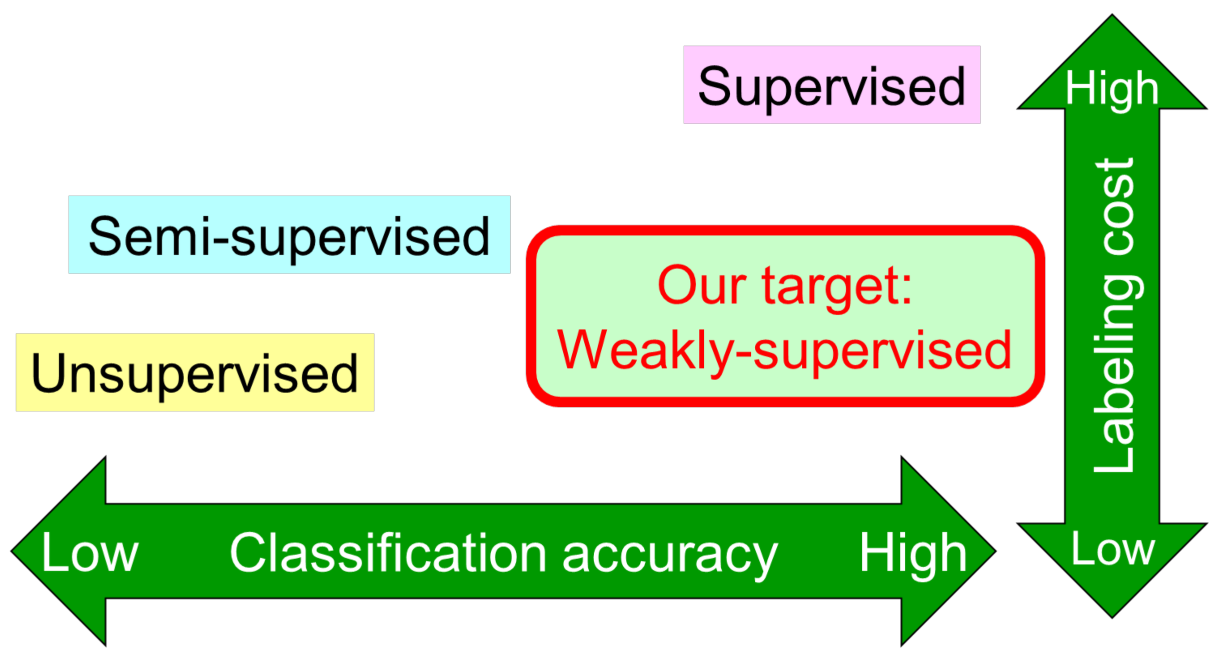
不完全情報学習チーム　杉山 将
Imperfect Information Learning Team
Masashi Sugiyama

Team's Vision and Social Impact:

- Develop novel algorithms that allow accurate learning from data with limited information.
- Enable machine learning in applications with imperfect and limited data, such as health care and natural disasters.



Research Activities:

- Develop practical algorithms that have theoretical support.
- Help applied researchers use our algorithms in their problems.

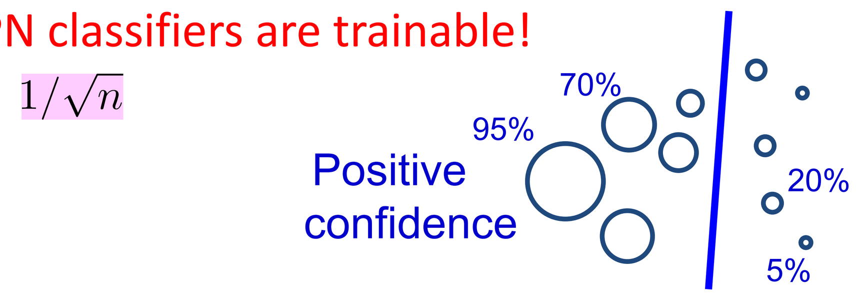
Weakly-Supervised Classification

- Real-world data often contains only limited information.
- We develop algorithms to learn from such data.

