Quasibinary Classifier for Images with Zero and Multiple Labels

Shuai Liao, Efstratios Gavves, ChangYong Oh, Cees Snoek

QUVA Deep Vision Lab, University of Amsterdam

ICPR 2020, 25th International Conference on Pattern Recognition, ONLINE 10 | 15 January 2021

Motivation

Consider a classification problem with K=3 classes: { *Bird*, *Cat*, *Dog* }.

One-vs-rest problem

[✓]Bird [ ]Cat [ ]Dog

Multi-label problem

[ ]Bird [✓]Cat [✓]Dog

Zero-label problem

[ ]Bird [ ]Cat [ ]Dog

(I) Softmax classifier

Use the prior #label=1.
Scale-up to a large number of classes.
Stable to train.
✗ Unable handle 0/N-label classification.

Softmax activation:

$$p_k = \frac{\exp(z_k)}{\sum_j \exp(z_j)}$$

(II) Ensemble of binary classifiers

Flexible to handle 0/N-label classification.
✗ Lack of correlation: independently trained.
✗ Do not scale-up to a large number of classes.
✗ Unstable to train.

Sigmoid activation:

$$p_k = \frac{\exp(z_k)}{1+\exp(z_k)}$$

Observation: Binary classifier and softmax have similar form. The difference is on the denominator (normalization factor).

Method

Goal: Learn a normalization function C(X) that is shared across entire dataset X={ $x^{(i)}$ } and classes.

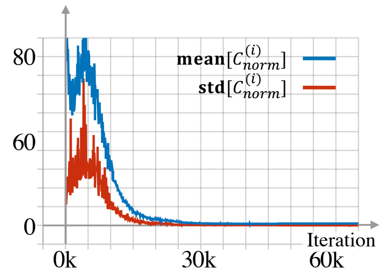
Probability for image x in being k-th class has the form: $\boxed{p(y=k|x) = q_k = \frac{\exp(z_k)}{\mathrm{C(X)}}}$, that is constrained by:

- Prior knowledge $\sum_{k=1}^{K} q_k = \#\text{lable}$ for each image (*See proof in paper.*)
- Valid probability range $q_k^{(i)} \in [0,1]$.

Training: Maximize Likelihood Estimation, while

1) Computing C(X) over a mini-batch $\sum_i^B \sum_k^K \exp(z_k^{(i)})/C = \sum_i \#\text{label}^{(i)}$
2) Minimizing $-\max(\log q_k^{(i)}, 0)$ (Penalize any violation of $q_k^{(i)} > 1$).

Test: Use the learned constant C: $q_k = \min(\exp(z_k)/\mathrm{C}, 1)$



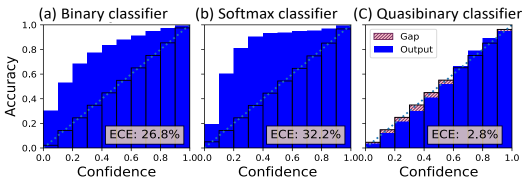
C(X) converges to a constant during training

Experiment

Multi-label image classification

Setup: Follow Li *etal. CVPR2017*.

	MS-COCO		NUS-WIDE	
	$F_1\uparrow$	ECE(%)↓	$F_1\uparrow$	ECE(%)↓
Binary classifier [16], [18], [30]	51.2	26.8	40.7	23.6
Softmax [31]	**54.7**	32.2	43.2	25.8
Softmax w/ temperature [32]	**54.7**	31.4	43.2	24.6
Quasibinary classifier	**54.7**	**2.8**	**43.5**	**3.3**



I **Reliability diagrams for multi-label classification**

Conclusion: Quasibinary classifier is both accurate and credible.

One-vs.-rest image classification (single-label)

Setup: Resnet18 with 32x32 and 224x224 input.

	CIFAR10	CIFAR100	Tiny-ImageNet	ImageNet
Binary classifier	**4.8**	35.4	×	×
Quasibinary classifier (*Ours*)	4.9	**21.9**	**42.9**	25.4
Softmax classifier	5.2	22.2	43.3	**23.9**

Conclusion: Quasibinary classifier is better than binary classifier, and comparable with softmax classifier.

Zero-label image classification

Setup: CIFAR60+40 dataset, with 40 classes of images from original CIFAR100 being treated as 0-label.

	IN Accuracy ↑	OUT MMC ↓	BOTH AU-ROC↑
Binary classifiers [16], [18], [30]	77.8 %	14.7 %	0.901
Softmax + $\mathcal{L}_{\text{Uniform}}$ [6]	**80.7** %	59.8 %	0.800
Softmax + $\mathcal{L}_{\text{MaxConf}}$ [11]	45.2 %	7.4 %	0.764
Quasibinary classifier	80.6 %	**6.9** %	**0.913**

Conclusion: Quasibinary classifier achieves good performance on all measures.