Adversarial Outlier Detection for Health Care Fraud

Tahir Ekin
Department of Information Systems and Analytics , McCoy College of Business, Texas State University

Introduction

- U.S. health care spending: $4.3 trillion ($12,914 per person) in 2021 corresponding to 18.3 % of Gross Domestic Product
- 3 to 10 percent of spending lost to overpayments
- Global health care overpayment loss: > $450 billion (6 %)
- Overpayments in the form of fraud, waste and abuse
- Losses due to provider, patient and insurer
- Examples: submission of false claims, kickback payments, self-referrals and **upcoding** (overcharging, overbilling)

Current Practice & Issues

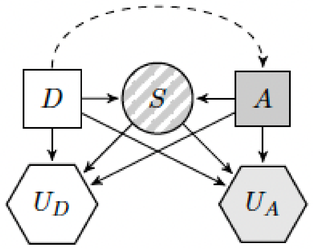
- Various statistical, data analytics and decision methods (Ekin, 2019)
- Global health care fraud detection market revenue at $1.5 billion, projected to reach $5.0 billion by 2026
- Isolated methods for data analysis and decisions
- Assumption of clean and legitimate data streams
- Some health care providers gaming threshold based methods

Motivation

- Address cases where adversaries may attempt to influence data which in turn may impact the outlier designations.
- Presentation of a decision theoretic approach for outlier detection in adversarial environments.
- Handle incomplete information inherent in adversarial interaction
- Illustration on health care claim billings subject to threshold based reviews/audits

Adversarial Risk Analysis

- Adversarial risk analysis (ARA) (Banks et al., 2015) models games as a decision-theoretic problem from the expected utility maximizing perspective of a given player.
- The player of interest takes into account their beliefs of the associated aleatory, epistemic, and solution-concept uncertainties.
- Bayesian models are constructed for the goals, capabilities, and strategies of the opponents by placing subjective distributions on all unknown quantities; which may be informed via expert judgments.
- For ex: Influence Diagram for a sequential game between a Defender (D) and an Attacker (A) given an uncertain event (S)



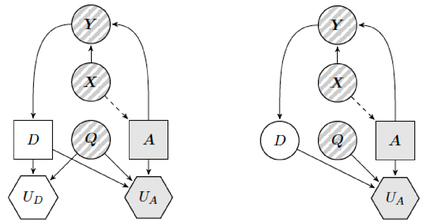
$$d^* = \text{argmax}_{d\in\mathcal{D}} \mathbb{E}_{\pi_D, p_D}[u_D(d, A, S)]$$
$$= \text{argmax}_{d\in\mathcal{D}} \int_{s\in\mathcal{S}} \int_{a\in\mathcal{A}} u_D(d, a, s) p_D(s \mid d, a) \pi_D(a)\, da\, ds$$

- First, we solve the Attacker's problem given a Defender decision alternative considering the distribution of the uncertain event outcome, and represent our uncertainty via p(a |d).
- Then, we incorporate our uncertainty about Attacker's decision to the Defender's decision problem and solve the expected utility maximization problem to compute the optimal decision, d*.
- ARA has been used in adversarial machine learning as in classification (Naveiro et al., 2019), reinforcement learning (Gallego et al., 2019) and hidden Markov models (Caballero et al., 2022)
- We aim to fill the methodological gap in the adaptation of ARA for unsupervised learning, in particular outlier detection.

Adversarial Outlier Detection

In the following, we adapt ARA for poisoning an outlier detection method.

- X: actual amount of a health care claim to be billed for payment
- Y: manipulated billing amount for the health care claim
- Q: risk and resource levels
- D: Defender decision involves using an outlier detection method output to determine the claims to audit
- A: Attacker decision involves determining if the claim payment should be manipulated
- Attacker utility function, U_A to be minimized includes the total amount of detected illegitimate claims and manipulation (including reputation) cost.
- Defender utility function, U_D to be maximized includes the balanced accuracy of the outlier detection method



Overall bi-agent influence diagram (BAID) (left) and
Attacker's reduced BAID (right) for the data-fiddler adversarial problem

- We utilize a concept drift idea that results in detection method misclassifying the actual attacks as legitimate (Bhargava, 2019).

Example

- Assume the automatic review threshold for a claim is $1,000. In addition to this threshold, an outlier detection method allows the auditor to identify suspicious claims for pre-payment reviews.
- The grand objective of the fraudster is to manipulate enough of his billings including even the legitimate ones to change the overall distribution of billings, and have the fraudulent claims not to be flagged as suspicious.

Conclusion

- It is crucial to understand the non-standard billing activities while generating leads for health care audits for the grand goal of payment integrity.
- Standard practice does not consider health care billings subject to intentional data manipulations.
- Proposed adversarial risk analysis based framework allows incomplete information and adversarial perturbations on the data inputs in an outlier detection method.
- Ongoing work: Comprehensive health care fraud example
- Limitations: accurate modeling of incomplete information
- Extension: Proactive defender strategies to limit the manipulation impact

Acknowledgments

This research is partially supported by the Air Force Scientific Office of Research (AFOSR) award FA-9550-21-1-0239, AFOSR European Office of Aerospace Research and Development award FA8655-21-1-7042, and Texas State University McCoy College of Business through Steven R. "Steve" Gregg Endowed Professorship.

References

- Banks, D.L., Aliaga, J.M.R., Insua, D.R. (2015). Adversarial Risk Analysis. CRC Press.
- Bhargava, R. (2019). Adversarial anomaly detection (Doctoral dissertation, Purdue University Graduate School).
- Caballero, W.N., Camacho, J.M., Ekin, T., Naveiro, R., 2022. Manipulating hidden-Markov-model inferences by poisoning batch data. Under Review.
- Ekin, T. (2019). Statistics and health care fraud: How to save billions. Chapman and Hall/CRC.
- Gallego, V., Naveiro, R., & Insua, D. R. (2019, July). Reinforcement learning under threats. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 9939-9940.
- Naveiro, R., Redondo, A., Insua, D. R., & Ruggeri, F. (2019). Adversarial classification: An adversarial risk analysis approach. International Journal of Approximate Reasoning, 113, 133-1.

TEXAS STATE UNIVERSITY
MEMBER THE TEXAS STATE UNIVERSITY SYSTEM