

2025 年度 高等学院同窓会学術研究奨励金 研究成果報告書概要（WEB 公開用）

高等学院長
高等学院同窓会理事長 殿

研究代表者氏名 [蒲地 樹]

学年・組・番号 [2 年 F 組 22 番]

研究課題： 見えないバイアス：AI 画像分類精度におけるデータの不均衡と公平性介入の影響
(英文) Invisible Bias: The Impact of Data Imbalance and Fairness Interventions on AI Image Classification Performance

研究概要：

(研究課題を選んだ動機、達成するための計画・目的・方法等について 200～400 字で記入してください)

近年、AI システムは採用や医療、司法など重要な意思決定に利用されているが、学習データに含まれる特定の属性の偏り（不均衡）が、AI の判断に「見えないバイアス」を生じさせ、公平性を損なうリスクがある。しかし、開発現場ではこの不均衡が看過されることも多い。そこで私たちは、データの不均衡が AI の画像分類精度に与える悪影響を定量的に実証し、さらにそのバイアスを解消するための緩和手法（介入）の効果を比較検証することを目的とした。方法としては、手書き数字画像データセット「MNIST」を使用し、特定のクラスのデータ数を意図的に減らした 3 段階の不均衡比率（5:1、10:1、20:1）のデータセットを作成した。これに対し、決定木やニューラルネットワーク（MLP）など 7 種類の機械学習アルゴリズムを用いて、①介入なし、②オーバーサンプリング、③アンダーサンプリング、④SMOTE（合成データ生成）の 4 条件で学習を行い、正解率とマクロ平均 F1 スコアを用いて性能の変化を分析した。

研究成果：

(研究の結果概要、結果に対するフィードバックや感想等について 200～400 字で記入してください)

実験の結果、データの不均衡は 7 つ全てのアルゴリズムにおいて著しい性能低下を招くことが確認された。一方で、検証した 3 つの緩和手法はいずれも「介入なし」と比較して性能を向上させた。具体的には、SMOTE 法が広範に高い効果を示したが、不均衡が深刻な場合（20:1）や特定のモデル（ロジスティック回帰等）ではアンダーサンプリングの方が有効であるなど、万能な単一手法は存在せず、モデルと不均衡比率に応じた使い分けが不可欠であることが判明した。ただし、どの手法を用いても、元からデータ数が均等な「完全バランスデータ」の精度には及ばなかったため、AI 開発の初期段階におけるデータ収集の公平性が最重要であると結論付けた。また、この成果を基にハナ国際シンポジウムで発表を行った。

研究者：(以下の、代表者・分担者は学年・組・氏名を明記する)

