

<https://doi.org/10.7236/JIIBC.2024.24.1.49>

JIIBC 2024-1-8

불균형 블랙박스 동영상 데이터에서 충돌 상황의 다중 분류를 위한 손실 함수 비교

Comparison of Loss Function for Multi-Class Classification of Collision Events in Imbalanced Black-Box Video Data

이의상*, 한석민**

Euisang Lee*, Seokmin Han**

요약 데이터 불균형은 분류 문제에서 흔히 마주치는 문제로, 데이터셋 내의 클래스간 샘플 수의 현저한 차이에서 기인한다. 이러한 데이터 불균형은 일반적으로 분류 모델에서 과적합, 과소적합, 성능 지표의 오해 등의 문제를 야기한다. 이를 해결하기 위한 방법으로는 Resampling, Augmentation, 규제 기법, 손실 함수 조정 등이 있다. 본 논문에서는 손실 함수 조정에 대해 다루며 특히, 불균형 문제를 가진 Multi-Class 블랙박스 동영상 데이터에서 여러 구성의 손실 함수(Cross Entropy, Balanced Cross Entropy, 두 가지 Focal Loss 설정: $\alpha = 1$ 및 $\alpha = \text{Balanced}$, Asymmetric Loss)의 성능을 I3D, R3D_18 모델을 활용하여 비교하였다.

Abstract Data imbalance is a common issue encountered in classification problems, stemming from a significant disparity in the number of samples between classes within the dataset. Such data imbalance typically leads to problems in classification models, including overfitting, underfitting, and misinterpretation of performance metrics. Methods to address this issue include resampling, augmentation, regularization techniques, and adjustment of loss functions. In this paper, we focus on loss function adjustment, particularly comparing the performance of various configurations of loss functions (Cross Entropy, Balanced Cross Entropy, two settings of Focal Loss: $\alpha = 1$ and $\alpha = \text{Balanced}$, Asymmetric Loss) on Multi-Class black-box video data with imbalance issues. The comparison is conducted using the I3D, and R3D_18 models.

Key Words : deep learning, data imbalance, loss function, metric, CNN.

1. 서론

데이터 불균형은 데이터셋 내의 각 클래스 샘플의 수가 현저하게 다른 경우를 의미한다. 특히 자동차 블랙박

스 영상 데이터의 경우 충돌 사건 관련 영상 데이터가 일반 주행 사건에 비해 상당히 부족한 경향이 있다. 이러한 경우, 충돌 사건 데이터는 소수 클래스로 분류되며, 일반 주행 사건 데이터는 다수 클래스로 분류된다. 또 다른 예

*정회원, 한국교통대학교 철도차량시스템공학과

**정회원, 한국교통대학교 데이터사이언스전공

접수일자 2023년 12월 27일, 수정완료 2024년 1월 27일

계재확정일자 2024년 2월 9일

Received: 27 December, 2023 / Revised: 27 January, 2024 /

Accepted: 9 February, 2024

*Corresponding Author: seokmin.han@ut.ac.kr

Dept. of Data Science, Korea National University of Transportation, Korea.

시로 암 스크리닝^[1, 2] 문제에서 종양이 있다는 판정을 받은 환자의 수가 종양이 없다는 판정을 받은 환자의 수에 비해 현저히 적으므로, 종양이 있는 경우가 소수 클래스, 종양이 없는 경우가 다수 클래스가 된다. 두 예시를 통해 볼 때, 대다수 불균형 데이터셋에서 소수 클래스가 연구 및 분석의 주요 관심 대상이 되는 경우가 많다. 하지만 데이터 불균형이 존재할 경우 모델이 다수 클래스의 특징을 과도하게 학습하여, 소수 클래스의 중요한 패턴이나 특징을 간과하게 될 위험이 존재한다. 즉 다수 클래스에 대해 과적합, 소수 클래스에 대해 과소적합이 발생한다는 의미이다. 또한 불균형 데이터에서는 Accuracy와 같은 전통적인 성능 지표가 오도될 수 있다. 예를 들어 다수 클래스가 전체의 90%를 차지한다면, 모델이 다수 클래스만 정확하게 예측하여도 90%의 Accuracy를 달성하기에 모델의 객관적 성능을 판단하는 데 어려움을 야기한다. 이러한 데이터 불균형을 해결하기 위한 대책은, 소수 클래스를 추가 수집하여 다수 클래스와의 균형을 맞추는 것이다. 하지만 현실 세계에서 소수 클래스 데이터를 추가 수집하는 것은 여러 가지 제약이 존재한다. 이를 위한 차선책으로는 Oversampling, Undersampling과 같은 Resampling 기법^[3], Augmentation^[4], 규제 기법^[5], 손실 함수 조정^[6] 등^[7]이 있다.

