ISSN 2383-630X(Print) / ISSN 2383-6296(Online)
Journal of KIISE, Vol. 53, No. 9, pp. 913-919, 2026. 9
https://doi.org/10.5626/JOK.2026.53.9.913

계층적 클래스 인식 디퓨전 모델 기반 불균형 금융 테이블 데이터 합성 연구

(A Hierarchical Class-Aware Diffusion Framework for Imbalanced Financial Tabular Data Synthesis)

최 현 기 † 김 영 재 ††
(Hyungee Choi) (Youngjae Kim)

요 약 금융산업의 AI 활용은 규제로 인한 데이터 접근이 제한되어 합성 데이터가 대안으로 부상하나, 금융 데이터의 클래스 불균형으로 기존 모델은 소수 클래스 정보 보존에 한계를 지닌다. 본 연구는 TabDiff[1]의 클래스 무관 순방향 노이즈가 소수 신호를 훼손하는 점에 주목하여, 디퓨전 학습의 샘플링·순방향·손실 단계에서 소수 클래스를 보존하는 프레임워크를 제안한다. 소수 클래스 비율을 3%로 조정한 UCI Default(카드연체)·Adult Income(소득예측) 및 소수 클래스 비율 1.10% BAF(계좌사기) 데이터 분석 결과, 소수 클래스 분포 충실도 개선과 함께 Minority F1 스코어를 각각 +126.7%($p<.001$), +23.1%($p<.01$), +32.4% 향상시켰다.

키워드: 합성 데이터, 디퓨전 모델, 클래스 불균형, 금융 데이터, 소수 클래스 보존

Abstract The adoption of AI in the financial industry is constrained by regulations that limit access to real data, leading to the emergence of synthetic data as an alternative. However, the class imbalance inherent in financial data limits existing models' ability to preserve minority-class information. This study proposes a framework that selectively preserves the minority class across the sampling, forward, and loss stages of diffusion training, motivated by the observation that the class-agnostic forward noise of TabDiff[1] corrupts minority-class signals. On UCI Default (card-payment delinquency) and Adult Income (income prediction), both adjusted to a 3% minority ratio, and BAF (account fraud) at a 1.10% minority ratio, the proposed method improves the Minority F1 score by +126.7% ($p<.001$), +23.1% ($p<.01$), and +32.4%, respectively. Additionally, it enhances the distributional fidelity of the minority class.

Keywords: synthetic data, diffusion model, class imbalance, financial data, minority-class preservation

· 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단(RS-2024-00416666) 및 정보통신기획평가원의 인공지능혁신인재양성사업(IITP-RS-2026-25547954)의 지원을 받아 수행된 연구임

† 정 회 원 : 서강대학교 데이터사이언스·인공지능전공 학생
hyungee1212@naver.com
†† 정 회 원 : 서강대학교 컴퓨터공학과 교수(Sogang Univ.)
youkim@sogang.ac.kr
(Corresponding author임)

논문접수 : 2026년 5월 6일
(Received 6 May 2026)
논문수정 : 2026년 6월 22일
(Revised 22 June 2026)
심사완료 : 2026년 6월 24일
(Accepted 24 June 2026)

정보과학회논문지 제53권 제9호(2026. 9)

1. 서 론

금융산업의 AI 적용은 금융 서비스 개발 및 금융 범죄 예방 영역에서 활용되고 있으나, 국내 「개인정보보호법」 등의 규제로 데이터 활용에 제약이 따른다. 이에, 원본 데이터의 특성을 반영한 데이터 합성 기법이 부상하고 있으며, 데이터 생성 시 소수 클래스의 특성을 보존하여 데이터의 유용성을 확보하는 전략이 필요하다.

테이블 데이터 합성에서 디퓨전 모델은 우수한 분포 재현 성능을 보이나, 기본적으로 학습 데이터의 클래스 불균형 상황을 가정하지 않은 채로 설계되었다. 특히 순방향 노이즈 주입 과정이 클래스와 무관하게 작동할 경우, 소수 클래스와 다수 클래스에 동일한 노이즈 강도가 적용되어 표본 수가 절대적으로 부족한 소수 클래스의 정보 손실이 크게 나타나며, 불균형 정도가 심해질수록 그 영향이 급격히 악화된다.

