

Social Simulation: The Need of Data-Driven Agent-Based Modelling Approach

Mazhar Sajjad^{*, **}, Karandeep Singh^{*, **}, Euihyun Paik^{**}, Chang-Won Ahn^{*, **}

^{*}Department of Computer Software^{1,2}

Korea University of Science and Technology (UST), Korea

^{**}Human Computing Research Lab

Electronics and Telecommunication Research Institute (ETRI), Korea

Daejeon, Korea

{sajjad,karandeep.singh,eupaik,ahn}@etri.re.kr

Abstract—Agent-based modeling and simulation (ABMS) has attracted social scientists and demographers in the field of social simulation. Due to large number of computer simulation technologies, ABMS approaches have been proposed with majority applications. ABMS composed of heterogeneous interacting agents, with several features which turn them into a significantly attractive modeling approach to simulate complex social systems. In this paper, first we explore the underlying social theories for ABMS, its simulation and modeling techniques, and computational frameworks. Second, our paper concentrate on the potential need of ABM techniques in the context of social simulation. An alternative ABM approach that is getting popularity is to inject data into agent-based simulation. To validate our model, we compare our results with actual-data. Our results closely matched with actual-data results in the case of mean-age at the demographic transition of first childbirth. Our work encourage ABM modelers to promote this trend while designing their models. Further, our paper is an attempt to merge the microsimulation approach into the agent-based simulation through injecting data into ABM approach

Keywords— Agent-based modelling, social simulation, individual-based modelling

I. INTRODUCTION

To explore social systems, people always remain at the centre of study. Various behaviours of people during their life stages will inevitably have an important impact on the society and environment that they are living in. At the same time, such individual changes at micro-level will also lead to an evolution of the whole population over time. It is such complexity within our social system that keeps attracting research interest. Indeed, people and societies, along with their past, present, and future magnitudes and compositions, have never failed to capture the fascination of researchers and modellers [1].

New innovations in technologies and new tools often bring new visions to our research and development. Computational models have now been extensively used to model complex social systems, not only because they can provide valuable background to experiment in reality, but also new research methodologies of modern computers can wholly transform human ability to complex social systems. In today world, this has become increasingly important, because we confronts rapid

and potentially profound transitions driven by social, economic, environmental and technological changes [1]. Traditionally the macroscopic approach has been adopted to model populations. Macroscopic models such as the cohort-component model have been widely applied in demographic studies. However, macroscopic models have limits in representing microscopic occurrences, discontinuity and heterogeneity within a system [2].

Modern studies of social Systems now often require detailed information about the behaviours and characteristics at the level of individual decision making units. The macroscopic approach struggles in such studies. On the other hand, the microscopic approach can provide valuable insight into the behaviours of a system under a range of conditions through the provision of rich details about individuals. As an Individual Based Model (IBM), a MSM can provide valuable information over a wide range of circumstances. Therefore it has become not only an excellent knowledge discovery tool, but also provided indispensable assistance to strategic decision making [3], [4], [5]. Now, advances in computing processing and techniques of analysis brought sophistication in many applications areas that MSM may address. Using agent-based simulation, BDI architecture is another alternative approach that explicitly take the agent's heterogeneity and explores how changes at micro-level transmitted into people perception at global level [13]. This paper proposes a model that feed empirical data into simulations, enables us to test demographic dynamics at the transition of childbirth. Our proposed ABM method will have more potential to analyse closely the changing patterns of childbirth process. The study will give a meaningful insights by combining both microsimulation and agent-based methods. Once the population has been created in this way, the model is used for some purpose such as policy analysis [17].

II. AGENT-BASED SOCIAL SIMULATION

Agent-based social simulation (ABSS) is an alternative form of computer modelling of complex systems, in which the agents within such a system are represented explicitly and individually within the model. The agents in the model represent human individuals, non-human animals, or human collectivises such as firms or states depends upon the application. They interact with

each other, and with environment, take place according to rules which may adaptive between agents and environment, and change as agents learning continuously. ABSS models are generally created using an object-oriented programming language like Java, Python etc., allowing each agent to be implemented as an object [6].

The ABSS is the research area that that formed by the intersection of three scientific fields, namely, agent-based computing, the social sciences, and computer simulation as shown in figure 1. Agent-based computing is a part of computer science that includes, agent-based modelling (ABM), design, and programming. The social sciences are a combination of many sciences that study the interaction among social sciences for example social psychology, management, policy, and some areas of biology. Finally, computer simulation study various techniques for simulating the events on the computer. These simulated events behave as an artificial system, which could be either existing or non-existing at the time of the simulation [6].

