

# Affinity Loss を用いた 深層学習による不均衡テキストデータ分類

川崎雄大<sup>†</sup> 櫻井義尚<sup>‡</sup>明治大学 先端数理科学研究科<sup>†</sup> 明治大学 総合数理学部<sup>‡</sup>

## 1. 研究背景

近年、機械学習の応用分野は多岐に渡り、画像データやテキストデータの分類はその代表的な応用例である。しかし、実際に機械学習を応用する際には十全なデータを得られないことは多く、分類タスクにおける不均衡データ問題もその一つである。

不均衡データへの対策は、不均衡を是正するデータのサンプリング手法や Cost-Sensitive Learning[1]などの様々な手法が提案されている。その中で、Hayat らは深層学習の損失関数に不均衡データを分類するための正則化項を加えた Affinity Loss を提案している[2]。

しかし、不均衡データ対策は特定のデータセットを対象としていることが多く、データの特性によって効果は左右されることが確認されている[3]。Hayat らの提案手法は特定の画像データセットに対してのみ性能が確認されており、本稿では Affinity Loss をテキストデータの不均衡分類に応用し、その性能とふるまいを確かめた。

## 2. Affinity Loss

本章では、深層学習の不均衡データ対策手法である Affinity Loss について説明する[2]。

Affinity Loss は、各クラスの特徴量空間上でのマージンを最大化する損失関数 $L_{mm}$ と、分類におけるデータ数不均衡を是正する正則化項 $R(w)$ から成る。

マージン最大化の働きをもつ損失関数 $L_{mm}$ ：

損失関数 $L_{mm}$ を説明するためにまず、あるデータ $i$ のクラス $j$ との類似度指標として  $d(f, w)$  を定義する。

$$d(f_i, w_j) = \exp\left(-\frac{\|f_i - w_j\|^2}{\sigma}\right) \quad (1)$$

ここで、 $f_i$ は $i$ 番目のデータを順方向計算した後の特徴量、 $w_j$ は $j$ 番目のクラスのベクトル、 $\sigma$ はハイパーパラメータとなる。

この類似度 $d(f_i, w_j)$ を用いて、損失関数 $L_{mm}$ は以下のように定義される。

$$L_{mm} = \sum_j \max(0, \lambda + d(f_i, w_j) - d(f_i, w_{y_i})) \quad : j \neq y_i \quad (2)$$

ここで $y_i$ は正解クラスを意味し、 $L_{mm}$ は正解クラスとの類似度に対する、他クラスとの類似度の大きさとなる。また、 $\lambda$ はクラス間のマージンを決定するハイパーパラメータとなる。

分類領域を均一にする正則化項  $R(w)$ ：

前述した損失関数に加えて、データ数の不均衡によるクラスの分類領域の不均衡を補正するための正則化項 $R(w)$ を以下のように定義する。

$$R(w) = \mathbb{E}\left[\left(\|w_j - w_k\|^2 - \mu^2\right)\right], s.t. j < k \quad (3)$$

$$\mu = \frac{2}{C^2 - C} \sum_{j < k} \|w_j - w_k\|^2 \quad (4)$$

ここで、 $C$ はクラス数であり、 $\mu$ は各クラスの重心間の距離の平均を表す。

上記を踏まえて、Affinity Loss は以下のように定義される。

$$L = L_{mm} + R(w) \quad (5)$$

## 3. 実験

テキストデータの分類問題において、Affinity Loss のふるまいを確かめる実験を行った。実験条件は以下の通りである。

データ： 分類に使うデータには深層学習フレームワーク Keras の datasets.imdb クラスより、IMDB 映画レビューデータセットを用いた[4]。訓練データ・テストデータがそれぞれ 25,000 件ずつ存在し、教師ラベルとしてレビューのポジティブ(1)、ネガティブ(0)のクラスが付与されたデータが 12,500 件ずつ存在する。今回は、訓練データ内のポジティブクラスデ

Imbalanced text data classification by deep learning using Affinity Loss

<sup>†</sup> Yuta Kawasaki, Meiji University

<sup>‡</sup> Yoshitaka Sakurai, Meiji University

ータをサンプリングして， [12,500(100%), 9,375(75%), 6,250(50%), 3,125(25%), 1,250(10%)] 件の不均衡データを作成した。

