Combinatorial Inference against Label Noise

Paul Hongsuck Seo, Geeho Kim, Bohyung Han
hsseo@postech.ac.kr, {snow1234, bhhan}@snu.ac.kr
POSTECH POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY
SEOUL NATIONAL UNIVERSITY
Google AI

Background and Motivation

▪ **Inevitable label noise in training data**
Data collection from crowdsourcing or Internet search engine

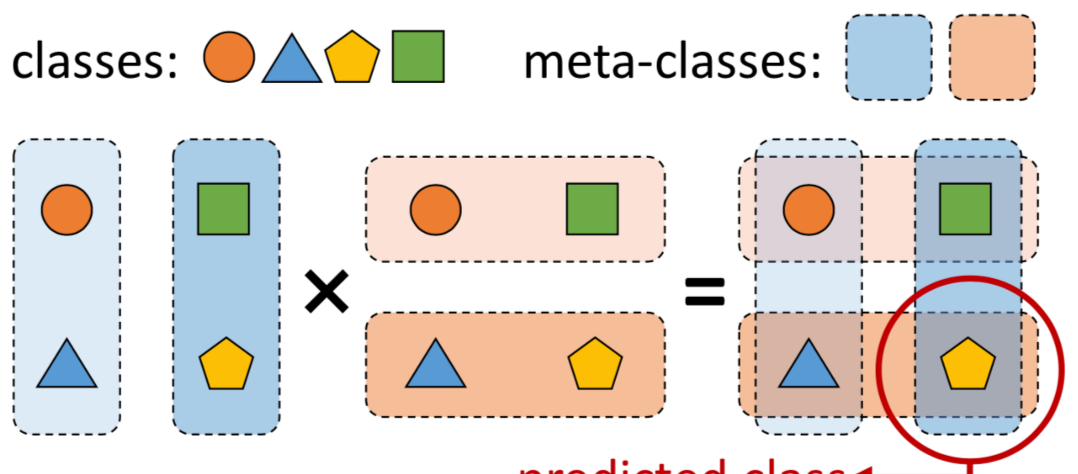
Figure 1: Example images of 259-pomeranian class in WebVision [4]. In addition to clean samples, dataset contains closed- and open-set noises where closed-set noises belong to known classes (correct classes marked at corner) and open-set noises are not associated with any known classes.

▪ **Existing mainstream approaches**
Focus on noise-resistant training algorithms **given level of noise**

▪ **Our contribution**
1) Reducing amount of label noise by constructing meta-classes
2) Proposing a novel combinatorial classification framework
3) Demonstrate noise-robustness of the proposed method

Combinatorial Classification

▪ **Definitions**
• K original classes
$$\mathcal{C} = \{c_1, \dots, c_K\}$$
• M meta-class sets
$$\mathcal{C}^m = \{c_1^m, \dots, c_{K_m}^m\} \quad (K \gg K_m)$$

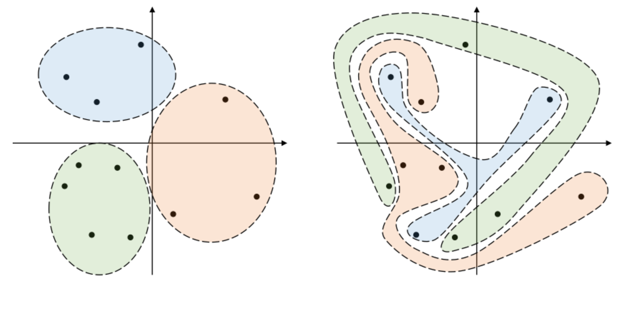


▪ **Inference and training**
Inference: $\dfrac{\prod_{m=1}^{M} P(\text{meta}(c_k; m)|x)}{\sum_{j=1}^{K} \prod_{m=1}^{M} P(\text{meta}(c_j; m)|x)}$ Training: $-\sum_{m=1}^{M} \log P(y^m|x)$

