Teaching Agent-Based Modelling and Machine Learning in an integrated way

Ellen-Wien Augustijn *[1], Rania Kounadi[1], Tatjana Kuznecova[1], Raul Zurita-Milla[1]

[1] Department of Geo-Information Process (GIP), Faculty of Geo-Information Science and Earth Observation (ITC),
University of Twente, The Netherlands
*Email: p.w.m.augustijn@utwente.nl

Abstract

The integration of Agent-Based Modelling (ABM) and Machine Learning (ML) provides many promising opportunities, yet this research field is underdeveloped. Different reasons are given for this lack of integration, including a shortage of behavioural data and technical implementation difficulties. However, we think that one crucial problem is being overlooked. In our educational system, we teach topics one by one and do not explicitly focus on the integration of various modelling paradigms. This is a missed opportunity that should be addressed, to prepare our students for a world where models are increasingly complex and where data and model integration becomes inevitable. In this paper, we share our experiences in a course in Geoinformatics, where integrated ABM and ML modelling is central. In our class, we use the Living Textbook to work on interlinked concept maps, and we have an overarching case study assignment. Preliminary outcomes show that students' learning and project work could benefit from simplifying the case study assignment and introducing the parallel teaching of ABM and ML. In general, different teaching methods and setups still need to be explored, to ensure that our future model designers are well equipped for their task.

Keywords: Integrated Modelling, Agent-Based model, Machine Learning, Education

1. Introduction

Although agent learning has always been regarded as one of the main motivations for the implementation of Agent-Based Models (ABMs), the number of ABMs that contain learning based on Machine Learning (ML) algorithms is still small. Most ABMs are process-driven rather than data-driven. Although empirically grounded models are gaining popularity, there are still many issues limiting the integration of ABM and ML. A better understanding of these limitations can help to push this research area forward. It is equally important to make sure that the new generations of scientists and modellers are well prepared for the challenges of the increasing complexity of modelling endeavours. In this work, we explore how changes in education can help to promote integrated modelling approaches.

1.1. Use of ML in ABMs

ABMs use empirical data in many different ways. Although the behaviour of the agents may be rule-driven, empirical data can be used to structure the environments that agents move in, or to define

global environmental variables like climate, or for calibration and validation of the model. The use of data-driven machine learning techniques in ABMs seems to be triggered by two different problem areas: the wish to create smarter agents/models, and the problem of calibrating/validating complex models due to a vast variable space.

The term agent-mining was introduced by Cao (2009) to indicate the merge between the two scientific domains of ABM and ML. Cao identifies challenges in both fields that can be resolved by their integration. For ABM, these challenges are in the line of steering of agent behaviour: agent-awareness (observation of changes in environment states) and agent-learning. For ML, they include: involving domain and human intelligence, data fusion and preparation, and adaptive learning. Baqueiro et al. (2009) studied the application of ML techniques in ABMs and how to utilize ABMs in ML research. They concluded that from an ABM perspective, the use of ML for validation is most promising and that for ML, ABMs are a way of creating datasets that otherwise would not be available.

Generally speaking, ML techniques can be applied in three different stages in the modelling process: pre-processing, agent behaviour and decision making, and post-processing of simulation output.

> **Pre-processing:** based on human behavioural datasets, an ML algorithm is trained to model which behaviour agents should display under a range of circumstances. The trained model is used as input to the ABM to replace rule-based modelling.

> **Agent behaviour:** Every time an agent has to make a decision, the ML algorithm is consulted to predict the agent behaviour. This can be done using a pre-trained ML algorithm, or include training on the fly, e.g. for reinforcement learning. Agents can learn from their own experiences or the experience of others.

> **Post-processing:** After running the ABM, the data can be mapped back to a trained ML algorithm to calibrate or validate the model.

Several examples exist of linked or integrated ABM-ML models. Buchmanns et al. (2016) used Random Forest during the parameterization of their model to account for the variety of relationships between household properties and relocation aspects in their residential mobility model. Zhang et al. (2015) applied a step-wise linear regression to forecast individual and aggregate residential rooftop solar adoption in San Diego County. The use of ML learning to steer agent behaviour was studied by Abdulkareem et al. (2018), who used Bayesian Networks to steer agent behaviour in a cholera ABM. The use of ML to calibrate ABMs was studied by Lamperti et al. (2017), and Zhang et al. (2018) used agent-based modelling to create a dataset to train an ML model to enhance performance time for an interactive urban decision-making model. Finally, the use of ML to validate ABMs has been investigated by Baqueiro et al. (2009).