**深層学習モデル：** 分類を行う深層学習モデルには，畳み込みニューラルネットワーク (CNN) を用いた [5]。構成は入力層，埋め込み層，畳み込み層，プーリング層，全結合層，出力層とした。Affinity Loss を用いると明記した場合以外は，損失関数には Cross-Entropy(CE) を用いた。

**比較手法：** 不均衡データへの対策手法として，次のものを比較した。Over-sampling, Under-sampling はデータレベル，Cost-Sensitive Learning, Affinity Loss はアルゴリズムレベルでの不均衡対策となる。

**Baseline：** 不均衡データ対策を用いないモデル。

**Over-sampling：** 少数クラスのデータ数が多数クラスと等しくなるように，少数クラスを重複ありでサンプリングする。

**Under-sampling：** 多数クラスのデータ数が少数クラスと等しくなるように，多数クラスをサンプリングする。

**Cost-Sensitive Learning(CoSen)：** 少数クラスの Loss 計算時に，Loss の値を{(最多クラスのデータ数)/(正解クラスのデータ数)}倍する。

**Affinity Loss：** 今回検証する手法。ハイパーパラメータは  $\sigma = 50$ ,  $\lambda = 0.1 \sim 0.75$  とした。

**評価指標：** 評価にはテストデータに対する精度(Accuracy)，適合率(Precision)，再現率(Recall)，適合率と再現率の調和平均(F1-score)を用いた。結果には，異なる乱数でサンプリング・訓練した平均値を使用している。

#### 4. 結果

実験の結果，それぞれのデータ不均衡率の際の F1-score は以下ようになった。

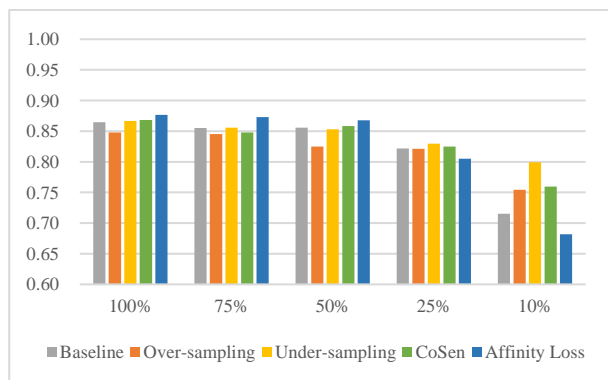


図1 ポジティブクラス数ごとの F1-score 比較

また，今回実験を行った中で最も不均衡によるポジティブクラスとネガティブクラスの差が大きい10%の際の各指標は以下ようになった。

表1 ポジティブクラス数 10%における各指標

Methods	Accuracy	Precision	Recall	F1-score
Baseline	0.757	0.862	0.611	0.715
Over-sampling	0.782	0.864	0.671	0.754
Under-sampling	<b>0.793</b>	0.779	<b>0.828</b>	<b>0.799</b>
CoSen	0.784	0.856	0.682	0.759
Affinity Loss	0.744	<b>0.931</b>	0.529	0.672

#### 5. まとめ

結果から，今回の実験条件ではクラス不均衡の割合が大きい際に Affinity Loss が性能を発揮できないことがわかった。

原因としては，データのクラス数が2クラスと少ないことや，Under-sampling の性能が高かったことから少ないサンプル数でも分類精度が大きく変化しない分類タスクであったことが考えられる。

今後は，異なるデータセットや深層学習モデルに対しても実験を行い，データやモデルの特性から効果的な不均衡データ対策手法を選択する基準を明確化することを目指す。

#### 参考文献

- [1] C. Elkan, “The Foundations of Cost-Sensitive Learning,” Proc. Int’l Joint Conf. Artificial Intelligence, pp. 973-978, 2001.
- [2] Munawar Hayat, Salman Khan, Waqas Zamir, Jianbing Shen and Ling Shao, “Max-margin Class Imbalanced Learning with Gaussian Affinity” URL : <https://arxiv.org/abs/1901.07711> (2019)
- [3] He, Haibo, and Edwardo A. Garcia. “Learning from imbalanced data.” IEEE Transactions on knowledge and data engineering, pp. 1263-1284, 2009.
- [4] Keras Documentation. 「IMDB 映画レビュー感情分類」 URL : <https://keras.io/ja/datasets/#imdb> (2020.1 参照)
- [5] Yoon Kim, “Convolutional Neural Networks for Sentence Classification”, Conference on Empirical Methods in Natural Language Processing, pp.1746-1751, 2014.