Gain-Some-Lose-Some: Reliable Quantification Under General Dataset Shift

Benjamin Denham, Edmund M-K Lai,
Roopak Sinha, and M. Asif Naeem

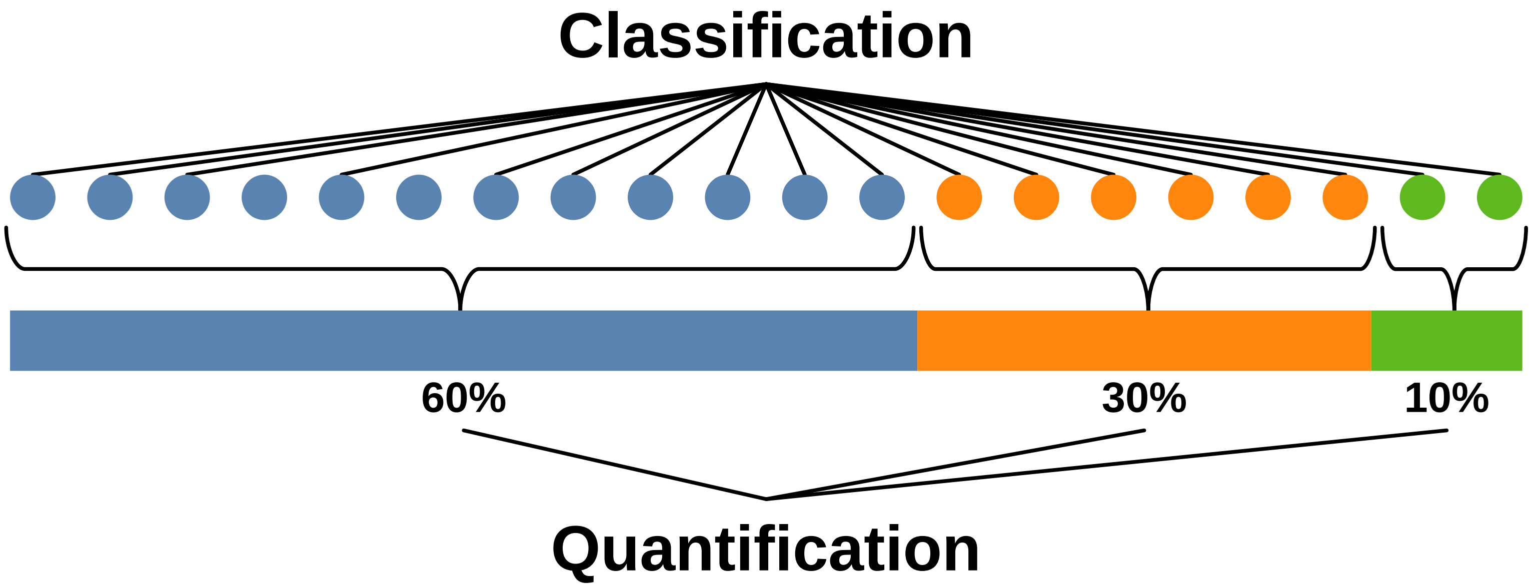
Funded by a Callaghan Innovation R&D Fellowship Grant FPAP1902 for Fisher & Paykel Appliances and a Doctoral Fees Scholarship from Auckland University of Technology

Presented by: Benjamin Denham (ben.denham@aut.ac.nz)

Learning Task: Quantification

Supervision: Training set of labelled instances with features (X) and class labels (Y)

Prediction target: Proportions of classes in unlabelled test sets



Dataset Shift in Quantification

Train/Source Dist. ❯ **Test/Target Dist.**

$$P^S(X,Y) \qquad P^T(X,Y)$$

Dataset Shift in Quantification

Train/Source Dist. ❯ **Test/Target Dist.**

$$P^S(X,Y) \qquad P^T(X,Y)$$

Shift Type	Assumptions	Methods
No shift	$P^S(X,Y) = P^T(X,Y)$	CC, PCC [1]

1. K. Keith and B. O'Connor, "Uncertainty-aware generative models for inferring document class prevalence," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 4575–4585.

Dataset Shift in Quantification

Train/Source Dist. $P^S(X,Y)$ ❯ **Test/Target Dist.** $P^T(X,Y)$

Shift Type	Assumptions	Methods
No shift	$P^S(X,Y) = P^T(X,Y)$	CC, PCC [1]
Prior shift	$P^S(Y) \neq P^T(Y)$ $P^S(X\|Y) = P^T(X\|Y)$	EM [2]

1. K. Keith and B. O'Connor, "Uncertainty-aware generative models for inferring document class prevalence," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 4575–4585.
2. M. Saerens, P. Latinne, and C. Decaestecker, "Adjusting the outputs of a classifier to new a priori probabilities: a simple procedure," *Neural computation*, vol. 14, no. 1, pp. 21–41, 2002.

