Automated Mechanism Design for Strategic Classification: Abstract for KDD'21 Keynote Talk

Vincent Conitzer
Duke University
Durham, NC, USA
conitzer@cs.duke.edu

ABSTRACT

AI is increasingly making decisions, not only for us, but also about us – from whether we are invited for an interview, to whether we are proposed as a match for someone looking for a date, to whether we are released on bail. Often, we have some control over the information that is available to the algorithm; we can self-report some information, and other information we can choose to withhold. This creates a potential circularity: the classifier used, mapping submitted information to outcomes, depends on the (training) data that people provide, but the (test) data depend on the classifier, because people will reveal their information strategically to obtain a more favorable outcome. This setting is not adversarial, but it is also not fully cooperative. Mechanism design provides a framework for making good decisions based on strategically reported information, and it is commonly applied to the design of auctions and matching mechanisms. However, the setting above is unlike these common applications, because in it, preferences tend to be similar across agents, but agents are restricted in what they can report. This creates both new challenges and new opportunities, as we demonstrate in our theoretical work and our initial experiments. This is joint work with Hanrui Zhang, Andrew Kephart, Yu Cheng, Anilesh Krishnaswamy, Haoming Li, and David Rein [1–8].

KEYWORDS

automated mechanism design, strategic classification

ACM Reference Format:
Vincent Conitzer. 2021. Automated Mechanism Design for Strategic Classification: Abstract for KDD'21 Keynote Talk. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3447548.3469650

BIOGRAPHY

Vincent Conitzer is the Kimberly J. Jenkins Distinguished University Professor of New Technologies and Professor of Computer Science, Professor of Economics, and Professor of Philosophy at Duke University, as well as Head of Technical AI Engagement at the Institute for Ethics in AI, and Professor of Computer Science and Philosophy, at the University of Oxford. Much of his work has focused on AI and game theory. More recently, he has started to work on AI and ethics: how should we determine the objectives that AI systems pursue, when these objectives have complex effects on various stakeholders?

He received his doctorate from Carnegie Mellon University in 2006. He has received the 2021 ACM/SIGAI Autonomous Agents Research Award, the Social Choice and Welfare Prize, a Presidential Early Career Award for Scientists and Engineers (PECASE), the IJCAI Computers and Thought Award, and an honorable mention for the ACM dissertation award. He has also been named a Guggenheim Fellow, a Sloan Fellow, a Kavli Fellow, an ACM Fellow, a AAAI Fellow, and one of AI's Ten to Watch.

ACKNOWLEDGMENTS

This work was funded in part by NSF award IIS-1814056.

REFERENCES

[1] Andrew Kephart and Vincent Conitzer. 2015. Complexity of Mechanism Design with Signaling Costs. In *Proceedings of the Fourteenth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).* Istanbul, Turkey, 357–365.
[2] Andrew Kephart and Vincent Conitzer. 2021. The Revelation Principle for Mechanism Design with Signaling Costs. *ACM Transactions on Economics and Computation (TEAC)* 9, 1 (2021), Article 6, 1–35.
[3] Anilesh Krishnaswamy, Haoming Li, David Rein, Hanrui Zhang, and Vincent Conitzer. 2021. Classification with Strategically Withheld Data. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence.* Virtual conference, 5514–5522.
[4] Hanrui Zhang, Yu Cheng, and Vincent Conitzer. 2019. Distinguishing Distributions When Samples Are Strategically Transformed. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS).* Vancouver, Canada, 3187–3195.
[5] Hanrui Zhang, Yu Cheng, and Vincent Conitzer. 2019. When Samples Are Strategically Selected. In *Proceedings of the 36th International Conference on Machine Learning (ICML-19).* Long Beach, CA, USA, 7345–7353.
[6] Hanrui Zhang, Yu Cheng, and Vincent Conitzer. 2021. Automated Mechanism Design for Classification with Partial Verification. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence.* Virtual conference, 5789–5796.
[7] Hanrui Zhang, Yu Cheng, and Vincent Conitzer. 2021. Classification with Few Tests through Self-Selection. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence.* Virtual conference, 5805–5812.
[8] Hanrui Zhang and Vincent Conitzer. 2021. Incentive-Aware PAC Learning. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence.* Virtual conference, 5797–5804.

KDD '21, August 14–18, 2021, Virtual Event, Singapore.

ACM ISBN 978-1-4503-8332-5/21/08.
https://doi.org/10.1145/3447548.3469650