Improving Pairwise Ranking for Multi-label Image Classification

Yuncheng Li, Yale Song, and Jiebo Luo

Code Available: https://goo.gl/Fi5hiG

YAHOO! RESEARCH · UNIVERSITY of ROCHESTER

Introduction

CC Image courtesy of *ReflectedSerendipity* on Flickr.

Highlights

- A novel loss function based on a log-sum-exp function.
- A novel label decision module to infer the exact labels.
- Achieved the state of the art performance on the VOC2007, NUS-WIDE, and MS-COCO for the multi-label task.

LSEP Loss Function

Log-Sum-Exp Pairwise Loss Function (LSEP Loss)

$$l_{\text{lsep}} = log\left(1 + \sum_{\phi(Y_i;t)} exp\left(f_v(x_i) - f_u(x_i)\right)\right)$$

- ✓ Smooth everywhere
- ✓ Nice theoretical properties (Bayes consistency)
- ✓ Margin enforcing
- ✓ Scalable w.r.t. vocabulary size

Gradients

$$\frac{\partial l_{\text{lsep}}}{\partial f(x_i)} = -\frac{1}{l_{\text{lsep}}} \sum_{\phi(Y_i;t)} \Delta Y_{i,u,v} e^{(-f(x_i)\Delta Y_{i,u,v})}$$

LSEP Loss Analysis

Nice Theoretical Properties (Bayes Consistency)

Theorem 1. *If $f^*(x)$ is the minimizer of Eqn.(7), then*

$$f_u^*(x) = \log P(u \in Y|x) + c, \quad \forall u \in \mathcal{Y}$$

Lower loss means better average performance

Margin Enforcing

$$l_{\text{lsep}}^{\text{asym}} = \sum_{v \notin Y_i} \sum_{u \in Y_i} max\left(0, \alpha_i + f_v(x_i) - f_u(x_i)\right)$$

Puts focus on violating pairs

Negative Sampling

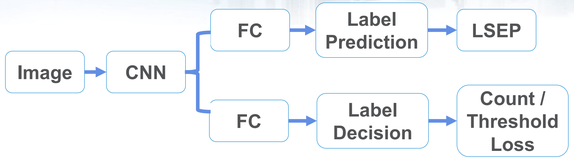
$$\phi(Y_i;t) \subseteq Y_i \otimes (\mathcal{Y} - Y_i), |\phi(Y_i;t)| = t \approx 1000$$

Scalable for large vocabulary

Loss Function Comparison

	Bayesian Consistency	Margin Enforcing	Large Vocabulary	Smooth Everywhere
Ours (LSEP)	Y	Y	Y	Y
Softmax	N	Y	Y	Y
Ranking	Y	Y	N	N
WARP	N	Y	N	N
BP-MLL	Y	N	N	Y

Label Decision



Label Count Prediction

$$l_{\text{count}} = -\log\left(\frac{\exp(g_{k_i}(f'(x_i)))}{\sum_{j=1}^{n}\exp(g_j(f'(x_i)))}\right)$$

- Cap label count at n (n=4)
- Classify the label count

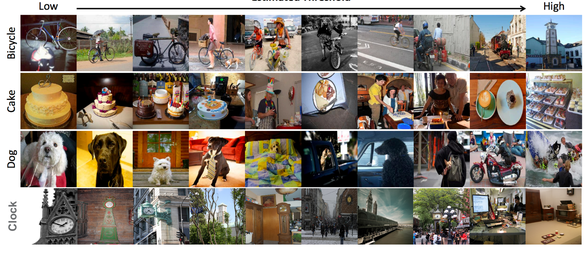
Adaptive Thresholding

$$\hat{Y} = \{l | f_k(x) > \theta_k, \ \forall k \in [1, K]\}$$

$$l_{thresh} = -\sum_{k=1}^{K} Y_{i,k} log(\delta_\theta^k) + (1 - Y_{i,k}) log(1 - \delta_\theta^k)$$

- Predict label set given adaptive thresholds
- Adaptive thresholds, w.r.t. the image and the label
- Cross entropy loss as relaxed objective

Examples of Estimated Threshold



Experiment Results

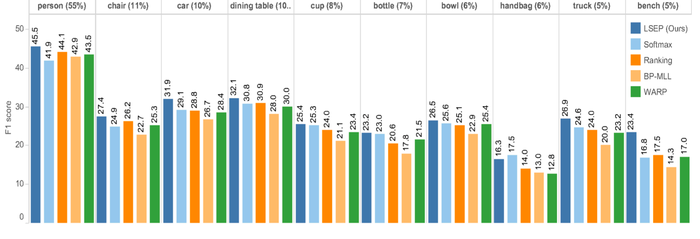
Comparing Loss Function

Method	NUS-WIDE						MS-COCO						VOC2007	
	PC-P	PC-R	OV-P	OV-R	F_1	0-1	PC-P	PC-R	OV-P	OV-R	F_1	0-1	F_1	0-1
Softmax (K)	42.7	52.5	54.2	67.5	43.2	5.02	56.2	56.8	59.7	61.7	54.8	5.63	73.2	56.6
Ranking (K)	42.6	56.3	54.7	68.2	45.1	5.31	57.0	57.8	60.2	62.2	55.4	5.71	70.8	56.3
BP-MLL (K)	40.9	56.8	53.9	67.1	44.0	4.89	55.8	56.0	58.9	60.8	53.6	5.22	65.3	54.0
WARP (K)	43.8	57.1	54.5	67.9	45.5	5.13	55.5	57.4	59.6	61.5	54.8	5.48	71.9	56.9
Softmax (θ)	50.6	**57.8**	62.2	**76.0**	52.1	26.1	58.4	**59.0**	59.5	63.6	57.2	16.6	74.1	53.4
Ranking (θ)	51.3	56.5	64.6	70.8	52.5	25.6	60.7	57.9	64.0	62.6	58.0	17.3	75.2	52.0
BP-MLL (θ)	36.7	48.2	49.4	57.0	39.2	17.5	50.1	56.6	52.7	61.6	51.6	14.5	68.1	42.5
WARP (θ)	48.4	53.1	59.8	64.6	48.5	21.3	57.3	58.9	60.7	63.5	56.9	15.9	74.7	47.5
Wang *et al.* [29]	40.5	30.4	49.9	61.7	-	-	66.0	55.6	69.2	**66.4**	-	-	-	-
LSEP (ours)	**66.7**	45.9	**76.8**	65.7	**52.9**	**33.5**	**73.5**	56.4	**76.3**	61.8	**62.9**	**30.6**	**79.1**	**64.6**

Comparing Label Decision

Method	NUS-WIDE						MS-COCO						VOC2007	
	PC-P	PC-R	OV-P	OV-R	F_1	0-1	PC-P	PC-R	OV-P	OV-R	F_1	0-1	F_1	0-1
Top-k	44.8	55.6	54.8	68.3	45.5	5.39	56.2	58.6	60.5	62.4	55.8	6.19	72.5	57.6
Threshold	55.0	**57.0**	67.2	**73.4**	**55.0**	29.3	59.0	**63.4**	61.5	**67.1**	59.8	23.2	76.4	54.5
Label count est.	61.4	46.1	73.7	64.7	50.9	33.4	67.7	57.6	72.0	62.2	61.4	30.4	78.3	**66.3**
Thresh. est. (ours)	**66.7**	45.9	**76.8**	65.7	52.9	**33.5**	**73.5**	56.4	**76.3**	61.8	**62.9**	**30.6**	**79.1**	64.6

Per Label Performance on MS-COCO



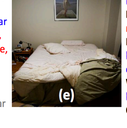
Comparing Loss Function Qualitatively

Softmax: person, frisbee, sports ball
Rank: person, frisbee, sportsball
BP-MLL: person, frisbee, sportsball, kite
WARP: person, car, frisbee, sportsball
Ours: person, sportsball

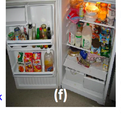
Softmax: person, kite, surfboard
Rank: person, kite, surfboard
BP-MLL: person, boat, kite, surfboard
WARP: person, boat, surfboard
Ours: person, surfboard

Softmax: bottle, bowl, banana, apple, orange
Rank: banana, apple, orange
BP-MLL: bowl, banana, apple, orange
WARP: banana, apple, orange
Ours: apple, orange

Softmax: cat, dog, couch, bed, book, teddybear
Rank: cat, bed, teddybear
BP-MLL: cat, bottle, cup, chair, couch, bed, remote, book, clock, teddybear
WARP: cat, dog, couch, bed, book, teddybear
Ours: cat, bed, teddybear



Softmax: person, bed, book, clock
Rank: person, bed, remote, book, clock
BP-MLL: bottle, chair, bed, remote, book, clock, teddybear
WARP: person, bottle, bed, remote, book, clock
Ours: bed, book, clock



Softmax: bottle, bowl, apple, orange, fridge
Rank: bottle, cup, bowl, apple, orange, fridge
BP-MLL: bottle, bowl, apple, orange, broccoli, carrot, fridge
WARP: bottle, bowl, apple, orange, fridge
Ours: bottle, fridge, bowl