


Confusion Matrix Design for Downstream Decision-Making

Yiding Feng   

Department of Industrial Engineering and Decision Analytics,
Hong Kong University of Science and Technology, Hong Kong

Wei Tang   

Department of Decisions, Operations and Technology,
Chinese University of Hong Kong, Hong Kong

Abstract

We initiate the study of *confusion matrix design*. In this problem, an algorithm designer needs to generate a machine learning model (for a classification task from contexts to labels) which makes predictions for a population of downstream decision makers. The prediction accuracy of the machine learning model is characterized by its confusion matrix, which is a stochastic matrix where each entry encodes the probability of predicting the true label to another label. Each downstream decision maker faces a separate optimization task and will decide his binary action based on his own context, realized prediction given his context, and the confusion matrix selected by the algorithm designer. Decision makers are heterogeneous, as they may hold different contexts. Both the decision makers and the algorithm designer will obtain utilities that are determined by the actions the decision makers take, and their true labels. The goal of the algorithm designer is to design a public confusion matrix that is used for all decision makers subject to some feasibility constraints in order to maximize her net utility. We consider a general class of net utility functions, which could be a combination of both decision makers' utilities and the algorithm designer's utility. Classic outcome-independent utility and utilitarian/Nash/egalitarian social welfare are all special cases of our net utility formulation.

We study the above problem through an information design framework, where we view training machine learning model as designing an information structure (signaling scheme) subject to some specific constraints motivated by the machine learning literature. By building the connection to the *public persuasion with heterogeneous priors*, we design convex programming-based algorithms that compute the optimal confusion matrix subject to (i) post-processing constraints and (ii) receiver operating characteristic (ROC) constraints in polynomial time, respectively. Besides the computational results, we also obtain analytical structural results and numerical results for the special cases of outcome-independent utility and social-aware utility, by utilizing the convex programming-based characterization of the optimal confusion matrix.

2012 ACM Subject Classification Theory of computation → Algorithmic game theory; Theory of computation → Machine learning theory

Keywords and phrases Confusion Matrix, Downstream Decision-making, Public Persuasion, Human-AI Interaction, Information Design

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.49

Category Extended Abstract

Related Version *Full Version*: <https://doi.org/10.2139/ssrn.4950758> [1]

Acknowledgements We thank the ITCS reviewers for the helpful and constructive comments.

References

- 1 Yiding Feng and Wei Tang. Confusion matrix design for downstream decision-making. *SSRN*, 2024. Full version of this abstract.



© Yiding Feng and Wei Tang;
licensed under Creative Commons License CC-BY 4.0
16th Innovations in Theoretical Computer Science Conference (ITCS 2025).
Editor: Raghu Meka; Article No. 49; pp. 49:1–49:1



Leibniz International Proceedings in Informatics
LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany