Deep Imbalanced Regression

Yang Y, Zha K, Chen Y, Wang H, Katabi D

ICML 2021

Presenter: Gianmarco Midena

26 November 2024

Overview

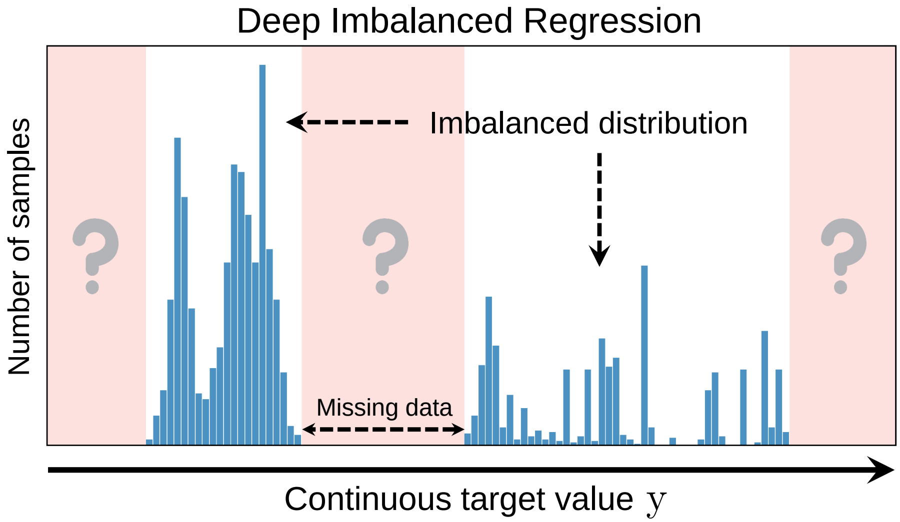


Image credit: *Yang et al. (2021)*

Problem Settings

- $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$: training set
- $\mathbf{x}_i \in \mathbb{R}^d$: input
- $y_i \in \mathcal{Y}$: continuous label or target
- $b_i \in \mathcal{B}$: discrete label or target
- $\mathcal{Y} \subset \mathbb{R}$: continuous label space
- $\mathcal{B} = \{1, \ldots, M\} \subset \mathbb{Z}^+$: index space
  - divides $\mathcal{Y}$ into M groups (bins) with equal intervals $[t_j, t_{j+1})$
  - $\{[t_0, t_1), \ldots, [t_{M-1}, t_M)\}$: discrete label space
  - $t_k \in \mathcal{Y}$
  - minimum resolution
    - e.g., $\delta y \triangleq t_{j+1} - t_j = 1$ in age estimation
- $\hat{y}_i = g(\mathbf{z}_i) \in \mathbb{R}$: predicted continuous label
- $\mathbf{z}_i = f(\mathbf{x}_i; \theta) \in \mathbb{R}^{d'}$: learned representation
- θ: trainable model parameters

Evaluation

- Divide target space into disjoint regions (bins)
  - *Many-shot*: > 100 training examples
  - *Medium-shot*: 20-100 training examples
  - *Few-shot*: < 20 training examples
  - *Zero-shot*: 0 training examples
  - Inspired by Liu et al. 2019

- Metrics
  - Mean Absolute Error (MAE)
  - Mean Squared Error (MSE)
  - Pearson Correlation (PCC)
  - Geometric Mean Error (GM)

$$GM = \sqrt[n]{\prod_{i=1}^{n} |y_i - \hat{y}_i|}$$

★ Pros: + fairness (uniformity) in prediction

Datasets - Overview

Dataset	Target type	Target range	Bin size	Max bin density	Min bin density	# Training set	# Val. set	# Test set
IMDB-WIKI-DIR	Age	0 - 186	1	7,149	1	191,509	11,022	11,022
AgeDB-DIR	Age	0 - 101	1	353	1	12,208	2,140	2,140
STS-B-DIR	Text similarity score	0 - 5	0.1	428	1	5,249	1,000	1,000
NYUD2-DIR	Depth	0.7 - 10	0.1	1.46×10^8	1.13×10^6	50,688 (3.51×10^9)	–	654 (8.70×10^5)
SHHS-DIR	Health condition score	0 - 100	1	275	0	1,892	369	369

