AGENT-BASED RULE-DRIVEN DISEASE MODELING AND RESPONSE EVALUATION FOR HEALTHCARE SYSTEM

Farrukh Arslan
Distributed Multimedia Systems Lab, Department of Electrical and Computer Engineering, School of Engineering, Purdue University, 475 Stadium Mall Drive, West Lafayette, IN 47907, USA

ABSTRACT:
With the growing threat of spread of diseases, thereby forming epidemic - and resulting in loss of life in major populated areas - many scientists and domain experts have become interested to find different ways to predict the spreading of epidemic such that the future course of action can be formulated. In this way they can better prepare the stakeholders to confine the disease in small regions; even totally prevent it from happening and therefore reduce the loss of human lives. Therefore the spread of an epidemic and predicting its course in a city is an important aspect of epidemic surveillance and thus the main subject of this study. This study helps generate various prevention and control techniques known as Visual Analytic Decision Support Environment to build a better and a much safer living for people. This is accomplished by using a combination of geographic information system tools and environmental controls to create a scalable, two layered; spatio-temporal Agent-Based Model that predicts the disease spread throughout any region by showing the interaction between agents/individuals. The first layer is to formulate probabilistic models to create a population, synthesizing a population of around two hundred thousand agents that would interact in a geospatial context. The second layer uses another set of probability distributions to predict a disease spread model based on a-priori information that has been provided by the city's health officials. The simulation results show the efficacy of this rule-based disease modeling and response system.

Key words: agent based model, decision environment, spatio-temporal epidemic, scenario simulation

INTRODUCTION
The control of outbreaks of infections and infectious diseases such as influenza and AIDS has become a major priority in today’s public health sector. Key research in this area has advanced the ability of public health officials to prevent and control these types of infectious disease outbreaks. Purification of the extracts, the use of the spectroscopic and photometric methods to identify extracts content and the quantity in which they are presents in the extract. A problem is to choose between some extracts containing same substances but in different amounts. Another problem is to consider the extracts containing no or less possible frequent allergens.

Previous research efforts show that modeling infectious disease spread will pinpoint the areas that need to be controlled, vacated and/or vaccinated. Two types of model approaches have been used throughout the preceding research work; individual-based models and computational epidemiology models. Individual (agent) based models have come to be more popular. The reason is that visualizing certain factors can be seen more explicitly in specific structures. Agent based modeling also gives researchers a way to connect the spatial and temporal aspects of disease spread. Previous research defines “an agent” in different ways. One uses agents as both locations and people, while others define an agent as the behavior of a person towards a certain disease. In this study “the agent” is defined as one individual and the connections are for defining interactions between these individuals.

RELATED WORK
S. Eubank et al. [1] studied the algorithmic and structural properties of very large social contact networks for the city of Portland, Oregon. A bipartite graph is formed consisting of people and locations. The nodes represent people;
whereas locations are represented by edges. However, this bipartite setup does not give users details about the interaction between nodes. They use a CL-model which utilizes the random graph approach. The CL-model simulates the basic characteristics of the dataset well. In addition, fast approximation algorithms were developed to compute basic structural properties. In this particular study, methods for generating random networks in near-linear time were investigated. These results were used to study the impact of policy decisions for large scale epidemic control in urban environments. This study shows the efficiency of different algorithms that deal with the disease spread model structure but did not visualize their results.

L. Perez and S. Dragicevic [2] developed an agent-based modeling approach which brought together geographic information systems and the simulation of communicable disease spread in urban environments through the results of individual interactions in a geospatial context. The GIS-agent based model design for this study can be easily customized to study the disease spread dynamics of any other communicable disease by simply adjusting the modeled disease timeline and/or the infection model and modifying the transmission process [3].

J. Wang et al. [4] proposed an epidemiologic model of flu to understand the disease’s spatial diffusion through a human contact network. This study’s objective is to develop an ABM approach that connects GIS and urban environments to simulate the spread of influenza. The model was developed using Repast Symphony platform with the application of JAVA and GIS tools. A prototype was developed to carry out agent-based simulation of influenza transmission and control in a specified region. In the model, influenza was described with a mathematical relation between transmission probability, the distance of two individuals, latent period, the time from infection to death, cure rate, and other parameters [5]. Users can change the simulation results through the adjustment of the value of time to the hospital, for instance.

The studies and models explained above provide valuable insight that will aid health officials to make better decisions when it comes to epidemic protocol. Nevertheless, each model described above is missing a certain feature; whether it is efficiency of computation, specific variables that are vital in the calculation of an accurate disease spread model, scalability, or the inability to compute more than one type of disease spread.