Dataset Shift in Quantification

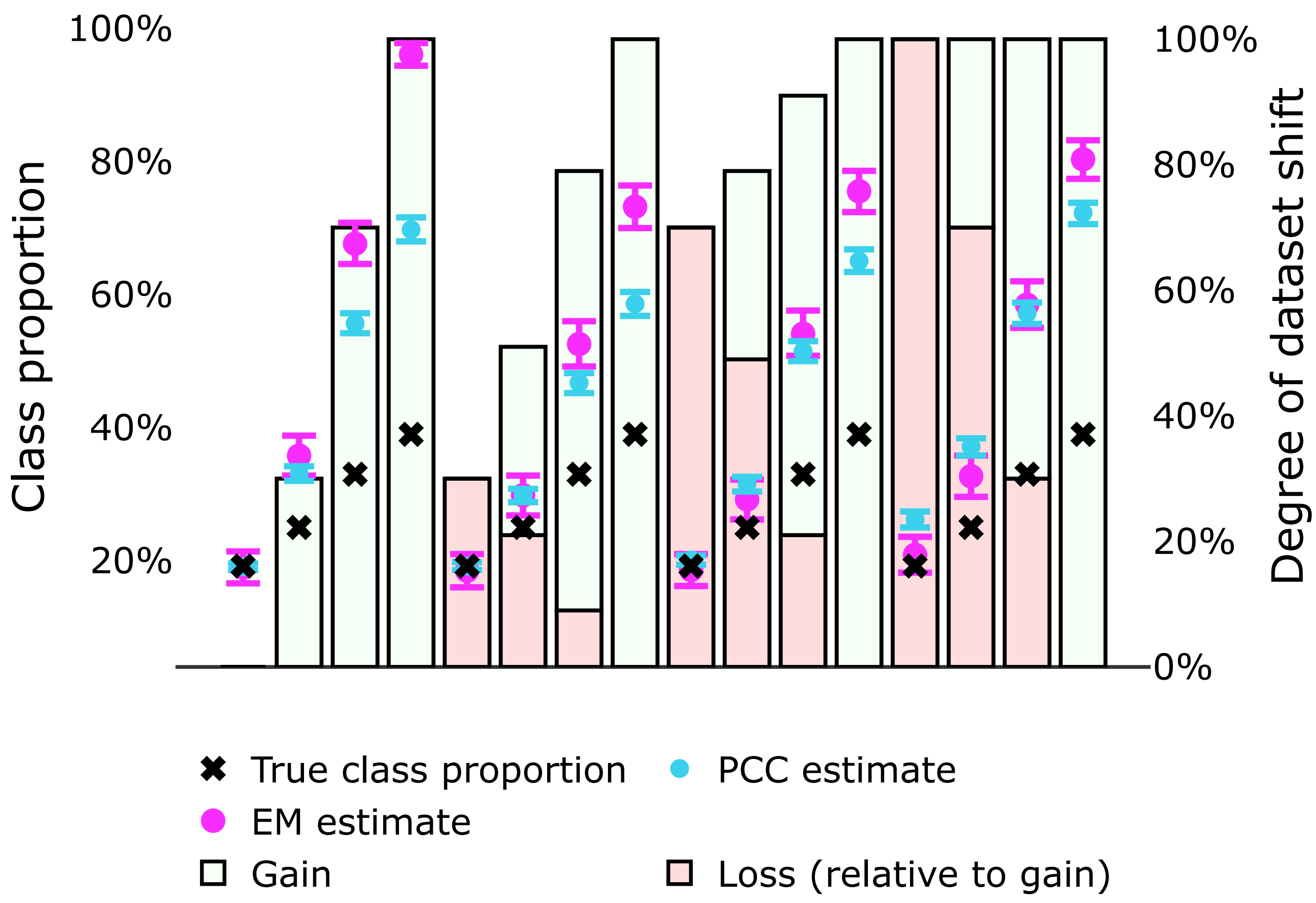
Train/Source Dist. ❯ **Test/Target Dist.**