研究代表者 2 年 F 組 22 番 蒲地 樹

研究分担者 2 年 J 組 17 番 松田 龍昌 1 年 A 組 39 番 堀瀬 善人

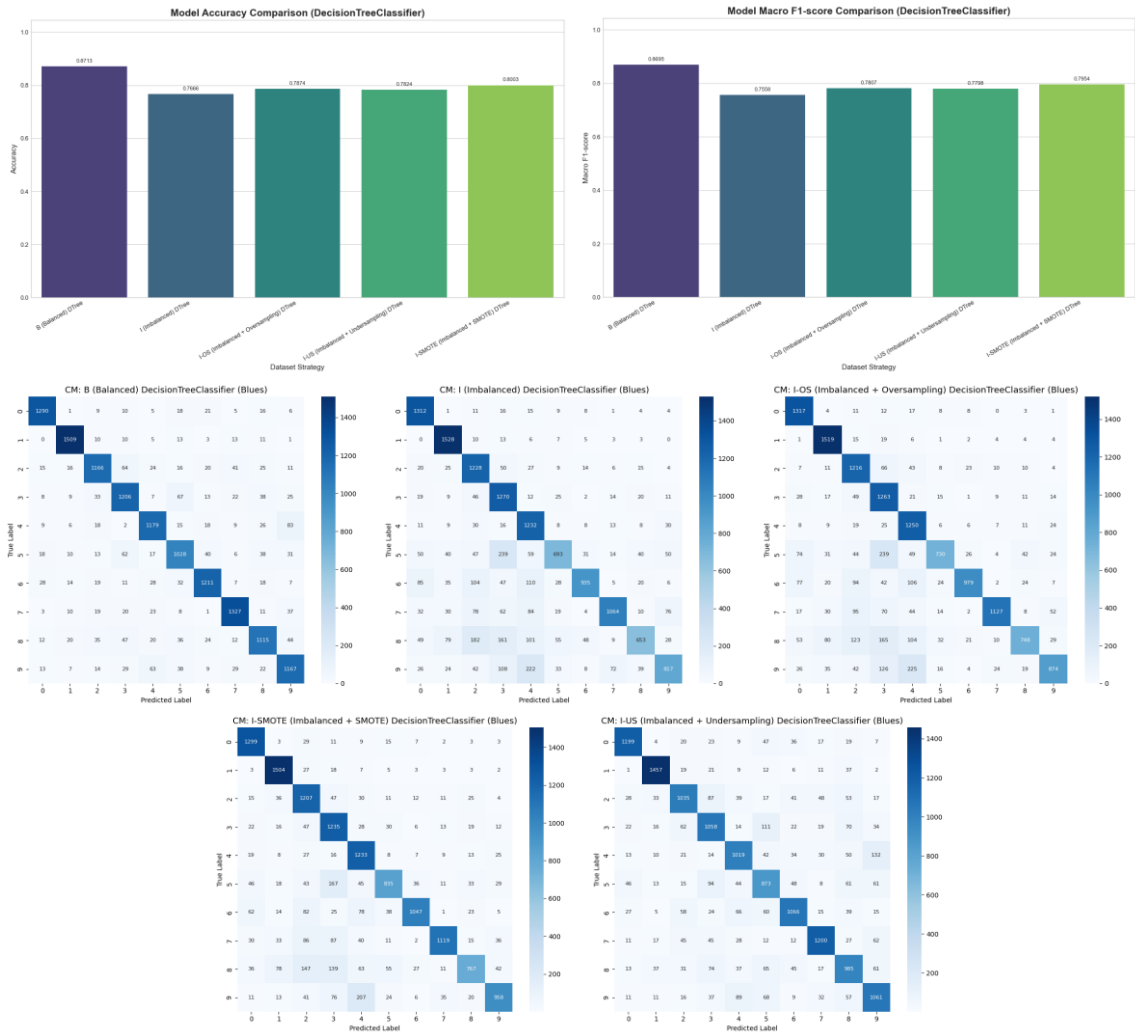
担当教諭 榎本隆之先生 (受給額： 23,000 円)

※研究課題、研究概要、研究成果、研究代表者名が WEB ページ上で公開されることに同意します

(次のページに続きます)

研究成果写真：
(研究過程がわかる写真や、研究結果がわかる写真などを数点貼り付けてください)

