NTT

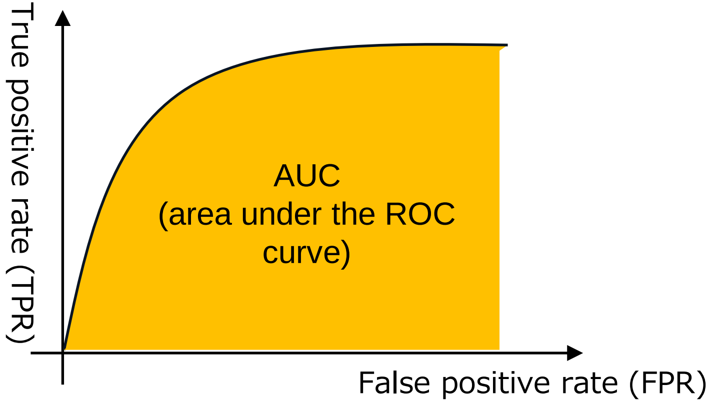
AUC Maximization under Positive Distribution Shift

Atsutoshi Kumagai[†], Tomoharu Iwata[†], Hiroshi Takahashi[†], Taishi Nishiyama [*][†], Yasuhiro Fujiwara[†]

[†] NTT
[*] NTT Security Holdings

AUC Maximization

- AUC is a standard evaluation metric for imbalanced binary classification.
- By maximizing the AUC with positive and negative training data, good classifiers can be learned even from imbalanced data.



$$\mathrm{AUC}(s) = \mathbb{E}_{\mathbf{x}^{\mathrm{p}} \sim p^{\mathrm{p}}(\mathbf{x})} \mathbb{E}_{\mathbf{x}^{\mathrm{n}} \sim p^{\mathrm{n}}(\mathbf{x})} \left[I(s(\mathbf{x}^{\mathrm{p}}) > s(\mathbf{x}^{\mathrm{n}})) \right],$$

AUC is the probability that a classifier *s* will rank a randomly drawn positive data higher than a randomly drawn negative data.

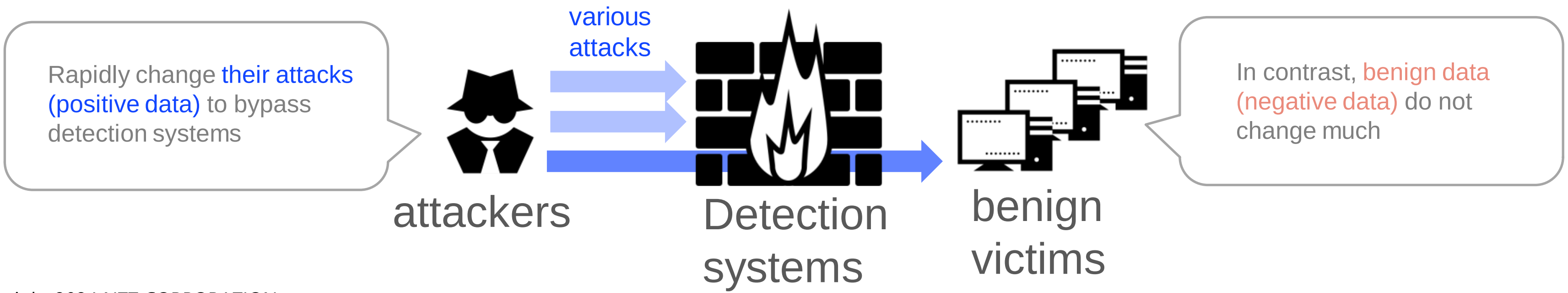
- However, AUC maximization methods usually do not consider distribution shifts.

Positive Distribution Shift

- We consider **Positive distribution shift :**
  - Negative-conditional density does not change, but positive-conditional density can vary between the training and testing phase.

$$p^{\mathrm{n}}_{\mathrm{tr}}(\mathbf{x}) = p^{\mathrm{n}}_{\mathrm{te}}(\mathbf{x}), \quad p^{\mathrm{p}}_{\mathrm{tr}}(\mathbf{x}) \neq p^{\mathrm{p}}_{\mathrm{te}}(\mathbf{x}).$$

- Positive distribution shift often occurs in real-world imbalanced classification.
  - **Cyber security**, medical care, visual inspection etc.



