in S} h(D_{i,j} - \tau_1) \\ & + \sum_{(i,j) \in D} h(\tau_2 - D_{i,j}) \end{aligned} \quad (1)$$

$$D_{i,j} = \|f(i) - f(j)\|_2 \quad (2)$$

여기서 f 는 학습된 공간 변환 함수이며, $f(i)$, $f(j)$ 는 i 번째와 j 번째 원 데이터를 공간 변환한 벡터이며, D 는 두 데이터의 변환된 공간에서의 거리이다.

식 (1)과 (2)는 Siamese Network에서 사용되는 손실 함수를 나타낸 것으로 hinge loss를 사용한다[14]. 또한, 같은 부류의 데이터가 가까워지기 위해 τ_1 의 값이 τ_2 의 값보다 작아야 한다[14].

Triplet Network는 anchor, positive, negative의 3개의 데이터를 한 개의 훈련 데이터로 간주하여 공간 변환 함수를 학습하는 기법인데, anchor는 기준이 되는 데이터이고 positive 데이터는 anchor와 같은 부류에 속하는 데이터이며 negative 데이터는 anchor와 다른 부류의 데이터이다.

$$\text{triplet loss} = \sum_i h(\tau + D_{i,i^+} - D_{i,i^-}) \quad (2)$$

식 (3)은 triplet loss를 정의하는 것으로 i , i^+ , i^- 는 각각 anchor, positive, negative 데이터이다[14]. 그리고 Triplet Network 역시 hinge loss를 사용하며 face verification이나 이미지의 유사도 순위를 결정하는 deep ranking 등의 분야에서 사용되고 있다[14, 20, 21].

Siamese Network나 Triplet Network는 훈련 데이터가 2개의 쌍(2 pair) 또는 3개의 쌍(3 tuple)으로 만들지기 때문에 학습 결과가 데이터 쌍이 어떻게 생성되는지에 많은 영향을 받으므로 제한된 정보를 학습한다는 단점이 있다[16].

Lifted Structure 기법은 훈련데이터가 2개의 쌍 또는 3개의 쌍으로 제한되지 않고 다중 쌍(batch)으로 구성되기 때문에 여러 부류에 속하는 데이터들이 동시에 학습되어 전체적인 데이터의 관계가 효율적으로 학습될 수 있다는 장점이 있다[16].

$$\text{Lifted Structure loss} = \frac{1}{2|P|} \sum_{i,j \in P} \max(0, J_{i,j})^2 \quad (4)$$

$$J_{i,j} = \log(\sum_{k \in N} \exp\{\tau - D_{i,k}\}) + D_{i,j} \quad (5)$$

식 (4)와 (5)는 이러한 Lifted Structure loss를 나타내는 식으로 P 는 positive pair 집합으로서 i 와 j 는 같은 부류에 속하는 데이터이고, N 은 negative pair 집합으로서 k 는 i, j 데이터와 다른 부류에 속하는 데이터이다 [16].

제 3 장 연구 동기 및 제안 방법

3.1 연구 동기

신용 예측 데이터의 경우 정상적으로 대출을 상환한 정상 class에 속하는 데이터의 수가 연체된 class보다 많은 특징을 가지고 있다. 이를 해결하기 위해 resampling 방법을 사용하여 데이터 imbalance의 영향을 줄이려는 시도가 다수 존재한다[10]. 하지만 데이터 resampling 방법의 경우 원본 데이터의 통계적 특성을 고려하지 않고 데이터의 수를 변형시키기 때문에 데이터의 특성을 훼손시킬 수 있다는 단점이 존재한다[11]. 따라서 본 논문에서는 거리에 따라서 데이터를 분류하기 때문에 데이터 imbalance 문제에 장점을 가진 deep metric을 신용 예측에 적용하였다.

3.2 제안 방법

3.2.1 Deep Metric을 신용 예측에 적용

신용 예측에 사용되는 훈련 데이터는 정상 데이터와 연체 데이터들의 비율에 현저히 큰 차이가 존재하기 때문에 모델의 성능을 향상하기 위한 데이터 불균형 문제 해소 방법이 필요하다. 그러나 기존 신용 예측 기법들은 여러 가지 방법을 적용하여 데이터 불균형 문제에 대처하였음에도 불구하고 여전히 성능의 저하를 충분히 개선시키지 못하였다[4, 5, 9].