$P^S(X, Y)$ ❯ $P^T(X, Y)$

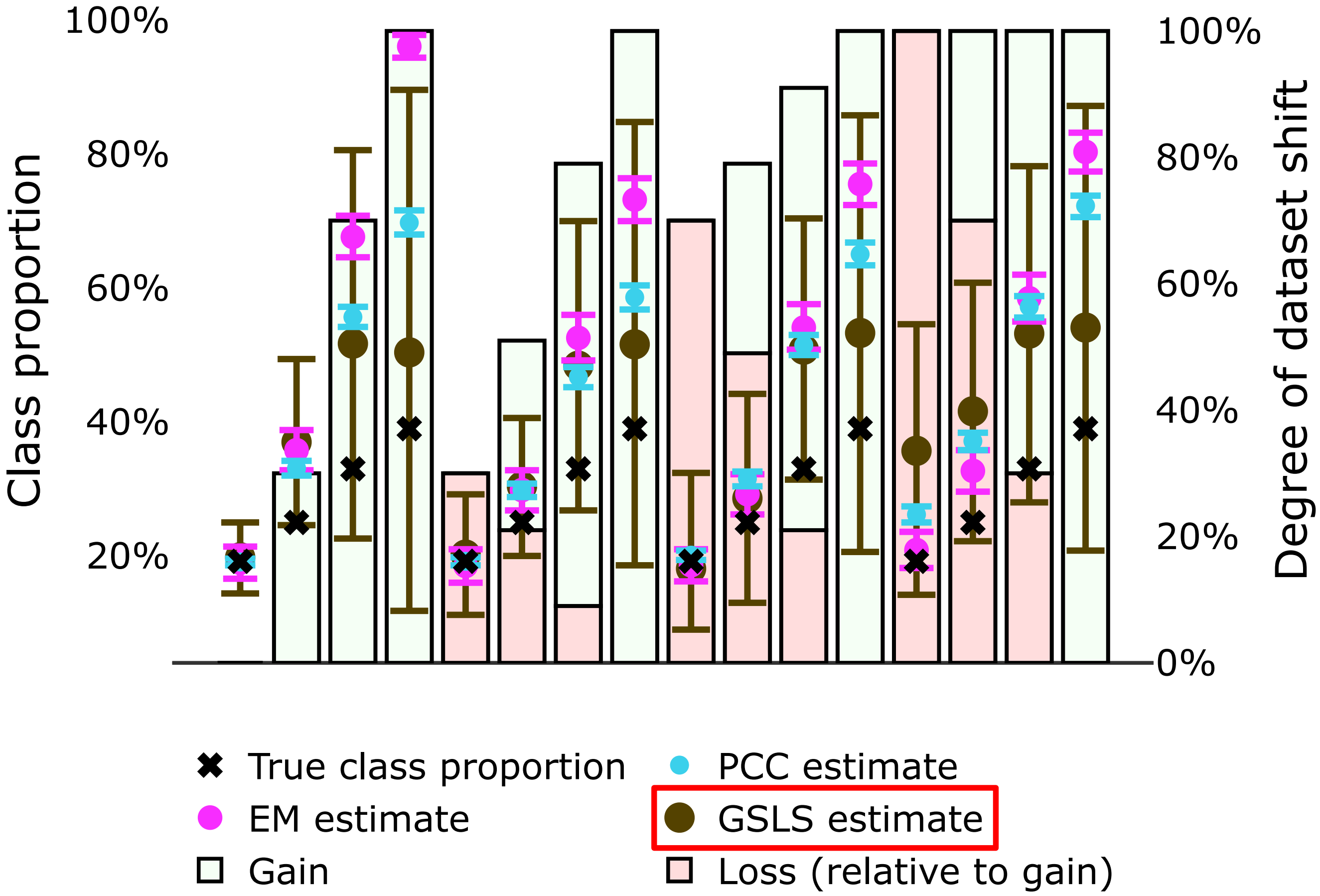
Shift Type	Assumptions	Methods
No shift	$P^S(X,Y) = P^T(X,Y)$	CC, PCC [1]
Prior shift	$P^S(Y) \neq P^T(Y)$ $P^S(X\|Y) = P^T(X\|Y)$	EM [2]
General shift	$P^S(X,Y) \neq P^T(X,Y)$	Proposed GSLS

1. K. Keith and B. O'Connor, "Uncertainty-aware generative models for inferring document class prevalence," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 4575–4585.
2. M. Saerens, P. Latinne, and C. Decaestecker, "Adjusting the outputs of a classifier to new a priori probabilities: a simple procedure," *Neural computation*, vol. 14, no. 1, pp. 21–41, 2002.

