RankUp: Boosting Semi-Supervised Regression with an Auxiliary Ranking Classifier

Pin-Yen Huang
Academia Sinica

Szu-Wei Fu
NVIDIA

Yu Tsao
Academia Sinica

[NeurIPS 2024]

Labeled Data is Costly and Time-Consuming to Acquire

1. 🩺 **Medical Data**: Requires experts, such as doctors, to label data.
2. 🗃️ **Large-scale Datasets**: Even simple labeling tasks become costly with millions of data points.
3. . . . (Many More)

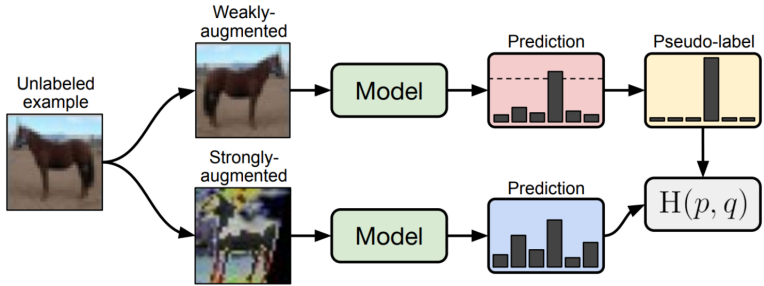
Labeled Data is Costly and Time-Consuming to Acquire

1. 🩺 **Medical Data**: Requires experts, such as doctors, to label data.
2. 🗃️ **Large-scale Datasets**: Even simple labeling tasks become costly with millions of data points.
3. ... (Many More)

📉 Reducing Labeling Costs: Semi-Supervised Learning (SSL)

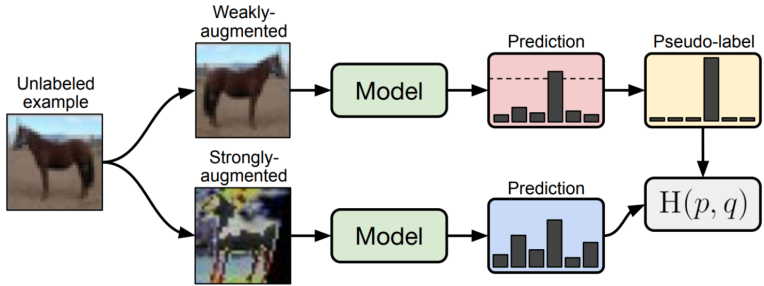
- **Reduces reliance for labeled data** by leveraging **unlabeled data** during training.
- Unlabeled data is usually **easier** and **less costly** to obtain than labeled data.

Background: FixMatch-Based SSL Methods



- **Image Source:** Sohn, Kihyuk, et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." *Advances in neural information processing systems* 33 (2020): 596-608.

Background: FixMatch-Based SSL Methods

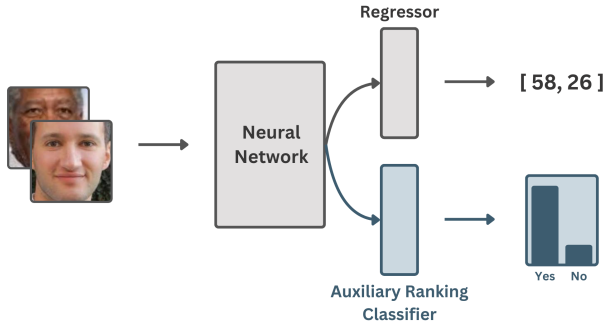


- **Image Source:** Sohn, Kihyuk, et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." *Advances in neural information processing systems* 33 (2020): 596-608.

⚠ Challenges for Regression Tasks

- Regression models lack **confidence measurement**, making FixMatch's **confidence thresholding** and **pseudo-labeling** unfeasible.

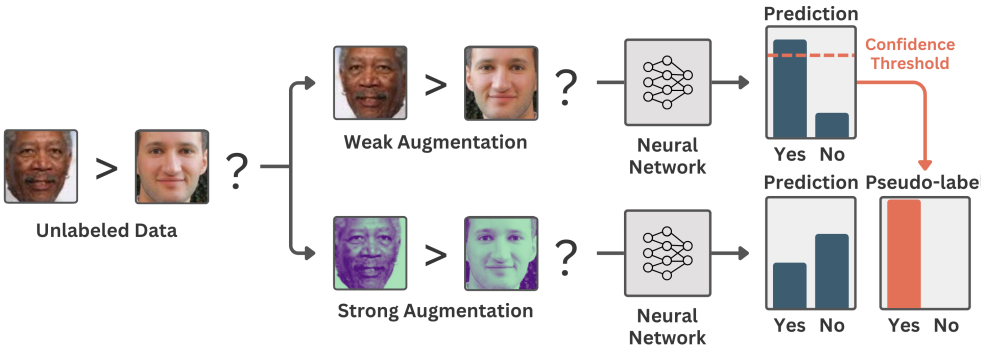
Proposed Method: RankUp



Auxiliary Ranking Classifier (ARC)

- ARC integrates with existing regression models to predict the **relative ranking** of data pairs (e.g., which data point has a higher regression value).