Deep Metric 기법이 본질적으로 데이터 불균형 문제에 강건하여 추가적인 resampling 적용이 필요하지 않다는 장점이 있고, feature representation learning의 일종으로서 원본 데이터의 특징을 효율적으로 학습하는 특징이 있다[16]. 따라서 본 논문에서는 신용 예측 모델로써 Deep Metric 기법을 적용하는 방법을 제안한다.

이러한 장점을 보유한 Deep Metric 모델은 Siamese Network, Triplet Network, Lifted Structure 등 3가지 기법이 있는데, 이 중에서 데이터의 상호 관계를 전체적으로 잘 파악할 수 있게 해주는 Lifted Structure 모델을 사용하여 신용 예측을 한다.

3.2.2 Deep Metric 학습 모델 구조 변경

기존 Deep Metric 모델은 대부분 이미지 학습에 사용되고 있기 때문에, CNN (Convolutional Neural Network)이 사용되고 있으며 그 내부 구조에는 2d convolution과 max pooling이 사용된다[12, 13]. 그러나 신용 데이터는 이미지 형태가 아닌 텍스트형(tabular)이기 때문에 신용 예측에 적용하기 위해서는 Deep Metric의 구조에 약간의 변형이 필요하다.

Deep Metric을 신용 예측에 적용하기 위해 그림 2와 같이 input을 데이터의 형태에 따라 범주형(categorical)데이터와 수치형(numerical)데이터로 분리하였다. 범주형 데이터를 학습시키는 방법에는 embedding을 사용하는 방법과 one-hot encoding을 적용하는 방법이 사용된다.

one-hot encoding은 범주 값의 개수만큼 bit수가 존재하며 범주 값에 대응되는 위치의 한개의 bit만 1이 되어 범주값을 나타낸다. 따라서 one-hot encoding은 범주값의 개수가 늘어나면 차원이 늘어나게 되어 많은 자원을 필요로 하며 학습 효율이 저하되는 단점이 있다[22]. 반면에 embedding은 차원을 고정시킬 수 있기 때문에, 범주형 데이터 학습을 위해서는 embedding 기법을 적용하여 학습시키는 것이 더 효율적이다.

또한, 신용 예측 데이터는 텍스트형의 데이터이므로 2d convolution 대신에 1d convolution을 사용하였다. 최근 1d convolution과 fully connected layer를 같이 사용하여 텍스트형 데이터를 분류할 때 기존 분류 모델들보다 뛰어난 성능을 나타내었으며[23] 이를 신용 예측에 적용하여 기존의 다른 모델들과 비교하여 향상된 성능을 나타내었다[24].

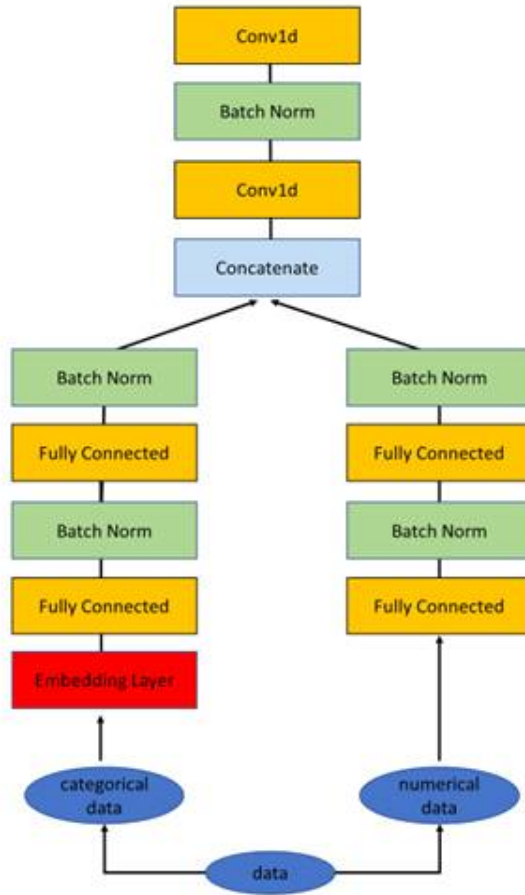


그림 2. 적용된 network 구조

추가적으로 overfitting 방지를 위해 Batch Normalization과 L2 regularization를 적용하였으며 활성화 함수로는 식 (6)와 같은 mish 활성화 함수를 사용하였다[25].

