Agent-based modeling of human education data

Yuqing Tang
Department of Computer Science
Graduate Center, City University of New York
365, 5th Avenue
New York, NY 10016, USA
ytang@gc.cuny.edu

Simon Parsons and Elizabeth Sklar
Department of Computer & Information Science
Brooklyn College, City University of New York
2900 Bedford Avenue
Brooklyn, NY 11210, USA
{parsons,sklar}@sci.brooklyn.cuny.edu

ABSTRACT

Agent-based simulation is increasingly used to analyze the performance of complex systems. There are two main ways agent-based models are built — from equation-based models and directly from data. We are building models in both ways, investigating approaches for creating them and for validating them. In this paper we describe results of our work on one specific agent-based model, showing how it can be validated against the equation-based model from which it was derived, and the extent to which it can be used to derive additional results over and above those that the equation-based model can provide.

Categories and Subject Descriptors

I.6.6 [**Simulation and Modeling**]: Model Validation and Analysis

General Terms

Experimentation, Verification

Keywords

Agent-based modeling, simulation, education

1. INTRODUCTION

Agent-based modeling contrasts with traditional approaches to simulation, which typically involve building sets of interrelated differential equations. Such traditional models, commonly called *equation-based models* (EBMs), have been widely applied and generate useful predictions about the behavior of populations. So why use agent-based models? There seem to be four main answers [2]:

- agent-based models are a natural way to describe systems comprised of interacting entities;
- agent-based models are flexible;
- agent-based models capture emergent phenomena; and
- agent-based models provide access to a greater level of useful detail.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.

In particular, modeling interactions between entities can be much easier in agent-based systems than in EBMs, even when one is comfortable with the concepts of partial differential equations, and can reveal details of the system being modeled that are obscured by EBMs.

In this paper, we describe results of our work on one specific agent-based model, showing how it can be validated against the more traditional model from which it was derived, and highlighting the extent to which it can be used to derive additional results over and above those that the traditional model can provide.

2. A MODEL OF HUMAN CAPITAL

The model that we consider in this paper is drawn from [3], an article that derives a linear model from US census data, and analyzes the aggregate behavior of the model. The model is derived in [3] to identify the effect of the tendency for human societies to stratify by level of education — so-called *human capital*. The model from [3] gives the level of human capital $z_{i,t+1}$ of members of the $t+1$th generation of the ith dynasty as being:

$$z_{i,t+1} = k_{t+1} + \alpha \left(\frac{z_{i,t} + z'_{i,t}}{2} \right) + \beta \left(\frac{\sum_{j=1}^{n} z_{j,t}}{n} \right) + \epsilon_{i,t+1} \quad (1)$$

The first term measures the effect on the level of education ("education" and "human capital" are used more or less interchangeably in this model) of the $t+1$th generation of the education of its parents in the t-th generation. The second term does something similar based upon the neighbors of the parents. The third term is constant across dynasties, but may vary in time to capture exogenous trends in education — for example legislation that requires a certain number of years of additional schooling for given generations. The final term captures a specific "shock" to the human capital in a specific generation of a specific dynasty — for example the early death of a parent, requiring the children to curtail their education.

The notion of "dynasty" and "generation" used here require a little explanation. Each generation of the ith dynasty has two children, one male and one female. Each is assumed to then become the spouse of an opposite sex member of another dynasty, forming a family which in turn produces one male and one female child. One family from a given generation of the ith dynasty remains in the ith dynasty, and one becomes part of another dynasty (the dynasty of the corresponding non-ith partner). Thus there is a constant number of members of each generation, and of each dynasty at each generation.

Based on this model, we have developed an agent-based simulation with a fixed number of agents, n in each generation, with $n/2$ dynasties, and 2 children per family. The basic simulation loop, which executes once for each generation, has three steps:

1. Establish level of z based on:
   (a) Parents
   (b) Neighbors of parents
2. Establish factors that influence z for children
   (a) Spouse
   (b) Neighbors
3. Generate children

Step 1 is fixed by (1), and Step 3 is fixed by the requirement to produce one male and one female child in each generation. Clearly the results are going to depend on the way in which Step 2 is implemented, and our model includes a number of variations described in full in [9].

3. EXPERIMENTS

An agent-based version of the model described in the previous section was implemented in REPAST [5], a Java-based Swarm-like [8] tool developed at the University of Chicago for agent-based modeling in social science applications. We handled the geographic aspects by placing agents on an $n \times n$ grid, where at most one dynasty "lives" in a single grid-square. By varying the size of the grid and number of agents we can create environments of differing population density and have modeled communities of up to 10,000 dynasties.

