Chapter 1
Prescriptive Analytics Using Synthetic Information

Madhav V. Marathe
Virginia Polytechnic Institute and State University, USA

Henning S. Mortveit
Virginia Polytechnic Institute and State University, USA

Nidhi Parikh
Virginia Polytechnic Institute and State University, USA

Samarth Swarup
Virginia Polytechnic Institute and State University, USA

ABSTRACT

In this chapter, the authors describe the use of synthetic information for doing prescriptive and predictive analytics. They discuss in detail how synthetic information is created by combining data from multiple sources and then describe its role in an ongoing disaster resilience study where they simulate the aftermath of a hypothetical nuclear detonation in Washington DC.

1. INTRODUCTION

Policy-making, by its nature, requires reasoning about counterfactual or what-if scenarios. In disaster planning, for instance, policy-makers have to come up with plans for evacuation, search and rescue, relief efforts, and rehabilitation. It is often the case that the data on which these plans are based come from similar disasters in other locations, and from prospective surveys where people are asked what they would do in certain circumstances.

Policy-makers, therefore, have to attempt to translate these data from related sources to their context of interest by asking what-if questions. For example, what if a hurricane of the magnitude of hurricane Sandy struck Washington DC? What if it caused a flood? What if it made the cell phone network inoperative? What if it caused thousands of people to be displaced from their homes?

In order to answer these questions, they might, for example, look to data from the aftermath of prior hurricanes in other cities, like Sandy, Katrina, and Irene. However, the structure of Washington DC is different in its physical layout, demographics, infrastructure, and available resources. To understand what would happen in Washington DC in a similar disaster, deep knowledge of the interactions between human behavior, infrastructure, and the disaster conditions is required. If we simply

DOI: 10.4018/978-1-4666-5063-3.ch001
extract patterns from the past data and expect the same things to happen in the new scenario, we will likely be incorrect (Haas, Maglio, Selinger, & Tan, 2011).

The situation is even more extreme when we consider disasters that have never taken place in the US, such as the detonation of a (relatively) small nuclear device at ground level. In this case, there are data available from nuclear tests that can be extrapolated and combined with data on the built environment to perform calculations of the physical impact of the event. However, disaster planning requires much more than that; it requires an understanding of the human response, and its interaction with and dependence upon the infrastructural and technological environment. Data analytics and careful, detailed, model-building can help to answer these questions, but these analytics need to be deep predictive analytics that can support what-if reasoning. This kind of analytics has been termed prescriptive analytics (Robinson, Levis, & Bennett, 2010).

In this chapter we describe our approach to prescriptive analytics, which is based on generating synthetic information, and then performing large-scale simulations using this synthetic information.

A synthetic information system in our context comprises of synthetic individuals, infrastructure elements and networks as well as behavioral representations of the agents and the infrastructure. We call it a synthetic information system since: (i) such information is synthesized from diverse sources of information that is often available at various levels of aggregation, (ii) individuals comprising the populations are not real but based on aggregate data and constructed to be statistically indistinguishable from the real population. Both these aspects are important. Furthermore, it is important that individual and collective behaviors are an integral part of the synthetic information. This is procedural data and is an important component when developing dynamic models.

Our approach, thus, is to build a very detailed, synthetic representation of the population and infrastructure of the region of interest, including procedural descriptions of behavior and infrastructure operation. This information can then be used to conduct simulations of multiple scenarios, interventions, and policies, which in turn can support policy-making.

The rest of this chapter is organized as follows. First we give a detailed account of how synthetic population and infrastructure models are constructed. Then we describe the application of these methods to the simulation of the aftermath of the detonation of an improvised nuclear device in Washington DC. We show how this kind of analytics can be used to prescribe a response strategy that focuses on rapid restoration of the communication infrastructure, which in turn could help to save lives by channeling natural human instincts (to seek information, to seek family members, to help others, etc.) in beneficial directions.

This is work that has been developed and carried out over a number of years by many people at the Network Dynamics and Simulation Science Laboratory and their collaborators. This overview chapter owes a debt of gratitude to all their contributions.

2. SYNTHETIC INFORMATION

Realistic simulation models depend closely on data. Obtaining such data, however, is often infeasible. The cost of collecting the data may be very high or the time to collect the data renders this option unrealistic. In many cases the data may simply not exist, or, if it does exist, may not be available due to privacy concerns.

Synthetic information is a term used for data that is generated from multiple unstructured data sources to adequately capture, construct and represent this type of data. Here the term data may mean traditional data, but can also include procedural data, that is, an algorithmic representation of a system component. Synthetic information can be generated in several ways with examples including...