Learning Natural Thresholds for Image Ranking

Somayeh Keshavarz, Quang Nhat Tran, Richard Souvenir

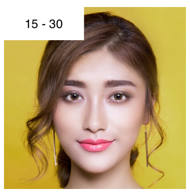
Computer and Information Sciences Department, Temple University

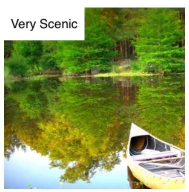
ICPR '20

Problem

Find Natural grouping in data- driven way to improve image ranking

- Ranking images with continuous label
  - Coarse prediction for data like age and scenicness which naturally are suited for regression analysis.
- Limitation of using classification methods
  - standard classification algorithms discard ordinal information in the class attribute.
- Limitation of k-rank methods
  - k-rank methods do consider ordinal information in the class attribute, but range of each class are predefined.
- Benefit of learning thresholds
  - Finding natural groupings in data will result in more discriminative features





Objective

- Simultaneous representation learning and label discretization

representation learning parameters

$$\underset{\Theta,\,\beta}{\arg\min}\ \mathcal{L}(\Theta, \beta\,|\mathcal{X}, \Upsilon).$$

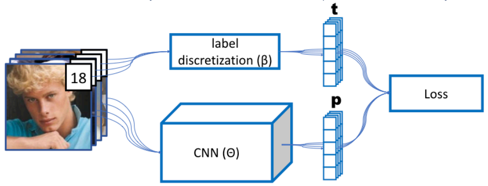
label discretization parameters   real label

Proposed Method

Using soft classification label to simultaneously learn thresholds and classification using our proposed conjugate histogram loss.

OCLET (Ordinal classification with learnable threshold)

OCLET combines ordinal classification and representation learning to provide a data-driven approach for label discretization for problems with naturally continuous output.



- Representation learning

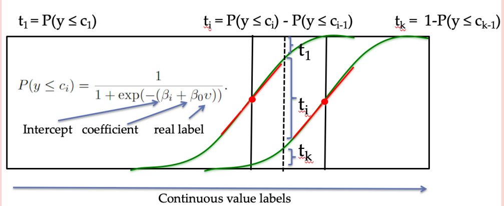
- Can adopt any pre-trained CNN classification model.

- Label Discretization

- Logistic Regression is used to model target label probability.

- Target probability is the difference between consecutive cumulative probabilities.

- Threshold is where the probability of adjacent classes are equal.



- Using alternative minimization

1- Representation learning

$$\underset{\beta}{\arg\min}\ \mathcal{L}(\beta|\Theta, \mathcal{X}, \Upsilon)$$

2- Label discretization

$$\underset{\Theta}{\arg\min}\ \mathcal{L}(\Theta|\beta, \mathcal{X}, \Upsilon)$$

Loss Function

Loss is the sum of our proposed conjugate histogram loss, margin loss and regularization loss.

conjugate histogram loss   tradeoff parameters   regularization loss

$$\mathcal{L}(\Theta, \beta|\mathcal{X}, \Upsilon) = \mathcal{L}_D + \lambda_M \mathcal{L}_M + \lambda_R \mathcal{L}_R$$

margin Loss

$$\mathcal{L}_M = \sum_{i=2}^{K} \max(0, (\Delta - (\Phi_i - \Phi_{i-1}))$$

Conjugate Histogram Loss

- Extended version of Histogram loss [1], which updates both "feature representation" parameters and "label discretization" parameters.

$$\mathcal{L}_D = \sum_{i=1}^{K} \mathcal{L}_{\mathcal{H}}\left(\{p_i^j\}, \{t_i^j\}\right) + \mathcal{L}_{\mathcal{H}}\left(\{t_i^j\}, \{p_i^j\}\right)$$

$\mathcal{L}_{\mathcal{H}}\left(\{p_i^j\}, \{t_i^j\}\right)$: Histogram loss to adjust predicted class posterior to match soft target vector.

$\mathcal{L}_{\mathcal{H}}\left(\{t_i^j\}, \{p_i^j\}\right)$: Histogram loss to adjust soft target vector to match predicted class posterior.

Results

Age Estimation

Discretize output of a pre-trained regression model using:

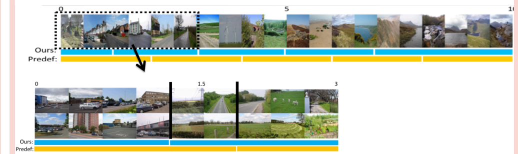
- OCLET thresholds
- Even thresholds
- Predefined thresholds

Model	Train Data	1-off ↑			m-MAE ↓		
		Even	Predef	OCLET	Even	Predef	OCLET
DEX[2]	IMDB-WIKI	69.2	71.6	76.9	1.5	1.6	1.5

- predictions using OCLET (bottom-left)
- predefined thresholds (bottom-right)

Scenicness Estimation

- OCLET groups images by semantic similarity.

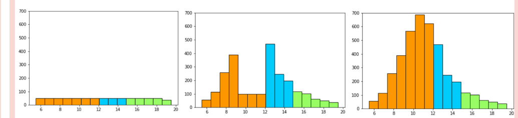


- OCLET thresholds improves classification.

Method	F-score
Workman et al.[3]	.48
Workman et al.[3] + OCLET	.53
OCLET	.54

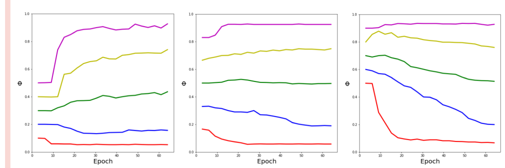
Data Distribution

- OCLET learns the same label discretization for different distributions of input data.



Initial threshold

OCLET does not depended on initial thresholds.



Reference

[1] E. Ustinova and V. Lempitsky, "Learning deep embeddings with histogram loss," in Advances in Neural Information Processing Systems, 2016

[2] on Computer Vision, 2017. S. Workman, R. Souvenir, and N. Jacobs, "Understanding and mapping natural beauty," in Proc. International Conference

[3] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision, 2018.