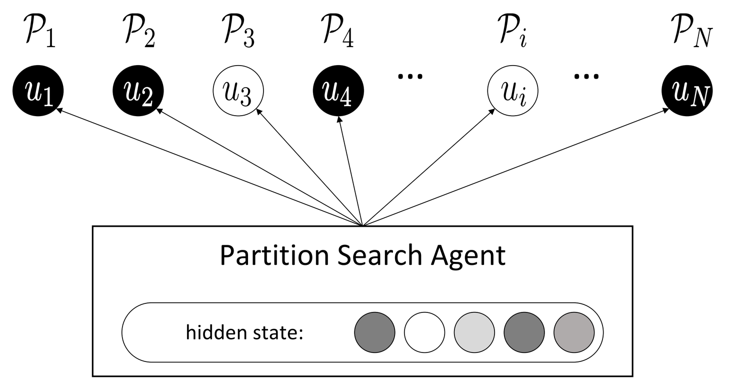
▪ **Deep combinatorial classifier**
Single shared feature extractor + M classification layers

Meta-class Sets Configurations

▪ **Clustering based meta-class sets configurations**
1) Use class features from weights from classification layer of pretrained network.
2) Select Q sub-dimensions of class features.
3) Perform k-means clustering.



▪ **Reinforced meta-class sets search**
Train search agent that selects optimal set of meta-class sets using policy gradients:



$$\nabla_\theta \mathcal{J}(\theta) \approx \frac{1}{S} \sum_{s=1}^{S} \sum_{i=1}^{N} \nabla_\theta \log P\left(u_i^{(s)} \middle| h; \theta\right) \left(\mathcal{R}^{(s)} - B\right)$$

where reward $\mathcal{R} = \mathcal{R}_{\text{acc}} - \alpha \sum_{i=1}^{N} u_i$

In-batch sample accuracy   Number of selected meta-class sets

Noise-robustness

▪ **Label noise reduction**
Meta-class representation directly reduces noise level:
$$\eta(\hat{\mathcal{D}}) = \mathbb{E}_{\hat{\mathcal{D}}}[\mathbf{1}(y \neq \hat{y})] = \frac{1}{N} \sum_{i}^{N} \mathbf{1}(y \neq \hat{y})$$
(y: clean label, $\hat{y}$: corrupted label)

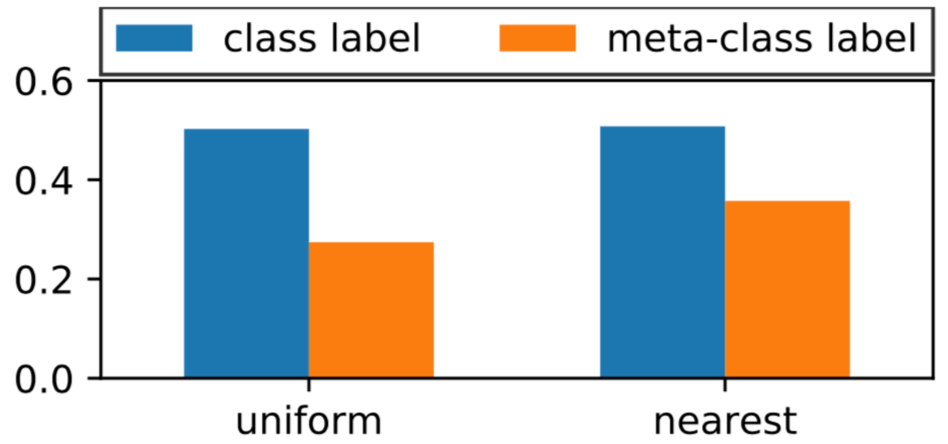
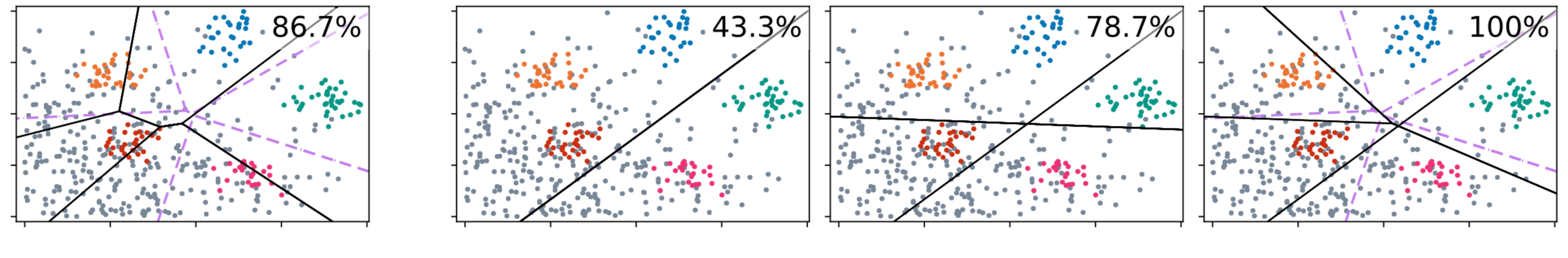


