Big Data, Social and Behavioral Sciences, and National Security

Fred S. Roberts, CCICADA Center, Rutgers University, Piscataway, NJ
froberts@dimacs.rutgers.edu

1. Introduction

The age of big data has revolutionized the social and behavioral sciences, making it possible to draw conclusions from relevant information in ways heretofore impossible. This makes it possible to address challenges from many problems facing our society, in particular national security, where issues of individual and group behavior and social and economic forces play a central role in evolving threats and potential responses. However, in order to benefit from the availability of big data to bring social and behavioral phenomena to bear on national security, we face a variety of research challenges. This paper explores a selection of such challenges.

2. Algorithmic Decision Theory[1]

The theory of decision-making is heavily based in methods developed in the social sciences. Today's decision-makers have available remarkable new technologies, huge amounts of information, and ability to share information at unprecedented speeds and quantities. These tools and resources should lead to better decisions. However, traditional social-science-based tools of decision theory are often inadequate in this new “big data” environment.

New tools of decision theory bring with them daunting new problems and the need to apply them to issues arising in national security presents complex new challenges. Among the issues are: The massive amounts of data available are often incomplete, unreliable, distributed, or uncertain; interoperating/distributed decision-makers and decision-making devices need to be coordinated; many sources of data need to be fused into a good decision. When faced with such issues, there are few efficient algorithms available to support decisions. There is a long tradition of algorithmic methods in logistics and planning. But algorithms to automate, speed up and improve real-time decision-making are much less common. Algorithms for decision support, especially algorithms that can approximate good analytic solutions, are needed. Such algorithms should improve ability of decision-makers (human or automated). These considerations have given rise to the field called Algorithmic Decision Theory (ADT) ([5]).

To give one ADT example: With the increasing amount of data and increasing speed with which decisions need to be made, it is often necessary to make decisions depending on feedback from earlier ones before having access to all relevant data. Such sequential decisions arise e.g., in numerous inspection processes, where the outcome of one inspection determines which inspections to do next. Sequential decision-making needs new models and algorithms as traditional methods do not scale. For instance, in inspections of containers at a port, a decision-maker has to decide how to inspect them, which to subject to further inspection and which to pass through. Stroud and Saeger [33] developed a decision logic in which containers are classified using a Boolean decision function (BDF). Different binary tree representations for a BDF have different associated inspection costs and one seeks an efficient decision tree representation. While such problems are generally computationally challenging, one can hope for efficient solutions in certain cases. For relevant work on container inspection see [4,19,23,24].

ADT is also relevant to protection of critical infrastructure. Stable and reliable operation of the electric power grid is an example. Decisions about design, operation, and repair of grid systems must be made

[1] This section is heavily based on [29].

rapidly using massive amounts of data available about the state of the system. ADT challenges in management and control of the grid include scale of the problem, intrinsic uncertainties in monitoring its “state," and the potential to cause cascading blackouts. Research is needed to develop algorithmic solutions leading to real-time predictions and decisions for controlling and securing the grid. Issues include: fusing multiple data streams into forms suitable for action and human interpretation; quantifying and balancing risk from acting on insufficient information vs. delaying action to get more data; determining when is it advisable to “break" the grid and isolate instability to avoid cascading failures.

3. Game Theory[2]

Game theory, widely developed in Economics, is a well-established tool for studying adversarial behavior. Milind Tambe and colleagues have done extensive work using Stackelberg games, leading to a range of deployed applications: scheduling checkpoints and canine patrols at LAX; deploying air marshals on international flights; scheduling randomized Coast Guard patrols near ports; deploying CG boats to protect ferries; scheduling multi-operation patrolling (fare evasion, counter-terrorism and crime) on LA area metro trains; preventing illegal, unreported, and unregulated fishing; etc. See e.g., [1,28].

Classical work in game theory provides rich equilibrium concepts. However, to allow game theory concepts to scale up to large, complex systems, computational and representational insights are required. The emerging field of computational game theory addresses such issues [22]. We need to develop new methods for dealing with huge problems, e.g., methods for computing the solution to a game with a huge and possibly changing number of players.

The study of repeated games is a challenging area of research [13,27]. A player might use early rounds to learn about an opponent’s strategy or defenses or preferences; e.g., a terrorist might spend early “rounds” observing. Repeated games allow players to modify their strategies based on results of earlier rounds, a common approach of today’s transnational criminal organizations.

Another interesting area of research is behavioral game theory. Its methods challenge basic notions of rationality [6]. Experimental work in behavioral economics has repeatedly shown that human subjects will frequently deviate from traditionally assumed notions of self-interest and greed [17]. Moreover, they seem to do so in ways that are predictable, repeatable, and amenable to mathematical modeling and analysis. Concepts such as altruism, revenge, and envy have been quantified and fruitfully measured [7,11].

