SAND2022-9327C

Using Social Simulations to Train, Test, and Evaluate Models for Social Prediction

Kiran Lakkaraju[1], Asmeret Naugle[1], Matthew Sweitzer[1], Steven Wiryadinata[1], and Griffin Lehrer[1]

[1]Sandia National Laboratories, Albuquerque, NM, USA

Introduction

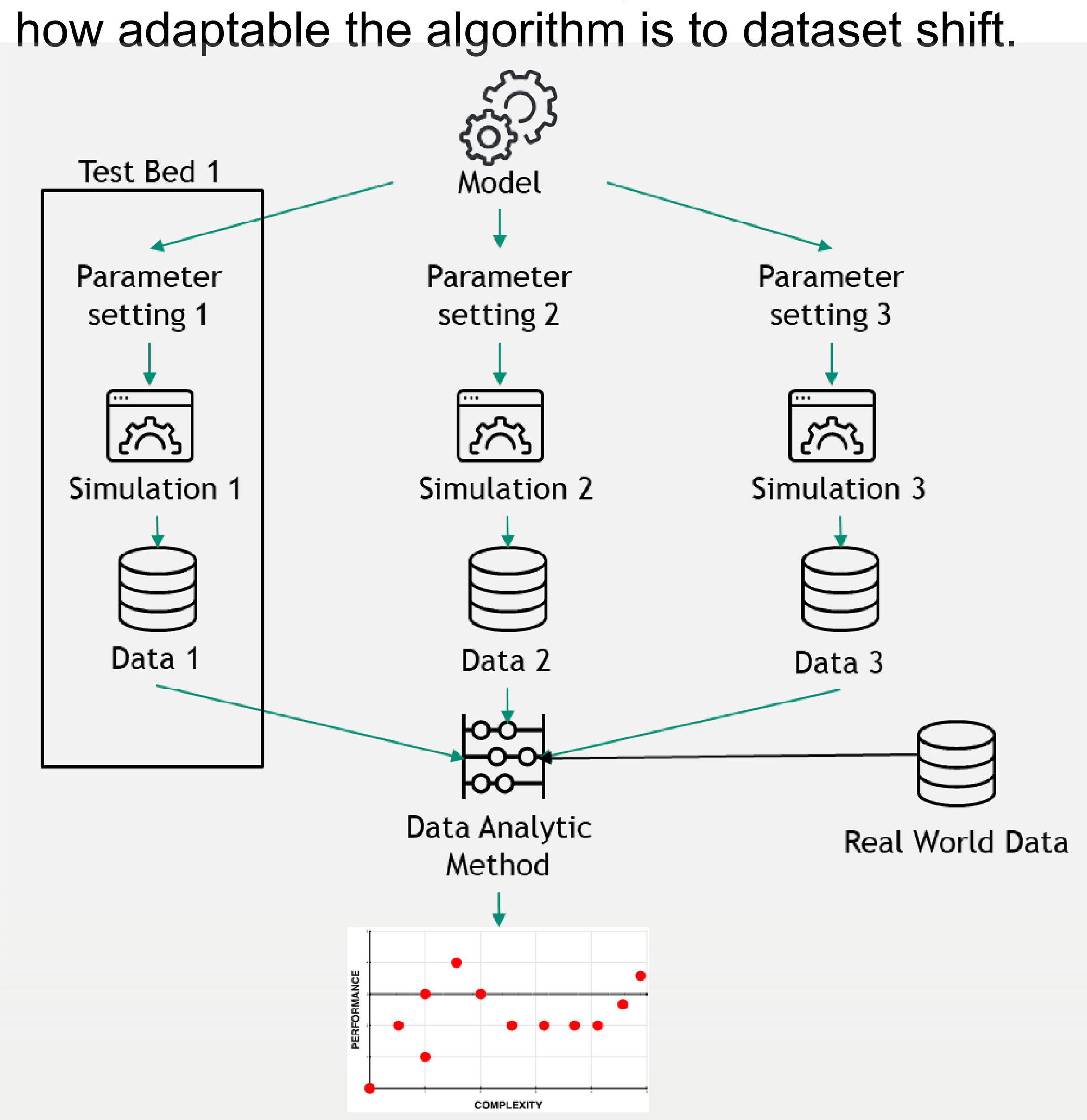
Disinformation and social media manipulation campaigns are a large and growing problem that is increasingly affecting national security[1], and their effects are accelerated by the use of online social media. Predictive algorithms, based on artificial intelligence and machine learning (ML) can provide a powerful tool to combat disinformation through identification of false information by content analysis, assessment of the reach of disinformation, and identification of ways to stop influence. These powerful methods require enormous amounts of data to identify patterns of behavior.

