Pilot Study: Agent-Based Exploration of Complex Data in a Hospital Environment

Ted Carmichael¹, Mirsad Hadzikadic¹, and Ognjen Gajic²

¹ University of North Carolina at Charlotte, College of Computing and Informatics, Charlotte, North Carolina, USA  
{mirsad,tdcarmic}@uncc.edu  
² Mayo Clinic  
Rochester, Minnesota  
gajic.ognjen@mayo.edu

Abstract. We present the results of a pilot study created to explore a sub-set of complex data, utilizing an agent-based model simulation tool. These results center on data taken from hospital admission records, tracking patient attributes and how they relate to patient outcomes. The focus of this work is to highlight three design principles: 1) using an iterative process between the modeling of a system and the grounding of that simulation with real-world data; 2) a focus on agent primitives, emphasizing bottom-up emergence of effects, rather than top-down control; and 3) integration of various “theories” of patient care and hospital effectiveness as a method for experimentation with Complex Adaptive System-based data mining.

Keywords: Complex Adaptive Systems, Agents, Data Mining, Machine Learning.

1 Introduction

Hospitals represent a complex environment for patient care, with many variables affecting patient outcomes. These variables can often interact in surprising ways, producing non-linear effects in a dynamic environment. Complex Adaptive Systems (CAS) techniques can be used to design an Agent-based Model (ABM) for simulating these systems, allowing for a meaningful exploration of large datasets related to patient outcomes.

1.1 Design Principles

In building an ABM for a complex environment, such as that found in a hospital setting, our design plan incorporates three distinct principles, designed to lead to a robust and useful simulation environment. The first of these is that the initial simulation is not intended to be the final product. Building a meaningful simulation is by necessity an iterative process, and the first stage is to have a model that is both flexible enough to allow calibration to real-world data, yet powerful enough to make meaningful predictions that can be used to suggest what additional data needs to be collected.
The second principle we follow relates to both of these points, in that the model emphasizes the use of agent primitives. By focusing on primitives that use simple rules of interaction, bottom-up emergence of salient phenomena is favored over other methods that explicitly incorporate such phenomena in a top-down manner. Often, models that attempt to prescribe too many effects lack the flexibility to adequately handle or predict unexpected outcomes.

The third principle is that the model should not be intended to replicate a single hospital environment, but rather be generally applicable to a variety of environments. One way to accomplish this is to incorporate additional flexibility in the simulation by allowing operator control over key attributes. This allows for more general applicability by allowing various scenarios – in this case, theories of patient care and patient tracking – to be individually initialized for repeated and varied experimental simulation runs. This is similar to the approach used by Whitmeyer et al. [6] allowing for multiple social theories to be incorporated in a simulation test-bed environment.

1.2 CAS Data Mining

Complex adaptive systems differ from machine learning and data mining tools in that CAS simulations use all patients (or agents) to play out (one could call it “predict”) what will happen over time with this particular population. Even the time component of each simulation is important, as it indicates how long it will take to reach a specific conclusion. In CAS simulations, insights emerge over time. For example, as we will see in the examples below, the fact that the patient population formed three clusters in the first experiment was not predefined and the varied shapes/cohesiveness of these clusters emerged over time. We can run numerous simulations just to see whether differing random distributions of patients or parameter settings would make any difference in the final outcome.

Data mining and machine learning tools work differently. They require that the study designer divide the data set into two disjoint groups: the training group and the testing group. The algorithm then uses the training set to produce, say, the classification tree, while the testing set is used subsequently to evaluate the predictive accuracy of this tree. We can combine the two approaches in the following way: use CAS to produce emergent properties of the population, which are then evaluated using the testing set.

1.3 Outline

Here we present the results from the first stage of this long-term research plan for building such a system. We identify the five patient attributes used in this pilot study, and detail the simulation tool used to experiment with these attributes. Experimental results are presented, which are used to highlight the inherent flexibility of the CAS paradigm, as well as the three design principles mentioned above. First, we perform a CAS analysis of the population. Then, we combine a CAS and a data mining method. These results are discussed and generalized in order to present a template for the applicability of agent-based CAS models to a wide variety of complex data.