3.1 Verification

Having constructed an agent-based model of human capital from the equation-based model in [3], we first need to "complete the loop" by performing a statistical analysis of the results from the agent-based model, obtained when using the parameter values assumed in the paper, to show that our agent-based model will achieve the same results as the equation-based model we started with. This verification step is needed in order to justify the further experimental results with the model.

The central result of [3], and the only quantitative result that we can use to check the model against, is the prediction that increasing sorting — which the paper takes to mean increasing the correlation between the human capital values of the parent agents of a generation — will only cause an increase in inequality — which the paper takes to mean that the standard deviation of the human capital distribution grows generation by generation — when the value of α is large. [3] demonstrates this by showing the effect of changing correlation from 0.6 to 0.8 for various values of α. Running experiments on a 50×50 grid — which allows us to deal with a population that is considerably larger than the 1500 individuals analyzed in [3] — we find that our model gives good agreement with the predictions made in [3].

For example, we can plot the effect of parental choice of spouse in terms of the percentage change in inequality (as defined in [3]) against α, the parameter that mediates the effect that parents have on their children's human capital. This gives Figure 1.

3.2 Identifying new features

As we discussed above, one of the advantages that agent-based models have over equation-based models is that one can examine the model in greater detail. Whereas equation-based models can only really be studied in terms of broad statistical features — such as the results from [3] examined above — we can probe agent-based models in considerable detail, discovering what happens to

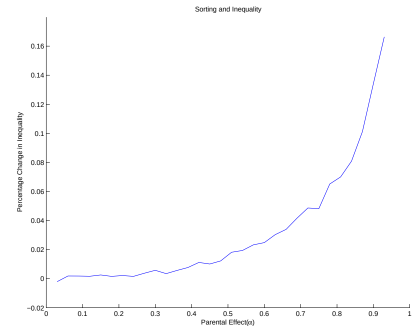

Figure 1: The relationship between the parental effect α and the percentage change in inequality.

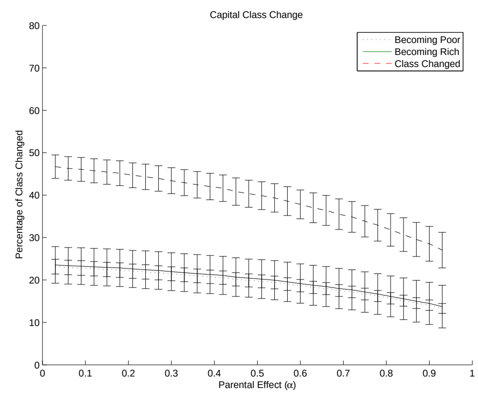

Figure 2: The relationship between the percentage of dynasties that change "class" and α.

individuals as well as to classes of individual. We have carried out such an investigation into the human capital model.

The headline result from [3], replicated by our agent-based model is that *on average* inequality in terms of human capital grows over generations. The widening standard deviation of the human capital distribution suggests that rich dynasties get richer and poor dynasties get poorer. However true this may be at a population level, it is interesting to ask whether it is true for all (or even most) individual dynasties, or whether there is some mobility between dynasties with different levels of human capital.

It turns out that such mobility exists.

We divided our dynasties up into three "classes" — the quotes reminding us that this terminology, while convenient, conflates human capital, basically years of formal schooling, with monetary capital. We call dynasties that fall within one standard deviation above or below the average human capital for the population *middle class*, we call those more than one standard deviation below average *poor*, and those more than one standard deviation above average *rich*. We then examined whether dynasties moved between classes.

The results are given in Figures 2 and 3. These show the way that the number of dynasties that are mobile in this sense changes for two different values of α and β, respectively. When α changes, β is held constant and vice-versa. The graphs plot three values, the total percentage of dynasties that move, the percentage that become richer, and the percentage that become poorer, all data being presented along with standard deviations. The results show that,

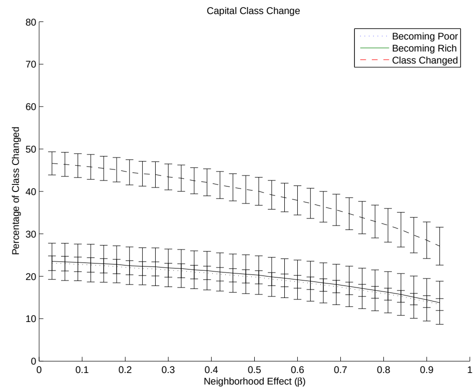

Figure 3: The relationship between the percentage of dynasties that change "class" and β