이러한 한계는 학습 가능 노이즈 스케줄을 도입한 TabDiff[1]에서도 동일하게 관찰된다. 대만 신용카드 연체 여부 이진 분류 데이터인 'UCI Default'에 TabDiff를 적용하여, 원본 데이터로 학습·평가한 분류기(TRTR, Train on Real, Test on Real)의 소수 클래스 F1 점수(Minority F1 score) 대비 합성 데이터로 학습하고 원본으로 평가한 분류기(TSTR, Train on Synthetic, Test on Real)의 회복률을 측정하였다. 그 결과 원본 비율(22.12%)에서는 TSTR이 TRTR의 96.02% 수준을 유지하였으나, 소수 비율을 3%로 조정하면 38.42%로 급락하였다(표 1). 이는 불균형이 심화될수록 합성 모델이 소수 클래스 정보를 보존하지 못함을 보인다.

본 연구는 이러한 한계를 보완하고자, 디퓨전 모델 생성 과정에서 클래스 정보를 인식하여 소수 클래스의 정보를 단계적으로 보존하는 계층적 접근을 제시한다.

2. 연구배경

2.1 선행연구

데이터 부족 및 클래스 불균형을 개선하기 위해 SMOTE와 같은 고전 증강 방식부터 GAN·디퓨전 등 기법이 연구되었다. SMOTE[2]는 소수 클래스 샘플의 k-최근접 이웃을 선형 보간 하는 방식으로 실용적이나, 샘플 간의 직선 결합에 의존하여 데이터에 내재된 새로운 분포 모드를 포착하지 못하고, 고차원 혼합형 데이터에서 비현실적 피처 조합을 생성하는 한계가 있다. 이에 본 연구는 SMOTE 계열을 분포 자체를 생성하는 접근이 아닌 보간 기반 증강으로 보아 직접 비교 대상에서 제외한다.

표 1 클래스 비율에 따른 TabDiff F1 회복률(n=5)
Table 1 TabDiff F1 Recovery by Class Ratio (n=5)

Class Ratio	TRTR Minority F1	TSTR Minority F1	TSTR/TRTR (%)
Original (22.12%)	0.4668	0.4482	96.02
Adjusted (3%)	0.0760	0.0292	**38.42**

GAN 계열에서는 테이블 데이터에 특화된 CTGAN[3]이 모드별 정규화와 조건부 생성기를 이용해 연속형·범주형이 혼재된 분포를 처리하였다. 그러나 GAN 특유의 모드 붕괴(mode collapse)와 학습 불안정성으로 인해, 클래스 불균형 환경에서 소수 클래스의 모드가 소실되는 한계가 보고되었다.

이후 연구는 입력 데이터에 노이즈를 주입한 뒤 역복원하여 표본을 생성하는 디퓨전 계열로 발전하였다. 디퓨전 모델은 적대적 학습이 불필요하여 GAN 대비 학습이 안정적이고 모델 학습 과정에서 모드 붕괴를 완화한다.

수치형과 범주형 피처를 단일 모델에서 처리하는 TabDDPM[4]은 노이즈 스케줄이 사전 정의된 함수로 고정되어 피처의 분포 특성을 반영하기 어렵고, 잠재 공간(latent space)에서 스코어 기반 디퓨전을 수행하는 TabSyn[5]은 VAE를 별도로 사전 학습해야 하며 합성 품질이 VAE의 잠재 표현력에 종속된다는 한계가 있다. 피처별 학습 가능 노이즈 스케줄을 도입한 TabDiff[1]는 피처별 최적 노이즈 강도를 데이터로부터 학습하나, 노이즈 스케줄이 클래스 무관하게 작동하여 불균형 환경에서 소수 클래스 신호가 구조적으로 손실된다. 이처럼 세 모델 모두 클래스 불균형 환경에서 소수 클래스 보존을 고려하여 설계되지 않았다.