4. Economic Epidemiology[3]

The 2014/15 Ebola outbreak in West Africa reminded us that the world is ill-prepared for a severe disease epidemic or similar global sustained public emergency. Because of locally severe disease risks, global interconnectedness through transportation, increasing migration, tourism and trade, infectious diseases emerge and re-emerge more frequently; spread greater distances; pass more easily between humans and animals; and evolve into new and more virulent strains. Analysis of the Ebola outbreak provided valuable insights into the role of funeral practices, hospitals, social contact, and population mobility. Social and behavioral science challenges abound here.

The application of economics to the study of infectious disease has taken on a new meaning with the development of a sub-discipline called “Economic Epidemiology” [19,35]. Economic epidemiology deals

[2] Parts of this section are based on [29].

[3] This section borrows heavily from [30].

with the mathematical conceptualization of the interplay among economics, human behavior, and disease ecology to improve our understanding of the emergence, persistence, and spread of infectious agents [2,3]. To correctly evaluate health interventions and public policies, models of disease spread must incorporate both group and individual behaviors (e.g., will people comply with directions).

Economic Epidemiology issues of importance are developing models to understand the cost and impact of alternative disease treatment strategies, the allocation of resources between prevention and treatment, the development of incentives to achieve desired individual and group behavior.

5. Radicalization

An increasing challenge in homeland security is the radicalization of individuals outside terrorist groups. There is interesting work on modeling the spread of ideologies analogously to the spread of disease [9, 31]. Among the things to understand to build such models are characteristics of people vulnerable to spread of violent ideologies, structural and social characteristics of systems enabling such spread, and the impact of terrorist group extreme behavior on the vulnerable population [32]. A great deal of work on radicalization, aided by modern data science, has been done by sociologists and social psychologists [21,36].

Galem and Javarone [15] suggest borrowing from the field of sociophysics [14]: physics-inspired models study a large spectrum of social behaviors including opinion dynamics [34], crowd behavior [8], criminal activities [12], cultural dynamics [16], and spread of radicalization [15,26].

6. Responses to Disasters[4]

Models of disaster response behavior generally assume a fixed social landscape with passive bystanders and rational actors who comply with authorities, whether disasters are natural or man-made. This assumption may depend on effective communication by authorities and media. Past examples suggest that episodes of mass panic or hysteria are rare and localized, while actions based on perceived self-interest (evacuation, queries from the worried-well, antibiotic stockpiling) are widespread. Acts of spontaneous altruism and mutual aid, as well as criminal opportunism and civil disruption, also occur. Models are needed to understand how changes in social behavior under stress affect success of interventions.

Social media play an increasingly important role in allowing authorities to gain situational awareness in disasters. [37,38] have studied how to determine when an "event" is occurring and how it develops by following the Twitter stream from the 2011 Japanese Earthquake and Tsunami and 2010 Haitian Earthquake, discovering how to quickly summarize the evolution of keywords and facets and distribution of users. They learned "topic signatures" indicating when an event of given type occurs and monitored the pattern as an event unfolds. This and other work led to the discovery that people follow typical sequences when communicating in emergency situations. Such work involves large amounts of data. For example, a study of Hurricane Sandy [38] analyzed 6.5 million geo-tagged Twitter posts.

Resilience is another topic that should have gotten its own section. Community resilience under natural disasters is critical. Anecdotal evidence suggests that the spike in social media complaints/emergency calls lasts for a shorter time once community training (improved resilience) is in place.

[4] This section depends heavily on [25]

In disaster science, social responses that need to be studied and made precise include movement; compliance (quarantine, resistance, trust); rumor; herd mentality; role of differences in geography or social group; behavior of first responders; individual altruism.

7. Randomization and National Security

After the 2015 Paris attacks, security professionals put increased emphasis on making decision-making more difficult for terrorists, and specifically on randomization. Randomization is intended to confuse the adversary, make them work harder to understand defensive tactics, and make attacks more expensive and risky. In some cases, e.g., when secondary inspection resources are limited, randomization can provide increased security.

The vulnerability of stadiums and arenas to attacks was underscored by the 2015 and 2017 attacks on the Stade de France and the Manchester Arena. Randomization has been described as a "best practice" for a variety of aspects of stadium security [10]. For example, patron screening can include random selection for more (or less) rigorous inspection; security officers can be deployed using randomized schedules; employees can be randomly chosen for background re-checks.