본 논문에서는 데이터 불균형 문제를 가진 Multi-Class 블랙박스 약 2500개의 train 데이터와 1800개의 test 데이터를 기반으로 복수의 손실 함수들(Cross Entropy, Balanced Cross Entropy, 두 가지 Focal Loss 설정: $\alpha = 1$ 및 $\alpha = \text{Balanced}$, Asymmetric Loss)의 성능을 비교하였다.

II. 본 론

1. Dataset설명

본 논문에서 사용한 데이터는 데이콘(dacon.io)^[8]에서 제공한 것을 활용하였다. 데이터셋의 클래스는 총 3개로 구성되어 있으며, 0번 클래스는 어떠한 충돌도 일어나지 않은 사건을 나타낸다. 1번 클래스는 주체 차량이 충돌에 관련된 사건을 의미하며, 2번 클래스는 주체 차량을 제외한 다른 차량들 사이에서 충돌이 발생하는 사건을 의미한다. 또한 데이터셋은 2482개의 train 데이터와 1800개의 test 데이터로 구성되어 있으며, train 데이터의 경우 0번 클래스가 1596개, 1번 클래스가 489개, 2번 클래스가 424개로 데이터 불균형 문제를 가지고

있다. 본 연구에서는 224x224로 크기를 조절하였다. 그림 1에는 사용된 데이터의 예시가 나타나있다. 그림 2에는, 각 클래스별 비율을 알아보기 쉽도록 원형그래프로 표시하였다.



그림 1. 충돌클래스의 예시
Fig. 1. Examples of Collision Classes

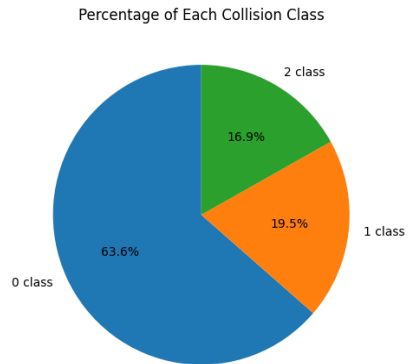


그림 2. train 데이터에서 각 클래스의 비율
Fig. 2. Proportion of Each Class in the Training Data

2. 손실 함수 관련 연구

본 절에서는 데이터 불균형 해소를 위해 개발된 손실 함수를 다룬다. 손실 함수는 머신러닝이나 딥러닝에서 모델의 예측이 목표 정답으로부터 얼마나 벗어나 있는지

를 측정하는 함수이다. 모델이 완벽하게 예측할수록 손실 함수의 값은 낮아지며, 부정확하게 예측할수록 값은 높아진다. 학습 과정에서 모델은 경사 하강법(SGD)^[9]과 같은 Optimizer를 사용하여 이 손실 함수의 값을 최소화하는 방향으로 가중치를 탐색한다. 불균형 데이터를 가지고 있는 분류 문제의 경우 Cross Entropy 같은 전통적인 손실 함수는 클래스의 불균형을 고려하지 않기에 편향된 학습을 유도할 위험이 있다.

이러한 문제를 해결하기 위해 각 클래스에 가중치를 주거나, 모델이 잘못 분류하는 샘플에 더 많은 관심을 기울이게 하는 focusing parameter를 활용하는 손실 함수를 사용하기도 한다. 다음은 본 연구에서 비교한 손실 함수로, 전통적인 함수부터 불균형 데이터 해소를 위해 개발된 함수에 대한 설명이다.

가. Cross Entropy Loss

Cross Entropy Loss(CE)는 분류 문제에서 가장 기본이며 핵심이 되는 손실 함수이다.

$$L = - \sum_i y_i \log(p_i) \dots (1)$$

해당 수식에서 y_i 는 실제 레이블, p_i 는 모델이 예측한 확률을 의미한다. 수식이 의미하는 바는 실제 레이블 y_i 와 예측 확률 p_i 사이의 불일치를 측정하는 것이다. 즉 실제 레이블과 예측 레이블이 같을수록 함수의 값이 낮아진다. CE의 경우 불균형을 고려하지 않고 각 클래스 샘플의 오류를 단순 합산하기에 다수 클래스가 모델 학습에 더 큰 영향을 미치며, 소수 클래스의 중요성은 상대적으로 무시되는 경향이 있다.