$$mish(x) = x \tanh(\zeta(x)) \quad (6)$$

$$\zeta(x) = \ln(1 + e^x) \quad (7)$$

제 4 장 실험 방법 및 결과

4.1 데이터

실험 데이터로는 A 저축은행 실제 대출 데이터와 UCI Machine Learning Repository에서 제공하는 default of credit card clients Data Set(일명 Taiwan credit card data)을 사용하였다. A 저축은행 데이터의 개수는 2569개이고 210개의 attribute로 구성되어 있으며 정상과 연체 데이터의 비율이 각각 95%, 5%로 구성되어 있어 심한 불균형을 보여주고 있다. UCI에서 제공하는 데이터의 개수는 30,000개이고 attribute는 24개로 구성되어 있으며 정상과 연체 데이터의 비율이 각각 78%, 22%로 되어 있다.

4.2 실험 개요

본 논문에서 제안된 방법과의 실험 결과의 비교를 위해서, 기존 기법 중에서 좋은 성능을 나타낸 것으로 알려진 XGB와 RF 모델을 사용하였다. XGB와 RF 모델에 resampling을 적용하지 않았을 때의 실험결과들과 SMOTE와 ADASYN 등의 resampling을 적용한 결과들을 비교하였다[4, 5, 7, 8, 9]. 훈련 데이터와 시험 데이터의 비율을 각각 60%, 40%로 나누었다. A 저축은행 데이터를 사용한 실험은 100개의 모델을 학습시킨 후 예측을 통해 결과를 비교하였으며 default of credit card clients Data Set을 사용한 실험은 50개의 모델을 학습시킨 후 예측 결과를 비교하였다.

4.3 성능 평가 및 비교

4.3.1 모델별 실험 결과 비교

그림 3과 같이 평가 지표로는 AUC와 신용 예측에서 사용되는 평가 기준인 승인율과 연체율을 사용하였다.

		예측 class		AUC (정상재현율 + 연체재현율) / 2
		연체 (Positive)	정상 (Negative)	
실제 class	연체 (Positive)	TP	FN	연체재현율 $TP / (TP + FN)$
	정상 (Negative)	FP	TN	정상재현율 $TN / (TN + FP)$
		연체율 $FN / (TN + FN)$		승인율 $(TN + FN) / (TN + FN + TP + FP)$

그림 4. confusion matrix

그림 4는 A 저축은행 데이터를 Deep Metric 모델을 사용하여 학습하여 변환된 공간으로 embedding한 결과를 t-sne를 사용해 2차원으로 시각화하였다 [26]. 그림 4에서 보여지는 바와 같이 원 데이터 공간에서의 데이터분포는 정상과 연체가 뒤섞여 있어서 분류하기에 적합하지 않으나 Deep Metric으로 변환된 공간에서의 데이터 분포는 같은 부류의 데이터들은 가깝게 위치하고 다른 부류의 데이터들은 서로 멀리 떨어져 있는 특징을 잘 보여준다.

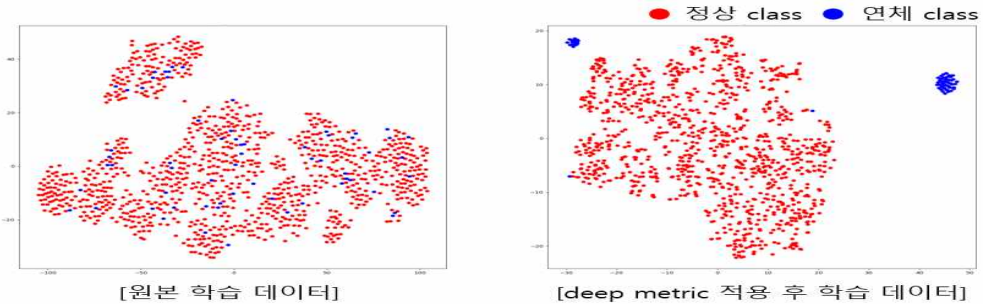


그림 5. A 저축은행 데이터에 Deep Metric 적용 시각화

표 1. Deep Metric 모델별 AUC 비교

모델	Lifted Structure	Triplet Network	Siamese Network
평균 AUC	0.548	0.514	0.401

표 2. Taiwan 데이터 Deep Metric 모델별 비교

모델	Lifted Structure	Triplet Network	Siamese Network
평균 AUC	0.695	0.642	0.556

표 1과 표 2는 Deep Metric의 3가지 기법들을 A 저축은행 데이터와 Taiwan credit card 데이터를 가지고 실험하여 그 성능들을 비교한 결과이다. 표1과 표 2에서 보는 바와 같이 Lifted Structure(LS)를 사용한 결과가 Triplet Network와 Siamese Network 등의 다른 기법들을 사용한 결과보다 좋은 성능을 나타낸다는 것을 확인하였다.

