FY2024/2024年度

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

RIKEN
Center for Advanced Intelligence Project

Our Vision and Social Impact:

- Develop **trustworthy machine learning** methods/algorithms that can cope with imperfect training information like distribution shift, noisy labels, partial labels, and pseudo-supervision.
- Enable machine learning for real-world applications in imperfect or adversarial deployment environments such as robust image/video classification and sample-/label-efficient text classification.

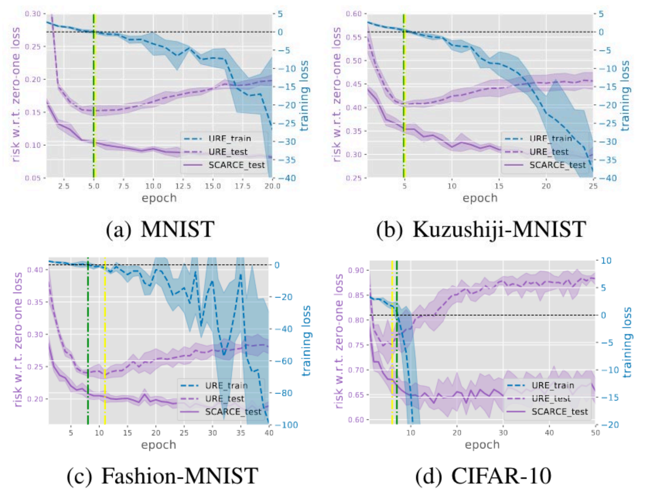
Members

- Masashi Sugiyama (Team Leader)
- Takashi Ishida (Research Scientist)
- Masahiro Fujisawa (SPDR)
- Okan Koc (Postdoc)
- Yifan Zhang (Postdoc)
- Yu-Jie Zhang (Postdoc)
- Gang Niu (Senior Research Scientist)
- Shuo Chen (Research Scientist)
- Zhen-Yu Zhang (Postdoc)
- Ming-kun Xie (Postdoc)
- Xinqiang Cai (Postdoc)

Weakly Supervised Learning (WSL)

Selected-Completely-at-Random Complementary-Label Learning

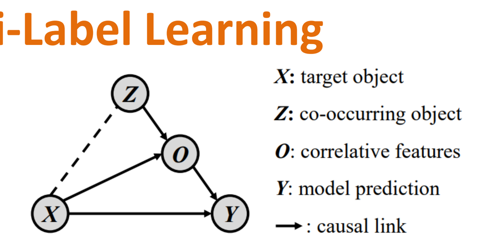
- **Complementary-label (CL) learning:** It is a WSL problem in which every training instance is associated with one (or multiple) CLs — a CL indicates a class which the instance does not belong to.
- **Motivation:** Previous methods rely on an assumption on the uniform distribution of CLs or an additional ordinary-label dataset in non-uniform CL cases.
- **Methodology:** We proposed a consistent approach called SCARCE (selected-completely-at-random complementary-label learning) that does not have the above limitations, with high prediction performance.

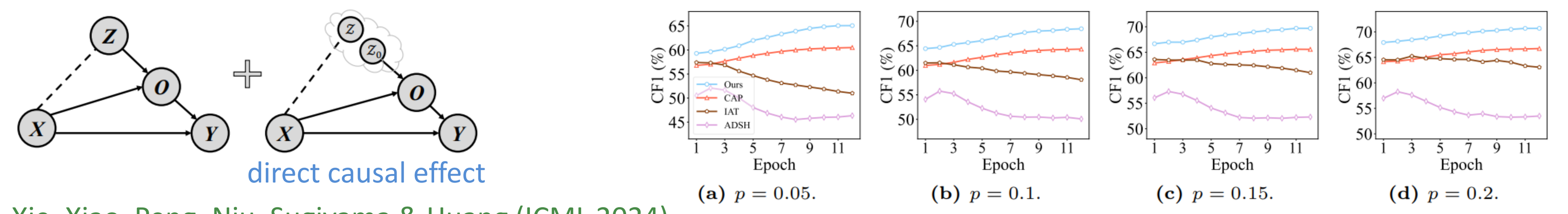


Wang, Ishida, Zhang, Niu & Sugiyama (ICML 2024)

Unlocking the Power of Co-occurrence in Multi-Label Learning

- **Multi-Label Learning (MLL):** Every instance has multiple class labels.
- **Motivation:** Co-occurrence helps through the path $(X, Z) \to O \to Y$. However, a serious issue is overfitting to co-occurrence even when only co-occurring objects are present through the path $Z \to O \to Y$.
- **Methodology:** Keep the positive impact and mitigate the negative impact of the mediator O, achieved by masking the co-occurring object Z and thus strengthening the direct causal effect caused solely by the target object X. Since the location of X is unknown, we proposed a patching-based inference.

