Cognitive Bias-Aware Dissemination Strategies for Opinion Dynamics with External Information Sources

Extended Abstract

Abdullah Al Maruf
University of Washington
Seattle, USA
maruf3e@uw.edu

Luyao Niu
University of Washington
Seattle, USA
luyaoniu@uw.edu

Bhaskar Ramasubramanian
Western Washington University
Bellingham, USA
ramasub@wwu.edu

Andrew Clark
Washington University in St. Louis
St. Louis, USA
andrewclark@wustl.edu

Radha Poovendran
University of Washington
Seattle, USA
rp3@uw.edu

ABSTRACT

The opinions of members of a population are influenced by opinions of their peers, their own internal predispositions, and information from external sources such as the media. Agents might perceive the received information differently due to various cognitive biases. In this paper, we propose a model of opinion evolution that uses *prospect theory* to represent the perception of information provided by an external source. Using the proposed model, we study the problem of selecting dissemination strategies for the external source to adopt in order to drive the opinions of individuals toward a desired value. As the initial predispositions of agents and functions characterizing agents' perceptions of information disseminated might be unknown to the source, we estimate the unknown terms in the dynamics and find the optimal strategy by leveraging Gaussian process learning. Our simulations on three different widely-used large graph networks demonstrate that the external source can effectively drive a larger fraction of opinions towards a desired value by using a prospect-theory-based dissemination strategies.

KEYWORDS

Opinion dynamics; external information source; prospect theory

ACM Reference Format:

Abdullah Al Maruf, Luyao Niu, Bhaskar Ramasubramanian, Andrew Clark, and Radha Poovendran. 2023. Cognitive Bias-Aware Dissemination Strategies for Opinion Dynamics with External Information Sources: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Understanding the evolution of opinions of members in a population in shaping individual behavior has led to significant research in multiple domains, including biology and social networks [9]. The dynamics of opinion evolution is typically modeled as a *weighted average* update [1, 3, 4, 7], where a weight quantifies the importance an agent assigns to an opinion of another agent. Initial predispositions [5, 6] and influence of external sources [10, 11] can be incorporated as additive terms in the model. However, agents might have cognitive biases including (i) having different perspectives on losses and gains, (ii) unconsciously assigning inflated values to low probability events and deflated values to high probability events, and (iii) evaluating outcomes relative to individually adopted reference points [14]. Prospect theory, introduced in [8], has been shown to effectively model cognitively biased behaviors (i) - (iii) in empirical evaluations on single individuals. However, the role of prospect theory in understanding opinion evolution among multiple interacting agents with cognitive biases has not been studied.

We propose a model of opinion evolution that uses prospect theory to represent perception of information provided by an external source when agents have cognitive biases. The external source could be an advertiser who wishes to market a product to a population, and is interested in nudging individuals towards showing interest in the product. In our model, the external source broadcasts information as a probability distribution over a random variable that represents possible outcomes of a phenomenon. Our model provides a computational framework to reason about opinion formation when agents exhibit the behaviors in (i)-(iii), and enables design of information dissemination strategies for the external source to steer agents' opinions towards a desired value.

2 MODEL

Agents are represented as a set of nodes $\mathcal{V}$ of a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $|\mathcal{V}| = N$. The weight w_{ij} of an edge quantifies the importance of agent j's opinion on agent i's opinion. We will assume that $w_{ij} \geq 0, \sum_j w_{ij} = 1$ for all i, j, and that $\mathcal{G}$ is strongly connected. The external source chooses a dissemination strategy from the set $Q = \{q_1(\theta), \cdots, q_r(\theta)\}$ and broadcasts it to all agents. Each strategy $q \equiv q(\theta) \in Q$ is a probability distribution over a discrete random variable θ. We use insights from *prospect theory* [8, 14] to characterize each agent's perception of information from the external source. The perception of agent i for the distribution $q(\theta) \in Q$ broadcast by the source is characterized by a *prospect*, $u_i^q := \sum_\theta p_i(q(\theta))v_i(\theta)$. Here, $v_i : \mathbb{R} \to \mathbb{R}$ and $p_i : [0, 1] \to [0, 1]$ (known as *value function* and *probability weighting function*, respectively) are nonlinear functions taken from empirical models of human cognitive biases (i)-(iii) developed in the social sciences [8, 14].