1.2. Factors hindering the integration

In literature, a range of factors are mentioned that limit the use of ML algorithms in ABMs. According to van der Ploeg et al. (2014), one of these limitations is that the implementation of intelligence often requires extensive training data. They tested a range of ML algorithms and found that they needed ten times as much data compared to traditional methods. Kennedy (2012) also mentions that the main

factor that hinders the implementation of ML in ABM is the lack of behavioural data, and missing quality data at suitable abstraction levels is mentioned as a limiting factor by Bruch and Atwell (2015).

Another potential factor hindering the integration, as mentioned by Rand (2006), is that the modeller needs to have expertise in both scientific fields. Besides building the ABM, the modeller needs to select a proper learning algorithm and train it.

We think that there is another factor, which is often overlooked. This is the fact that ABMs and ML are often taught in separate courses. This does not motivate students to look for an integrated application of these two modelling paradigms. We try to remediate this problem by redesigning an existing course to focus on teaching ABM and ML in an integrated way. The setup of this experiment is explained in section two.

2. Teaching integrated modelling

We redesigned an elective course of our Geo-Informatics programme to teach ABM and ML in an integrated fashion. Our objective was to teach our students to make transparent choices in selecting a modelling approach and to show them how model integration can be realized. Two teaching methods were selected to meet this objective: a joint Living Textbook implementation (2.2) and an assignment based on model integration (2.3).

2.1. Course structure

The course is an elective in the last semester of the first year of the Geo-Informatics two-year MSc programme. It is open to all ITC students, including students from other specializations or programmes. The course is taught in parallel to the so-called "track courses", with a study weight of 2 days (16 hours) a week over 12 weeks.
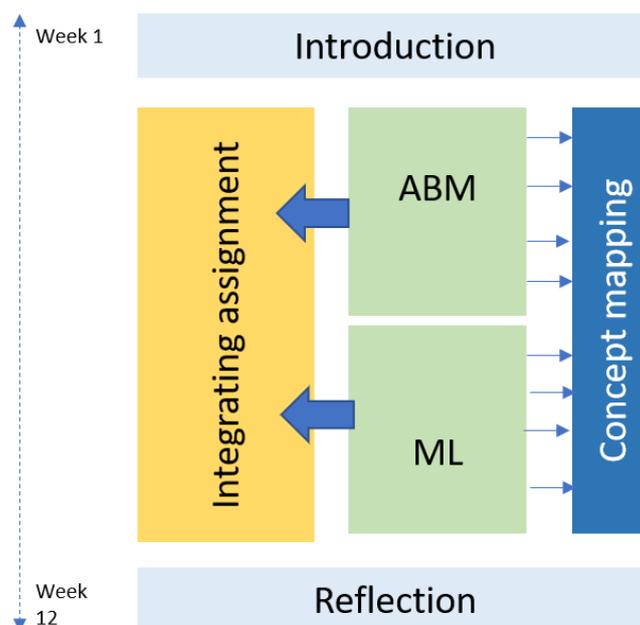
Figure 1: Course structure

After a short introduction, we sequentially teach ABMs, followed by ML (Figure 1). Main learning activities include lectures, practical exercises and group assignment work. Eighteen students participated in the first edition of this course. All are international students.

2.2. Living Textbook assignment

The Living Textbook is a digital ontology-based textbook developed by the Faculty ITC of the University of Twente (Augustijn et al., 2018). It consists of a concept map, showing all concepts and relationships, and of a wiki that contains the descriptions of these concepts. The tool provides functionality as a digital textbook but also as an environment where students can create their own concept maps.

As an ungraded formative assignment, students were asked to add the concepts taught during the course to a common concept map. This concept map would gradually grow to include all concepts, and relationships between concepts, that were discussed during the course. Although the assignment was ungraded, students can use the concept map and the corresponding descriptions during their open book exam at the end of the course.

After every lecture, new concepts and relationships were identified during a plenary discussion. Two students were assigned to add these concepts and relationships to the concept map. The idea behind this assignment is that students explore the relationships between the different concepts. This leads to a higher level of learning, according to Bloom's taxonomy (Bloom, 1956) – analyse level (relate and compare). During the modelling assignment, this should make students more aware of possible integration options.