표 3과 표 4는 기존 신용 예측 모델들과 Deep Metric LS를 사용한 실험 결과들의 비교를 보여준다. Deep Metric을 적용한 모델이 기존의 다른 모델보다 개선된 성능을 보였다. 특히 데이터의 비율이 낮은 연체 부류의 데이터들을 예측하는 성능이 다른 모델보다 높음을 확인하였다. 또한, 각 부류의 데이터들을 예측하는 성능을 측정하는 AUC 또한 기존 모델들과 비교하였을 때 높은 결과를 보였다. 그리고 그림 5와 그림 6에서 볼 수 있듯이 제안한 모델을 사용하였을 때 결과에 대한 AUC 분포가 다른 모델들에 비해 높은 것을 확인하였다.

표 3. A 저축은행 데이터 모델별 실험 결과 비교

모델	Deep metric(LS)	XGB	XGB + SMOTE	XGB + ADASYN	RF	RF + SMOTE	RF + ADASYN
평균 정상재현율	0.879	0.973	0.947	0.948	0.98	0.948	0.947
평균 연체재현율	0.218	0.047	0.103	0.102	0.029	0.088	0.092
평균 승인율	0.873	0.972	0.944	0.945	0.979	0.946	0.944
평균 연체율	0.052	0.057	0.055	0.056	0.057	0.056	0.056
평균 AUC	0.548	0.51	0.525	0.524	0.504	0.518	0.52

표 4. default of credit card 데이터 실험 결과 비교

모델	Deep metric(LS)	XGB	XGB + SMOTE	XGB + ADASYN	RF	RF + SMOTE	RF + ADASYN
평균 정상재현율	0.756	0.935	0.917	0.913	0.941	0.883	0.87
평균 연체재현율	0.634	0.368	0.403	0.405	0.37	0.484	0.493
평균 승인율	0.868	0.846	0.846	0.849	0.872	0.801	0.789
평균 연체율	0.121	0.161	0.156	0.157	0.16	0.143	0.142
평균 AUC	0.695	0.652	0.66	0.659	0.656	0.683	0.681

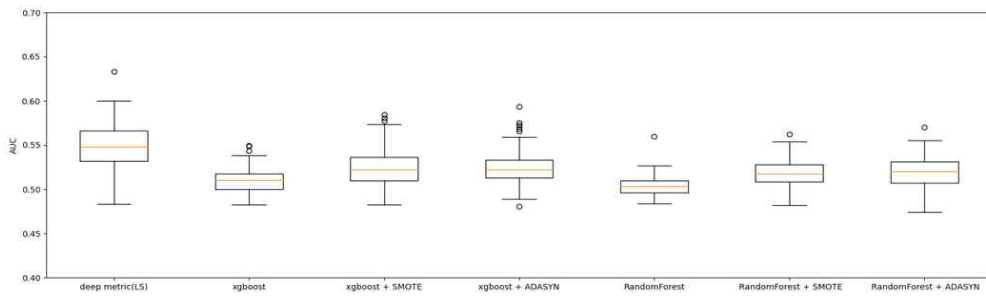


그림 6. A 저축은행 데이터 AUC 분포 비교

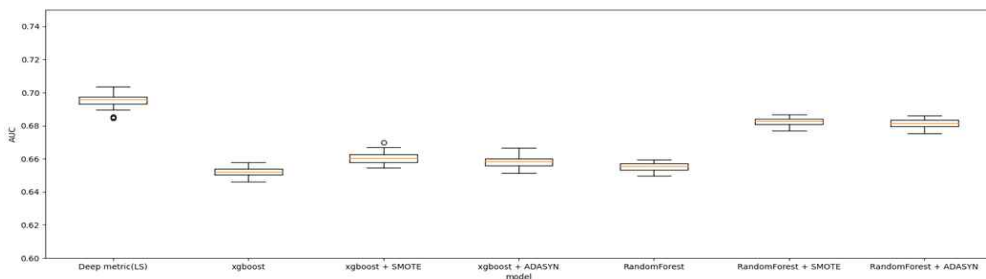


그림 7. default of credit card 데이터 AUC 분포 비교

제 5 장 결론

해당 논문에서는 deep metric 모델 중 lifted structure를 사용하였으며 image 중심의 deep metric 구조를 tabular 데이터 학습을 위해 변경하여 신용 예측에 적용하였다. 또한, 기존 신용 예측에서 좋은 성능을 보인 XGBoost과 Random Forest 및 resampling 기법을 적용한 모델과 비교를 하였다. 제안한 방법과 기존 신용 예측 모델을 비교하였을 때 제안한 방법이 향상된 성능을 나타냄을 확인하였다.