클래스 불균형 학습은 개입 단계에 따라 데이터·손실·모델 구조로 구분된다. 데이터 단계에서 미니배치의 클래스 구성비를 통제하는 Class-Balanced Batch Sampling[6]은 주로 분류 학습의 편향 완화 기법으로 제안되었으며, 본 연구는 이를 소수 클래스 보존을 위한 디퓨전 학습에 차용한다. 손실 단계의 Focal Loss[7]·Class-Balanced Loss[8]는 단일 시점의 손실에 클래스·샘플 가중을 부여하고, 같은 손실 단계의 Min-SNR[9]·P2 weighting[10]은 timestep별 손실 기여도를 조정하나 클래스 차원을 다루지 않는다. 모델 구조 단계의 CBDM[11]은 이미지 도메인에서 클래스 간 생성 분포를 균형화하는 정규화 손실을 추가하여 디퓨전 학습에 클래스 인식을 도입하였으나, 순방향 노이즈 스케줄 자체는 클래스 무관하게 유지된다.

금융 도메인에서는 FinDiff[12]가 혼합형 금융 데이터를 통합 처리하는 디퓨전 모델을 제안하였고, 이를 확장한 Imb-FinDiff[13]는 클래스 라벨 예측을 결합한 이중

학습 목적(dual learning objective)으로 불균형을 다루었으나, 모두 순방향 노이즈 스케줄은 클래스 무관하게 유지된다. 종합하면 기존 연구는 소수 클래스 보존을 데이터·손실·모델 구조 중 단일 단계에 결합하였으며, 디퓨전의 순방향 노이즈 주입 단계부터 클래스 인식을 적용한 사례는 확인되지 않는다.

2.2 본 연구의 차별점

본 연구는 TabDiff[1]의 피처별 차등 노이즈 스케줄이 클래스별로 이질적 분포를 갖는 금융 데이터의 구조적 특성을 보존하는 데 적합하다는 점에 주목하여, 이를 베이스라인으로 샘플링·순방향·손실 함수 세 단계 모두에 소수 클래스 인식(minority-aware)을 결합한다. 특히 순방향 단계에서는 클래스 라벨과 피처 기여도를 노이즈 스케줄에 함께 반영하여 소수 클래스의 주요 피처 신호를 차등적으로 보존한다.

제안 모델은 (1) 샘플링 단계의 Class-Balanced Batch Sampling(CBBS), (2) 순방향 단계의 Class-Conditional Noise Schedule(CC), (3) 손실 함수 단계의 Timestep-Stratified Class Weighting(TSCW)의 3중 모듈로 구성된다. 이 중 CC는 테이블 구조 디퓨전 모델의 순방향 단계에 클래스 라벨과 피처 기여도를 주입하여 클래스에 따라 노이즈를 차등 적용하는 역할을 수행하며, CBBS와 TSCW는 각각 샘플링·손실 단계에서 소수 클래스 학습 신호를 보강한다.

3. 제안방법

3.1 제안 구조

제안 모델은 그림 1과 같이 TabDiff[1]의 학습 파이프라인을 따르며, 샘플링·순방향·손실 세 단계에 클래스 인식 모듈(CBBS, CC, TSCW)을 결합하여 각 단계에서 소수 클래스 정보를 보존한다.

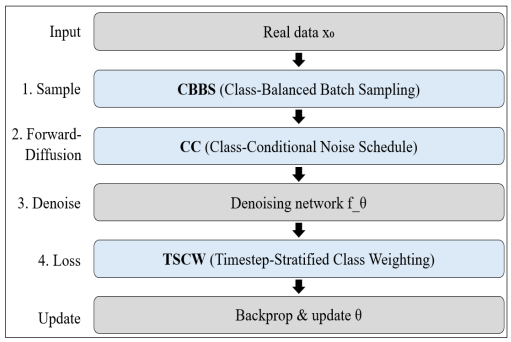

그림 1 제안 프레임워크 학습 파이프라인 구조

Fig. 1 Training Pipeline of the Proposed Framework

3.2 CBBS – Class-Balanced Batch Sampling

$$P(y = 1) = \rho$$

표준 샘플링에서는 배치 내 소수 클래스 샘플 수가 자연 비율을 따라 확률적으로 변동하며, 극단 불균형 환경에서는 일부 배치의 소수 샘플 수가 0에 가까워져 손실 기여가 다수 클래스에 지배된다. 이는 소수 클래스 학습 신호 부족으로 이어져 생성 시 소수 클래스가 과소 생성되는 문제로 이어진다. 이에, CBBS(Class-Balanced Batch Sampling)를 샘플링 단계에 적용하여 매 배치에서 소수 클래스 추출 확률을 사전 지정한 값 ρ로 고정하고 소수 클래스 학습 신호를 안정적으로 확보한다. 본 연구는 ρ를 자연 비율을 약간 상회하는 값으로 두어, 클래스 불균형 상황에서 소수 클래스가 과소 생성되는 문제를 해소하면서 합성 분포가 원본을 벗어나지 않는 균형 잡힌 지점을 확보하는 것을 목적으로 한다.