The reason behind doing these computer simulations is usually to achieve an in-depth understanding of the phenomenon, e.g., predicting future behaviour, and performing experiments that cannot be carried out in reality for some reason or another. Although computer simulation is considered a computer science sub-area, but a lot of development has been done in application areas, such as physics, mechanical engineering, biology, and to some extent even within the social sciences [1], [6], [11]. The main role of ABSS is to provide models and tools for agent-based simulation of social phenomena, and to apply these in different areas. However, through its inter-disciplinary flavour, ABSS has a unique potential for providing cross-fertilisation between the participating fields of research. (See Figure 3.) In this section we will study these possibilities.

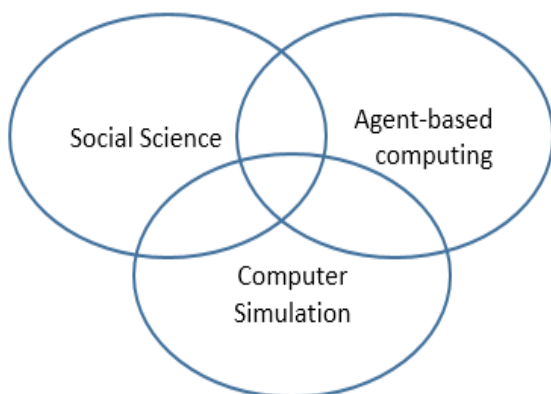


Figure 1. The three research areas that formed ABSS and their interrelationship [6]

III. DATA-DRIVEN AGENT-BASED MODELLING APPROACH

Our aim here is to propose an alternative design approach adopted in agent-based model. The method could be especially

useful in situations where we have some sources of empirical data or can collect data like in the form surveys etc. Micro-simulation [7] has traditionally been used in areas where it is easy to obtain empirical data, in the form of surveys and censuses and equations or rules to define the agent's behaviours. Micro-simulation has been proven to be a successful approach in some applications such as traffic modelling and financial modelling etc., but it has been difficult to apply in social modelling that are not so well structured or where there are important dependencies between agents. Micro-simulation requires high quality and very specific details about data, difficult calibration even if detailed demographic data are available. Second, micro-simulation is computing intensive, can take long time and considerable efforts to accomplish the task, Micro-simulation is unable to model interactions between agents, there is one-direction interaction only i.e. the impact of the policy on the individuals, but not the impact of individuals on the policy and difficult to validate these models.

Agent based modelling is an alternative approach that models social life as interactions among intelligent agents and is related to traditional sociological systems that model social processes as interactions among variables. ABMs originate from a sub-field of Artificial Intelligence and it is concerned with the behaviour of multiple agents working together to solve a given problem that is beyond the capability or knowledge of individuals [8]. Agent-based models usually follow an event-based rules approach rather than using transition probabilities. However, the limitations of modelling or the lack of sufficient data frequently make it difficult to implement explicit rules and therefore they have to turn to other solutions, one of which is to use transition probabilities, which represent implicit rules. Qualitative information, although rarely used in ABM, can also be introduced [9], [12].

Adopting this approach, it is possible to reformulate the classical stages of the logic of agent-based simulation, which uses empirical data only for validation [10]. In order to guide the model with data and initialisation, the modellers needs to collect data from people social behaviours, second design the model that would be guided by empirical data, equations or any qualitative information, third initialize the model by some census or survey data and fourth validate the simulation result by comparing it with collected data. During validation, the data used for simulation of the model should be different to guarantee validation test from the design of the model. The data-driven ABM model gives a new angle to the classical approach of ABM where the researcher needs to achieve consistency with the world, to represent processes on various spatial and temporal scale at one time, to enable agents to follow internal and external rules at the same time and to integrate observable and postulate behaviour while preserving achievability of endogenous emergence [1]

IV. FERTILITY MODEL OF DEMOGRAPHIC DYNAMICS: A CASE STUDY

The fertility model implement the transition to motherhood module of demographic transitions. The factors that may affect the transition would be age, marital status, number of children, education and employment status. All agents are heterogeneous by age, income level and education. In order to reflect the behaviour of agents in Korean context, we drive the model with Korean 1990 census data. The census data gives individual parameters such as age, education, gender, education levels and age at first marriage, number of children and the education level of each agent. The fertility model can be measured in a variety of ways, such as the crude fertility rate, age-specific fertility rate, total fertility rate, gross reproduction rate, and net reproduction rate.

