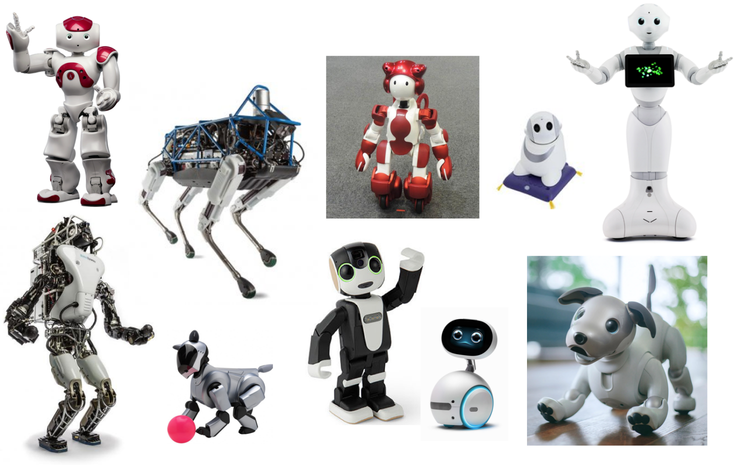
Complementary-Label Learning for Arbitrary Losses and Models

Takashi Ishida[1,2] **Gang Niu**[2]
Aditya Krishna Menon[3] **Masashi Sugiyama**[2,1]

[1] The University of Tokyo [2] RIKEN [3] Google

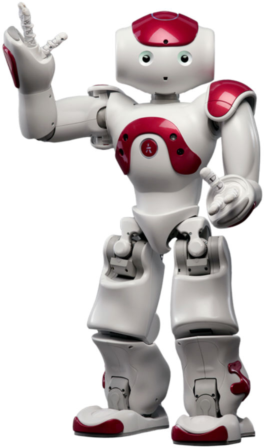
ICML 2019, Long Beach, June 13, 2019

Classify Robot images into 100 classes!



www.bostondynamics.com/robots, www.kisspng.com/png-nao-humanoid-robot-robotics-pepper-robots-716455/, japanese.engadget.com/2017/11/03/aibo/, www.sankei.com/economy/photos/160408/ecn1604080030-p4.html gpad.tv/develop/sharp-robohon-browser-program-tool-sr-b04at/, www.uni-info.co.jp/news/2017/0118_2.html www.theverge.com/2014/2/4/5378874/sonys-new-aibo-is-a-french-bulldog-named-boss, https://zenbo.asus.com/

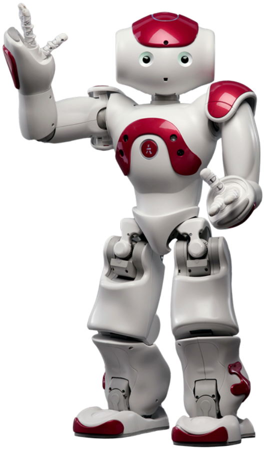
What is the name of this robot?



Class candidates

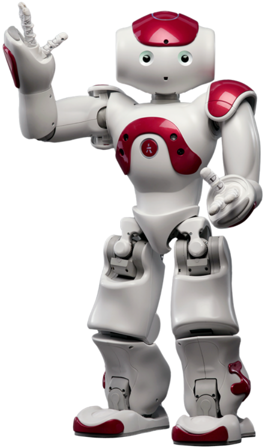
1	RoBoHoN
2	EMIEW
3	Pepper
4	Aibo
5	Atlas
	⋮
100	Spot Mini

The difficulty of labeling images



Class candidates

1 RoBoHoN
2 EMIEW
3 Pepper
4 Aibo
5 Atlas
⋮
83 Nao (Correct!)
⋮
100 Spot Mini



Class candidates

1 RoBoHoN
2 EMIEW
3 Pepper (Wrong!)
4 Aibo
5 Atlas
⋮
83 Nao
⋮
100 Spot Mini

Complementary Label:
Pepper

Goal of Our Paper

- Can we train with only complementary labels? → Yes!
  - Ishida, Niu, Hu, & Sugiyama [NeurIPS 2017]
  - Yu, Liu, Gong, & Tao [ECCV 2018]

☹ However, previous works on complementary-label learning,

→ had restrictions on losses,

→ had restrictions on models,

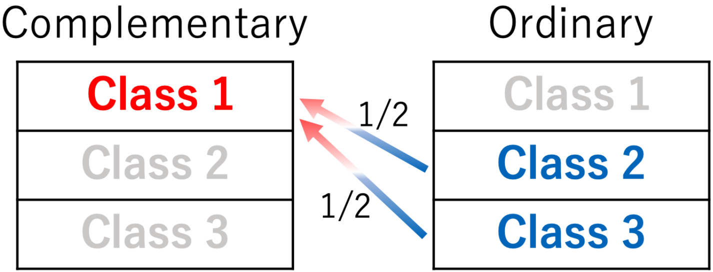
→ or did not derive an unbiased estimator

- We propose an **unbiased classification risk estimator** for complementary-label learning for arbitrary losses and models!

Main Idea

- Regard complementary-label learning as a noisy-label problem and apply noise correction!
  - Cid-Sueiro, García-García, & Santos-Rodríguez [ECML-PKDD 2014]
  - Natarajan, Dhillon, Ravikumar, & Tewari [NeurIPS 2013]

→ *Complementary labels are noisy labels with uniform transition from other (true) classes*



Main Discovery

- Unbiased risk estimation is possible w/o loss/model restrictions:

$$\mathbb{E}_{p(x,y)}[\ell(y,\boldsymbol{g}(x))] = \mathbb{E}_{\overline{p}(x,\overline{y})}\Big[-(K-1)\cdot\ell(\overline{y},\boldsymbol{g}(x))+\textstyle\sum_{j=1}^{K}\ell(j,\boldsymbol{g}(x))\Big]$$

  - Assumption: $\overline{p}(\overline{y}|x) = \sum_{y\neq\overline{y}} p(y|x)/(K-1)$
  - $\ell : [K] \times \mathbb{R}^K \to \mathbb{R}_+$ is loss function
  - $\boldsymbol{g} : x \to \mathbb{R}^K$: decision function
  - $\mathbb{E}$ denotes the expectation
  - x: pattern, y: true class label, $\overline{y}$: complementary class label
  - $p(x,y)$: joint ordinary distribution
  - $\overline{p}(x,\overline{y})$: joint complementary distribution

Conclusions

- Proposed general risk estimator for learning from complementary labels.
- Does not have restrictions on loss function or the model.

Come see our poster @ Pacific Ballroom #181 for **more**!

→ Further correction schemes of the learning objective, experiments, etc.