나. Balanced Cross Entropy Loss

Balanced Cross Entropy Loss(BCE)는 CE를 확장한 개념으로 2.가 절에서 언급한 문제점을 해결하기 위해 가중치를 부여한 손실 함수이다.

$$L = - \sum_i w_i [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \dots (2)$$

w_i 는 가중치를 의미한다. BCE는 소수 클래스에 더 높은 가중치를 할당함으로써, 소수 클래스 예측 오류에 더 큰 패널티를 부여한다. 이는 모델이 소수 클래스를 더 잘 인식하고 학습할 수 있도록 돕는다. 본 연구에선 w_i 를 클래스 빈도 가중치로 설정하여 진행하였다.

다. Focal Loss

Focal Loss(FL)^[10]는 BCE를 확장한 개념의 손실 함수이다.

$$L = - \sum_i \alpha_i (1 - p_i)^\gamma y_i \log(p_i) \dots (3)$$

α_i 는 가중치를 의미하며 γ 는 손실을 조정하는 focusing parameter이다. γ 의 역할은 쉬운 샘플은 어려운 샘플에 비해 손실을 크게 감소시키고, 어려운 샘플은 쉬운 샘플에 비해 손실을 적게 감소시킴으로써 모델이 어려운 샘플에 더 집중하게 만드는 파라미터이다.

예를 들어, 쉬운 샘플은 모델의 예측 확률이 높기에 높은 p_i 를 가지게 된다. 이로 인해 $(1 - p_i)$ 는 작은 값을 가지게 되고 $(1 - p_i)^\gamma$ 는 더욱 작은 값을 가지게 된다. 반면 어려운 샘플은 낮은 p_i 로 인해 쉬운 샘플보다 높은 $(1 - p_i)$ 를 가지게 되며, $(1 - p_i)^\gamma$ 가 쉬운 샘플에 비해 크게 감소하지 않아 어려운 샘플이 상대적으로 큰 손실을 가지게 된다. 이러한 메커니즘으로 FL은 불균형 데이터에서 유용하게 동작한다. 본 연구에선 γ 는^[8]에서 제안한 2로 설정하고 α_i 값을 1과 클래스 빈도 가중치(Balanced weight)로 설정하여(BFL) 비교하였다.

라. Asymmetric Loss

Asymmetric Loss(ASL)는 FL을 확장한 개념의 손실 함수이다. FL의 경우 γ 를 너무 높게 설정하면 쉬운 샘플의 기여도를 충분히 낮추지만, 동시에 어려운 샘플의 Gradient 또한 사라질 수 있는 문제가 있다. 이에 따라 각 양성 샘플과 음성 샘플의 focusing parameter를 분리하여 정의하는 것이 ASL의 주요 개념이다.^[11]

$$L = - \sum_i [y_i (1 - p_i)^{\gamma_{pos}} \log(p_i) + (1 - y_i) p_i^{\gamma_{neg}} \log(1 - p_i)] \dots (4)$$

$\gamma_{pos}, \gamma_{neg}$ 는 각각 양성 샘플과 음성 샘플의 focusing parameter이다.

각각의 focusing parameter를 분리함으로 세밀한 조정이 가능해 클래스 불균형을 더 효과적으로 다룰 수 있다. 본 연구에선 $(\gamma_{pos}, \gamma_{neg})$ 는^[11]에서 제안한 0과 4로 설정하여 진행했다.

3. 평가 지표

서론에서 언급하였듯 불균형 데이터의 문제 중 하나는

성능 지표의 해석에 오해를 야기하는 것이다. 이러한 이유로 불균형 데이터에서는 평가 지표를 선택할 때 주의를 기울여야 한다.^[12]

III. 실험 및 결과

1. F1 score 비교

본 논문에서는, 3D image를 처리할 수 있는 I3D모델과 R3D_18모델^[13, 14, 15]을 이용하였다. 본 연구에 사용된 데이터에 대한 모델별, 손실함수별 판별성능을 표 1에 정리하여 비교하였다.