There are a variety of research challenges arising from randomization. For example, how to: evaluate the effectiveness of randomization; find simple randomization designs for selecting patrons for various types of screening; design implementation procedures that minimize the possibility of being accused of profiling.

Insider threats are a challenge for infrastructure protection. Routine background rechecks for all employees can be costly, so randomly selecting some employees for recheck is an attractive alternative. But would random rechecks be acceptable to unions? What percentage of employees should be chosen each time period? What incentives can be designed for employees to self-report issues before being randomly discovered?

8. Information Sharing Environments

The 9/11 Commission report emphasized information sharing, leading to research on formal "information sharing environments" (ISE's) [18]. To be successful, an ISE must include agreed-upon and rigorously defined technological components, e.g., interoperability. However, it must also address human elements that often block effective sharing of information.

Some research questions here are[5]: What processes best identify and resolve conflicts within an ISE? What cultural impediments keep stakeholders from implementing an ISE? What stakeholder organizational commitments are required for success? What privacy and safeguarding processes are needed? What governance policies make stakeholders feel represented?

A fundamental purpose of an ISE involves collecting, analyzing, and disseminating data across jurisdictional and disciplinary boundaries. While some issues of data stewardship are technical, others stem from defining policies and practices concerning data quality, costs of data collection, and privacy protection. Best practices for such policies and practices need to be developed, keeping human factors at the forefront.

[5] These research questions derive from ideas in [18]

References

1. An, B., et al., “PROTECT – A deployed game theoretic system for strategic security allocation for the United States Coast Guard,” *AI Magazine*, **4** (2012) 96-110.

2. Barrett, S., “Eradication versus control: The economics of global infectious disease policies," *Bull. World Health Organ*., **82** (2004), 683-688.

3. Bauch, C.T., Earn, D.J., “Vaccination and the theory of games," *Proc. Natl. Acad. Sci. USA*, **101** (2004), 13391-13394.

4. Boros, E., Fedzhora, L., Kantor, P.B., Saeger, K., Stroud, P., “Large scale LP model for finding optimal container inspection strategies," *Naval Research Logistics,* **56** (2009) 404-420.

5. Brafman, R.I., Roberts, F.S., and Tsoukias, A. (Eds.), *Algorithmic Decision Theory, Proc. Second Intl. Conf. ADT 2011*, Lecture Notes in Computer Science book series (LNCS, volume 6992), Springer, 2011.

6. Camerer, C.F., *Behavioral Game Theory*, Princeton University Press, 2003.

7. Camerer, C.F., Ho, T.H., and Chong, K., “Models of thinking, learning and teaching in games," *American Econ. Review Papers and Proceedings,* **93** (2003), 192-195.

8. Castellano C, Fortunato S, Loreto V., “Statistical physics of social dynamics,” *Rev. Mod. Phys*., **81** (2009) 591–646.

9. Castillo-Chávez, C., Song, B., “Models for the transmission dynamics of fanatic behaviors,” in: H.T. Banks, C. Castillo-Chávez (Eds.), *Bioterrorism: Mathematical Modeling Applications in Homeland Security*, SIAM Frontiers in Applied Mathematics, vol. 28, SIAM, Philadelphia, 2003, pp. 155–172.

10. CCICADA Center, *Best Practices in Anti-terrorism Security for Sports and Entertainment Venues: Resource Guide,* CCICADA Center, Rutgers University, July 2013, available at https://www.safetyact.gov/externalRes/refdoc/CCICADA%20BPATS.pdf

11. Costa Gomes, M., Crawford, V. Broseta, B., “Experimental studies of strategic sophistication and cognition in normal-form games," *Econometrica*, **69** (2001), 1193-1235.

12. D’Orsogna M, Perc M., “Statistical physics of crime: A review,” *Phys. Life Rev.* **12** (2015) 1–21. pmid:25468514

13. Fudenberg, D., Tirole, J., *Game Theory*, MIT Press, 1991.

14. Galam, S. “Sociophysics: A review of Galam models,” *International Journal of Modern Physics C,*. **19** (2008) 409–440.

15. Galam, S., Javarone, M.A., “Modeling radicalization phenomena in heterogeneous populations,” *PLoS ONE* **11** (2016): e0155407. https://doi.org/10.1371/journal.pone.0155407

16. Gracia-Lazaro, C., Quijandria, F., Hernandez, L., Floria, L.M., Moreno, Y., “Co-evolutionary network approach to cultural dynamics controlled by intolerance,” *Phys. Rev. E.,* **84** (2011) 067101.

17. Harstad, R.M., “Dominant strategy adoption, efficiency, and bidder's experience with pricing rules," *Experimental Economics*, **3** (1990), 261-280.