2.2. Set-up of the case study assignment

During the graded assignment, students work in groups of three. Initially, students were provided with a simplistic baseline model that they could implement following step-by-step practical exercises. The exercises were complemented with additional questions to facilitate and guide students' creativity and promote independent thinking. Each group was asked to extend the ABM on Tick bites risk that they developed during the exercises and to enhance this model using one of the techniques offered during the ML element of the course. The implementation of the ML part could be either as a pre-processing step or as a post-processing step. Students can choose different ML algorithms, yet the chosen algorithm should be meaningful and applied correctly. During the assignment, regular feedback sessions were organised in which groups could discuss their ideas with the teaching team.

3. Preliminary results and further outlook

At the time of the submission of this paper, the course was drawing at its close but it was not finished yet. Therefore, all results are preliminary.

For the Living Textbook assignment, the students developed the concept map shown in Figure 2.
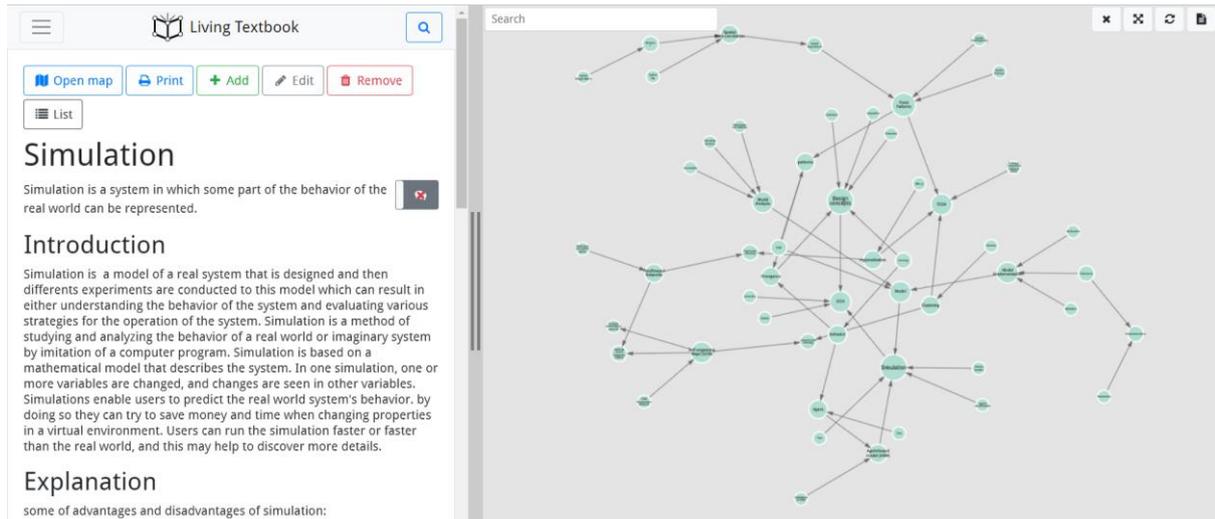
Figure 2: The concept description (left) and the student concept map (right)

This concept map contains 53 concepts, 56 relationships and 19 external resources. Although the map seems to have a reasonable degree of interconnectedness, the level of integration of the topics is limited. We classified concepts as belonging to either ABM (gold), ML (red) or other (grey) and re-projected the concept map (Figure 3). From this new visualisation, it is immediately clear that ML and ABM are only connected via one shared link.
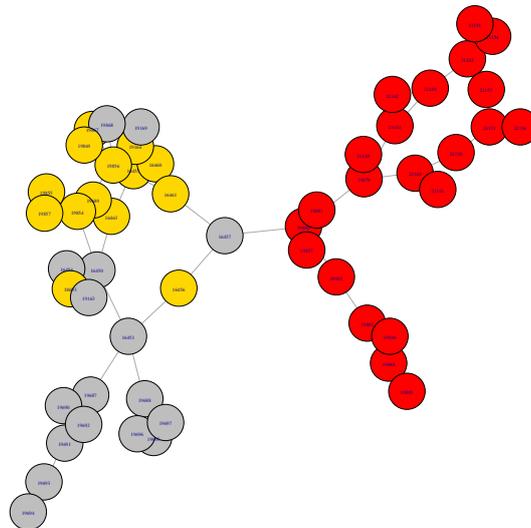
Figure 3: Reprojection of the concept map with ABM concepts as gold, Machine Learning concepts as red and other concepts in grey colour.