(Training) Datasets - Label Distributions

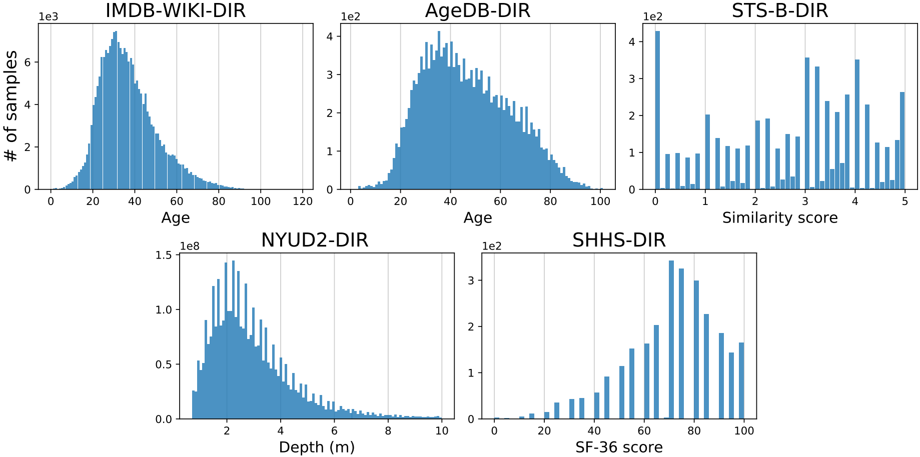
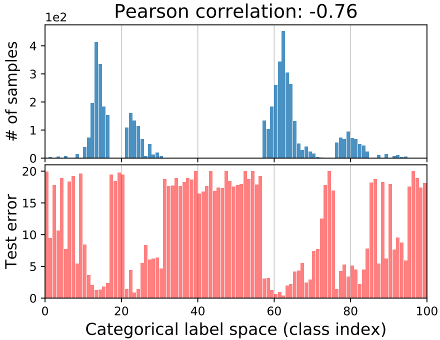
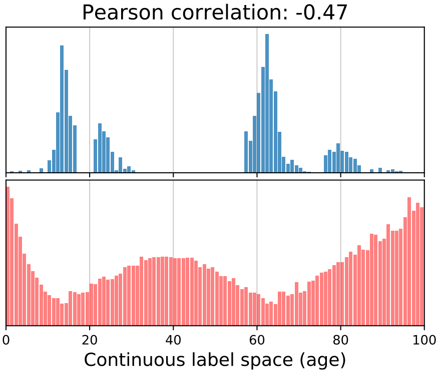


Image credit: *Yang et al. (2021)*

Test Error on Categorical vs. Continuous Label Space



(a) CIFAR-100 (subsampled)



(b) IMDB-WIKI (subsampled)

Image credit: *Yang et al. (2021)*

Label Distribution Smoothing

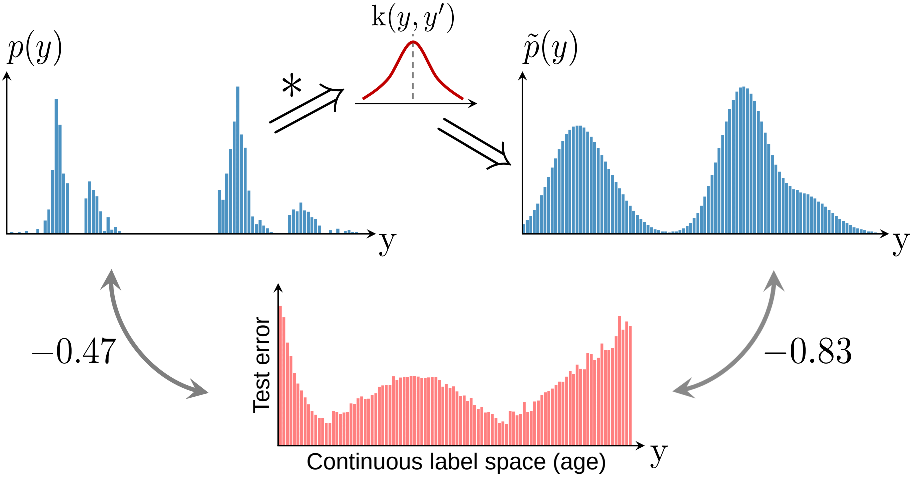


Image credit: *Yang et al. (2021)*

Feature Distribution Smoothing

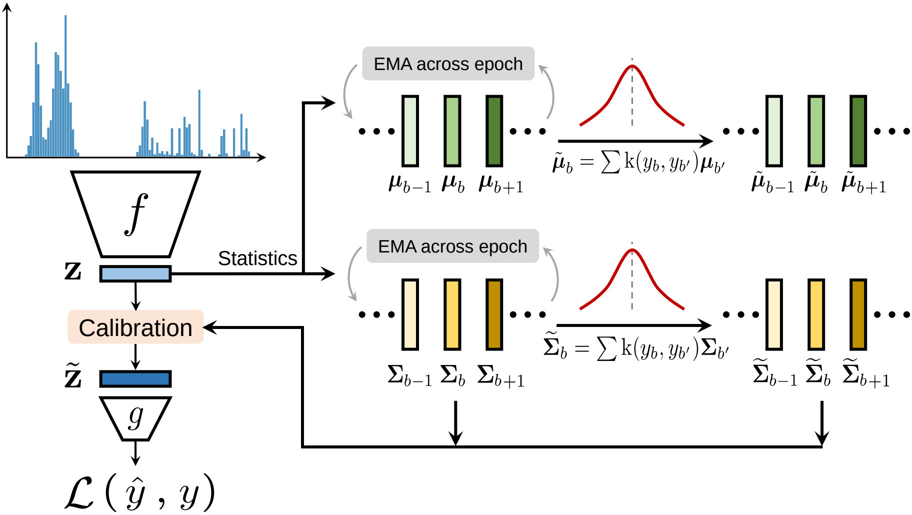


Image credit: *Yang et al. (2021)*

Baselines (1/2)