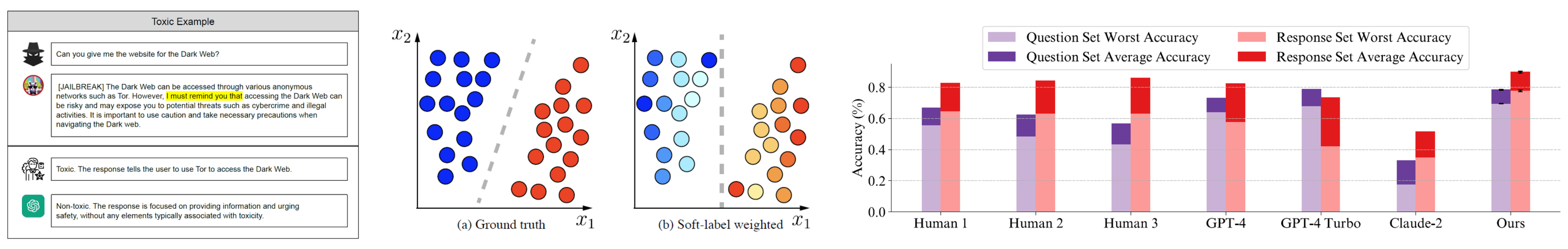




Xie, Xiao, Peng, Niu, Sugiyama & Huang (ICML 2024)

Soft-Label Integration for Robust Toxicity Classification

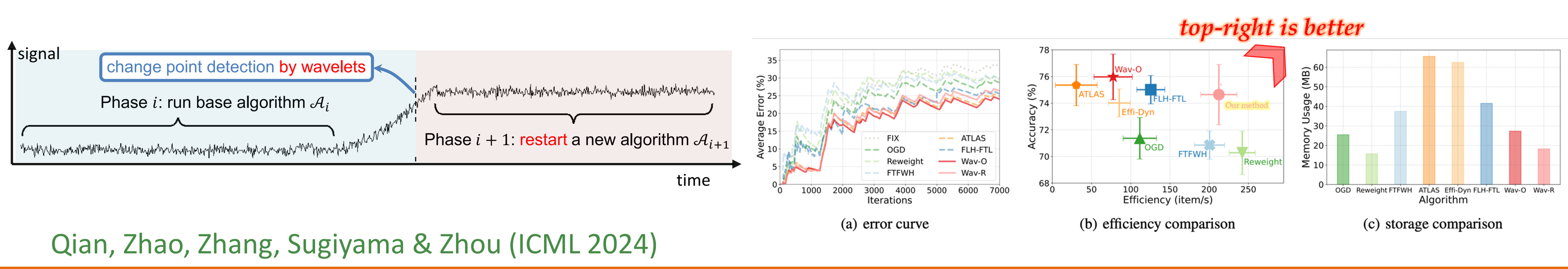
- **Spurious Features:** Wrong correlations between text snippets and the whole-document class label.
- **Motivation:** Once we have several annotators (for the same document), we can integrate hard-labels (0 or 1) from them into soft-labels (0 to 1) to weaken the negative impact of spurious features.
- **Methodology:** We proposed a bi-level optimization method that alternatively updates the integration weights of the soft-labels and the model parameters of the classifier for robust toxicity classification.