Our Main Results

- Under the positive distribution shift, **AUC on the test distribution can be maximized** by using **positive** and **unlabeled data** in the training distribution and **unlabeled data** in the test distribution.

$$f(\mathbf{x}^{\mathrm{p}}, \mathbf{x}^{\mathrm{n}}) := \sigma(s(\mathbf{x}^{\mathrm{p}}) - s(\mathbf{x}^{\mathrm{n}}))$$

Sigmoid func.

Score func. (to be learned)

$$\mathrm{AUC}_\sigma(s) = \mathbb{E}_{\mathbf{x}^{\mathrm{p}} \sim p_{\mathrm{te}}^{\mathrm{p}}(\mathbf{x})} \mathbb{E}_{\mathbf{x}^{\mathrm{n}} \sim p_{\mathrm{te}}^{\mathrm{n}}(\mathbf{x})} \left[f(\mathbf{x}^{\mathrm{p}}, \mathbf{x}^{\mathrm{n}}) \right]$$

$$\propto \left[\mathbb{E}_{\mathbf{x} \sim p_{\mathrm{te}}(\mathbf{x})} \mathbb{E}_{\bar{\mathbf{x}} \sim p_{\mathrm{tr}}(\mathbf{x})} \left[f(\mathbf{x}, \bar{\mathbf{x}}) \right] - \pi_{\mathrm{tr}} \mathbb{E}_{\mathbf{x} \sim p_{\mathrm{te}}(\mathbf{x})} \mathbb{E}_{\mathbf{x}^{\mathrm{p}} \sim p_{\mathrm{tr}}^{\mathrm{p}}(\mathbf{x})} \left[f(\mathbf{x}, \mathbf{x}^{\mathrm{p}}) \right] \right] + C$$

Positive class-prior in the training dist. (can be estimated)

Advantages of the derived AUC (objective func.)

- Requires **only positive labels in the training dist.** as supervision.
- Is a simple form, which is **easy to implement**.
- **Can use any differentiable score functions** such as linear models and neural networks.
- **Does not require the positive class-prior in the test dist.,** which is difficult to obtain.

Experimental Results

NTT

Our method outperforms various existing methods in imbalanced classification

			Use PU data in the train dist.				Use PU data in the train dist. & U data in the test dist.				
Data	π_{te}	Ours	CE	nnPU	puAUC	PULA	AnnPU	ApuAUC	APULA	CpuAUC	PURR
MNIST	0.1	**0.723**	**0.796**	**0.782**	**0.801**	**0.802**	**0.782**	**0.798**	**0.793**	**0.804**	**0.815**
	0.2	**0.854**	0.796	0.784	0.801	0.802	0.785	0.793	0.769	0.803	**0.871**
	0.3	**0.902**	0.796	0.783	0.800	0.802	0.780	0.786	0.745	0.802	**0.914**
Fashion MNIST	0.1	0.787	0.932	0.825	0.937	**0.954**	0.772	0.920	**0.939**	0.943	0.920
	0.2	**0.891**	0.930	0.825	0.937	**0.955**	0.771	0.890	**0.911**	0.943	**0.929**
	0.3	**0.960**	0.930	0.825	**0.938**	**0.955**	0.866	0.870	0.863	**0.944**	**0.961**
SVHN	0.1	**0.554**	0.504	0.501	0.518	0.494	0.501	0.512	0.492	0.524	0.511
	0.2	**0.660**	0.503	0.501	0.518	0.494	0.501	0.507	0.490	0.523	0.522
	0.3	**0.736**	0.503	0.501	0.518	0.494	0.501	0.507	0.490	0.523	0.519
CIFAR10	0.1	**0.727**	**0.682**	0.455	**0.749**	**0.751**	0.489	**0.739**	**0.750**	**0.750**	0.434
	0.2	**0.825**	0.679	0.467	0.741	0.739	0.491	0.732	0.742	0.744	0.536
	0.3	**0.874**	0.682	0.451	0.750	0.738	0.545	0.722	0.750	0.749	0.710
# best		11	2	1	3	5	1	2	4	3	5

Table: Average Test AUCs

When the class-prior in the test dist. is available, we can further boost our method's performance by using loss corrections. Please check the paper for details!