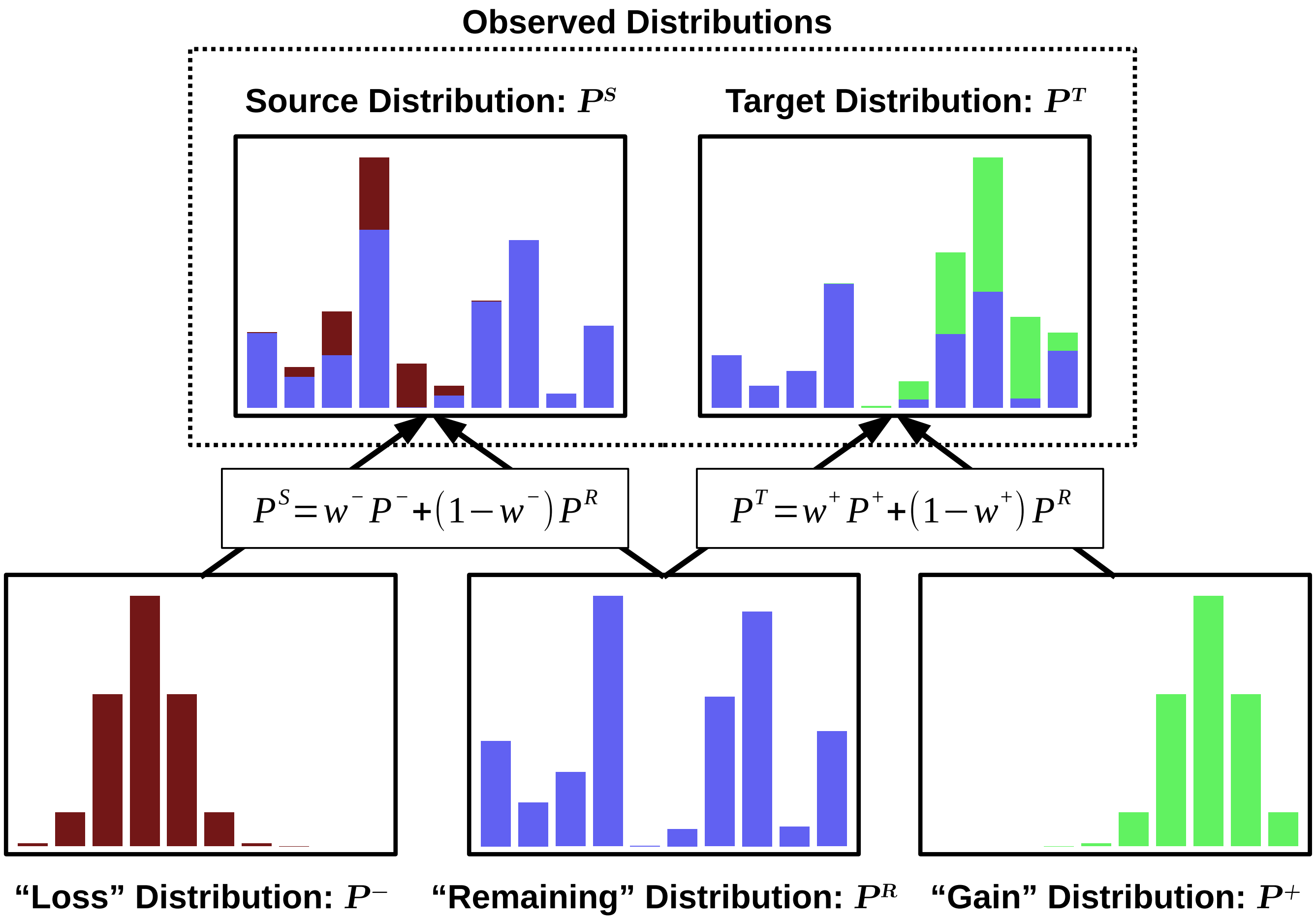
The Impact of Dataset Shift



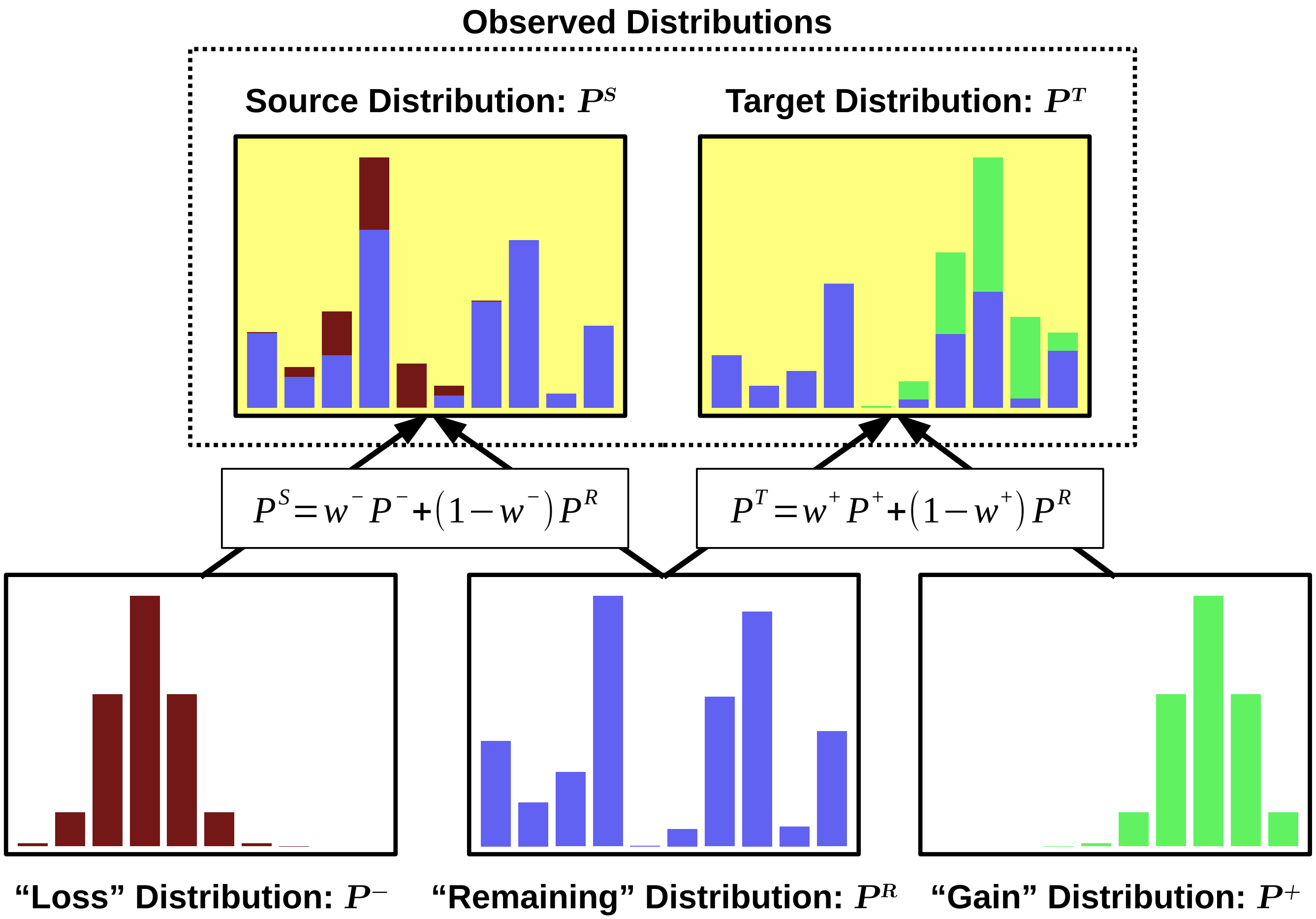
The Impact of Dataset Shift



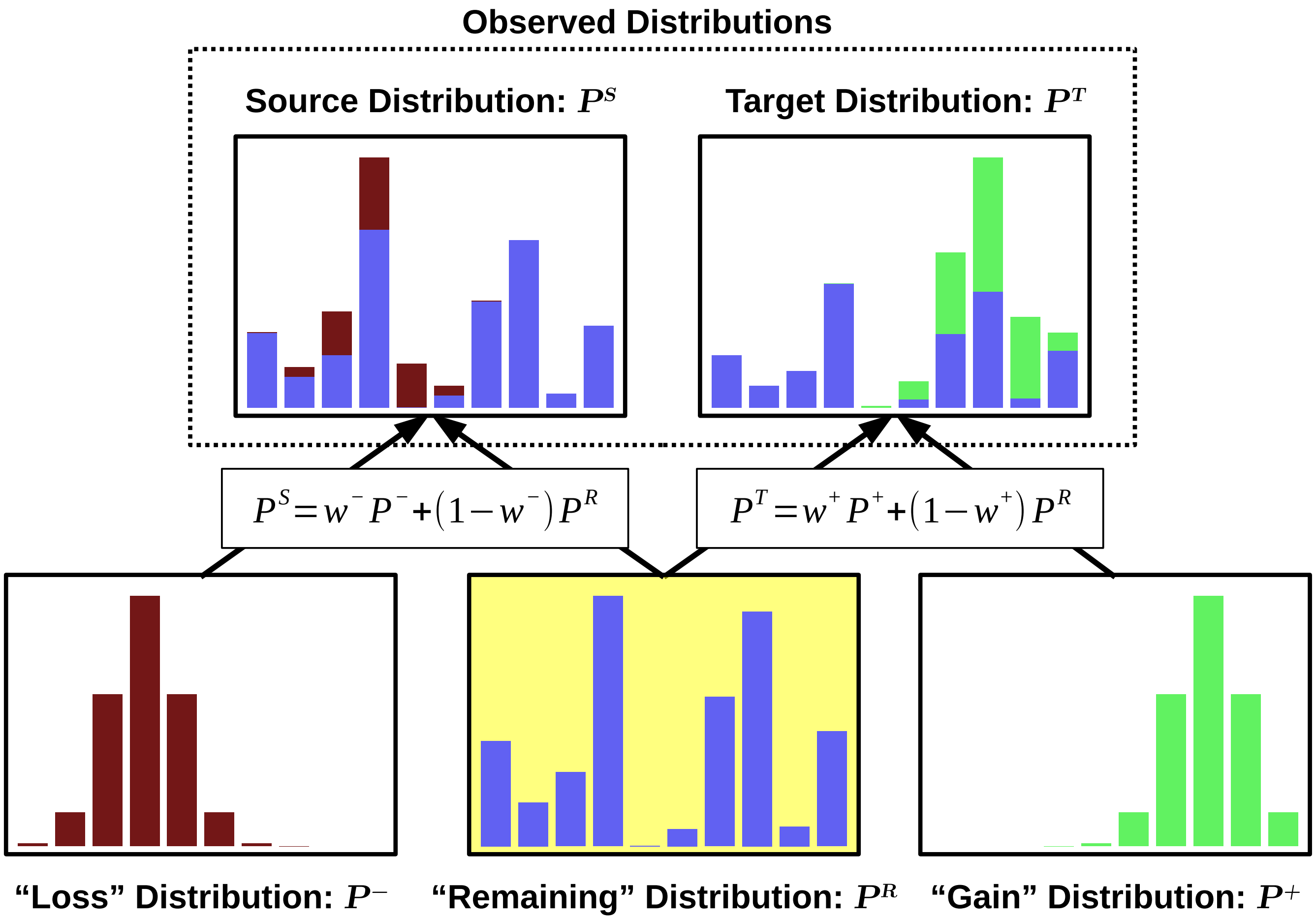
Our Model: Gain-Some-Lose-Some



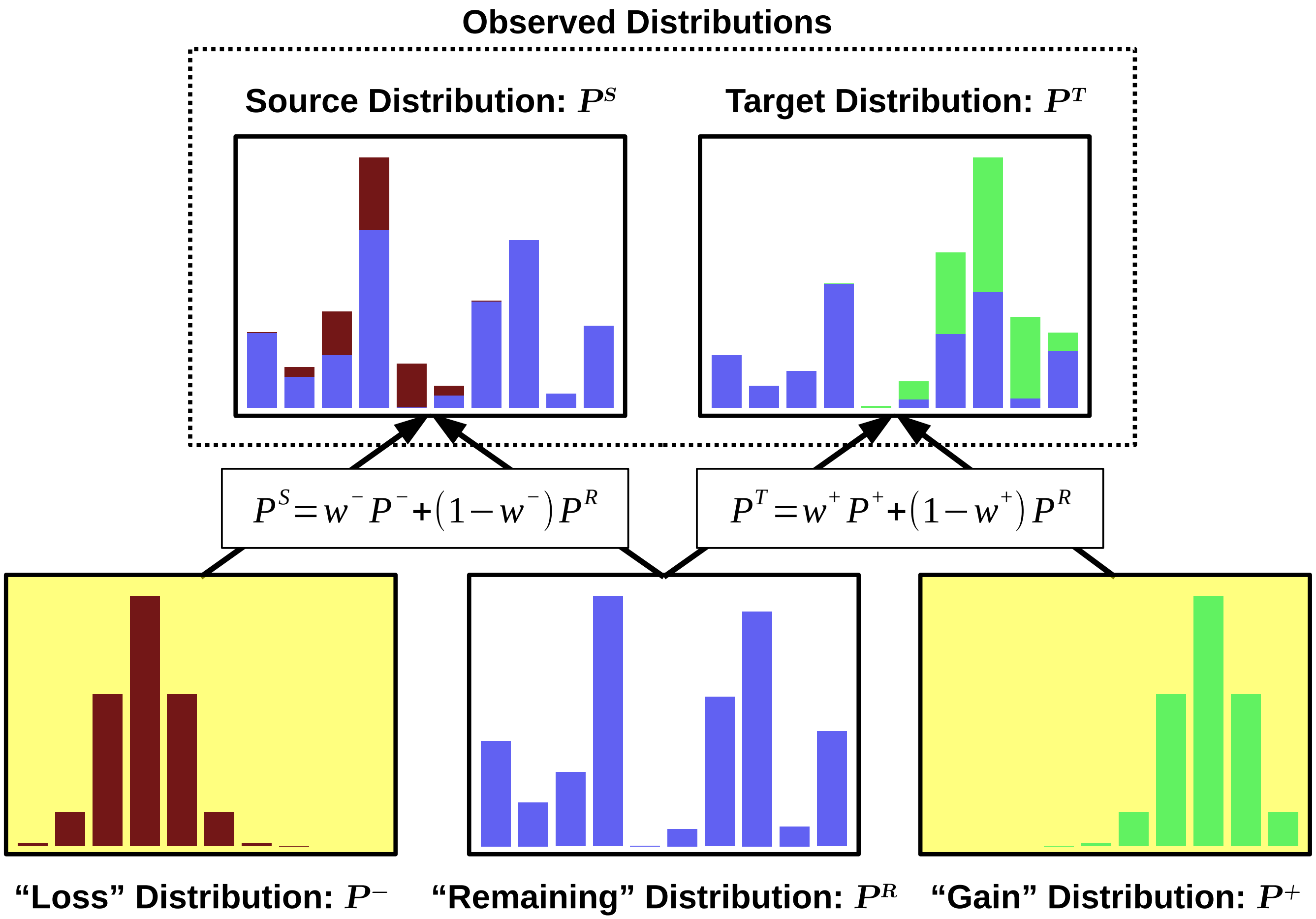
Our Model: Gain-Some-Lose-Some



Our Model: Gain-Some-Lose-Some



Our Model: Gain-Some-Lose-Some



Experimental Comparison

Coverage of true class proportions by 80% prediction intervals

Dataset	$\{w^+, w^-\}$	$\{0,0\}$	$\{0,.3\}$	$\{0,.7\}$	$\{0,1\}$	$\{.3,0\}$	$\{.3,.3\}$	$\{.3,.7\}$	$\{.3,1\}$	$\{.7,0\}$	$\{.7,.3\}$	$\{.7,.7\}$	$\{.7,1\}$	$\{1,0\}$	$\{1,.3\}$	$\{1,.7\}$	$\{1,1\}$	PS