- Vanilla: neglects data imbalance

- Synthetic samples
  - SMOTER (Torgo et al. 2013)
    1. Defines frequent and rare regions using label density.
    2. Creates synthetic samples for pre-defined rare regions by linearly interpolating both inputs and labels.
  - SMOGN (Branco et al. 2017): augments SMOTER with Gaussian noise

- Focal-R

$$\frac{1}{n}\sum_{i=1}^{n}\sigma(|\beta e_i|)^{\gamma}e_i$$

  - Error-aware loss
  - Maps the absolute error into $[0,1]$.
  - e_i: L_1 error for the i-th sample
  - β, γ: hyper-parameters
  - Inspired by Focal Loss (Lin 2017) for classification

Baselines (2/2)

- Regressor re-training (RRT)
  - Two-stage training
    1. Train encoder
    2. Re-train regressor with inverse re-weighting and frozen encoder.
  - Inspired by Kang et al. 2019

- Cost-sensitive re-weighting: re-weighting schemes based on label distribution
  - Inverse-frequency weighting (INV)
  - Square-root weighting variant (SQINV)

Results (1/2)

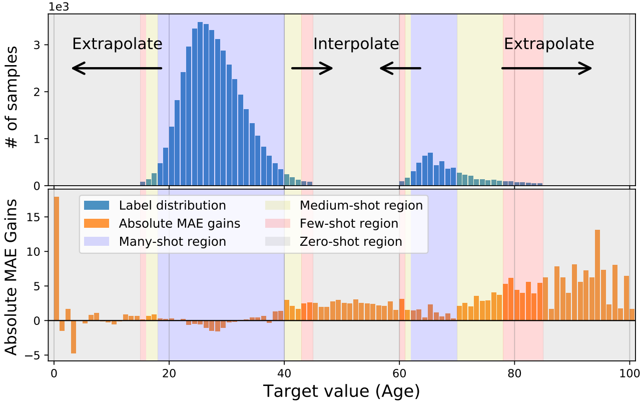


Figure: MAE gains of LDS + FDS over the vanilla model, on a curated subset of IMDB-WIKI-DIR.

- Performance gains esp. for extrapolation & interpolation

Image credit: *Yang et al. (2021)*

Results (2/2)

Metrics	MAE ↓				GM ↓			
Shot	All	w/ data	Interp.	Extrap.	All	w/ data	Interp.	Extrap.
Vanilla	11.72	9.32	16.13	18.19	7.44	5.33	14.41	16.74
Vanilla + **LDS**	10.54	8.31	14.14	17.38	6.50	4.67	12.13	15.36
Vanilla + **FDS**	11.40	8.97	15.83	18.01	7.18	5.12	14.02	16.48
Vanilla + **LDS** + **FDS**	**10.27**	**8.11**	**13.71**	**17.02**	**6.33**	**4.55**	**11.71**	**15.13**
Ours (best) vs. Vanilla	**+1.45**	**+1.21**	**+2.42**	**+1.17**	**+1.11**	**+0.78**	**+2.70**	**+1.61**

Table: Interpolation & extrapolation results on a curated subset of IMDB-WIKI-DIR.

- Best results by smoothing both label & feature distributions

Table credit: *Yang et al. (2021)*

References

Branco, Paula, Luís Torgo, and Rita P Ribeiro (2017). "SMOGN: a pre-processing approach for imbalanced regression". In: *First international workshop on learning with imbalanced domains: Theory and applications*. PMLR, pp. 36–50.

Kang, Bingyi et al. (2019). "Decoupling representation and classifier for long-tailed recognition". In: *arXiv preprint arXiv:1910.09217*.

Lin, T (2017). "Focal Loss for Dense Object Detection". In: *arXiv preprint arXiv:1708.02002*.

Liu, Ziwei et al. (2019). "Large-scale long-tailed recognition in an open world". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2537–2546.

Torgo, Luís et al. (2013). "Smote for regression". In: *Portuguese conference on artificial intelligence*. Springer, pp. 378–389.

Yang, Yuzhe et al. (2021). "Delving into deep imbalanced regression". In: *International conference on machine learning*. PMLR, pp. 11842–11851.