표 1. F1 score 비교

Table 1. Comparison of F1 Scores

	CE[%]	BCE[%]	FL[%]	BFL[%]	ASL[%]
I3D	82	81	79	79	81
R3D_18	80	81	77	74	80

실험 결과, I3D와 R3D_18^[13, 14, 15] 모델에 대한 평가에서는 FL을 사용했을 때 상대적으로 낮은 성능을 보였다. 이는 FL의 $\alpha = \text{Balanced}$ 로 설정했을 때도 마찬가지였다. 반면 CE, BCE, ASL을 적용하였을 때는 두 모델 모두 유사한 수준의 성능을 나타냈다. 각 모델에 대한 PR curve는 그림 3, 그림 4에 그래프로 표시하였다.

2. 모델별, 클래스별 PR curve 비교

I3D 모델의 경우 다수 클래스(0번)에 대해 모든 손실 함수가 준수한 성능을 보였다. 소수 클래스(1번, 2번)에서는 뜻밖에도 불균형 해소를 목적으로 개발된 손실 함수가 아닌 CE가 가장 좋은 성능을 나타냈으며, 1번 클래스에서 FL($\alpha = \text{Balanced}$), 2번 클래스에서는 ASL이 CE 다음의 우수한 성능을 보였다.

R3D_18 모델에서는 I3D 사례와 동일하게 다수 클래스(0번)에서 모든 손실 함수가 준수한 성능을 보였다.

반면 소수 클래스(1번, 2번)에서 모두 BCE가 다른 손실 함수에 비해 좋은 성능을 보였으며, 1번 클래스에서는 ASL이, 2번 클래스에서는 CE가 그 뒤를 이어 좋은 성능을 보였다.

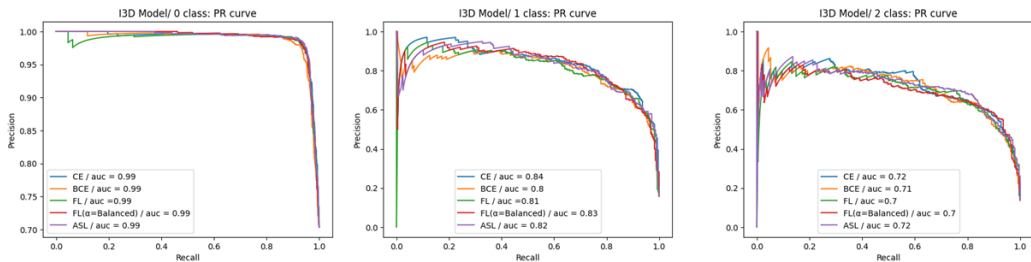


그림 3. I3D모델 PR curve 클래스의 비율

Fig. 3. Proportion of Classes in the PR Curve of the I3D Model

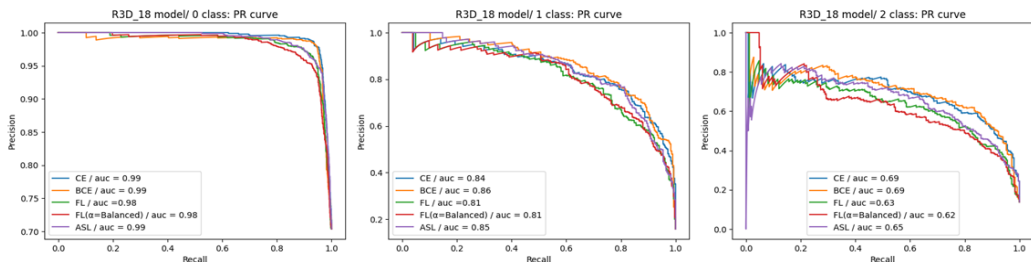


그림 4. R3D_18모델 PR curve 클래스의 비율

Fig. 4. Proportion of Classes in the PR Curve of the R3D_18 Model

IV. 결 론

본 연구는 복수의 손실 함수들을 I3D와 R3D_18 모델에 적용하여 비교 분석하였다. 분석 결과 BCE, ASL 그리고 예상 외로 CE가 F1 score, PR curve, AUC 지표에서 일관되게 우수한 성능을 보이는 것으로 나타났다. 반면 FL은 분석된 사례에서 다른 손실 함수들에 비해 떨어지는 성능을 보였다.