Pconf Classification

- Problem: Only positive (P) data is available, not negative (N) and unlabeled (U) data:
  - Data from rival companies cannot be obtained.
  - Only positive results are reported (publication bias).
- "Only-P learning" is unsupervised.
- Results: From Positive-conf (Pconf) data, PN classifiers are trainable!

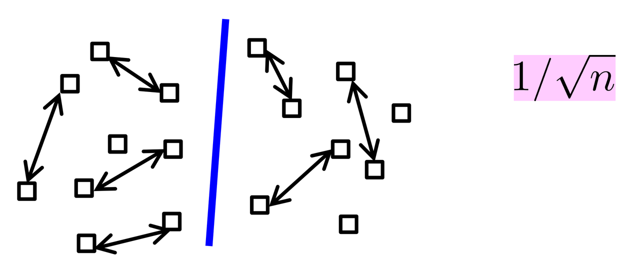
$1/\sqrt{n}$



Ishida, Niu & Sugiyama (NeurIPS2018)

SU Classification

- Problem: Delicate classification (salary, religion…):
  - Highly hesitant to directly answer questions.
  - Less reluctant to just say "same as him/her".
- Results: From similar and unlabeled (SU) data, PN classifiers are trainable!



$1/\sqrt{n}$

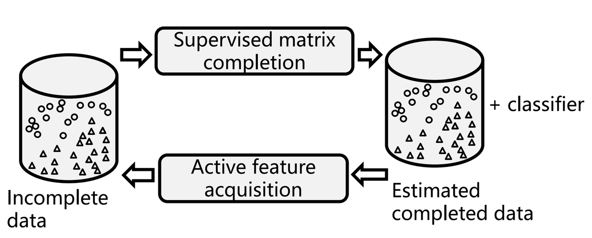
Bao, Niu & Sugiyama (ICML2018)

Active Feature Acquisition

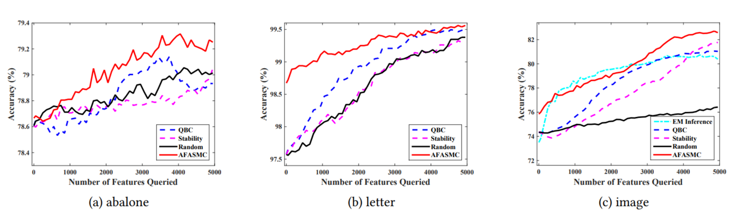
- Problem: features are missing in many real world applications

Patient	Weight	Blood Pressure	CT	…	Diagnosis
No.1	54.1	?	N	…	Yes
No.2	84.9	114	?	…	No
No.3	?	138	P	…	Yes
…	…	…	…	…	…
No.N	76.3	?	?	…	No

- Goal: better diagnosis + lower feature acquisition cost
- Method:
  - Supervised matrix completion
  - Active feature acquisition



- Results:
  - Lower acquisition cost + better accuracy



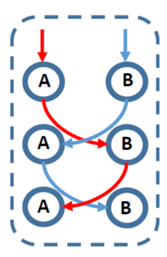
Huang, Xu, Xie, Sugiyama, Niu and Chen (KDD2018)

Robust Learning

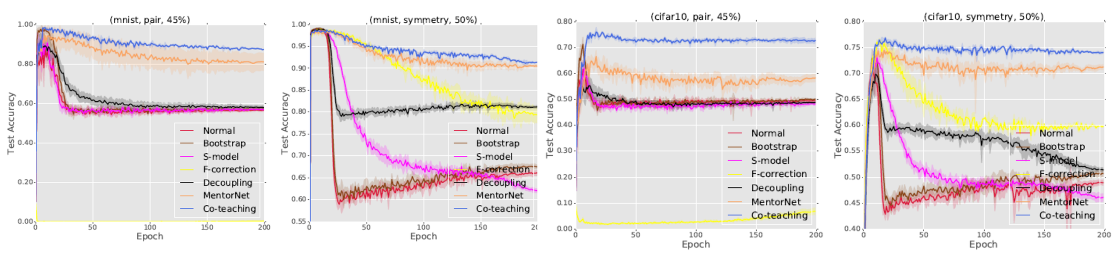
- Real-world data can be highly noisy.
- We develop robust algorithms that learn accurately from extremely noisy data.