Figure 2:

▪ **Experiments on toy example**



(a) Standard classifier (b) One meta-class set (c) Two meta-class sets (d) Three meta-class sets

Figure 3: Visualizations of decision boundaries of (a) standard classifier and (b-d) combinatorial classifiers on two-dimensional Gaussian mixture model with five components. Accuracies are at the top-right corner for each classifier. For combinatorial classifiers, we gradually add classifiers one-by-one from (b) to (d). Gray dots show noisy samples with random labels while black solid and purple dashed lines represent decision boundaries drawn by clean and noisy datasets, respectively.

Experimental Results

▪ **Datasets**
• *CUB-200*: controlled experiments with synthetic noise
1) Open- vs. closed-set noise   2) Uniform vs. nearest noise
• *WebVision*: large-scale experiments with real-world noise

▪ **Results on CUB-200**

Table 1: Accuracies [%] on CUB=200 with different levels and types of label noise.

Methods	Open-set noise					Closed-set noise				
	Clean dataset ($\eta = 0$)	Moderate noise level		Extreme noise level		Clean dataset ($\eta = 0$)	Moderate noise level		Extreme noise level	
		Uniform	Nearest	Uniform	Nearest		Uniform	Nearest	Unifrom	Nearest
Standard	80.57 ± 0.37	73.37 ± 0.34	77.14 ± 0.27	70.04 ± 0.71	75.45 ± 0.50	79.58 ± 0.18	63.65 ± 0.26	65.21 ± 0.42	42.35 ± 0.50	47.70 ± 0.41
Decoupling [6]	79.32 ± 0.83	71.42 ± 0.70	76.07 ± 0.40	66.79 ± 0.44	74.80 ±0.46	77.79 ± 0.23	62.52 ± 0.23	66.24 ± 0.53	43.91 ± 0.52	51.92 ± 0.18
F-correction [19]	80.66 ± 0.60	73.55 ± 0.70	77.03 ± 0.29	69.76 ± 0.59	75.52 ± 0.32	80.01 ± 0.42	63.81 ± 0.16	64.69 ± 0.21	42.23 ± 0.54	48.00 ± 0.46
S-model [18]	80.75 ± 0.37	73.52 ± 0.47	77.13 ± 0.97	70.06 ± 0.65	75.59 ± 0.33	79.42 ± 0.27	63.08 ± 0.74	64.90 ± 0.29	42.17 ± 0.70	48.01 ± 0.47
MentorNet [7]	80.39 ± 0.36	73.53 ± 0.56	77.27 ± 0.49	70.34 ± 0.42	75.75 ± 0.62	79.78 ± 0.20	68.03 ± 0.32	65.49 ± 0.14	47.74 ± 1.64	48.25 ± 0.39
q-loss ($q = 0.3$) [10]	81.55 ± 0.52	75.04 ± 0.46	78.09 ± 0.40	71.84 ± 0.95	76.32 ± 0.65	80.41 ± 0.36	68.52 ± 0.51	66.34 ± 0.25	53.18 ± 0.49	49.30 ± 0.35
q-loss ($q = 0.5$) [10]	82.19 ± 0.58	77.51 ± 0.53	78.58 ± 0.46	75.40 ± 0.66	76.47 ± 0.34	80.76 ± 0.38	75.24 ± 0.31	67.49 ± 0.56	60.89 ± 0.32	49.28 ± 0.57
q-loss ($q = 0.8$) [10]	75.15 ± 1.96	71.02 ± 1.68	47.40 ± 1.16	67.16 ± 1.29	32.60 ± 2.38	40.70 ± 2.25	29.31 ± 1.14	24.98 ± 1.61	17.67 ± 1.06	15.95 ± 0.65
Co-teaching [11]	80.90 ± 0.13	75.37 ± 0.54	77.41 ± 0.36	74.02 ± 0.29	75.57 ± 0.22	79.74 ± 0.14	68.21 ± 0.35	66.24 ± 0.30	52.72 ± 0.56	49.81 ± 0.19
CombCls	82.80 ± 0.36	79.38 ± 0.83	79.28 ± 0.52	79.19 ± 0.29	77.95 ± 0.53	81.36 ± 0.23	71.75 ± 0.24	68.35 ± 0.35	51.90 ± 0.35	52.00 ± 0.22
CombCls+Co-teaching	**82.86 ± 0.25**	**79.93 ± 0.44**	**80.22 ± 0.31**	**80.50 ± 0.61**	**78.26 ± 0.43**	**81.52 ± 0.47**	**75.30 ± 0.10**	**70.46 ± 0.31**	**62.77 ± 0.66**	**52.49 ± 0.79**

