

Title: [[20240812111622 data manipulation-based distribution generalization]]

Data generalization 기법의 세 가지 분류 계통 중의 하나로, 데이터들을 가공하여 모델이 Out-of-distribution 문제에 대해 좀 더 일반화된 특성을 학습할 수 있도록 만드는 방식이다. Data manipulation에 기반한 DG 기법이 OOD와 in-distribution 모두에서 유의미하게 예측할 수 있음을 실험을 통해 확인한 사례가 있다¹.

data manipulation을 통한 DG의 목표를 수식으로 표현하면 다음과 같다. 이때 $\mathcal{M}(\cdot)$ 은 manipulation function이다.

$$\min_h \mathbb{E}_{x,y}[\ell(h(x), y)] + \mathbb{E}_{x',y}[\ell(h(x'), y)]$$

$$x' = \mathcal{M}(x)$$

Created : 2024-08-12 11:16:22

Tags : [[out-of-distribution]] - [[generalization]] - [[preprocessing]]

Related note :

References :

- J. Wang et al., “Generalizing to Unseen Domains: A Survey on Domain Generalization,” IEEE Trans. Knowl. Data Eng., pp. 1 – 1, 2022, doi: 10.1109/TKDE.2022.3178128.
- D. Adila and D. Kang, “Understanding Out-of-distribution: A Perspective of Data Dynamics,” Nov. 29, 2021, arXiv: arXiv:2111.14730. Accessed: Aug. 12, 2024. [Online]. Available: <http://arxiv.org/abs/2111.14730>

¹참조문헌 2