Fertility rates are influenced by a number of factors, including intentional measures such as contraception and major social events. Possible demographic components and its relevant variables that affecting the fertility are: mortality, work status (full-time, part-time, un-employment,), daily wages, marital status, spouse alive, divorced, widow, education, elementary school, middle school, high school, college, university, age (16 to 45), gender, living place. Many factors effecting the fertility transition based on the context of the country, because the context of the society is mainly matters in demographic transitions. In Korea the women labour force participation and education are considered the key factors during fertility transition [6]. Figure 2 shows the various fertility transition states related to different demographic components.

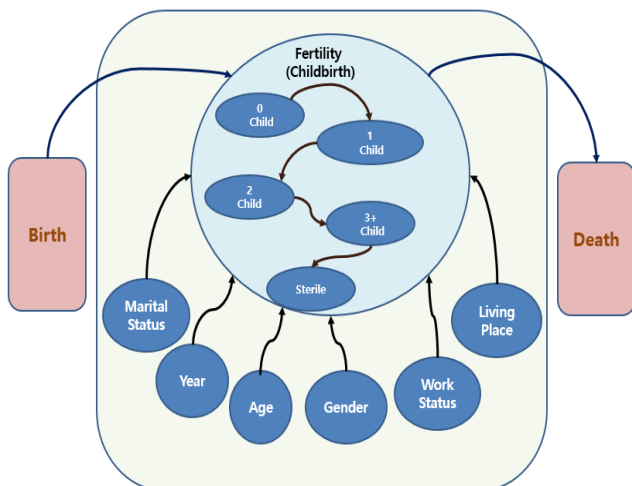


Figure 2. Fertility transition states with demographic components

We test our model to find the mean age at first childbirth. We are using family economic stability condition to take decision about first childbirth. The detail computational analysis of the model is not in the scope of this paper, because here our main emphasis is to validate our simulation result with actual-data result.

A. Model Initialization with Data

The model is injected by 100,000 number of agents taken from the census data of 1990. The actual size of 1990 is approximately 43 million, so this means that during simulation each agent represents the behaviour of 430 agents of the actual world. Further, we compare our simulation results with data-based fertility probability of 1990, 2000 and 2010 with our simulated probabilities of 10 and 20 years respectively. Table 2 shows the detail description of initial data taken from 1990 Korean census [14]. Out of total 100,000 agents, we have 53128 married agents between the ages of 16 and 45. In order to see the age-specific marriage patterns over time, we divide the number of acceptable agents among various groups. We can see that the total number of men's agents is greater than the number of women between the ages of 16 and 45. Once model is initialized with 1990 census data then we run our simulation for 30 years to check the mean age of women at first childbirth.

Table I. Initial data description taken from 1990 census

Age groups	Frequency of Men	Frequency of Women	Total Number of men and women
16-20	5000	4934	9934
21-25	5200	3972	9172
26-30	5670	5346	11016
31-35	4490	4946	9436
36-40	3430	3653	7083
41-45	2510	2954	5464
16-45	26300	20859	47159

B. Results

In this section, we implanted our designed model in Anylogic professionalⁱ 7.0. We simulate the starting population for 30 years and tested the mean-age at first childbirth. To verify our model, the results obtained from our model compare with actual data-based results. After initializing the model with 1990 census data, we run our model for next 20 years i.e. till 2010. We check the mean age at five years interval during simulation. Table-2 shows the validation comparison results of our simulation with actual data-based result. We can see that our results is closely matched with actual data. The percentage difference is quite low at all intervals. It shows that model initialize with data closely replicate the actual results. These results validate the authentication of our proposed data-driven ABM approach.

Table II. Results validation by comparing with actual-data

Analysis Factor	1990	1995	2000	2005	2010
Data-based Childbirth Age	25.90	26.55	27.23	28.10	29.12
Simulated Childbirth Age	25.20	26.05	26.70	28.50	29.50
%age Difference	2.7%	1.88%	1.72%	1.42%	1.30%

V. CONCLUSIONS

In this study, we aim to combine the strength of both MSM and ABM to provide the capacity to model heterogeneous movements, interactions and behaviours of a large number of individuals within a complex social system. Dynamic changes are captured within the model by simulating the discrete steps when individuals going through fertility process of demographic transitions. Our case study shows that injecting a model with empirical data can improve the replication between the ABM model and the observed social data in the form of census or survey. Our model suggests that it is worthy to compare different collections of data and conclusions from heterogeneous sources to improve authenticity of the model results. The ABM needs to be designed in such a way so that its output results can be compared directly with empirical data. Our paper motivate researchers to merge the microsimulation approach into the agent-based simulation by injecting data into ABM approach.