HLL	PCC	**82%**	**84%**	**82%**	40%	62%	71%	70%	39%	47%	58%	58%	40%	40%	52%	52%	40%	71%
	EM	**100%**	**100%**	**97%**	67%	**89%**	**93%**	**91%**	64%	74%	**84%**	**82%**	63%	64%	77%	76%	66%	**98%**
	GSLS	**99%**	**99%**	**99%**	**86%**	**94%**	**96%**	**97%**	**92%**	**92%**	**93%**	**94%**	**92%**	**85%**	**89%**	**88%**	**86%**	**95%**
HLA	PCC	**81%**	74%	63%	26%	60%	63%	57%	26%	40%	43%	42%	26%	26%	31%	30%	25%	62%
	EM	**95%**	**93%**	80%	67%	**80%**	**83%**	76%	64%	56%	64%	65%	63%	67%	69%	70%	67%	**89%**
	GSLS	**89%**	**83%**	76%	50%	78%	76%	73%	52%	66%	64%	63%	52%	49%	53%	52%	49%	77%
DIG	PCC	**96%**	**92%**	72%	21%	45%	53%	48%	19%	28%	33%	32%	18%	21%	26%	25%	20%	49%
	EM	**100%**	**96%**	76%	43%	78%	**85%**	66%	36%	52%	63%	55%	35%	41%	49%	49%	40%	**95%**
	GSLS	**99%**	**97%**	**95%**	75%	**94%**	**96%**	**97%**	**87%**	**89%**	**90%**	**92%**	**89%**	77%	77%	78%	76%	**87%**
ISX	PCC	80%	41%	25%	12%	24%	27%	19%	10%	14%	17%	16%	10%	10%	14%	11%	10%	20%
	EM	**99%**	79%	63%	42%	61%	63%	52%	33%	51%	53%	50%	32%	41%	44%	48%	44%	**97%**
	GSLS	**97%**	**92%**	**82%**	74%	**91%**	**94%**	**91%**	**85%**	**87%**	**87%**	**88%**	**87%**	76%	75%	75%	76%	**84%**
ISP	PCC	74%	19%	8%	4%	16%	16%	9%	4%	8%	9%	8%	4%	5%	6%	6%	5%	7%
	EM	**90%**	20%	11%	8%	14%	18%	10%	6%	6%	9%	9%	5%	8%	8%	8%	9%	74%
	GSLS	**96%**	64%	51%	52%	64%	61%	55%	54%	57%	52%	57%	58%	51%	47%	48%	52%	40%