참고 문헌

- [1] Chen, N., Ribeiro, B., & Chen, A. (2016). Financial credit risk assessment: a recent review. *Artificial Intelligence Review*, 45(1), 1–23
- [2] Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International journal of forecasting*, 16(2), 149–172
- [3] Namvar, A., Siami, M., Rabhi, F., & Naderpour, M. (2018). Credit risk prediction in an imbalanced social lending environment. *arXiv preprint arXiv:1805.00801*.
- [4] Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, 91, 106263.
- [5] Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446–3453.
- [6] Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), 1447–1465..
- [7] Qiu, W. (2019, July). Credit risk prediction in an imbalanced social lending environment based on XGBoost. In *2019 5th International Conference on Big Data and Information Analytics (BigDIA)* (pp. 150–156). IEEE.
- [8] Li, Y. (2019, August). Credit risk prediction based on machine learning methods. In *2019 14th International Conference on Computer Science & Education (ICCSE)* (pp. 1011–1013). IEEE.
- [9] Li, H., Cao, Y., Li, S., Zhao, J., & Sun, Y. (2020). XGBoost model and its application to personal credit evaluation. *IEEE Intelligent*

Systems, 35(3), 52–61

[10] Xiao, J., Wang, Y., Chen, J., Xie, L., & Huang, J. (2021). Impact of resampling methods and classification models on the imbalanced credit scoring problems. *Information Sciences*, 569, 508–526.

[11] Jing, X. Y., Zhang, X., Zhu, X., Wu, F., You, X., Gao, Y., ... & Yang, J. Y. (2019). Multiset feature learning for highly imbalanced data classification. *IEEE transactions on pattern analysis and machine intelligence*, 43(1), 139–156.

[12] Kulis, B. (2013). Metric learning: A survey. *Foundations and Trends® in Machine Learning*, 5(4), 287–364.

[13] Kaya, M., & Bilge, H. Ş. (2019). Deep metric learning: A survey. *Symmetry*, 11(9), 1066.

[14] Lu, J., Hu, J., & Zhou, J. (2017). Deep metric learning for visual understanding: An overview of recent advances. *IEEE Signal Processing Magazine*, 34(6), 76–84.

[15] Susan, S., & Kaushik, A. (2022). Localized Metric Learning for Large Multi-class Extremely Imbalanced Face Database. In *International Conference on Database Systems for Advanced Applications* (pp. 64–78). Springer, Cham.

[16] Oh Song, H., Xiang, Y., Jegelka, S., & Savarese, S. (2016). Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4004–4012).

[17] Hu, J., Lu, J., & Tan, Y. P. (2014). Discriminative deep metric learning for face verification in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1875–1882).

[18] Lu, J., Wang, G., Deng, W., Moulin, P., & Zhou, J. (2015). Multi-manifold deep metric learning for image set classification. In

Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1137–1145).

[19] Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification–verification. *Advances in neural information processing systems*, 27.

[20] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815–823).

[21] Wang, J., Song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J., ... & Wu, Y. (2014). Learning fine-grained image similarity with deep ranking. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1386–1393).

[22] Dahouda, M. K., & Joe, I. (2021). A deep-learned embedding technique for categorical features encoding. *IEEE Access*, 9, 114381–114391.

[23] Liu, K., Kang, G., Zhang, N., & Hou, B. (2018). Breast cancer classification based on fully-connected layer first convolutional neural networks. *IEEE Access*, 6, 23722–23732.

[24] Hayashi, Y., & Takano, N. (2020). One-dimensional convolutional neural networks with feature selection for highly concise rule extraction from credit scoring datasets with heterogeneous attributes. *Electronics*, 9(8), 1318.

[25] Misra, D. (2019). Mish: A self regularized non-monotonic neural activation function. *arXiv preprint arXiv:1908.08681*, 4(2), 10–48550.

[26] Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).