**RULE BASED DISEASE MODEL**

In this study, we will use a simple simulation model [6] as described in Figure 1. This simulation model uses the SIR disease model which defines the rules of infection in three stages. The first is the susceptible state (S) where an agent has the ability to get infected at any point in time. The second state is infection (I), the agent will move to this state when a sufficient amount of neighbors are also in this state. After a specified time period, an agent will move into the third state which is recovered (R). A deterministic SIR model was developed to analyze the numerous factors involved in disease spread. This would be a base tester before developing a probabilistic SIR model to track disease spread throughout a city. There are two SIR rule files that are being tested.

![Figure 1. SIR model](image)

**Rule #1:** This is a basic SIR model where the only parameters are the number of neighboring agents that are infected, the number of days until infection and the number of days until
recovery. An agent will become infected only if two or more of its neighbors are infected. The agent will only transition to the infection stage two days after the connection of infected neighbors. An agent will transition to the recovery state after five days of infection.

Rule #2: This SIR model uses age groups as a defining factor of disease spread. The synthesized population was split into five different age groups. Each age group has its rules of disease distribution. For example, as seen in Table 1, a child can only become infected if two children, one adult or one elderly agent is also infected. Whereas in Table 2, each age group is given a specific time frame to transition from the infection stage to the recovery stage.

<table>
<thead>
<tr>
<th>Table 1: Rules of infection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Group</td>
</tr>
<tr>
<td>CHILD</td>
</tr>
<tr>
<td>ADULT</td>
</tr>
<tr>
<td>ELDERLY</td>
</tr>
</tbody>
</table>

BRUNSWICK LENS MODEL

In this study, we will use the Brunswick Lens Model that can be expanded in several ways, depending on what information or data being used or tested. Figure 2 depicts the expanded Brunswick lens model which could be used to measure the judgment and quality of the simulated results. In previous developed versions of the model, the data was not split into simulation and operator cues. This is important in our study of epidemic since the final judgment that is being made is dependent on epidemiological base knowledge. The primary cues (X) will be what are seen as the output of the simulated SIR model. Whereas the secondary cues (U) will be the addition of the epidemiological knowledge to the primary cues. This will in turn lead to a judgment made by a user.

Table 2: Recovery Rule

<table>
<thead>
<tr>
<th>Age Group</th>
<th>CHILD</th>
<th>ADULT</th>
<th>ELDERLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovery</td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

Cues are events that are being monitored to make a sound judgment on any given situation. In our case, the judgment will be hindered due to access control.

In this epidemics simulation model access control will be introduced as the descriptor becomes a primary cue. The secondary cue will be created through the epidemiological knowledge of the area. Access Control can be executed in terms of both spatial and time measurements. In this specific case study, the spatial component of the data is divided into union councils (UC). Examples of cues that will be measured are:

- Union Council Density of Io (initial infection) Size of Io on day 1
- Number of infected days 1-3 based on gender
- Number of infected days 1-3 based on age group

\[
SS = \frac{((R_{0T})(V_{Rx})(G)(V_{ux})(R_{ux}))^2}{\left[\left(\tau_{x0} - \tau_{x0}/S_0\right)\right]^2 - \left[\left(\tau_{y0} - \tau_{y0}/S_0\right)\right]^2 - \left[\left(\tau_{y0} - \tau_{y0}/S_0\right)\right]^2}
\]

Where \( G = \frac{1+\sum_{i=1}^{\text{Io}}}{\sqrt{(1+\sum_{i=1}^{\text{Io}})^2}} \)
Prediction: Number of infected agents on day 6

The Quality of Judgment will be measured based on the following cues and more [6-8]. Experiments will be to change the time range to see when the cue gives the best quality of judgment (i.e., days 4-5 instead of 1-3). The experimentation of the quality of judgment will be run under a policy of different access control roles. These roles will help show which cues are more valuable and how the quality of judgment changes with each level of access control restraint. Roles one to five can only access days one through three of the epidemic data with specific restrictions. Whereas roles five and six can access days four and five also with specific restrictions. Below is the description of the information that can be accessed by each role.

1. Top five densely populated Union Councils
2. Top five sparsely populated Union Councils
3. Educated population
4. The route that has the median amount of population using it
5. Same as role 1
6. Same as role 2

![Rule Set #1](image-url)