The final exam will be used to test whether the construction of the concept map and the integrated project enhanced the understanding of students concerning integrated modelling. For this, we will use a set of questions to check if the student can, for a given problem:

- Identify if integration is required and if this should be done in the pre-processing, agent-behaviour or post-processing phase (selection of the correct implementation strategy)
- Select an appropriate ML Learning algorithm and briefly explain its main characteristics.

- Identify possible limiting factors for the implementation of the selected ML algorithm (data size or availability, technical complexity etc.)

During the meetings with the groups, we already noticed that integrated modelling is regarded as difficult. Some students trust a lot in one of the two modelling strategies and are drawn away from a truly integrated approach. By this moment, discussions with the assignment groups show that students are not yet prepared to see how the two modelling approaches can complement each other. Our task, as teachers, may be to devote more time and effort into demonstrating a wider variety of use cases and scenarios where integrated modelling is highly advantageous, while also educating our students on possible barriers and limitations of such approaches.

The case study is also regarded as being difficult. Students do not feel confident enough to deviate much from the baseline model ideas provided them in the exercises by the teachers. The fact that it was not possible to teach ABM and ML in parallel meant that the integration of ML could only be applied at the end of the case study. Also, students did not have much time to accomplish this part of the task. Possibly, parallel teaching should be considered for future editions of this course.

We believe that as teachers, we have an obligation to try to show students that it may be worthwhile to look over the boundaries of a single modelling paradigm. Although teaching builds up from simple to more complex courses, it is often not possible to schedule three consecutive courses (ABM, ML and integrated ABM/ML). We, therefore, need to develop teaching methods that reach higher-level learning objectives faster. In any case, it is clear that a world where the amount, role and complexity of models are ever-increasing, requires geoinformation professionals that can bridge the (ABM-ML) model integration gap.

4. Acknowledgements

5. References

Abdulkareem, S.A., Augustijn, E.-W., Mustafa, Y.T. and Filatova, T. 2018. Intelligent judgements over health risks in a spatial agent-based model. International Journal of Health Geographics. 17(1), p8.

Augustijn, E.-W., Lemmens, R., Verkroost, M.-J., Ronzhin, S. and Walsh, N. 2018. The Living Textbook: Towards a new way of Teaching Geo-Science. In: Agile, Lund, Sweden.

Baqueiro, O., Wang, Y.J., McBurney, P. and Coenen, F. 2009. Integrating Data Mining and Agent Based Modeling and Simulation. In: Berlin, Heidelberg. Springer Berlin Heidelberg, pp.220-231.

Bloom, B.S. 1956. Taxonomy of Educational Objectives, Handbook: The Cognitive Domain. New York: David McKay

Bruch, E. and Atwell, J. 2015. Agent-Based Models in Empirical Social Research. 44(2), pp.186-221.

Buchmann, C.M., Grossmann, K. and Schwarz, N. 2016. How agent heterogeneity, model structure and input data determine the performance of an empirical ABM – A real-world case study on residential mobility. Environmental Modelling & Software. 75, pp.77-93.

Cao, L. 2009. Introduction to Agent Mining Interaction and Integration. In: Cao, L. ed. Data Mining and Multi-agent Integration. Boston, MA: Springer US, pp.3-36.

Kennedy, W.G. 2012. Modelling Human Behaviour in Agent-Based Models. In: Heppenstall, A.J., et al. eds. Agent-Based Models of Geographical Systems. Dordrecht: Springer Netherlands, pp.167-179.

Lamperti, F., Roventini, A. and Sani, A. 2017. Agent-Based Model Calibration using Machine Learning Surrogates. In.

Rand, W. 2006. Machine Learning meets agent-based modeling: when not to go to a bar Unpublished Paper.

van der Ploeg, T., Austin, P.C. and Steyerberg, E.W.J.B.M.R.M. 2014. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. 14(1), p137.

Zhang, H., Vorobeychik, Y., Letchford, J. and Lakkaraju, K. 2015. Data-Driven Agent-Based Modeling, with Application to Rooftop Solar Adoption. In: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4-8, 2015, Istanbul, Turkey.

Zhang, Y., Grignard, A., Lyons, K., Aubuchon, A. and Larson, K. 2018. Real-time Machine Learning Prediction of an Agent-Based Model for Urban Decision-making (Extended Abstract). In: AAMAS 2018, July 10-15, 2018,, Stockholm, Sweden.