3.3 CC – Class-Conditional Noise Schedule

$$\text{x\_num: } \sigma\_\text{num}(t, y, c) = \sigma\_\text{base},c(t) \cdot \exp(-\alpha \cdot \mathbb{1}[y=1] \cdot \text{imp\_c})$$

$$\text{x\_cat: } t\_\text{eff}(t, y, c) = t \cdot \max(0.01, 1 - \alpha \cdot \mathbb{1}[y=1] \cdot \text{imp\_c})$$

순방향 단계에서는 클래스 라벨과 피처 기여도를 노이즈 스케줄에 함께 반영하여 소수 클래스 신호를 차등적으로 보존하는 CC(Class-Conditional Noise Schedule)를 제안한다.

피처 기여도(imp_c)는 XGBoost 분류기를 통해 학습 시작 전 학습 데이터에서 1회 산출하는 것으로 하며, [0,1]로 정규화한 뒤 학습 중 변화하지 않도록 고정한다.

이를 통해 데이터 클래스와 피처 기여도에 따라 차등 노이즈를 부여할 수 있다. 또한, 다수 클래스(y=0)는 노이즈 감쇠를 적용하지 않고, 소수 클래스(y=1)만을 대상으로 피처 기여도(imp_c)에 비례하는 노이즈 감쇠를 적용하여 전체 분포를 보존하는 상태에서 소수 클래스 신호를 차등적으로 보호한다. 피처별로는 수치형 피처에 σ를 직접 곱하는 지수 감쇠를 적용하고, 범주형 피처에는 timestep t를 t_eff로 변환하여 마스킹 확률을 조절하는 선형 감쇠를 적용한다.

α는 노이즈 감쇠 강도를 결정하는 하이퍼파라미터로, α=0이면 CC가 비활성화되어 TabDiff[1] 원본의 순방향 과정과 동일하게 작동하고, α가 커질수록 소수 클래스의 피처 기여도가 높은 피처에 대한 노이즈 감쇠가 강해진다.

3.4 TSCW – Timestep-Stratified Class Weighting

$$w(t, y) = 1 + \gamma \cdot (1 - t) \cdot \mathbb{1}[y=1]$$

TSCW(Timestep-Stratified Class Weighting)는

기존 timestep 가중 연구(Min-SNR[9], P2 weighting [10])의 timestep 가중에 데이터의 클래스 축(𝟙[y=1])을 곱하여 소수 클래스의 노이즈 복원 단계에 가중치를 둔다.

다수 클래스 가중은 1로 고정되며, 다수 클래스는 학습 데이터가 충분하여 추가 가중 없이 분포 패턴이 안정적으로 학습되므로 TabDiff[1] 원본의 합성 품질이 보존된다. 소수 클래스 가중은 (1−t)에 비례하여 노이즈 복원 후반(t→0)에서 최대화된다.

γ는 약한 노이즈 영역에서의 소수 클래스 가중 최댓값(1+γ)을 결정하는 하이퍼파라미터로, γ=0이면 TSCW가 비활성화되어 다수·소수 가중이 모든 timestep에서 동일하고, γ가 커질수록 소수 클래스 손실 증폭이 강해진다. 이는 강한 노이즈 영역에서 분포 차이가 노이즈에 묻혀 클래스 식별이 무의미하므로 가중을 부여하지 않고, 패턴이 식별되는 약한 노이즈 영역에서는 소수 클래스 손실을 선택적으로 증폭하여 학습 효율을 높이는 방식이다.

4. 실험개요

4.1 데이터셋

본 연구는 클래스 불균형 환경에서의 합성 데이터 유용성을 검증하기 위해, (i) 금융 도메인의 테이블 데이터, (ii) 이진 분류 타깃, (iii) 불균형 환경 합성을 평가하기에 충분한 표본 수를 기준으로 카드연체·소득예측·계좌사기 도메인의 공개 데이터셋 3종을 선정하였다(표 2).