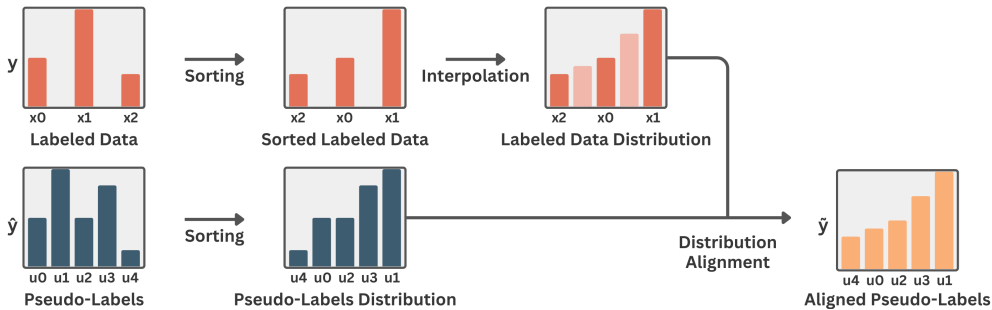
Proposed Method: RankUp



Integrating Semi-Supervised Classification Methods

- Existing semi-supervised **classification** methods can be used to train ARC.

Proposed Method: RDA



Regression Distribution Alignment (RDA)

- RDA improves the quality of regression pseudo-labels by **aligning the distribution** of **pseudo-labels** with the distribution of the **labeled data**.

Main Experiments Results

	UTKFace (Image Age Estimation)					
	Labels = 50			Labels = 250		
	MAE↓	R^2↑	SRCC↑	MAE↓	R^2↑	SRCC↑
Supervised	14.13±0.56	0.090±0.092	0.371±0.071	9.42±0.16	0.540±0.014	0.712±0.010
Π-Model	13.82±1.02	0.100±0.086	0.387±0.092	9.45±0.30	0.534±0.030	0.706±0.015
Mean Teacher	13.92±0.20	0.127±0.037	0.423±0.023	8.85±0.25	0.586±0.020	0.745±0.013
CLSS	13.61±0.92	0.138±0.101	0.447±0.074	9.10±0.15	0.586±0.016	0.737±0.014
UCVME	13.49±0.95	0.157±0.110	0.412±0.127	8.63±0.17	0.626±0.006	0.767±0.007
MixMatch	11.44±0.45	0.401±0.028	0.674±0.035	7.95±0.15	0.692±0.013	0.832±0.008
RankUp (Ours)	9.96±0.62	0.514±0.043	0.703±0.019	7.06±0.11	0.751±0.011	0.835±0.008
RankUp + RDA (Ours)	**9.33**±0.54	**0.552**±0.041	**0.770**±0.009	**6.57**±0.18	**0.782**±0.012	**0.856**±0.005
Fully-Supervised	4.85±0.01	0.875±0.000	0.910±0.001	4.85±0.01	0.875±0.000	0.910±0.001

Table: Comparison of RankUp against other methods on the UTKFace dataset.

Feature Representations

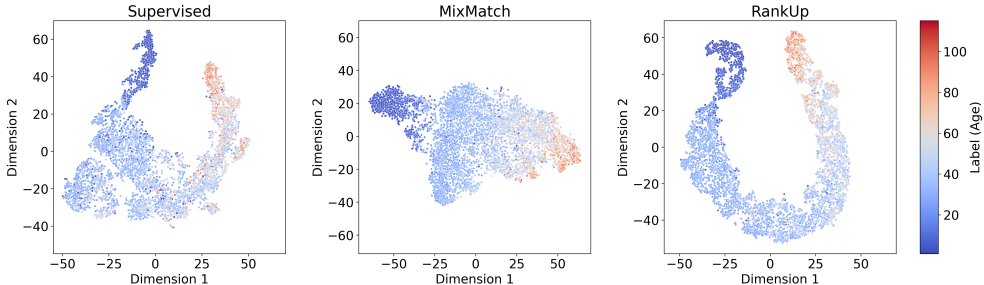
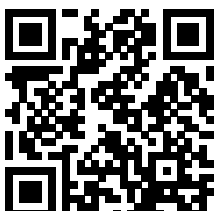


Figure: t-SNE visualizations of feature representations from different semi-supervised regression methods.

Thank you!

If you're interested in this work, feel free to explore the following resources:

Paper



https://arxiv.org/abs/2410.22124

Code Repository

https://github.com/pm25/semi-supervised-regression