Co-teaching

- Goal: Robust DNN training with noisy train data.
- Proposal: Co-teaching with two networks.
  - In each mini-batch data, each network samples its small-loss instances as the useful knowledge, and teaches such useful instances to another network.



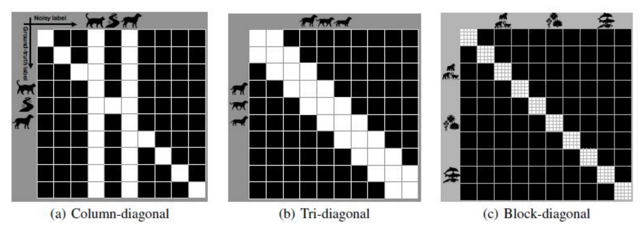
- Effectiveness: Works well (but no theory)!



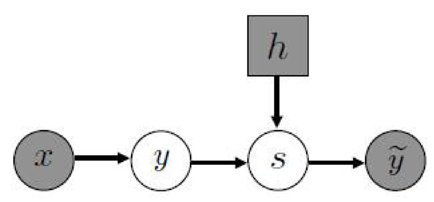
Han, Yao, Yu, Niu, Xu, Hu, Tsang & Sugiyama (NeurIPS2018)

Masking

- Goal:
  - Convey human cognition of invalid class transition



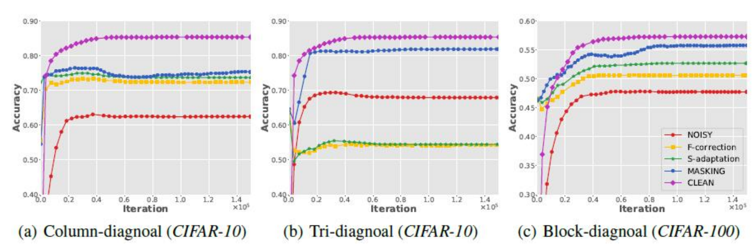
- Model:
  - Structure-aware probabilistic model



- Results:

Models	Performance(%)
NOISY	68.9
F-correction	69.8
S-adaptation	70.3
MASKING	71.1
CLEAN	75.2

Clothing1M dataset



Han, Yao, Niu, Zhou, Tsang, Zhang & Sugiyama (NeurIPS2018)

Distributionally Robust Learning

- Problem: Training and testing data distribution may differ.

- Distributionally robust learning:
  - Being robust to the worst test distribution.
- Our finding: In classification, this merely results in the same (non-robust) classifier.
  - Since the 0-1 loss is different from a surrogate loss.

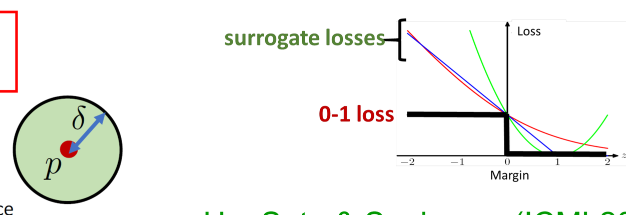
$$\min_{\theta} \sup_{q \in \mathcal{Q}_p} \mathbb{E}_{q(x,y)}[\ell(g_\theta(x), y)]$$

$$\mathcal{Q}_p = \{q \mid \mathrm{D}_f(q \| p) \leq \delta\}$$

"f-divergence ball"

[Bagnell 2005, Ben-Tal+ 2013, Namkoong+ 2016, 2017]

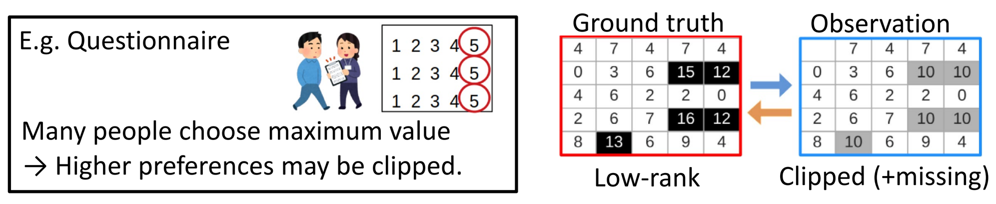
E.g. KL divergence, Chi-square divergence