Default[14]는 2005년 대만 신용카드 고객 30,000명을 대상으로 다음 달 카드대금 연체 여부를 예측하는 데이터셋이고, Adult[15]는 1994년 미국 인구조사 기반 48,842명의 연소득 5만 달러 초과 여부를 예측하는 데이터셋이다. 해당 두 데이터셋은 TabDiff[1], TabSyn[5], TabDDPM[4] 벤치마크 데이터로, 원본 소수 비율(22.12%, 24.08%)을 3%로 조정하여 불균형 환경을 조성했다.

BAF[16]는 Feedzai가 NeurIPS 2022에 공개한 은행 계좌 개설 사기 탐지 데이터셋으로, 익명화된 실제 데이터에 차분 프라이버시를 적용한 합성 벤치마크이다. 본 연구는 전체 데이터 중 100,000 건을 사용하며, 자연 발생 사기 비율 1.10%를 원본 비율 그대로 사용한다.

4.2 평가지표

합성 데이터의 품질은 유용성(Utility), 분포 충실도(Fidelity), 데이터 프라이버시(Privacy) 세 축으로 평가하며, 지표는 소수 클래스와 전체를 분리 측정한다(표 3).

4.3 평가 방법

유용성은 합성 데이터로 분류기를 학습하고 실제 테스트셋으로 평가하는 TSTR 방식으로 측정하며, 비교의 일관성을 위해 분류기는 XGBoost[17] 단일 모델을 사용한다. 합성 데이터는 비율 조정 없이 각 모델이 생성한 비율 그대로 평가한다. 베이스라인 TabDiff[1]를 기준으로 TabSyn[5], TabDDPM[4]을 함께 비교 평가한다.

생성 단계에서 5개 시드로 모델을 학습하고(n=5), 유용성 지표는 생성 시드마다 XGBoost 평가를 30회 반복(n=30)한 평균을 산출한 뒤 5개 시드 간 평균·표준편차

표 3 평가지표

Table 3 Evaluation Metrics

Category	Metric	Description
Utility	Minority F1 ↑	Minority-class classification performance
	MCC ↑	Correlation coefficient from confusion matrix
Fidelity	JSD (mean) ↓	Categorical marginal distribution (overall)
	JSD (minority) ↓	Categorical marginal distribution (minority)
	KS (mean) ↓	Numerical empirical distribution (overall)
	KS (minority) ↓	Numerical empirical distribution (minority)
	Corr (overall) ↓	Inter-column dependency (overall)
	Corr (minority) ↓	Inter-column dependency (minority)
Privacy	MIA (overall) → 0.5	Membership inference risk (overall)
	MIA (minority) → 0.5	Membership inference risk (minority)

표 2 데이터셋 개요

Table 2 Overview of Datasets

Dataset	Domain	Rows	Feat (num/cat)	Minority
Default	Card-payment delinquency	30,000	23 (14/9)	3% (adj.)
Adult	Income prediction	48,842	14 (6/8)	3% (adj.)
BAF	Account fraud	100,000	29 (18/11)	1.10% (orig.)

를 보고한다. 충실도·프라이버시 지표는 결정론적으로 산출되므로 분류기 반복 없이 생성 시드당 1회 산출 후 5개 시드 간 평균을 보고한다.

통계적 유의성은 제안 모델과 베이스라인 간 paired Student's t-test(양측, 유의수준: 0.05)로 검증한다. 유용성(Utility) 지표는 값이 클수록 우수한 성능을 나타내며, 분포 충실도는 값이 작을수록 분포를 더 잘 반영한 것이다. 프라이버시는 특정 레코드의 학습 데이터 포함

표 4 데이터셋별 평가지표 결과 (Vanilla → Ours, n=5)
Table 4 Evaluation Results per Dataset (Vanilla → Ours, n=5)