이 결과는 모든 손실 함수가 클래스 불균형을 가진 데이터셋에서 동등하게 효과적이지 않다는 사실을 강조한다. 더 나아가, 특정 손실 함수가 특정한 모델 구조나 데이터셋 특성에 맞추어 조정되어야 할 필요성을 보여준다. 따라서, 실제 응용 프로그램에 적합한 손실 함수를 선정할 때, 모델 아키텍처와 데이터 특성에 대한 심도 있는 고려가 필수적이다. 특히 FL의 상대적으로 낮은 성능에 대한 근본적인 이유를 파악하고 개선 방안을 모색하는 것이 앞으로의 연구에서 중요한 과제로 남는다. 마지막으로, 이 연구의 결과는 특정 조건 및 데이터셋에서 얻어진 것으로, 모든 상황에 보편적으로 적용될 수 있다고 보기에는 제한적이다. 또한, 하이퍼파라미터의 조정이 각 손실 함수의 성능에 미치는 영향을 고려할 때, 앞으로의 연구는 이러한 변수의 최적화에 더욱 집중해야 할 것이다. 손실 함수의 특성을 극대화하고, 모델 성능에 결정적인 영향을 미치는 하이퍼파라미터의 세심한 조정은, 특히 불균형 데이터셋을 다루는 데 있어서, 보다 정밀하고 효과적인 방법론을 제공할 것으로 기대된다.

References

- [1] L. Shen, et al, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," Scientific Reports, 2019.
DOI:<https://doi.org/10.1038/s41598-019-48995-4>
- [2] Y-G. Jung, et al, "Effective Diagnostic Method Of Breast Cancer Data Using Decision Tree," JIIBC, Vol. 10, No.5, pp.57-62, 2010.
- [3] R. Mohammed, et al, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," ICICS, 2020.
DOI:<https://doi.org/10.1109/ICICS49469.2020.239556>
- [4] M. Xu, et al, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," Pattern Recognition, Vol. 137, 2023.
DOI:<https://doi.org/10.1016/j.patcog.2023.109347>
- [5] C. Wang, et al, "Adaptive ensemble of classifiers with regularization for imbalanced data classification," Information Fusion, Vol. 69, pp.81-102, 2021.
DOI:<https://doi.org/10.1016/j.inffus.2020.10.017>
- [6] X. You, et al, "A study on loss function against data imbalance in deep learning correction of precipitation forecasts," Atmospheric Research, Vol. 281, 2023.
DOI:<https://doi.org/10.1016/j.atmosres.2022.106500>
- [7] J-H. Park, et al, "Improving Fashion Style Classification Accuracy using VAE in Class Imbalance Problem," The Journal of KIIT, Vol. 19, No. 2, pp. 1-10, Feb 2021.
DOI:<https://doi.org/10.14801/jkiit.2021.19.2.1>
- [8] <https://dacon.io/>
- [9] M-J. Kang, "Comparison of Gradient Descent for Deep Learning," Journal of the Korea Academia-Industrial cooperation Society (JKAIS), Vol. 21, No.2, pp. 189-194, Feb 2020
DOI:<https://doi.org/10.5762/KAIS.2020.21.2.189>
- [10] T. Lin, et al, "Focal Loss for Dense Object Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, pp.318-327, 2018.
DOI:<https://doi.org/10.1109/TPAMI.2018.2858826>
- [11] T. Ridnik, et al, "Asymmetric Loss For Multi-Label Classification," ICCV, 2021.
DOI:<https://doi.org/10.1109/ICCV48922.2021.00015>
- [12] J. M.Johnson and T. M.Khoshgoftaar, "Survey on deep learning with class imbalance," Journal of Big Data, Vol. 6, No.27, 2019.
DOI:<https://doi.org/10.1186/s40537-019-0192-5>
- [13] J. Carreira and A. Zisserman, "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset," arXiv, 2017.
DOI:<https://doi.org/10.48550/arXiv.1705.07750>
- [14] D. Tran, et al, "A Closer Look at Spatiotemporal Convolutions for Action Recognition," arXiv, 2017.
DOI:<https://doi.org/10.48550/arXiv.1711.11248>
- [15] K. Hara, et al, "Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition," ICCVW, 2017.
DOI:<https://doi.org/10.1109/ICCVW.2017.373>

저 자 소 개

이 의 상(정회원)



- 2024년 : 한국교통대 철도차량시스템 공학과 (졸업예정)

한 석 민(정회원)



- 2001년 : 서울대학교 전기공학부(공학사)
- 2003년 : 서울대학교 전기컴퓨터공학부(공학석사)
- 2008년 : 서울대학교 전기컴퓨터공학부(공학박사)
- 2017 ~ 현재 : 한국교통대학교 데이터사이언스전공 교수

※ 이 연구는 한국연구재단(National Research Foundation of Korea)의 지원으로 수행되었음. (NRF-2021R1F1A1049467).