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

Initial Opinion	WS: % of Final Opinions > 0.5		BA: % of Final Opinions > 0.5		FB: % of Final Opinions > 0.5	
	PT (Eqn. (1))	Exp. (Eqn. (2))	PT (Eqn. (1))	Exp. (Eqn. (2))	PT (Eqn. (1))	Exp. (Eqn. (2))
$Unif(-1, 0)$	**18.5%**	0	**12.1%**	0	**7.91%**	0
$Unif(0, 1)$	**21.6%**	0.51%	**22.1%**	0.65%	**15.26%**	0
$Unif(-1, 0.5)$	**14.5%**	0	**14.7%**	0.1%	**9.49%**	0
$Unif(-1, -0.5)$ OR $Unif(0.5, 1)$	**17.7%**	6.5%	**17.9%**	5.6%	**12.13%**	0

Table 1: This Table compares the final opinions of agents when the external source aims to drive opinions towards $x^* = 1$ in the Watts-Strogatz (WS), Barabasi-Albert (BA), and Facebook social network (FB) graphs. We compare opinion evolution under our proposed prospect-theoretic model (PT) with the expectation-based update model (Exp). For different distributions of initial opinions, we examine the fraction of agents whose final opinions are in *strong agreement* ($x > 0.5$) with x^*. We observe that our PT model consistently results in a significantly larger fraction of agents whose final opinions are in strong agreement with x^*.

Network	PT (Eqn. (1))	Exp. (Eqn. (2))
Watts-Strogatz	**1.488**	2.223
Barabási-Albert	**1.709**	2.457
Facebook	**2.343**	2.625

Table 2: This Table compares the average distance, $L(q, x^*)$, of the final opinions of agents from the desired opinion x^*. We observe that for all three networks, $L(q, x^*)$ is lower when the external source computes an optimal dissemination strategy that considers prospect-theoretic agents (*Column 2*) compared to a strategy that does not consider such behavior (expectation based-update, *Column 3*).

For each agent i, let $x_i(k) \in \mathbb{R}$, $x_i^b \in \mathbb{R}$ and $T_i > 0$ denote the opinion at time-step k, initial predisposition, and level of trust on the external source, respectively. Assume the parameters $(\lambda_{i1}, \lambda_{i2}, \lambda_{i3})$ represent the relative importance that agent i ascribes to opinions of its peers, its own initial predisposition, and the information from the source, respectively. Let $x(k) := [x_1(k) \cdots x_N(k)]^T$, $x^b := [x_1^b \cdots x_N^b]^T$, $u^q := [u_1^q \cdots u_N^q]^T$, W be an $N \times N$ matrix with entries w_{ij}. We define Λ_1, Λ_2, Λ_3 and T as diagonal matrices with entries λ_{i1}, λ_{i2}, λ_{i3} and T_i, respectively, where $\Lambda_1 + \Lambda_2 + \Lambda_3 = I$ and $\Lambda_2 + \Lambda_3 \neq 0$. Then, the opinion dynamics for strategy q is given by:

$$x(k+1) = \Lambda_1 W x(k) + \Lambda_2 x^b + \Lambda_3 T u^q. \quad (1)$$

3 ANALYSIS

The following result gives conditions under which the opinion dynamics in Eqn. (1) converges to a unique steady-state value.

Theorem 3.1. *Assume $\mathcal{G}$ is strongly connected, and let $\Lambda_1 W \neq I$. Suppose the strategy $q \in Q$ selected by the external source is fixed from a certain time-step. Then, the dynamics in Eqn.* (1) *will converge to a unique value $x_{ss}^q := \lim_{k\to\infty} x(k)$. Moreover, x_{ss}^q will be independent of initial values of opinions of agents, $x(0)$, and any strategies the source uses prior to using the strategy q.*

The external source aims to select an optimal dissemination strategy q^* to drive agents' opinions x_{ss}^q towards a desired value x^*. However, initial predispositions of agents and functions characterizing the perceptions of information broadcast by the source may not be known. Specifically, the term $h^q := \Lambda_2 x^b + \Lambda_3 T u^q$ in Eqn. (1) is unknown to the external source.

The external source is assumed to have access to a set $\mathcal{D}$ of noisy observations of Eqn. (1) corresponding to each strategy. The term h^q is a random vector with mean $\mu^q = [\mu_{1,D}^q, \cdots, \mu_{N,D}^q]^T$ and covariance $\Sigma^q = diag\ [(\sigma_{1,D}^q)^2, \cdots, (\sigma_{N,D}^q)^2]$. Then the external source can use Gaussian process learning [13] to estimate (μ^q, Σ^q). These can be used to find the optimal dissemination strategy q^* that minimizes $L(q, x^*) := \frac{1}{N}\,\mathbb{E}[(x_{ss}^q - x^*)^T (x_{ss}^q - x^*)]$.