*Figure 3. Daily number of infections with Rule #1 and initial infection in a low density UC*
RESULTS AND DISCUSSION
Two scenarios were run under the SIR rules mentioned above. [9-11] For both scenarios, the initial infection started in a low density populated area. In Figure 3, the rule of infection was that at least two neighbors would be infected in order for the current individual to get infected. Whereas as shown in Figure 4, there needs to be four neighbors infected for the current agent to get infected. Looking at figure 3, we observe that there is a slight increase in infections that is throughout the multiple days and there is not a decrease since the recovery rate is not fast enough to beat the infection rate. On the other hand and as described by Figure 4, the rule set changes and the recovery rate can compete with the infection and that is why we observe a decrease in infection between days 3 and 4. It was expected in rule set #2 that the number of infected individuals will go back to zero at a given time since the amount of neighbors needed to infect is larger than rule set #2. Increasing the number of neighbors needed to infect shows that the rate of infection decreases and this has been confirmed in both graphs of Figures 3 and 4. Figure 3 shows that there is a sharp increase in infected individuals after day 12. Since on day 12, there is a significant amount of agents infected and then going into days 13 and 14 (weekend) there is a larger number of agents that are now eligible to become infected. Having the low infected neighbor rule (neighbors infected to be 2 or more) creates a very large number of people infected on days 13 and 14. The major difference in rule sets is that in figure 3, the entire population of agents gets infected at some point in time. Whereas at the end of the simulation in figure 4, there will be a portion of the population of agents that have not gone through the infection and recovery state of the SIR model. After looking at the simulation results, the next portion of the study was to look at the access control portion of the system with respect to the users of the system. The results of the simulation were taken for each role that was discussed previously. In Table 3, the correlations between each cue and the roles were calculated in a correlation coefficient table using Minitab. This table will introduce the beginning stages of judgment quality analysis. Each cue has a corresponding correlation coefficient with the judgment (Ys). Cues M and F are the gender cues for all the roles and cues 1-5 are the age group cues (corresponding to the five age groups) for the roles stated previously.
Ye is the actual number of people infected on day 6 and Ys is the judgment made by each role on day 6. Correlations of each cue in comparison to the judgment giving cue (M) the most quality in judgment. Seeing as the correlation between Ys and cue (M) is the highest out of the rest of the cues in terms of access control. Although if we look at the Ye row, it can be observed that cue (3) is the cue with the most correlation towards the actual data. This observation shows that there is a great loss in information when access control is introduced for cue (3) and for the quality of judgment to increase the results on cue (3) needs to be preserved. Also that cue (M) will maintain comparable if not all information results between actual and access control results.

CONCLUSION AND FUTURE WORK
This research study was to evaluate the simulation system of an SIR model and to introduce judgment quality in access controlled results on the specific experiments run on the simulated agent based model. Two rule sets of SIR experiments were run on the agent-based model and the resulting data was analyzed to show that the intuitive hypothesis made matched the results in Figures 3 and 4. Showing that the more neighbors needed to infect any given agent, there would be a lower infection rate giving the recovery rate a competitive edge. Using the first experiment (rule set #1) cues were created in terms of the access control policy defined. Cues were divided in terms of gender and age giving a total of seven cues to evaluate. The correlation coefficient table created illustrated which cues are significant with regards to judgment quality for both the access controlled data and the actual data. Looking at both sides gives a better understanding on how each cues’ importance changes with respect to access control. The results of Table 3 conclude that cue (M) and (3) are responsible for increasing and decreasing judgment quality, respectively. Future experimentation will be to address different types of access control policies and introducing the knowledge equation (G) into the mix.

ACKNOWLEDGMENTS
This work was done when author was working in distributed multimedia system lab at Purdue University. Acknowledgment is due to the director of distributed multimedia systems lab for his guidance and also to the people of lab for sharing data and resources.

Biography
Farrukh Arslan - a research assistant at Purdue University - formulated this paper in black and white when he was working in distributed multimedia systems lab in ece department, Purdue University. He is also an independent researcher and consultant with the focus on IT applications. He is a member of IEEE and ACM.

REFERENCES
[4] Jiasheng Wang, Jianhong Xiong, Kun Yang, Shuangyun Peng, Quanli Xu, “Use of GIS and Agent-Based Modeling to Simulate the Spread of Influenza,” No information of where and when it was published.


Appendix

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>F</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Ys</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.675</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.929</td>
<td>0.851</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.846</td>
<td>0.955</td>
<td>0.928</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.85</td>
<td>0.844</td>
<td>0.829</td>
<td>0.932</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.7</td>
<td>0.149</td>
<td>0.63</td>
<td>0.309</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.631</td>
<td>0.72</td>
<td>0.616</td>
<td>0.743</td>
<td>0.765</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ys</td>
<td>-0.491</td>
<td>0.066</td>
<td>-0.242</td>
<td>-0.195</td>
<td>-0.372</td>
<td>-0.396</td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td>Ys</td>
<td>-0.489</td>
<td>-0.234</td>
<td>-0.33</td>
<td>-0.439</td>
<td>-0.549</td>
<td>-0.081</td>
<td>-0.09</td>
<td>0.867</td>
</tr>
<tr>
<td></td>
<td>1.319892</td>
<td>0.177419</td>
<td>0.650538</td>
<td>0.524194</td>
<td>1.064516</td>
<td>-0.32527</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.89071</td>
<td>0.42623</td>
<td>0.601093</td>
<td>0.799636</td>
<td>0.147541</td>
<td>0.163934</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>