1st International Conference of Science of Science & Innovation, June 7-9, 2022
National Academy of Sciences, Washington, D.C., USA

A Synthetic Prediction Market for Estimating Confidence in Published Work

Sarah Rajtmajer, Christopher Griffin, Jian Wu, Robert Fraleigh, Laxmaan Balaji, Anna Squicciarini, Anthony Kwasnica, David Pennock, Michael McLaughlin, Timothy Fritton, Nishanth Nakshatri, Arjun Menon, Sai Ajay Modukuri, Rajal Nivargi, Xin Wei, C. Lee Giles

Keywords: Replication, Synthetic prediction markets, Machine learning

Extended Abstract

Concerns about the replicability, robustness and reproducibility of findings in scientific literature have gained widespread attention over the last decade in the social sciences and beyond. This attention has been catalyzed by and has likewise motivated a number of large-scale replication projects which have reported successful replication rates between 36% and 78%. Given the challenges and resources required to run high-powered replication studies, researchers have sought other approaches to assess confidence in published claims. Initial evidence has supported the promise of prediction markets in this context. However, they require the coordinated, sustained effort of collections of human experts and typically rely on availability of a ground truth. They are limited by human participants' narrow view of the literature and cognitive biases, the compounded effects of which are poorly understood in market settings.

We suggest that markets populated by artificial agents provide an opportunity to overcome or mitigate many of these limitations. Our talk will describe a fully synthetic market for replication prediction wherein algorithmic agents (trader bots) are trained and tested on proxy ground truth pulled from existing replication studies. Our work is complementary to recent efforts using machine learning for reproducibility prediction [1, 6, 3]. Unlike prior approaches, the market scores only a subset of the papers in our test set but accuracy on that subset is very high. The market affords explainability by way of the record of trades and relevant features. Our prototype system was demo'ed at AAAI 2022 [4].

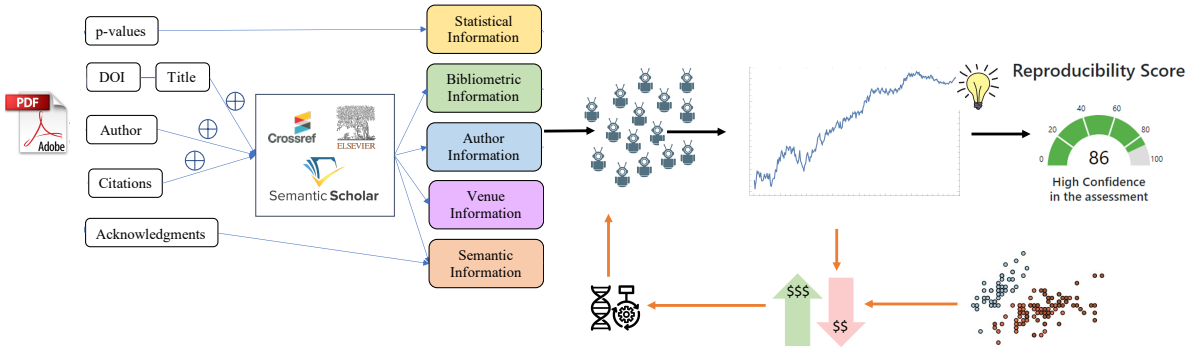

Figure 1: (Black arrows) A scientific paper is processed through the FEXRep feature extraction framework. Features are shared with the agents who purchase assets corresponding to binary outcomes of a notional replication study of the primary claim of that paper. The price of these assets at market close is an indicator of confidence in the claim. (Orange arrows) During training, agents purchase assets corresponding to claims drawn from prior replication projects for which ground truth is known. An evolutionary algorithm is used to update the population.

The model. Our approach has two modules: feature extraction pipeline and synthetic prediction market. Extracted features are provided to bot traders during train and test (Figure 1).
Feature extraction pipeline. The Feature EXtraction framework for Replicability prediction (FEXRep) extracts five categories of features related to a given scholarly preprint or published paper and its metadata: biblometric, venue-related, author-related, statistical and semantic information. At present, 41 total features are extracted, ranging from p values and sample size to number of authors and acknowledgement of funding (see [5]). In the prototype system, all features represent paper-level information. Our talk will discuss ongoing efforts to extract features at the claim-level to support assessment of multiple claims within each paper.
Synthetic market. Agents in the market are initialized with a fixed amount of cash and provided with the set of extracted features representing a paper in question. Agents may purchase assets corresponding to *will replicate* or *will not replicate* outcomes of a notional replication study of the primary claim of that paper. Agent purchase logic is defined using a sigmoid transformation of a convex semi-algebraic set defined in feature space. Time-varying asset prices affect the structure of the semi-algebraic sets leading to time-varying agent purchase rules (see [2] for further detail including theoretical properties of the market). The price of a *will replicate* asset at market close is taken as proxy for confidence in the primary claim of the paper. During training, parameters that define agent purchase logic are identified using an evolutionary algorithm. Explanations of outputs derive from the record of agents participating and trades made.

Initial testing of our market used a collection of known 192 known replication outcomes from the literature. Our talk will detail training data and experimental settings.

Results on scored papers. Our system provides confidence scores for 68 of 192 (35%) of the papers in our set. On the set of scored papers, accuracy is 0.894, precision is 0.917, recall is 0.903, and **F1** is **0.903** (macro averages). A sizeable un-scored subset of data (65%) is the trade-off for high accuracy on the scored subset of the data.
System non-scoring. Like its human-populated counterparts, the market is vulnerable to lack of participation. Agents will not participate if they have not seen a sufficiently similar training point (paper). This is more common when the training dataset is small. Meaningful ways to increase agent participationare being explored and will be discussed in our talk.

Acknowledgements. We acknowledge support by DARPA W911NF-19-2- 0272. This work does not necessarily reflect the position or policy of DARPA and no official endorsement should be inferred.

References

[1] A. Altmejd, A. Dreber, E. Forsell, J. Huber, T. Imai, M. Johannesson, M. Kirchler, G. Nave, and C. Camerer. Predicting the replicability of social science lab experiments. *PloS one*, 14(12):e0225826, 2019.

[2] N. Nakshatri, A. Menon, C. L. Giles, S. Rajtmajer, and C. Griffin. Design and analysis of a synthetic prediction market using dynamic convex sets. *arXiv preprint arXiv:2101.01787*, 2021.

[3] S. Pawel and L. Held. Probabilistic forecasting of replication studies. *PloS one*, 15(4):e0231416, 2020.

[4] S. Rajtmajer, C. Griffin, J. Wu, R. Fraleigh, L. Balaji, A. Squicciarini, A. Kwasnica, D. Pennock, M. McLaughlin, T. Fritton, et al. A synthetic prediction market for estimating confidence in published work. *arXiv preprint arXiv:2201.06924*, 2021.

[5] J. Wu, R. Nivargi, S. S. T. Lanka, A. M. Menon, S. A. Modukuri, N. Nakshatri, X. Wei, Z. Wang, J. Caverlee, S. M. Rajtmajer, et al. Predicting the reproducibility of social and behavioral science papers using supervised learning models. *arXiv preprint arXiv:2104.04580*, 2021.

[6] Y. Yang, W. Youyou, and B. Uzzi. Estimating the deep replicability of scientific findings using human and artificial intelligence. *Proceedings of the National Academy of Sciences*, 117(20):10762–10768, 2020.