Proposition 3.2. *The optimal policy is*

$$q^* = \arg\min_{q \in Q} \Big(Tr[\Sigma(x_{ss}^q)] + (\mu(x_{ss}^q) - x^*)^T (\mu(x_{ss}^q) - x^*) \Big),$$

where $\mu(x_{ss}^q) := (I - \Lambda_1 W)^{-1} \mu^q$ and $\Sigma(x_{ss}^q) := (I - \Lambda_1 W)^{-1} \Sigma^q (I - W^T \Lambda_1)^{-1}$ and $Tr[\cdot]$ denotes the trace of a matrix.

4 EXPERIMENTS

We consider three graph networks: (i) **Watts-Strogatz small world graph** [15] (with $N = 1000$ nodes), (ii) **Barabási-Albert scale-free graph** [2] (with $N = 1000$ nodes), and (iii) **Facebook friendship social network graph** [12] (a subset of the Facebook friendship graph with $N = 2235$ nodes and 91000 edges). We assume the external source has 20 distinct dissemination strategies and wants to drive the opinions of agents toward a desired value $x^* = 1$. We compare our prospect-theoretic model of opinion evolution in Eqn. (1) with the following expectation-based update model:

$$x^E(k+1) = \Lambda_1 W x^E(k) + \Lambda_2 x^E(0) + \Lambda_3 T \Big(\sum_{\theta} q(\theta)\theta \Big) \mathbf{1}_N, \quad (2)$$

where x^E are agent opinions and $\mathbf{1}_N \in \mathbb{R}^N$ is a vector with all entries equal to 1. The expectation update model is a generalization of the model proposed in [10]. Given the initial opinion values of the agents, we compare their final values when opinions evolve following our model with prospect-theoretic information dissemination in Eqn. (1) and the expectation-based model in Eqn. (2) in Table 1. Table 2 compares the values of the average distance, $L(q, x^*)$, of agents' final opinions from the desired opinion when the external source takes into account prospect-theoretic perceptions of agents and when it does not during computation of an optimal dissemination strategy, with initial opinions of agents sampled from $Unif(-1, 0)$. We observe that the external source is more effective in driving opinions of individuals towards a desired value when using our prospect-theoretic model.

ACKNOWLEDGMENT

We gratefully acknowledge the support of the US National Science Foundation (Grant CNS-2153136), Air Force Office of Scientific Research (Grants FA9550-22-1-0054 and FA9550-20-1-0074), and the Office of Naval Research (Grant N00014-20-1-2636).

REFERENCES

[1] Claudio Altafini. 2012. Consensus problems on networks with antagonistic interactions. *IEEE Trans. Automat. Control* 58, 4 (2012), 935–946.

[2] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *science* 286, 5439 (1999), 509–512.

[3] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. 2000. Mixing beliefs among interacting agents. *Advances in Complex Systems* 3, 01n04 (2000), 87–98.

[4] Morris H DeGroot. 1974. Reaching a consensus. *J. Amer. Statist. Assoc.* 69, 345 (1974), 118–121.

[5] Noah E Friedkin and Eugene C Johnsen. 1990. Social influence and opinions. *Journal of Mathematical Sociology* 15, 3-4 (1990), 193–206.

[6] Noah E Friedkin and Eugene C Johnsen. 2011. *Social influence network theory: A sociological examination of small group dynamics.* Vol. 33. Cambridge University Press.

[7] Rainer Hegselmann, Ulrich Krause, et al. 2002. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation* 5, 3 (2002).

[8] Daniel Kahneman and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 2 (1979), 263–292.

[9] Naomi Ehrich Leonard and Simon A Levin. 2022. Collective intelligence as a public good. *Collective Intelligence* 1, 1 (2022), 26339137221083293.

[10] Tingyu Li and Hengmin Zhu. 2020. Effect of the media on the opinion dynamics in online social networks. *Physica A: Statistical Mechanics and its Applications* 551 (2020), 124117.

[11] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. 2014. Opinion dynamics on interacting networks: Media competition and social influence. *Scientific Reports* 4, 1 (2014), 1–7.

[12] Ryan A. Rossi and Nesreen K. Ahmed. 2015. The network data repository with interactive graph analytics and visualization. In *AAAI*. https://networkrepository.com

[13] Matthias Seeger. 2004. Gaussian processes for machine learning. *International Journal of Neural Systems* 14, 02 (2004), 69–106.

[14] Amos Tversky and Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 4 (1992), 297–323.

[15] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 6684 (1998), 440–442.