Metric	Vanilla	Ours	Δ%	p
Default (card-payment delinquency, 3%)				
Minority F1 ↑	0.0292	**0.0662**	+126.7	<.001***
MCC ↑	0.0474	**0.0804**	+69.6	<.001***
JSD (mean) ↓	0.0270	0.0273	+1.1	.238
JSD (minority) ↓	0.1281	**0.1079**	−15.8	.005**
KS (mean) ↓	0.0144	0.0146	+1.4	.778
KS (minority) ↓	0.0527	**0.0426**	−19.2	.011*
Corr (overall) ↓	0.4472	0.4646	+3.9	.579
Corr (minority) ↓	1.8292	1.6433	−10.2	.066
MIA (overall)	0.5402	0.5392	—	.468
MIA (minority)	0.5315	0.5313	—	.961
Adult (income prediction, 3%)				
Minority F1 ↑	0.2620	**0.3224**	+23.1	.004**
MCC ↑	0.3140	**0.3439**	+9.5	.012*
JSD (mean) ↓	0.0338	0.0333	−1.5	.557
JSD (minority) ↓	0.1411	**0.1290**	−8.6	.034*
KS (mean) ↓	0.0073	0.0093	+27.4	.038*
KS (minority) ↓	0.0362	0.0359	−0.8	.871
Corr (overall) ↓	0.1308	0.1382	+5.7	.147
Corr (minority) ↓	0.5181	0.4710	−9.1	.112
MIA (overall)	0.5159	0.5173	—	.363
MIA (minority)	0.5369	0.5442	—	.308
BAF (account fraud, 1.10%)				
Minority F1 ↑	0.0432	**0.0572**	+32.4	.223
MCC ↑	0.0777	0.0954	+22.8	.284
JSD (mean) ↓	0.0314	0.0276	−12.1	.400
JSD (minority) ↓	0.1415	**0.1194**	−15.6	.003**
KS (mean) ↓	0.0125	0.0085	−32.0	.361
KS (minority) ↓	0.0674	0.0438	−35.0	.171
Corr (overall) ↓	0.6335	0.2336	−63.1	.358
Corr (minority) ↓	1.5482	1.1073	−28.5	.052
MIA (overall)	0.5015	0.5036	—	.433
MIA (minority)	0.4904	0.4586	—	.127

표 5 모델별 유용성(Minority F1) 비교 (Mean±std, n=5)
Table 5 Utility Comparison Across Models by Minority F1 (Mean±std, n=5)

Model	Default	Adult	BAF
Ours	0.0662±0.0142	0.3224±0.0188	0.0572±0.0174
TabDiff (Vanilla)	0.0292±0.0074	0.2620±0.0148	0.0432±0.0262
TabSyn	0.0159±0.0102	0.2564±0.0391	0.0149±0.0149
TabDDPM	0.0444±0.0133	0.2942±0.0261	0.0532±0.0295

여부를 추론하는 MIA(Membership Inference Attack)[18]로 측정하며, 거리 기반 black-box 공격[19]의 AUROC가 0.5(무작위 추측)에 가까울수록 우수하다.

5. 실험결과

실험결과 데이터셋별 채택 하이퍼파라미터는 각 모듈에 대한 그리드 서치를 통해 Default·Adult에 (ρ, α, γ)=(0.037, 0.01, 1.0), BAF에 (0.017, 0.01, 0.5)를 적용하였으며, 추가적으로 BAF는 기울기 폭발이 발생하여 그래디언트 클리핑(max_norm=0.7)을 적용하였다.

표 4, 5의 수치는 5개 시드의 평균이고, 별표는 양측 paired t-test의 유의수준이다(*p<.05, **p<.01, ***p<.001). 데이터셋별 평가지표 결과는 표 4에, 모델 간 비교는 표 5에 제시한다.

유용성 지표에서 Default와 Adult의 제안 모델은 Minority F1(+126.7%, p<.001, +23.1%, p<.01)과 MCC (+69.6%, p<.001, +9.5%, p<.05)를 통계적으로 유의한 수준에서 개선하였다. BAF 데이터셋에서는 소수 클래스 비율이 1.10%로 시드 간 분산이 커 유의성에는 이르지 못했으나, Minority F1이 +32.4% 향상되어 동일한 방향의 유용성 개선 성능을 보였다.