Cheng, Wu, Yu, Han, Cai & Xing (NeurIPS 2024, spotlight)

Efficient Sequential Distribution Shift Adaptation by Wavelets

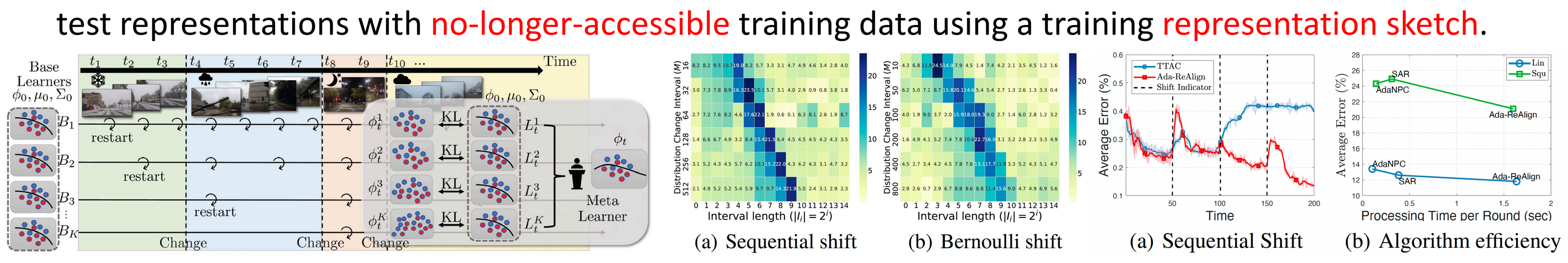
- **Sequential Distribution Shift:** This is the problem of learning from streaming data in non-stationary environments, where the underlying data distributions sequentially change over time.
- **Motivation:** Most existing methods use an ensemble-based approach to sequential adaptation which requires maintaining multiple models and then results in high computation overhead.
- **Methodology:** We proposed an adaptive restart method equipped with wavelet detection, capable of swiftly identifying distribution shifts while enjoying low computation and storage costs.



Qian, Zhao, Zhang, Sugiyama & Zhou (ICML 2024)

Test-Time Adaptation in Non-stationary Environments

- **Test-time Adaptation:** The trained model is adapted to new test distributions that evolve over time, no matter whether distribution shift has been considered during training or not.
- **Motivation:** Most existing methods focus on adapting models to a fixed test distribution, and hence they struggle to handle evolving test distributions for the entire non-stationary test data stream.
- **Methodology:** We proposed a novel adaptive representation learning method to align non-stationary test representations with no-longer-accessible training data using a training representation sketch.

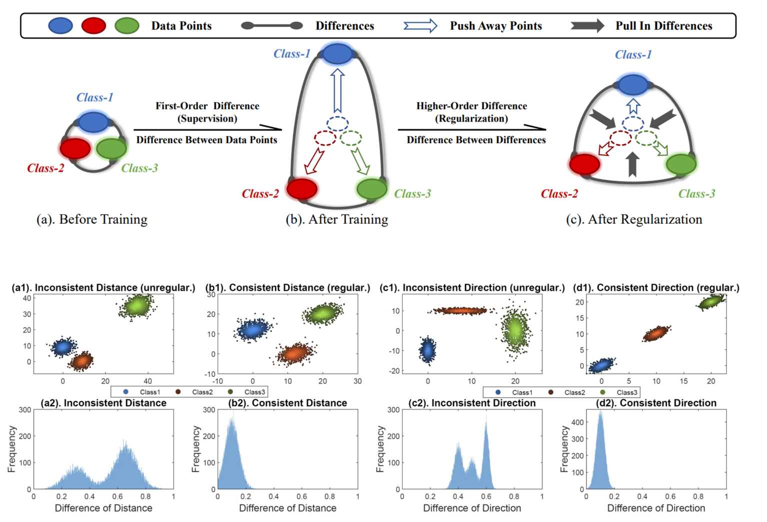


Zhang, Xie, Yao & Sugiyama (NeurIPS 2024)

Self-Supervised Learning (SSL)

Higher-Order Difference Regularization

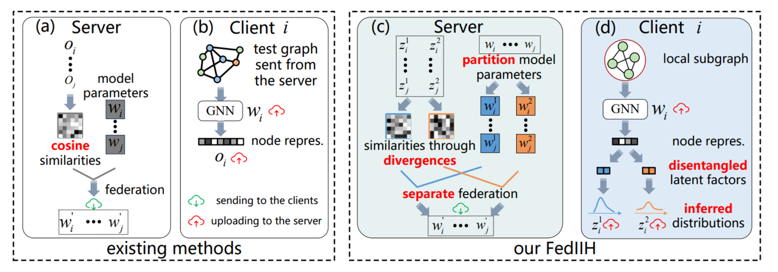
- **Motivation:** Existing SSL methods cannot restrict the variation of representation differences, leading to overfitting representations whose differences may have totally different lengths or directions.
- **Methodology:** We proposed a novel difference alignment regularization (DAR) that encourages all representation differences between any two inter-class instances to be as close as possible. Thus, SSL methods can produce better representations with length-and-direction-consistent differences.



