불균형 데이터 상황에서의 프로토타입 기법을 통한 해석 가능한 분류 모델의 학습

주형준, 이정우
서울대학교

joohj911@cml.snu.ac.kr, junglee@snu.ac.kr

Interpretable classification models through prototyping techniques in imbalanced data situations

Joo Hyung Jun, Lee Jung Woo
Seoul National Univ.

요 약

딥러닝의 성공은 대량의 고품질 데이터로부터 이루어졌습니다. 정확하게 분포되어 있는 주석 처리된 데이터를 통해 정확한 학습을 이루어 냈지만 이러한 데이터는 현실에서 얻기 어렵거나 높은 비용을 요구합니다. 실제로 현실에서의 데이터는 클래스간의 불균형이나 노이즈 혹은 사람에 의해 편향되어 있는 경우가 많습니다. 이러한 데이터에서는 과적합이나 과소적합으로 인해 분류 모델을 학습하는 것이 어렵습니다. 이로 인하여 해석 가능한 분류 모델에서는 설명이 편향되거나 정확도가 낮은 결과가 얻어지는 경우가 많습니다. 이 논문에서는 편향된 데이터셋에서 프로토타입 기법을 통한 해석 가능한 모델이 어떠한 방식으로 구성되어야 하는지에 대해 탐구합니다. 실제 간단한 제한을 통해 명확한 설명과 정확도를 얻을 수 있음을 실험으로 보입니다.

Ⅰ. 서 론

최근 몇 년 동안 딥러닝은 이미지 분류, 구간 분할, 물체 감지 등과 같은 컴퓨터 비전 분야의 작업에서 놀라운 성능을 보여왔습니다. 이러한 놀라운 성능은 클래스 간의 비율이 정확하게 분포되어 있는 깔끔한 주석의 훈련 데이터들을 대량으로 사용했기 때문에 얻어질 수 있었습니다. 하지만 대부분의 현실의 시나리오에서는 부정확한 주석 처리와[1] 현실에서의 물체들의 희귀성과[2] 같은 다양한 제한 사항으로 인해 정리된 데이터 셋을 얻기 힘든 경우가 많습니다. 이러한 이유로 인해 현실에서 다양한 경우에 모델을 사용하기 위해서는 불균형한 데이터셋에서도 학습을 잘 해야할 필요성이 있습니다. 특히 해석 가능한 분류 모델에서는 설명이 데이터 구성에 영향 받기 때문에 정확한 설명을 위해 더욱 잘 학습하는 기법이 필요 됩니다. 이 논문에서는 해석 가능한 분류 모델 중 프로토타입 기법을 사용한 방식에서 어떠한 방식을 통해야 불균형한 데이터셋에서 잘 학습할 수 있는지에 관하여 탐구합니다.

Ⅱ. 본론

본 논문에서 사용하는 프로토타입 기법은 Li et al.[3] 에서 연구된 모델을 기반으로 합니다. 이 모델은 예시 기반 설명 모델 중 프로토타입 관련 설명 모델의 기반이 되는 모델로 잠재 공간 상에서 프로토타입과 인코딩된 입력의 잠재 벡터 사이의 거리를 통해 분류를 진행합니다. 잠재 공간 상에서 학습된 프로토타입을 디코더를 통해 시각화 할 수 있기 때문에 프로토타입과 거리를 통해 설명을 제공할 수 있습니다. 이러한 모델에서 더 강력한 시각화와 입력데이터에 대한 잠재 벡터 위치의 안정성을 위해 잠재 공간 제한과 분포 임베딩을 사용하여 발전시킨 모델을 기준 모델로 사용합니다.

실제 실험은 0.01 수치를 통한 long-tail 불균형 세팅에서[4] MNIST 데이터 셋을 통해 진행되었다. 기준 모델과 여러 제한 방식을 통해 얻어진 결과는 Table 1 과 같다.

	Mean acc	Worst group acc
Baseline model	94.13	80.94
Proto num fixed	94.66	83.43
Orthogonal loss added	94.24	82.03
Last layer fixed	**95.42**	**85.97**

Table 1: Accuracy for different setting

기준 모델을 통해 얻어진 프로토타입을 확인하였을 때 학습된 프로토타입이 데이터의 불균형을 따라가 대부분 비율이 높은 클래스의 모습을 하고 있는 것을 확인할 수 있었다. 이를 해결하기 위해 분류기의 마지막 단인 fully connected layer 의 값을 1 과 -0.5 로 초기화 시켜

클래스당 프로토타입의 개수가 같도록 고정시켜 주었다. 또한 이렇게 학습을 진행하였을 때, Figure 1 과 같이 프로토타입의 모양이 겹치게 생성되는 경우가 많아 클래스 내의 다양성을 보장하기 위해 orthogonality loss 를 추가해 주었다. 그 후 마지막으로 학습의 안정성을 보장하기 위해 초기화 시켰던 fully connected layer 의 값을 고정시킨 후 학습을 진행해 보았다. 실험 결과 프로토타입의 개수를 클래스 별로 같게 지정했을 때 평균 정확도와 가장 낮은 클래스 그룹의 정확도가 향상하는 것을 확인할 수 있었고 다양성을 위한 loss term 을 추가해주고 안정성을 향상 시켰을 때 가장 낮은 클래스 그룹의 정확도가 월등히 향상됨을 확인할 수 있었다.

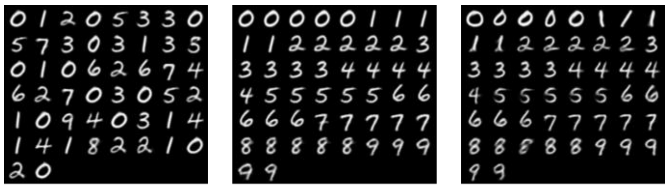

Figure 1: (a) Baseline model, (b) Proto num fixed, (c) Last layer fixed

Ⅲ. 결론

본 논문에서는 프로토타입 기법을 통한 해석 가능한 분류 모델이 데이터 불균형 상황에서 어떠한 방식으로 학습 되어야 하는지 탐구를 진행하였다. 불균형 상황을 해소하기 위해 프로토타입의 개수를 지정해주는 방식과 비율이 적은 클래스의 데이터를 잘 비교하기 위해 프로토타입의 다양성을 보장해주는 방식이 실제 실험에서 높은 효과를 보이며 방식의 효율성을 입증하였다.

ACKNOWLEDGMENT

This work is in part supported by National Research Foundation of Korea (NRF, 2021M3F3A2A02037893(30)), Institute of Information communications Technology Planning Evaluation (IITP, 2021-0-00106(40), 2021-0-02068(30)) grant funded by the Ministry of Science and ICT (MSIT), INMAC, and BK21-plus.

참 고 문 헌

[1] Han, B., Yao, Q., Liu, T., Niu, G., Tsang, I., Kwok, J. T., and Sugiyama, M. A survey of label-noise representation learning: Past, present and future. ArXiv, abs/2011.04406, 2020.

[2] Zhang, Y., Kang, B., Hooi, B., Yan, S., and Feng, J. Deep long-tailed learning: A survey. arXiv preprint arXiv:2110.04596, 2021c.

[3] Li, O., Liu, H., Chen, C., and Rudin, C. Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions. In AAAI, 2018.

[4] Cui, Y., Jia, M., Lin, T.-Y., Song, Y., and Belongie, S. Class- balanced loss based on effective number of samples. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9268- 9277, 2019.