본 연구의 소수 클래스 보존 학습 파이프라인은 소수 클래스의 분포 충실도를 집중적으로 개선하는 효과를 보였다. 소수 클래스 범주형 분포 지표 JSD(minority)는 세 데이터셋 모두 유의하게 개선되었다(Default −15.8% p<.01, Adult −8.6% p<.05, BAF −15.6% p<.01). 수치형 분포 KS(minority)와 상관 구조 Corr (minority)는 Default KS(minority)를 제외하면 시드 간 분산으로 통계적 유의성에는 이르지 못했으나, 세 데이터셋 모두 악화 없이 개선되는 일관된 경향을 보였다. 반면 전체(mean/overall) 차원에서는 유의한 개선이 없었으며, Adult KS(mean)은 오히려 유의하게 악화되었다(p<.05). 이는 본 방법의 효과가 소수 클래스에 선택적으로 작용하며 전체 분포 차원에서는 제한적 trade-off

가 존재함을 시사한다.

프라이버시 측면에서 전체 MIA(overall)은 세 데이터셋 모두 0.5 근방을 유지하였다. 소수 클래스 MIA(minority)에서는 Adult와 BAF에서 0.5로부터의 이탈도가 다소 증가하였으나, 통계적으로 유의하지 않다. 표 5에서는 본 연구에서 제안하는 모델과 대표적인 디퓨전 모델을 비교하였고, 세 데이터셋 모두에서 가장 높은 Minority F1을 달성하였다.

6. 결론 및 한계점

6.1 결 론

본 연구는 TabDiff의 클래스 무관 순방향 노이즈 구조가 극단 불균형 환경에서 소수 클래스 신호를 훼손함을 확인하고, 샘플링·순방향·손실 세 단계에 CBBS·CC·TSCW를 결합한 계층적 프레임워크를 제안하였다. 손실·모델 구조 단계에 국한되었던 기존 클래스 인식 접근을 순방향 노이즈 주입 단계로 확장한 것이 연구의 주요 차별점이다.

Default(카드연체)·Adult(소득예측)·BAF(계좌사기) 세 데이터셋의 5개 시드 검증에서, 제안 방법은 소수 클래스 유용성을 일관되게 향상시키면서 소수 클래스의 분포 충실도를 함께 개선하였다. 이는 합성 데이터가 유용성뿐 아니라 원본 금융 데이터의 분포 및 상관 구조를 보존하는 경향을 보임을 의미한다. 또한 유용성 개선에 따른 소수 클래스 식별 위험의 통계적으로 유의한 증가는 발견되지 않았다.

이러한 개선은 서론에서 제기한 다운스트림 손실 문제의 완화로 직접 이어진다. Default(3%)에서 원본 학습 분류기(TRTR)의 Minority F1(0.076) 대비 합성 데이터로 학습한 분류기(TSTR)의 회복률은 Vanilla TabDiff가 38.42%에 그친 반면 제안 모델은 87.11%를 기록하여, 불균형 환경에서의 급격한 성능 저하를 상당 부분 회복하였다.

6.2 한계점 및 후속 연구

본 연구의 방법론적 측면과 데이터 실증 측면에서의 후속 연구를 제언한다. 먼저 하이퍼파라미터 탐색이 단일 시드 기반의 탐색적 그리드 서치에 그쳐, 향후 다중 시드 기반의 그리드 서치 또는 베이지안 최적화 등 체계적 탐색이 필요하다. 또한 순방향 노이즈 강도를 조절하는 CC(Class-Conditional Noise Schedule)는 세 데이터셋에서 α가 낮은 수준(0.01)으로 채택되어, 강한 노이즈 차단을 통한 직접적 보존보다는 클래스 정보를 학습 중반 구간으로 안내하는 방식으로 작동한 것으로 해석된다.

학습 안정화 차원에서는 BAF(계좌사기) 데이터셋에서 학습 과정의 비정상 수렴에 대해 그래디언트 클리핑을 도입하여 정상 학습을 확보하였으나 이는 기울기 폭발을 사후 억제한 것일 뿐 학습 불안정의 근본 원인을 제거한 것은 아니며, 향후 모델 차원에서의 안정적인 학습에 대한 연구가 추가적으로 필요하다. 유용성 분류 평가에서는 XGBoost 단일 분류기로 측정하여, 향후 다양한 분류기로 확장하는 것이 고려된다. 끝으로 데이터 실증 측면에서는 극단 불균형 환경의 금융 공개 데이터 확보가 제한적이었기 때문에 BAF(계좌사기, 1.10%)를 활용하였고, 실제 운용 환경의 사기 분포와 차이가 있을 수 있으므로, 극단적 불균형성을 지니는 실제 금융 데이터에서의 검증과 시드 확대를 통한 일반화 검증이 요구된다.

