Generating an Agent Taxonomy using Topological Data Analysis

Extended Abstract

Samarth Swarup and Reza Rezazadegan
Biocomplexity Institute & Initiative
University of Virginia
Charlottesville, Virginia
[swarup,rr7nz]@virginia.edu

ABSTRACT

One of the challenges with the interpretability of large and complex multiagent simulations is understanding the kinds of agents that emerge from the interactions in the simulation, in terms of agent states and behaviors. We address one aspect of this challenge, which is to generate an agent taxonomy by analyzing the simulation outputs. We show that topological data analysis (TDA) can be used for this problem by applying it to agent trajectories, and present some promising results from the analysis of a large-scale disaster simulation. The results show a taxonomy of multiple types of agents that emerge, and which can be tracked over time through this taxonomical description.

KEYWORDS

multiagent simulation; topology; taxonomy; simulation analytics

ACM Reference Format:
Samarth Swarup and Reza Rezazadegan. 2019. Generating an Agent Taxonomy using Topological Data Analysis. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 2 pages.

1 INTRODUCTION

Large-scale simulations are starting to be used in modeling and understanding various social phenomena such as disaster response and evacuation [e.g., 1, 3]. One of the main benefits of this approach is that multiagent simulations offer a naturalistic representation where interventions are easy to evaluate and which can, hopefully, lead to operationalizable and actionable plans and policies. However, there are a number of interesting technical challenges in this area, not the least of which is that sense-making with such complex simulations is hard. New simulation analytics methods are needed for making simulations easy to understand and use, especially by end-users who may not be computer scientists. This is necessary if we wish to see widespread and mainstream adoption and application of multiagent simulation methods.

One such challenge is to extract a taxonomy of agents from the results of a multiagent simulation. This is useful because it is a helpful way to present information to end-users. For instance, operational end-users who actually have to implement response plans during a disaster would benefit greatly from this type of information. This is a hard problem, however, because simulations exhibit emergent effects due to interactions between agents and, thus, the kinds of agents that "emerge" from a simulation might be different from the types of agents that are designed into the simulation. The existence of emergent roles has long been recognized in disaster management and response [5], and a typology of emergent behaviors in disasters has also been suggested [4].

Our goal here is to devise a method for generating a *taxonomy* of agents from the results of a simulation. A taxonomy goes beyond a typology in that it not only identifies meaningful types from a data set, but also establishes relationships among those types. For example, a taxonomy of biological organisms generally groups them into "taxa" by shared morphological characteristics. It can also create a ranking by grouping the taxa, like a hierarchical clustering method. Our approach is to use topological data analysis (TDA) [e.g., 2]. To assess the method on a complex data set, we use sample trajectory data from a recent disaster simulation [1]. We show that a set of emergent categories of agents can be extracted from the results.

2 TOPOLOGICAL DATA ANALYSIS

Topology is the study of spaces equipped with a notion of neighborhood between their elements. TDA aims to find a hypothetical topological space to which a given data set belongs. For example, we can take the proximity graph of a data set, i.e., the graph obtained from the data set by connecting pairs of points whose distance is less than a given threshold. This graph G can then be enhanced to a higher dimensional topological space by taking its clique complex [6], i.e., the simplicial complex obtained by adding all the cliques in G as hyper-edges.

Lum et al. [2] recently introduced a simplified method to assign a graph to the data. The data set is equipped with a filtration function and one uses this function to divide the dataset into a set of overlapping bins. One then clusters the data in each bin. Each such cluster gives us a vertex of the output graph and two such vertices are connected by an edge if their corresponding clusters have elements in common.

They applied this method to several data sets such as gene expression from breast tumors, voting data from the United States House of Representatives and player performance data from the NBA to obtain new insight on associations in among data points. In each case, they found stratifications of the data which are more refined than those produced by traditional methods.

Though it uses clustering, TDA is solving a fundamentally different problem. It captures the shape of the data cloud as a graph, even if the data points are not in distinct clusters. It thus reveals the structure of the data set, and the edges in the graph can reveal meaningful relationships.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

3 ANALYSIS OF A DISASTER SIMULATION

We analyze a sample of agent trajectories from a recent disaster simulation [1]. In this section, we briefly describe the simulation before going on to show the results of our method applied to the data set.

In the scenario, an improvised nuclear device is detonated in Washington DC, USA. This hypothetical disaster is known as National Planning Scenario 1 (NPS-1). In the NPS-1 simulation [1], they modeled a detailed "synthetic" population of the region, including agent demographics, household structure, daily activity patterns, road networks, and various kinds of locations, such as workplaces, schools, government buildings, etc. The simulation also contained models of multiple infrastructures, including power, communication, transportation, and health. They also modeled multiple behaviors, and the interactions between human behaviors and the infrastructures.

We obtained a sample of 10,000 agents, out of a total of 730,833 agents modeled in the simulation. The variables included in the data set are *distance from ground zero* in meters, *level of radiation exposure* in centiGrays, *health state*, which is an integer in the range [0, 7], and *behavior*, which is nominal, indicating which of the above behaviors an agent is executing at each time step.