Colouring configurations where at least 80% of intervals covered the true class proportion

Experimental Comparison

Coverage of true class proportions by 80% prediction intervals

Dataset	$\{w^+, w^-\}$	$\{0, 0\}$	$\{0, .3\}$	$\{0, .7\}$	$\{0, 1\}$	$\{.3, 0\}$	$\{.3, .3\}$	$\{.3, .7\}$	$\{.3, 1\}$	$\{.7, 0\}$	$\{.7, .3\}$	$\{.7, .7\}$	$\{.7, 1\}$	$\{1, 0\}$	$\{1, .3\}$	$\{1, .7\}$	$\{1, 1\}$	PS
HLL	PCC	82%	84%	82%	40%	62%	71%	70%	39%	47%	58%	58%	40%	40%	52%	52%	40%	71%
	EM	100%	100%	97%	67%	89%	93%	91%	64%	74%	84%	82%	63%	64%	77%	76%	66%	98%
	GSLS	99%	99%	99%	86%	94%	96%	97%	92%	92%	93%	94%	92%	85%	89%	88%	86%	95%
HLA	PCC	81%	74%	63%	26%	60%	63%	57%	26%	40%	43%	42%	26%	26%	31%	30%	25%	62%
	EM	95%	93%	80%	67%	80%	83%	76%	64%	56%	64%	65%	63%	67%	69%	70%	67%	89%
	GSLS	89%	83%	76%	50%	78%	76%	73%	52%	66%	64%	63%	52%	49%	53%	52%	49%	77%
DIG	PCC	96%	92%	72%	21%	45%	53%	48%	19%	28%	33%	32%	18%	21%	26%	25%	20%	49%
	EM	100%	96%	76%	43%	78%	85%	66%	36%	52%	63%	55%	35%	41%	49%	49%	40%	95%
	GSLS	99%	97%	95%	75%	94%	96%	97%	87%	89%	90%	92%	89%	77%	77%	78%	76%	87%
ISX	PCC	80%	41%	25%	12%	24%	27%	19%	10%	14%	17%	16%	10%	10%	14%	11%	10%	20%
	EM	99%	79%	63%	42%	61%	63%	52%	33%	51%	53%	50%	32%	41%	44%	48%	44%	97%
	GSLS	97%	92%	82%	74%	91%	94%	91%	85%	87%	87%	88%	87%	76%	75%	75%	76%	84%
ISP	PCC	74%	19%	8%	4%	16%	16%	9%	4%	8%	9%	8%	4%	5%	6%	6%	5%	7%
	EM	90%	20%	11%	8%	14%	18%	10%	6%	6%	9%	9%	5%	8%	8%	8%	9%	74%
	GSLS	96%	64%	51%	52%	64%	61%	55%	54%	57%	52%	57%	58%	51%	47%	48%	52%	40%

Colouring configurations where at least 80% of intervals covered the true class proportion