References

[1] J. Shi, M. Xu, H. Hua, H. Zhang, S. Ermon, and J. Leskovec, "TabDiff: A Mixed-type Diffusion Model for Tabular Data Generation," *Proc. of the International Conference on Learning Representations (ICLR)*, 2025.

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321-357, 2002.

[3] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling Tabular Data using Conditional GAN," *Proc. of the Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

[4] A. Kotelnikov, D. Baranchuk, I. Rubachev, and A. Babenko, "TabDDPM: Modelling Tabular Data with Diffusion Models," *Proc. of the International Conference on Machine Learning (ICML)*, 2023.

[5] H. Zhang, J. Zhang, B. Srinivasan, Z. Shen, X. Qin, C. Faloutsos, H. Rangwala, and G. Karypis, "Mixed-Type Tabular Data Synthesis with Score-based Diffusion in Latent Space," *Proc. of the International Conference on Learning Representations (ICLR)*, 2024.

[6] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, "Decoupling Representation and Classifier for Long-Tailed Recognition," *Proc. of the International Conference on Learning Representations (ICLR)*, 2020.

[7] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2999-3007, 2017.

[8] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-Balanced Loss Based on Effective Number of Samples," *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9268-9277, 2019.

[9] T. Hang, S. Gu, C. Li, J. Bao, D. Chen, H. Hu, X.

Geng, and B. Guo, "Efficient Diffusion Training via Min-SNR Weighting Strategy," *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV),* pp. 7441-7451, 2023.
[10] J. Choi, J. Lee, C. Shin, S. Kim, H. Kim, and S. Yoon, "Perception Prioritized Training of Diffusion Models," *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11472-11481, 2022.
[11] Y. Qin, H. Zheng, J. Yao, M. Zhou, and Y. Zhang, "Class-Balancing Diffusion Models," *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18434-18443, 2023.
[12] T. Sattarov, M. Schreyer, and D. Borth, "FinDiff: Diffusion Models for Financial Tabular Data Generation," *Proc. of the ACM International Conference on AI in Finance (ICAIF),* pp. 64-72, 2023.
[13] M. Schreyer, T. Sattarov, A. Sim, and K. Wu, "Imb-FinDiff: Conditional Diffusion Models for Class Imbalance Synthesis of Financial Tabular Data," *Proc. of the ACM International Conference on AI in Finance (ICAIF)*, pp. 617-625, 2024.
[14] I.-C. Yeh and C.-H. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Systems with Applications,* Vol. 36, No. 2, pp. 2473-2480, 2009.
[15] R. Kohavi, "Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid," *Proc. of the International Conference on Knowledge Discovery and Data Mining (KDD),* pp. 202-207, 1996.
[16] S. Jesus, J. Pombal, D. Alves, A. Cruz, P. Saleiro, R. P. Ribeiro, J. Gama, and P. Bizarro, "Turning the Tables: Biased, Imbalanced, Dynamic Tabular Datasets for ML Evaluation," *Proc. of the Conference on Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track,* 2022.
[17] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD),* pp. 785-794, 2016.
[18] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," *Proc. of the IEEE Symposium on Security and Privacy (S&P)*, pp. 3-18, 2017.
[19] D. Chen, N. Yu, Y. Zhang, and M. Fritz, "GAN-Leaks: A Taxonomy of Membership Inference Attacks against Generative Models," *Proc. of the ACM SIGSAC Conference on Computer and Communications Security (CCS),* pp. 343-362, 2020.

최 현 기
2026년 서강대학교 AI·SW대학원 데이터사이언스·인공지능(석사). 관심분야는 금융 데이터, 금융 AI, 머신러닝, 데이터 사이언스

김 영 재
2001년 서강대학교 컴퓨터공학과(학사) 2003년 한국과학기술원 컴퓨터공학과(석사) 2009년 Pennsylvania State University 컴퓨터과학 및 공학(박사). 2009년~2015년 미국 에너지부 산하 Oak Ridge National Laboratory 연구개발 책임연구원. 2016년~현재 서강대학교 컴퓨터공학과 교수. 관심분야는 인공지능 시스템, 머신러닝, 분산 및 병렬 컴퓨팅