▪ **Ablation Studies**

Dataset used for meta-class set generation	Open-set	Closed-set
Clean	77.82	45.40
Moderately noisy	77.58	50.22
Extremely noisy	79.19	51.90

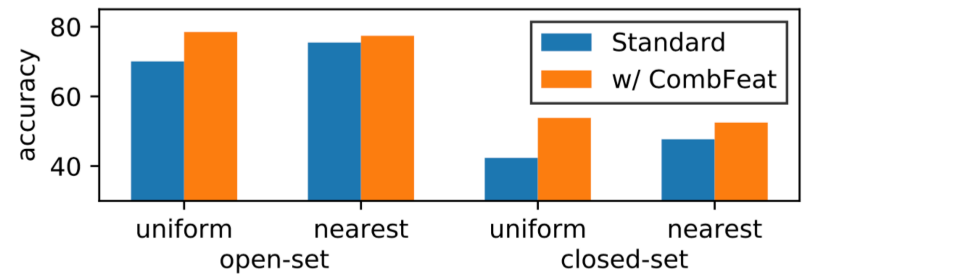


Figure 5: Results of ablation studies. Accuracies [%] of (left) combinatorial classifier trained on extremely noisy datasets with meta-class sets generated from datasets with different levels of uniform noise and (right) Standard with feature extractor of CombCls in various noise configurations.

Table 3: Results of combinatorial classification on extremely noisy datasets with different meta-class set configurations. Acc. means accuracy [%] and Param. represents ratio of model parameter count to that of Standard.

Methods	Meta-class set	Open-set noises				Closed-set noises			
		Uniform		Nearest		Unifrom		Nearest	
		Acc.	Param.	Acc.	Param.	Acc.	Param.	Acc.	Param.
Standard	N/A	70.04	1.00	75.45	1.00	42.35	1.00	42.84	1.00
CombCls	Random	78.66	1.00	76.75	1.00	46.98	0.50	48.55	0.50
	Clustering	79.19	1.00	77.95	1.00	51.90	0.50	52.00	0.50
	Clustering+Search	**79.98**	**0.42**	**78.35**	**0.44**	**54.52**	**0.41**	**52.43**	**0.29**

▪ **Model Analysis and Results on WebVision**

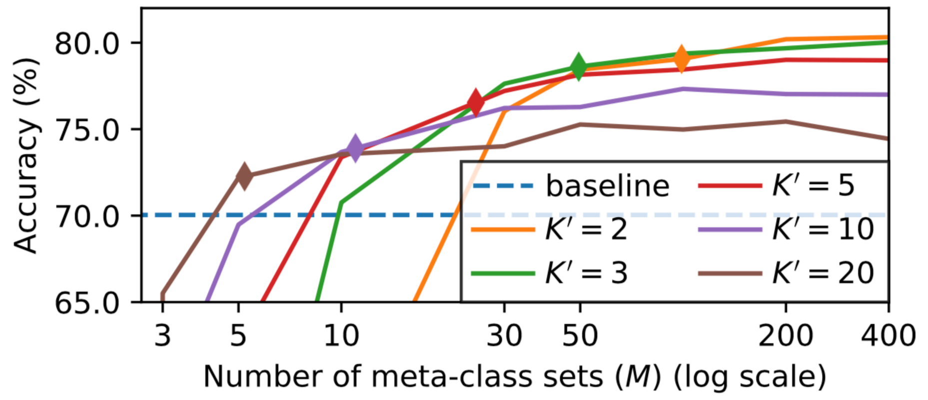


Figure 6: Accuracy [%] of combinatorial classifiers with open-set uniform noise ($\eta = 0.5$) while varying K' and M.

Table 4: Results on WebVision.

Methods	Acc. [%]
Standard	79.82
Decoupling [6]	79.38
F-correction [19]	80.96
S-model [18]	81.36
MentorNet [7]	80.46
q-loss ($q = 0.3$) [10]	82.18
Coteaching [11]	83.06
CombCls	83.14
CombCls+Search	83.26
CombCls+Coteaching	**84.14**