Experimental Comparison

Coverage of true class proportions by 80% prediction intervals

Dataset	$\{w^+, w^-\}$	$\{0,0\}$	$\{0,.3\}$	$\{0,.7\}$	$\{0,1\}$	$\{.3,0\}$	$\{.3,.3\}$	$\{.3,.7\}$	$\{.3,1\}$	$\{.7,0\}$	$\{.7,.3\}$	$\{.7,.7\}$	$\{.7,1\}$	$\{1,0\}$	$\{1,.3\}$	$\{1,.7\}$	$\{1,1\}$	PS
HLL	PCC	82%	84%	82%	40%	62%	71%	70%	39%	47%	58%	58%	40%	40%	52%	52%	40%	71%
	EM	100%	100%	97%	67%	89%	93%	91%	64%	74%	84%	82%	63%	64%	77%	76%	66%	98%
	GSLS	99%	99%	99%	86%	94%	96%	97%	92%	92%	93%	94%	92%	85%	89%	88%	86%	95%
HLA	PCC	81%	74%	63%	26%	60%	63%	57%	26%	40%	43%	42%	26%	26%	31%	30%	25%	62%
	EM	95%	93%	80%	67%	80%	83%	76%	64%	56%	64%	65%	63%	67%	69%	70%	67%	89%
	GSLS	89%	83%	76%	50%	78%	76%	73%	52%	66%	64%	63%	52%	49%	53%	52%	49%	77%
DIG	PCC	96%	92%	72%	21%	45%	53%	48%	19%	28%	33%	32%	18%	21%	26%	25%	20%	49%
	EM	100%	96%	76%	43%	78%	85%	66%	36%	52%	63%	55%	35%	41%	49%	49%	40%	95%
	GSLS	99%	97%	95%	75%	94%	96%	97%	87%	89%	90%	92%	89%	77%	77%	78%	76%	87%
ISX	PCC	80%	41%	25%	12%	24%	27%	19%	10%	14%	17%	16%	10%	10%	14%	11%	10%	20%
	EM	99%	79%	63%	42%	61%	63%	52%	33%	51%	53%	50%	32%	41%	44%	48%	44%	97%
	GSLS	97%	92%	82%	74%	91%	94%	91%	85%	87%	87%	88%	87%	76%	75%	75%	76%	84%
ISP	PCC	74%	19%	8%	4%	16%	16%	9%	4%	8%	9%	8%	4%	5%	6%	6%	5%	7%
	EM	90%	20%	11%	8%	14%	18%	10%	6%	6%	9%	9%	5%	8%	8%	8%	9%	74%
	GSLS	96%	64%	51%	52%	64%	61%	55%	54%	57%	52%	57%	58%	51%	47%	48%	52%	40%

Colouring configurations where at least 80% of intervals covered the true class proportion

Experimental Comparison

Coverage of true class proportions by 80% prediction intervals

Dataset	$\{w^+, w^-\}$	$\{0,0\}$	$\{0,.3\}$	$\{0,.7\}$	$\{0,1\}$	$\{.3,0\}$	$\{.3,.3\}$	$\{.3,.7\}$	$\{.3,1\}$	$\{.7,0\}$	$\{.7,.3\}$	$\{.7,.7\}$	$\{.7,1\}$	$\{1,0\}$	$\{1,.3\}$	$\{1,.7\}$	$\{1,1\}$	PS
HLL	PCC	82%	84%	82%	40%	62%	71%	70%	39%	47%	58%	58%	40%	40%	52%	52%	40%	71%
	EM	100%	100%	97%	67%	89%	93%	91%	64%	74%	84%	82%	63%	64%	77%	76%	66%	98%
	GSLS	99%	99%	99%	86%	94%	96%	97%	92%	92%	93%	94%	92%	85%	89%	88%	86%	95%
HLA	PCC	81%	74%	63%	26%	60%	63%	57%	26%	40%	43%	42%	26%	26%	31%	30%	25%	62%
	EM	95%	93%	80%	67%	80%	83%	76%	64%	56%	64%	65%	63%	67%	69%	70%	67%	89%
	GSLS	89%	83%	76%	50%	78%	76%	73%	52%	66%	64%	63%	52%	49%	53%	52%	49%	77%
DIG	PCC	96%	92%	72%	21%	45%	53%	48%	19%	28%	33%	32%	18%	21%	26%	25%	20%	49%
	EM	100%	96%	76%	43%	78%	85%	66%	36%	52%	63%	55%	35%	41%	49%	49%	40%	95%
	GSLS	99%	97%	95%	75%	94%	96%	97%	87%	89%	90%	92%	89%	77%	77%	78%	76%	87%
ISX	PCC	80%	41%	25%	12%	24%	27%	19%	10%	14%	17%	16%	10%	10%	14%	11%	10%	20%
	EM	99%	79%	63%	42%	61%	63%	52%	33%	51%	53%	50%	32%	41%	44%	48%	44%	97%
	GSLS	97%	92%	82%	74%	91%	94%	91%	85%	87%	87%	88%	87%	76%	75%	75%	76%	84%
ISP	PCC	74%	19%	8%	4%	16%	16%	9%	4%	8%	9%	8%	4%	5%	6%	6%	5%	7%
	EM	90%	20%	11%	8%	14%	18%	10%	6%	6%	9%	9%	5%	8%	8%	8%	9%	74%
	GSLS	96%	64%	51%	52%	64%	61%	55%	54%	57%	52%	57%	58%	51%	47%	48%	52%	40%

Colouring configurations where at least 80% of intervals covered the true class proportion

Conclusions and Future Work

- GSLS gives more **reliable prediction intervals** under more **general conditions of shift**.
- GSLS communicates the **degree of shift**, enabling users to take **proportionate corrective action**.
- In future work, we plan to:
  - Analyse specific shift conditions where GSLS fitting is suboptimal
  - Develop a framework for selecting the optimal quantification method for observed shift

Thanks for Watching

Source code: github.com/ben-denham/gsls