3.1 Results

To enable viewing the results in a 3D plot, we restrict our analysis to one pair of variables (plus time). We also limit our TDA graph construction to time step 20 because we found that agent states don't change very much after that. To run TDA for the full sample of 10,000 agents takes a few hours (on a MacBook Pro with 2.6 GHz Intel Core i7 and 16GB RAM), and results in a plot that is too cluttered to understand easily. Therefore, we demonstrate results with a random sample of 100 agents. We tried the analysis with multiple random samples of 100 agents, and the results are qualitatively similar each time.

Figure 1 shows the results. The actual trajectories are shown with thin gray lines, while the TDA graph is shown in blue. The graph has been simplified slightly by homeomorphic smoothing (edge contraction). The idea of homeomorphic smoothing is to simplify a graph by removing nodes of degree 2 and connecting their neighbors to each other. The graphs generated by TDA often exhibit long paths where it follows a trajectory of an agent that doesn't interact with other agents. For the purpose of generating a taxonomy, these intermediate nodes in paths in the TDA graph don't add any information and can be removed.

3.2 Taxonomy

The annotations in Figure 1 show some of the categories that emerge.

We can treat the graph nodes at the end of the simulation as the top level taxa in our emergent taxonomy. Thus, for example, in Figure 1, the top level taxa correspond to agents who are

(1) close to ground zero from the beginning, but have low exposure,
(2) far ground zero to begin with but move closer, and have slightly more exposure,
(3) close to ground zero and have high exposure.

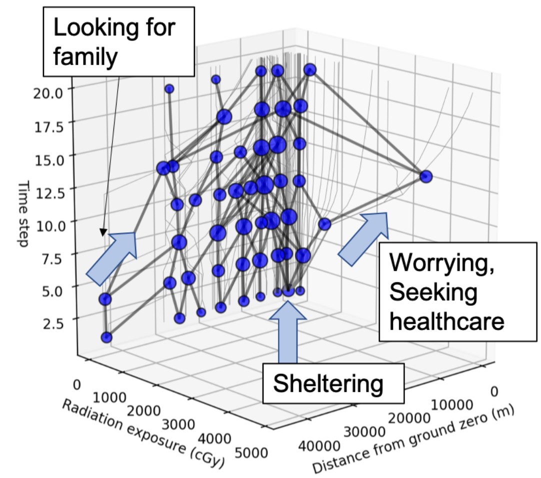

Figure 1: Topological data analysis of 100 randomly chosen agent trajectories in the disaster simulation, where the variables are distance from ground zero and level of radiation exposure.

As we follow the graph edges and move down from the top level, we can describe how the agents got to the states in the top level. As the annotation in the right panel of Figure 1 show, the first category broadly corresponds to people who are sheltering, the second to people who are seeking family members, and the third to people who are worrying or seeking healthcare.

Thus the method gives us a ranked classification, i.e., a taxonomy, not just a typology of agents in the simulation. Further, these categories are not designed into the simulation, but emerge from the interactions induced by agent movement, communication, and behavior.

ACKNOWLEDGMENTS

S.S. was supported in part by DTRA CNIMS Contract HDTRA1-17-0118.

REFERENCES

[1] Christopher Barrett, Keith Bisset, Shridhar Chandan, Jiangzhuo Chen, Youngyun Chungbaek, Stephen Eubank, Yaman Evrenosoğlu, Bryan Lewis, Kristian Lum, Achla Marathe, Madhav Marathe, Henning Mortveit, Nidhi Parikh, Arun Phadke, Jeffrey Reed, Caitlin Rivers, Sudip Saha, Paula Stretz, Samarth Swarup, James Thorp, Anil Vullikanti, and Dawen Xie. 2013. Planning and Response in the Aftermath of a Large Crisis: An Agent-based Informatics Framework. In *Proceedings of the 2013 Winter Simulation Conference*, R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl (Eds.). IEEE Press, Piscataway, NJ, USA, 1515–1526.

[2] P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson. 2013. Extracting insights from the shape of complex data using topology. *Scientific Reports* 3 (2013), Article 1236.

[3] Xiaoshan Pan, Charles S. Han, Ken Dauber, and Kincho H. Law. 2007. A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI & Society* 22 (2007), 113–132.

[4] Damienne Provitolo, Edwige Dubos-Paillard, and Jean-Pierre Müller. 2011. Emergent Human Behaviour During a Disaster: Thematic *Versus* Complex Systems Approaches. In *Proceedings of EPNACS 2011 within ECCS'11*. Vienna, Austria.

[5] Robert A. Stallings and E. L. Quarantelli. 1985. Emergent Citizen Groups and Emergency Management. *Public Administration Review* 45 (1985), 93–100. Special Issue: Emergency Management: A Challenge for Public Administration.

[6] Afra J. Zomorodian. 2005. *Topology for Computing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511546945