18. IJIS Institute, *Information Sharing and Safeguarding (IS&S) Playbook, Version 2*, IJIS Institute, Ashburn, VA, October 2016, http://www.standardscoordination.org/sites/default/files/docs/ISS_Environment_Playbook.pdf

19. Kantor, P., Boros, E., “Deceptive detection methods for effective security with inadequate budgets: The testing power,” *Risk Analysis,* **30** (2010), 663-673.

20. Klein, E., Laxminarayan, R., Smith, D.L., Gilligan, C.A., “Economic incentives and mathematical models of diseases," *Environment and Development Economics*, **12** (2007), 707-732.

21. Kruglanski, A.W., Gelfand, M.J., Belanger, J.J., Sheveland, A., Hetiarachchi, M., Gunaratna, R.., “The psychology of radicalization and deradicalization: How significance quest impacts violent extremism” *Advances in Political Psychology,* **35** (2014).

22. Linial, N., “Game-theoretic aspects of computing," in R.J. Aumann and S. Hart (eds.), *Handbook of Game Theory with Economic Applications*, II, chapter 38, (1994), 1340-1395.

23.Madigan, D., Mittal, S., Roberts, F.S., “Sequential decision making algorithms for port of entry inspection: Overcoming computational challenges," in G. Muresan, T. Altiok, B. Melamed, and D. Zeng (eds.), *Proceedings of IEEE International Conference on Intelligence and Security Informatics (ISI-2007)*, IEEE Press, Piscataway, NJ, May 2007, 1-7.

24. Madigan, D., Mittal, S., Roberts, F.S., “Efficient sequential decision making algorithms for container inspection operations,” *Naval Research Logistics*, 58 (2011), 637-654.

25. McKenzie, E., Roberts, F., *Modeling Social Responses to Bioterrorism Involving Infectious Agents*, Report, DIMACS Center, Rutgers University, July 2003.

26. McMillon, D., Simon, C.P., Morenoff, J., “Modeling the underlying dynamics of the spread of crime,” *PloS ONE,* **9** (2014); e88923. pmid:24694545

27. Myerson, R., *Game Theory*, Harvard University Press, Cambridge, MA, 1991.

28. Pita, J., et al., “Deployed ARMOR protection: The application of a game theoretic model for security at the Los Angeles International Airport,” in *Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2008.

29. Roberts, F.S., “Computer science and decision theory,” *Annals of Operations Research*, **163** (2008), 209-253.

30. Roberts, F.S., “Greedy algorithms in economic epidemiology,” in A. Gumel and S. Lenhart (eds.), *Modeling Paradigms and Analysis of Disease Transmission Models*, American Mathematical Society, Providence, RI, Vol. 75 (2010), 249-268.

31. Santonja, F.J., Tarazona, A.C., Villanueva, R.J., “A mathematical model of the pressure of an extreme ideology on a society,” *Computers and Mathematics with Applications,* **56** (2008) 836-846.

32. Shepherd, L.O.V, “Suicide terrorism: Modeling group dynamics and individual behavior,” in J.I. Victoroff (ed.), *Tangled Roots: Social and Psychological Factors in the Genesis of Terrorism*, 2006, pp. 410-430.

33. Stroud, P.D., Saeger, K.J., “Enumeration of increasing Boolean expressions and alternative digraph implementations for diagnostic applications," in H. Chu, J. Ferrer, T. Nguyen, and Y. Yu (eds), *Proceedings Volume IV, Computer, Communication and Control Technologies*: I, International Institute of Informatics and Systematics, Orlando, FL, 2003, 328-333.

34. Sznajd-Weron, K., Sznajd, J., “Opinion evolution in closed community,” *International Journal of Modern Physics C.*, **11** (2000) 1157.

35. Tanaka, M.M., Kumm, J., Feldman, M.W., “Coevolution of pathogens and cultural practices: A new look at behavioral heterogeneity in epidemics," *Theoretical Population Biology*, **62** (2002), 111-119.

36. Thompson R.L., “Radicalization and the use of social media,” *Journal of Strategic Security,* **4** (2011) 167–190.

37. Tyshchuk, Y., Li, H., Ji, H., Wallace, W.A., “Evolution of communities on Twitter and the role of their leaders during emergencies,” in R. Missaoui and I. Sarr (Eds.), *Social Network Analysis - Community Detection and Evolution*, Springer, New York, 2014.

38. Wang, H., Hovy, E., Dredze, M., “The Hurricane Sandy Twitter corpus,” *Proceedings AAAI Workshop: WWW and Pubic Health Intelligence*, 2015.