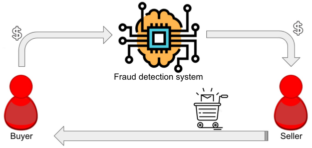
Stable Fraud Detection via Batch Balancing

Ruohan Zhan Advisor: Susan Athey Collaborators: Nitin Sharma, Vidyut Naware

Introduction

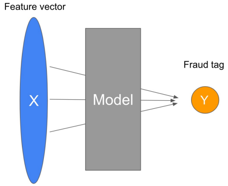
- Where: a digital two-sided marketplace of buyers and sellers;
- What: decide whether or not to decline a transaction;
- How: fraud detection algorithm with seller-buyer information as input and decision as output.



Setup

Current Fraud Detection System

- Buyer models: generate buyer risk using buyer data including **buyer funding instrument**
- Seller models: generate seller risk using merchant data including the associated **seller account**
- Rule-based model: decide total risk given buyer risk, seller risk, the current **channel** (that processes transaction).



Motivation

- Suppose the underlying structure model is $Y = f(X) + \epsilon$, where $X \sim P_X$.
- P_X can differ across environments (e.g. region, time period, population, etc.).
- f may be less variant.

Stable Learning and Covariate Balancing

Research Problem --- Generalizability

With data from one environment, learn a model that can be generalized to other environments.

Stable Learning --- Covariate balancing

Reweight samples such that other features are balanced with respect to each current feature.

Continuous features (see [2] for details)

Adjust sample weights to decorrelate features.

$$W = \arg\min_W \sum_{a=1}^{\infty}\sum_{b=1}^{\infty}\sum_{j=1}^{p} \|\mathbb{E}[(X^a_{\cdot,j})^T \Sigma_W X^b_{\cdot,-j}] - \mathbb{E}[(X^a_{\cdot,j})^T W]\mathbb{E}[W X^b_{\cdot,-j}]\|_2^2$$

Computationally, adopt the first moment approximation and choose a=b=1.

Batch Balancing

Challenge: original global balancing method is not computationally feasible for large volumes of transaction data

Global Balancing [1, 2] → Batch Balancing (ours)

- Compute balancing weights per batch;
- Use first order moment matching to ensure batch consistency.

$$W = \arg\min_W \overbrace{\sum_{j=1}^{p} \left\|\mathbb{E}[X^T_{\cdot,j}\Sigma_W X_{\cdot,-j}] - \mathbb{E}[X^T_{\cdot,j}W]\mathbb{E}[WX_{\cdot,-j}]\right\|_2^2}^{\text{Variable decorrelation}} + \underbrace{\eta \cdot \|WX - X_0\|_2^2}_{\text{Balanced distribution consistency}}$$

where X_0 is the moving average feature mean.

Data

PayPal 2018 transaction data.

- Buyer data: (X, Y) = (buyer features, buyer fraud tag)
- Feature size: 625.

Stratification of Environments

Time period
- Training data: transactions before Thanksgiving;
- Test data: transactions after Thanksgiving.

Buyer subsegment
Four subsegments in total
- Training data: three subsegments
- Test data: the fourth subsegment

	Thanksgiving data	Buyer segment 1	Buyer segment 2	Buyer segment 3	Buyer segment 4
Training data size	490415	534801	178842	442239	533718
Test data size	71761	28399	384358	120961	29482

Experiments

Evaluation metric: Logistic loss

$$-Y\log(p) - (1-Y)\log(1-p)$$

Compared models:

- Vanilla model: logistic model
- Balanced model (ours): logistic model + batch balancing

Test data	Thanksgiving data	Buyer segment 1	Buyer segment 2	Buyer segment 3	Buyer segment 4
Balanced model	0.0239	0.0209	0.0161	0.0202	0.0191
Vanilla model	0.0248	0.0308	0.0247	0.0257	0.0274

Acknowledgements

We are grateful to generous support from PayPal Innovation Fellowship.

References

1. Kuang, Kun, Peng Cui, Susan Athey, Ruoxuan Xiong, and Bo Li. "Stable prediction across unknown environments." In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1617-1626. 2018.
2. Kuang, Kun, Ruoxuan Xiong, Peng Cui, Susan Athey, and Bo Li. "Stable prediction with model misspecification and agnostic distribution shift." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 4485-4492. 2020.