Hu, Sato & Sugiyama (ICML2018)

Clipped Matrix Completion

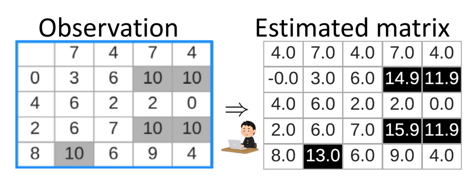
- Goal: Restore a low-rank matrix from clipped values.



- Method: Minimize square loss + squared-hinge loss.

$$\min_X \frac{1}{2}\Sigma_{ij:non-clipped}(obs_{ij} - X_{ij})^2 + \frac{1}{2}\Sigma_{ij:clipped}\max(0, obs_{ij} - X_{ij})^2 + R(X)$$

- Theoretical support:
  - The problem is solvable.
  - Recovery is accurate.



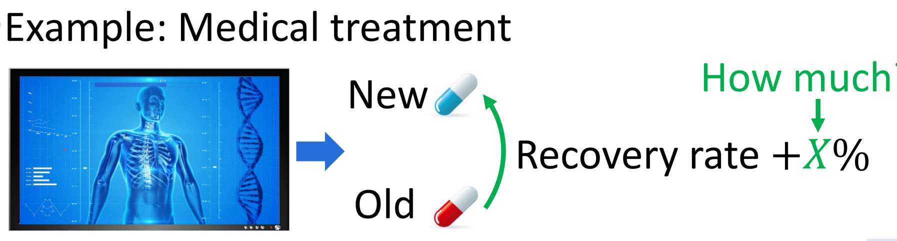
Teshima, Xu, Sato & Sugiyama (AAAI2019)

Decision Making

- Real-world interaction data can be expensive.
- We develop algorithms to learn an optimal decision maker from limited interaction data.

Uplift Modeling

- Goal: Estimate the relative impact of treatment.
  - Example: Medical treatment

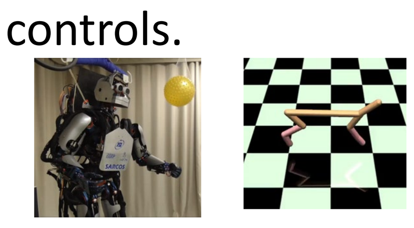


- Standard data: treatment t & outcome y. (t, y)
- Our setup: one of them is always missing
  - No complete data available. $(t, \cancel{y})$ or $(\cancel{t}, y)$
- Result: Estimation is still feasible under reasonable assumptions.
- Proposed method is fast & accurate.

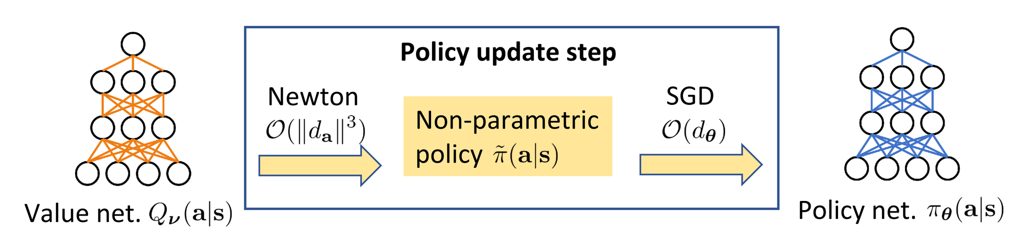
Yamane, Yger, Atif & Sugiyama (NeurIPS2018)

Deep Reinforcement Learning (DRL)

- Goal: DRL for high-dimensional controls.
  - Example: Robotics



- Issues:
  - Data is expensive to obtain.
  - Efficiency methods does not scale well.
- Our solution: 2nd-order DRL method with linear computational complexity.



Tangkaratt, Abdolmaleki & Sugiyama (ICLR2018)