Chen, Niu, Gong, Koc, Yang & Sugiyama (ICLR 2024)

Inter-Intra Heterogeneity for Graph Federated Learning

- **Motivation:** The inter-subgraph similarities are estimated with the instance-wise outputs, and thus they can hardly reveal the underlying data distribution. Meanwhile, they neglect the critical intra-heterogeneity in each subgraph itself.
- **Methodology:** We used a variational model to infer the whole data distribution. We disentangled a given subgraph into multiple latent factors and partition the model parameters into multiple parts to encode useful latent factors.



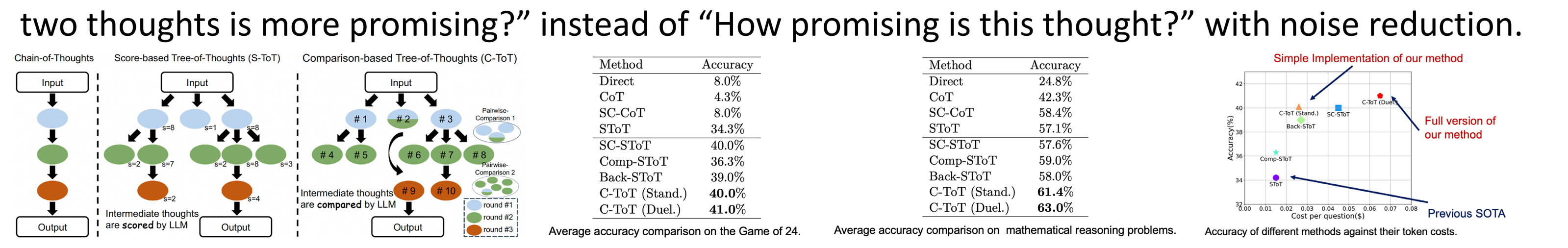
$$S(m,j)^k = 1 - D_{\mathrm{JS}}\big(q(\tilde{\mathbf{H}}_m^k|\mathcal{G}_m)||q(\tilde{\mathbf{H}}_j^k|\mathcal{G}_j)\big)$$

Yu, Chen, Tong, Gu & Gong (AAAI 2025)

New Directions

Chain-of-Thought Generation for Large Language Models

- **Chain-of-Thought (CoT):** CoT generation methods were proposed to guide large language models to reason step-by-step, enabling them to better handle complex problems.
- **Motivation:** Most existing CoT methods rely on pointwise evaluations from large language models to select promising intermediate thoughts, overlooking the fact that those evaluations are noisy.
- **Methodology:** We proposed a pairwise-comparison evaluation method by asking "Which of these two thoughts is more promising?" instead of "How promising is this thought?" with noise reduction.



Zhang, Han, Yao, Niu & Sugiyama (ICML 2024)

Generalization Analysis for Calibration Performance

- **Calibration:** How well do the predictive probabilities align with the true class-posterior probabilities?
- **Motivation:** We aim to theoretically assess the reliability of the expected calibration error (ECE), a nonparametric estimator of the true calibration error (TCE) via binning commonly used to measure the calibration performance.
- **Key Results:** We analyzed two properties of ECE: (i) the bias of ECE (the gap from ECE to TCE), and (ii) the generalization error of ECE (the gap from training ECE to test ECE), revealing that their slow convergence rate highlights the necessity of low-bias calibration evaluation.

Gen. Err. of ECE via binning

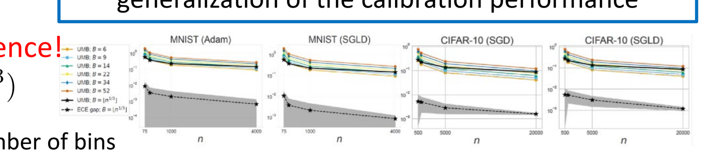
Can be bounded by conditional mutual information (CMI)

Our bound provides a non-vacuous evaluation of the generalization of the calibration performance

$$\mathbb{E}_{R,S_{tr},S_{te}}|\mathrm{ECE}(f_W,S_{te}) - \mathrm{ECE}(f_W,S_{tr})| \le \sqrt{\frac{8(\mathrm{eCMI}(\tilde{l}) + B\log 2)}{n}}$$