ケーススタディ：決定木分類の分析



実験結果：

Imbalance Ratio	Classifier	Scenario	Accuracy	Macro Avg F1-score
5:1	Decision Tree	Balanced	0.9359	0.8211
		Imbalanced	0.9377	0.8261
		Imbalanced + Oversampling	0.9368	0.8261
		Imbalanced + SMOTE	0.9357	0.8225
		Imbalanced + Undersampling	0.84	0.7056
	K-Nearest Neighbors	Balanced	0.978	0.9181
		Imbalanced	0.9789	0.9008
		Imbalanced + Oversampling	0.9731	0.9087
		Imbalanced + SMOTE	0.9671	0.9004
		Imbalanced + Undersampling	0.9409	0.848
	Logistic Regression	Balanced	0.9487	0.8593
		Imbalanced	0.9638	0.8565
		Imbalanced + Oversampling	0.9484	0.8471
		Imbalanced + SMOTE	0.95	0.8566
		Imbalanced + Undersampling	0.9322	0.8247
	MLP Classifier	Balanced	0.9738	0.9198
		Imbalanced	0.9751	0.9203
		Imbalanced + Oversampling	0.9771	0.9295
		Imbalanced + SMOTE	0.976	0.9258
		Imbalanced + Undersampling	0.9598	0.89
	Random Forest	Balanced	0.9711	0.8281
		Imbalanced	0.9731	0.8387
		Imbalanced + Oversampling	0.974	0.8155
		Imbalanced + SMOTE	0.9744	0.8053
		Imbalanced + Undersampling	0.9579	0.7855
	SGD Classifier	Balanced	0.8938	0.733
		Imbalanced	0.9238	0.751
		Imbalanced + Oversampling	0.8829	0.7122
		Imbalanced + SMOTE	0.8967	0.7356
		Imbalanced + Undersampling	0.8809	0.724
	SVC	Balanced	0.966	0.8466
		Imbalanced	0.9791	0.8829
		Imbalanced + Oversampling	0.9756	0.9187
		Imbalanced + SMOTE	0.9747	0.9148
		Imbalanced + Undersampling	0.9511	0.8632
10:1	DecisionTree	Balanced	0.8713	0.8695
		Imbalanced	0.7666	0.7558
		Imbalanced + Oversampling	0.7874	0.7807
		Imbalanced + Undersampling	0.7824	0.7798
		Imbalanced + SMOTE	0.8003	0.7954
	KNeighbors	Balanced	0.9431	0.9426
		Imbalanced	0.8579	0.8532
		Imbalanced + Oversampling	0.9092	0.9073
		Imbalanced + SMOTE	0.8904	0.8893
		Imbalanced + Undersampling	0.9214	0.9204
	LogisticRegression	Balanced	0.9152	0.9141
		Imbalanced	0.8441	0.8393
		Imbalanced + Oversampling	0.8613	0.8577
		Imbalanced + Undersampling	0.8818	0.8801
		Imbalanced + SMOTE	0.8567	0.8529
	MLPClassifier	Balanced	0.9614	0.9612
		Imbalanced	0.9346	0.9336
		Imbalanced + Oversampling	0.9383	0.9373
		Imbalanced + Undersampling	0.9295	0.9287
		Imbalanced + SMOTE	0.9371	0.9361
	RandomForest	Balanced	0.967	0.9667
		Imbalanced	0.8697	0.8659
		Imbalanced + Oversampling	0.9054	0.9038
		Imbalanced + Undersampling	0.9358	0.935
		Imbalanced + SMOTE	0.9236	0.9223
	SGDClassifier	Balanced	0.8962	0.898
		Imbalanced	0.8259	0.8226
		Imbalanced + Oversampling	0.8971	0.896
		Imbalanced + Undersampling	0.892	0.8908
		Imbalanced + SMOTE	0.8969	0.8953
	SVC	Balanced	0.9631	0.9629
		Imbalanced	0.9064	0.9053
		Imbalanced + Oversampling	0.9314	0.931
		Imbalanced + Undersampling	0.9244	0.924
		Imbalanced + SMOTE	0.9299	0.9295