ACKNOWLEDGEMENT

The work is supported by the Korea ICT R&D program of MSIP/IITP (10047117, Development of Distributed/Parallel Multi-Dimensional Demographic Micro Simulation Technologies for Population Dynamics and Socio-Economic Experimentation). The authors also extend their thanks to the whole social simulation team for their constant advice during the development of the work.

REFERENCES

- [1] Wu, Mingqing Belinda. *A hybrid microsimulation model for a UK city population with dynamic, spatial and agent based features*. University of Leeds, 2012.
- [2] Billari, Francesco C., and Alexia Prskawetz, eds. *Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour*. Springer Science & Business Media, 2012.
- [3] Orcutt, Guy H. "A new type of socio-economic system." *The review of economics and statistics*, 1957, p. 116-123.
- [4] Van Imhoff, Evert, and Wendy Post. "Microsimulation methods for population projection." *Population* 10, no. 1 (1998), p. 97-136.
- [5] Harding, Ann. "Challenges and opportunities of dynamic microsimulation modelling." In *Plenary paper presented to the 1st General Conference of the International Microsimulation Association, Vienna*, vol. 21. 2007.
- [6] Davidsson, Paul. "Agent based social simulation: A computer science view." *Journal of artificial societies and social simulation*, 2002.
- [7] Gupta, Anil, and Vishnu Kapur. *Microsimulation in government policy and forecasting*. Elsevier Science Inc., 2000.
- [8] Jennings, Nicholas R. "On agent-based software engineering." *Artificial intelligence*, 2000, p. 277-296.
- [9] Yang, Lu, and Nigel Gilbert. "Getting away from numbers: Using qualitative observation for agent-based modeling." *Advances in complex systems* 11, no. 02, 2008, p.175-185.
- [10] Gilbert, G. Nigel. *Agent-based models*. No. 153. Sage, 2008.
- [11] Macal, Charles M., and Michael J. North. "Agent-based modeling and simulation." In *Winter simulation conference*, pp. 86-98. Winter Simulation Conference, 2009.
- [12] Billari, Francesco C., and Alexia Prskawetz, eds. *Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour*. Springer Science & Business Media, 2012.
- [13] M.Sajjad, K.Singh and C.Won Ahn. "Multi-Agent Modeling for Match-Making using BDI Architecture." The 2015 International Conference on High Performance Computing & Simulation, (HPCS), 2015.
- [14] Korean statistics, KOSTAT, www.kostat.kr
- [15] Diaz, Belinda Aparicio, and Thomas Fent. "An agent-based simulation model of age-at-marriage norms." In *Agent-Based Computational Modelling*, pp. 85-116. Physica-Verlag HD, 2006.
- [16] Mazhar Sajjad, and Chang Won Ahn. "Agent-based modelling to analyze the impact of work status on childbirth: a case study of Korea", the 11th international multidisciplinary modeling and simulation (MAS), conference, Bordeaux, France, 2014.
- [17] Karandeep Singh, Mazhar Sajjad, and Chang Won Ahn. "Towards full scale population dynamics modelling with an agent based and micro-simulation based framework", the 17th International Conference on Advanced Communication Technology (ICACT), Phoenix park, Korea, 2015.



Sajjad Mazhar received his Bachelor from university of Peshawar, Pakistan and his Master degree from Dongguk University, Seoul, South Korea. He is currently a PhD student of University of Science and Technology (UST) at Electronics and Telecommunication Research Institute (ETRI), South Korea. His research interest include Distributed artificial intelligence, Social Computing, Complex System, Big Data Analysis and Processing.



Brar, Karandeep Singh received his Bachelor degree from Punjabi University, Punjab, India and his Master degree from PEC university of Technology, Chandigarh, India. He is currently a PhD student of University of Science and Technology (UST) at Electronics and Telecommunication Research Institute (ETRI), South Korea. His research interest include Social Computing, Big Data Analysis and Cognitive Machine Learning.



Dr. Paik Eui-Hyun is principal researcher at human computing research lab in Electronics and Telecommunication Research Institute (ETRI). He received his Ph.D in School of Computer Science at Soongsil University in 1997; and his B.E. from Computer Science at same University in 1984. His theoretic research focuses on the overlapping area of computer science, sociology and operations research. His practical research includes population and social security simulation, intelligence analysis, disaster management.



Dr. Ahn Chang-Won is principal researcher at human computing research lab in Electronics and Telecommunication Research Institute (ETRI) and Professor in Department of Computer Software at University of Science and Technology (UST). He completed his PhD in Industrial Engineering with specialty in stochastic processes and queueing theory from KAIST (Korea Advanced Institute of Science and Technology) in 1998. His research interest includes

ⁱ <http://www.anylogic.com/>