A fundamental assumption of the majority of ML algorithms is that the data on which the algorithm was trained has the same distribution as the data on which it is applied. This assumption, however, is violated often and severely in practice, a problem which is called "dataset shift"[2]. This is exacerbated in disinformation campaigns by shifts in social media algorithms, public attention, message characteristics, and the amount of competition in the information environment. It is thus unknown how well different ML cascade prediction methods perform in conditions where disinformation datasets differ in their underlying mechanisms.

Simulation Testbeds

We propose a novel solution method to address this: using artificial social simulations to generate data with varying underlying causal mechanisms, which can be used to test ML algorithms.

Agent-Based Models, where individuals, organizations, or groups are represented as autonomous entities, are a popular method for capturing the emergent complexity of social systems. Traditionally the use of social simulation has been to explore phenomena and test the impact of interventions. Our use of these simulations is instead as a synthetic data generator. By comparing ML model performance between simulated datasets, we can ascertain how adaptable the algorithm is to dataset shift.



Agent-Based Model

Our goal for this simulation is to create a simple, but adaptable framework for agent interactions that can scale the amount and complexity of features. N=1,000 agents are placed into a network and some are randomly seeded messages in their "inbox". On each model tick, agents interpret whether they will place the message in their "outbox". At the conclusion of the tick, all outbox messages are sent to the inbox of adjacent agents following the directed edges in the network. Additional messages are seeded by the model on subsequent ticks and the model terminates at a the conclusion of the 100th tick. The following parameters were varied between runs:

- **Network Type:** one of Barabasi-Albert (Scale -Free), Watts-Strogatz (Small-World), or Erdős–Rényi (random; "median" type)
- **Edge Density:** 0.6%, 1.0%, or 2.0% connected
- **Rewire Probability (SW):** 0.1, 0.2, or 0.4
- **Q_i:** agent i's subjective resend probability
  - *Mean*: 0.8, 1.0, 1.2; *SD*: 5%, 20%, or 50% of mean
- **ϕ_m:** messages' virality ($p_{resend} = Q_i * \phi_m$)
  - Random power distribution with α = 5, 10, or 50
- **K_i:** agent i's attention limit (# to read from inbox)
  - *Mean*: 1, 5, 10; *SD*: 5%, 20%, or 50% of mean
- **Message addition:**
  - **Number:** *Mean:* 100, 250, 500; *SD:* 20% of mean
  - **Frequency after initialization:** 1, 10, or 25 ticks
- **Number of agents seeded each message:** 1, 10, or 25

Parameters were explored independently; if one feature varied in a set of simulations, all others were held at their median value. In total, 45 unique parameter settings were run to simulate dataset shift.

Machine-Learning Algorithms

The prediction problem that the ML algorithms address is popularity prediction, or the reach of a message. For all *i*, *j* simulation pairs, we evaluate model performance as the relative prediction error (RMSE) of the model trained on T_i and tested on V_j. i.e.:

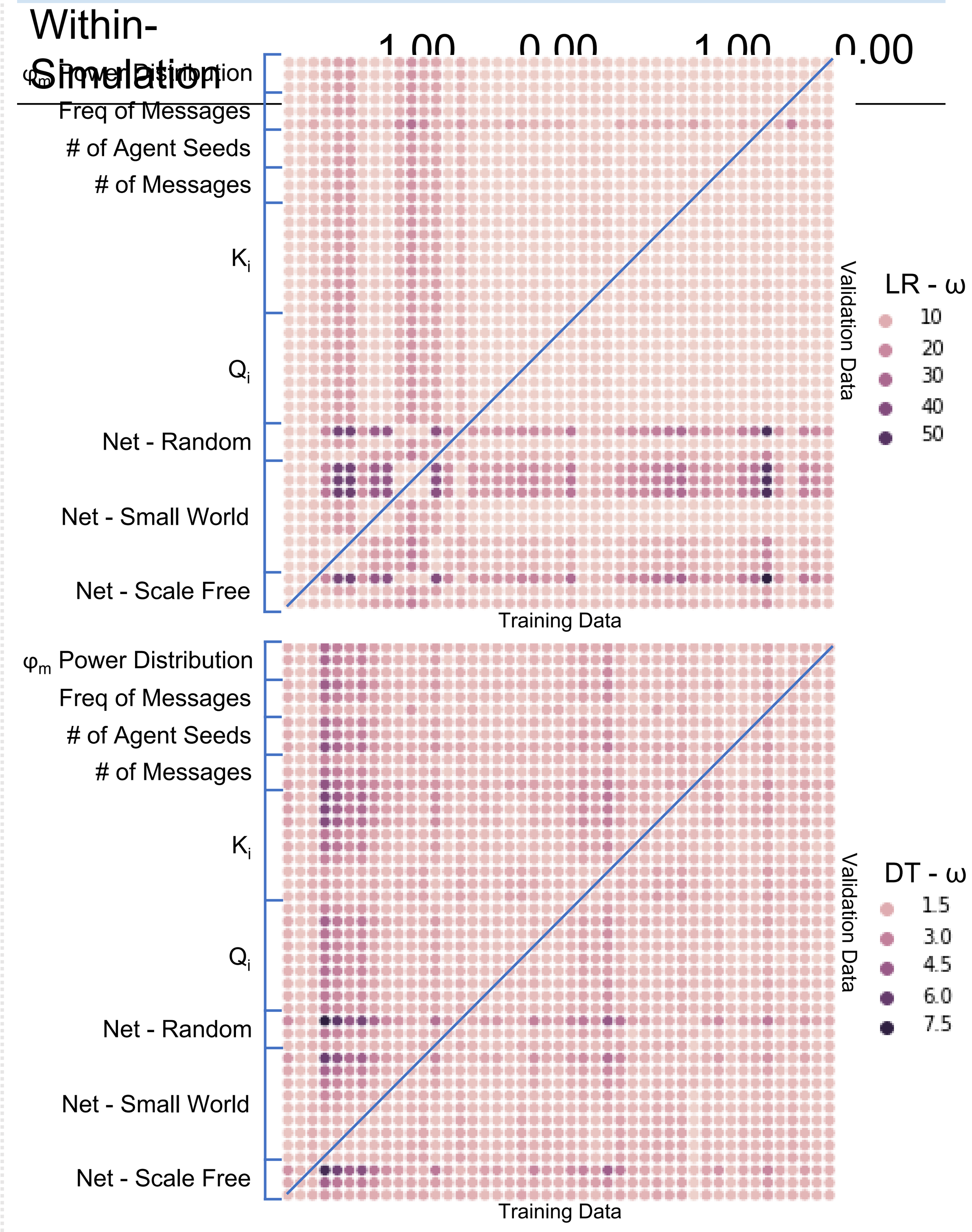
$$\omega_{i,j} = \frac{RMSE(T_i, V_j)}{RMSE(T_i, V_i)}$$

All simulations were randomly split on messages into training (80%), testing (10%) and validation (10%) sets. Independent features include the mean global centrality, in-, and out-degree of agents that had resent the message at tick 1. The dependent variable is the number of agents that resend the message by tick 5. The data analytic methods tested are Linear Regression and Decision Trees.

Results

	LR - RMSE		DT - RMSE	
	m	*sd*	*m*	*sd*
Out-of-Simulation	676.35	949.27	176.64	81.26
Within-Simulation	131.30	52.30	143.73	84.55

	LR - ω		DT - ω	
	m	*sd*	*m*	*sd*
Out-of-Simulation	5.44	7.40	1.38	0.71
Within-Simulation	1.00	0.00	1.00	0.00



Discussion

Machine learning models are able to form out-of-sample predictions about the spread of information over a social network, but the degree of accuracy is context- and model-dependent: In general, the decision tree out-performed linear regression under conditions of dataset shift by a factor of ~4. However, when the validation set came from the same simulation as the training set, the linear regression model performed slightly better.

Model performance decreases when the training set comes from a small world network – particularly low density. Conversely, both high-density networks and high message frequency (i.e., greater information competition) are particularly difficult scenarios to predict the spread of information. Decision trees tend to form better predictions when agent vary in their resending rates (Q_i), while linear regression formed better predictions in low-attention scenarios (K_i).

More research is needed to better understand how machine learning algorithms operate on varying social data. While the simulation technique applied here offers valuable insight into the complexities of prediction in different contexts, the generalizability of these results to real-world disinformation cascade datasets is so far unknown.

References

[1] Samantha Bradshaw, Hannah Bailey, and Philip N. Howard. Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation. Technical report, Programme on Democracy & Technology, Oxford, UK, 2021.

[2] Jose G. Moreno-Torres, Troy Raeder, Rocío Alaiz-Rodríguez, Nitesh V. Chawla, and Francisco Herrera. A unifying view on dataset shift in classification. *Pattern Recognition*, 45(1):521–530, 2012.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. ____________

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.