Imbalance Ratio	Classifier	Scenario	Accuracy	Macro Avg F1-score
20:1	Decision Tree	Balanced	0.9622	0.7308
		Imbalanced	0.9638	0.7501
		Imbalanced + Oversampling	0.9636	0.7478
		Imbalanced + SMOTE	0.9616	0.7274
	K-Nearest Neighbors	Imbalanced + Undersampling	0.8756	0.6523
		Balanced	0.9749	0.8592
		Imbalanced	0.9771	0.7344
		Imbalanced + Oversampling	0.9724	0.8397
	Logistic Regression	Imbalanced + SMOTE	0.9624	0.8077
		Imbalanced + Undersampling	0.9347	0.7207
		Balanced	0.95	0.7545
		Imbalanced	0.9762	0.7076
	MLP Classifier	Imbalanced + Oversampling	0.9484	0.7513
		Imbalanced + SMOTE	0.9522	0.7665
		Imbalanced + Undersampling	0.9284	0.703
		Balanced	0.9684	0.8096
10:1	Random Forest	Imbalanced	0.9767	0.7826
		Imbalanced + Oversampling	0.9753	0.8122
		Imbalanced + SMOTE	0.972	0.8142
		Imbalanced + Undersampling	0.95	0.7663
	SGD Classifier	Balanced	0.9771	0.7556
		Imbalanced	0.978	0.7195
		Imbalanced + Oversampling	0.9778	0.713
		Imbalanced + SMOTE	0.978	0.7228
	SVC	Imbalanced + Undersampling	0.9696	0.7139
		Balanced	0.8967	0.6429
		Imbalanced	0.9709	0.506
		Imbalanced + Oversampling	0.8758	0.6
	Decision Tree Classifier	Imbalanced + SMOTE	0.9013	0.6503
		Imbalanced + Undersampling	0.8933	0.651
		Balanced	0.9651	0.815
		Imbalanced	0.9802	0.495
20:1	SVC	Imbalanced + Oversampling	0.9664	0.8231
		Imbalanced + SMOTE	0.9638	0.8111
		Imbalanced + Undersampling	0.9358	0.7476
		Balanced	0.9644	0.8231

分析の結果：

Dataset	Algorithm	Best Intervention	Best Intervention vs. Balanced	Best Intervention vs. Imbalanced
5:1	Decision Tree Classifier	SMOTE	Worse	Better
5:1	K-Nearest Neighbors Classifier	SMOTE	Worse	Better
5:1	Logistic Regression	IOS	Worse	Better
5:1	MLP Classifier	IOS	Worse	Better
5:1	Random Forest Classifier	IUS	Worse	Better
5:1	SGD Classifier	SMOTE	Better	Better
5:1	SVC	IOS	Worse	Better
10:1	Decision Tree Classifier	SMOTE	Worse	Better
10:1	K-Nearest Neighbors Classifier	SMOTE	Worse	Better
10:1	Logistic Regression	IUS	Worse	Better
10:1	MLP Classifier	IOS	Worse	Better
10:1	Random Forest Classifier	IUS	Worse	Better
10:1	SGD Classifier	SMOTE	Worse	Better
10:1	SVC	IUS	Worse	Better
20:1	Decision Tree Classifier	SMOTE	Worse	Better
20:1	K-Nearest Neighbors Classifier	SMOTE	Worse	Better
20:1	Logistic Regression	IUS	Worse	Better
20:1	MLP Classifier	IOS	Worse	Better
20:1	Random Forest Classifier	IUS	Worse	Better
20:1	SGD Classifier	SMOTE	Worse	Better
20:1	SVC	IUS	Worse	Better

ハナ国際シンポジウム：



すべてのグラフファイル (Google Drive) :

URL: https://drive.google.com/drive/folders/1pcubsRM-FOLz1YiUDjnBA2MNfi9a2b6v?usp=drive_link



体験ウェブサイト (ロードに時間がかかる可能性があります) :

URL: <https://digit-recognizer-iuh7.onrender.com/>



Digit Recognition

Beta Version(MLPClassifier_Results_target5000)

Waseda University Senior High School Team_Bias

[▶ About the Experiment & Models](#)

Draw your digit clearly within the guide box for best results. Adjust line width below.



Model Input (28x28)


Line Width: 3.0

Predict All

Clear

☐ Show Detailed Probabilities

model_B:

Prediction: 4 (Confidence: 56.6%)

model_I:

Prediction: 4 (Confidence: 86.0%)

model_I-OS:

Prediction: 4 (Confidence: 85.9%)

model_I-US:

Prediction: 4 (Confidence: 51.7%)

model_I-SMOTE:

Prediction: 4 (Confidence: 86.4%)

Processing time: 2.71 ms

以上