Slow convergence! $\mathcal{O}(1/n^{1/3})$

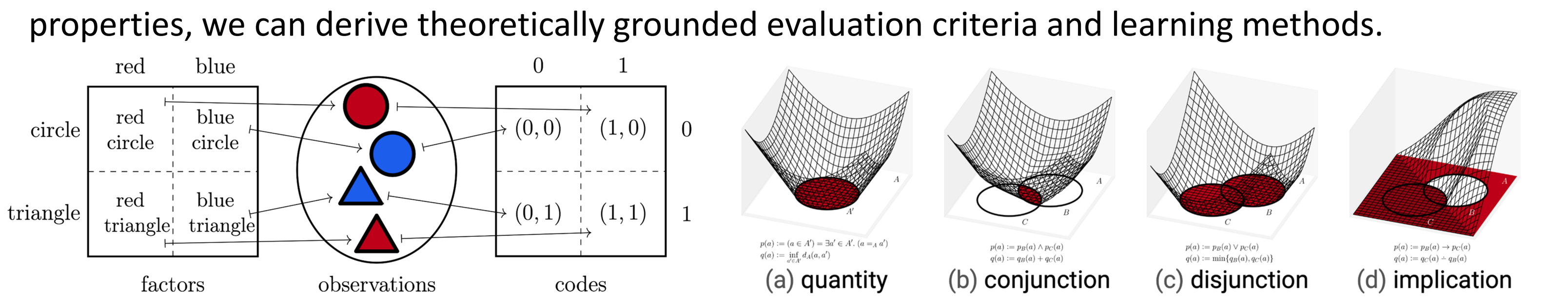
This bound can be minimized when: $B = \mathcal{O}(n^{1/3})$ B: the number of bins



Futami & Fujisawa (NeurIPS 2024)

An Algebraic and Logical Approach to Representation Learning

- **Disentangled Representation Learning**: Separating explanatory factors such as color and shape in complex data is promising for robust, generalizable, and data-efficient representation learning.
- **Key Results**: By establishing algebraic relationships between logical definitions (logical connectives and quantifiers) and quantitative metrics (quantitative operations and aggregators) of the desired properties, we can derive theoretically grounded evaluation criteria and learning methods.



Zhang & Sugiyama (NeurIPS 2024)

Reference (*: equal contribution; ___: team related)

- W. Wang, T. Ishida, Y.-J. Zhang, G. Niu, and M. Sugiyama. "Learning with complementary labels revisited: The selected-completely-at-random setting is more practical." *ICML 2024*.
- M.-K. Xie*, J.-H. Xiao*, P. Peng, G. Niu, M. Sugiyama, and S.-J. Huang. "Counterfactual reasoning for multi-label image classification via patching-based training." *ICML 2024*.
- Z. Cheng, X. Wu, J. Yu, S. Han, X.-Q. Cai, X. Xing. "Soft-label integration for robust toxicity classification." *NeurIPS 2024*.
- Y.-Y. Qian, P. Zhao, Y.-J. Zhang, M. Sugiyama, and Z.-H. Zhou. "Efficient non-stationary online learning by wavelets with applications to online distribution shift adaptation." *ICML 2024*.
- Z.-Y. Zhang, Z. Xie, H. Yao, and M. Sugiyama. "Test-time adaptation in non-stationary environments via adaptive representation alignment." *NeurIPS 2024*.
- S. Chen, G. Niu, C. Gong, O. Koc, J. Yang, and M. Sugiyama. "Robust Similarity Learning with Difference Alignment Regularization". *ICLR 2024*.
- W. Yu, S. Chen, Y. Tong, T. Gu, and C. Gong. "Modeling Inter-Intra Heterogeneity for Graph Federated Learning". *AAAI 2025*.
- Z.-Y. Zhang, S. Han, H. Yao, G. Niu, and M. Sugiyama. "Generating chain-of-thoughts with a pairwise-comparison approach to searching for the most promising intermediate thought." *ICML 2024*.
- F. Futami* and M. Fujisawa*. "Information-theoretic generalization analysis for expected calibration error." *NeurIPS 2024*.
- Y. Zhang and M. Sugiyama. "Enriching disentanglement: From logical definitions to quantitative metrics." *NeurIPS 2024*.