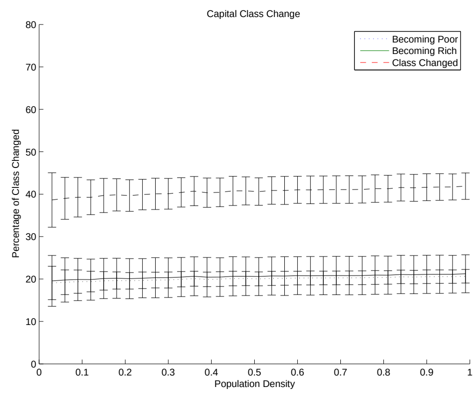

Figure 4: The relationship between the percentage of dynasties that change "class" and population density.

no matter what the value of α and β, there is some mobility (at least 25% of the population, and as much as 45% of the population changes class). Furthermore this change is symmetrical — more or less the same number of dynasties get richer as get poorer.

Note that this effect is separate from the growing inequality — because "middle class" is always defined in terms of the *current* standard deviation, if inequality was the only effect, the percentage of dynasties changing class would be lower than the figure we find. What we see here is the result of mixing, that is, individuals choosing partners or neighbors who are sufficiently far above or below them in human capital terms that their offspring move from one class to another.

We also checked that class mobility was unaffected by other parameters of the model. Since the neighbor effect is based upon a geographic notion of neighborhood, and since neighbors certainly have an effect on class mobility, then one might imagine that changing the density of the population might have some effect on class mobility as well. However, this is not the case. As Figure 4 shows, population density has no systematic effect on class mobility.

For further details of our results, see [9].

4. SUMMARY

This paper set out to show that it was possible to construct an agent-based model from a traditional, equation-based, model, and: (i) verify the agent-based model against the predictions made by the equation-based model; and (ii) use the agent-based model to identify new predictions that cannot be obtained directly from the equation-based model. Both these aims have been achieved.

This work fits into a wider effort that is attempting to model aspects of the education system [7], with the overall aim of being able to identify, and thus establish the impact of, changes in education policy (rather as [1] does for the case of rent control). As described in [7], we have developed a number of models, including a model of interactions in classrooms [6] — which allows us, for example, to model the effects of different pedagogical techniques to overcome absenteeism — and a model of school districts — which, for example, enables us to study the effect of policies like "No child left behind".

Our aim is to tie these models together, and, more ambitiously, to tie them into a comprehensive simulation of the way that education fits into the economy as a whole. One way to think of this is through the model provided in [4]. This describes individuals as making a choice about how much education to receive. This education, combined with their inherent aptitude, equips each agent with a certain level of productivity which, after its education is complete, allows it to pay off its education and purchase units of consumption. Thus education becomes a decision which both enables consumption and constrains it (because education has to be paid for). The kinds of model described in [7] tie in with this approach — the classroom model will affect the way that ability relates to productivity for example — and the human capital model described here provides a means of relating the education decisions made by one generation to those made by the next, giving us a means of projecting the effects of our models over generations.

5. ACKNOWLEDGMENTS

This work was partially supported by funding from NSF #REC-02-19347.

6. REFERENCES

[1] R. N. Bernard. Using adaptive agent-based simulation models to assist planners in policy development: The case of rent control. Working Paper 99-07-052, Sante Fe Institute, 1999.

[2] E. Bonabeau. Agent-based modelling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Science*, 99(3):7280–7287, May 2002.

[3] M. Kremer. How much does sorting increase inequality? *The Quarterly Journal of Economics*, 112(1):115–139, Feb. 1997.

[4] J. Laitner. Earnings within educational groups and overall productivity growth. *The Journal of Political Economy*, 108(4):807–832, August 2000.

[5] `http://repast.sourceforge.net`.

[6] E. Sklar and M. Davies. Multiagent Simulation of Learning Environments. In *Proceedings of the Fourth International Conference on Autonomous Agents and MultiAgent Systems (AAMAS-2005)*, 2005.

[7] E. Sklar, M. Davies, and M. S. T. Co. SimEd: Simulating Education as a MultiAgent System. In *Proceedings of the Third International Conference on Autonomous Agents and MultiAgent Systems (AAMAS-2004)*, pages 998–1005, 2004.

[8] `http://www.swarm.org/`.

[9] Y. Tang, S. Parsons, and E. Sklar. Agent-based modelling of human education data. Technical report, Department of Computer and Information Science, Brooklyn College of the City University of New York, 2005. (the full version of this paper).