Warszawa, 27.10.2023

dr hab. Mariusz Białecki, prof. IGF PAN
Department of Geophysical Imaging
Institute of Geophysics, Polish Academy of Sciences

Review of the PhD Thesis

"Identification of partially observed deterministic and stochastic cellular automata"

by Witold Bołt

The doctoral dissertation in the form of a thematically consistent series of publications was prepared in System Research Institute, Polish Academy of Sciences, under supervision prof. dr. Bernard De Baets from Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University and auxiliary supervision of dr. hab. inż. Jan W. Owsiński from System Research Institute, Polish Academy of Sciences. The description of the thesis is written in English on 37 pages. It consists of abstract, list of 5 presented articles with indication of co-authors involvement, short description of other scientific activity (including other 7 publications and list of 10 international conferences and seminars), detailed, precise and very clear presentation of main results on 17 pages, notes on future work on 3 pages, and bibliography (35 items).

The dissertation comprises of the following papers:

1. W. Bołt, J. M. Baetens, and B. De Baets, "An evolutionary approach to the identification of cellular automata based on partial observations," in 2015 IEEE Congress on Evolutionary Computation (CEC), 2015, pp. 2966–2972. doi: 10.1109/CEC.2015.7257258.
2. W. Bołt, B. Wolnik, J. M. Baetens, and B. De Baets, "On the identification of α–asynchronous cellular automata in the case of partial observations with spatially separated gaps," in Challenging Problems and Solutions in Intelligent Systems, G. d. Trė, P. Grzegorzewski, J. Kacprzyk, J. W. Owsiński, W. Penczek, and S. Zadrożny, Eds.

Springer International Publishing, 2016, pp. 23–36, isbn: 978-3-319-30165-5. doi: 10.1007/978-3-319-30165-5_2.

3. W. Bołt, A. Bołt, B. Wolnik, J. M. Baetens, and B. De Baets, “A statistical approach to the identification of diploid cellular automata,” in Theory and Practice of Natural Computing, C. Martín-Vide, R. Neruda, and M. A. Vega-Rodríguez, Eds., Springer International Publishing, 2017, pp. 37–48, isbn: 978-3-319-71069-3. doi: 10.1007/978-3-319-71069-3_3.
4. W. Bołt, A. Bołt, B. Wolnik, J. M. Baetens, and B. De Baets, “A statistical approach to the identification of diploid cellular automata based on incomplete observations,” Biosystems, vol. 186, p. 103 976, 2019. doi: 10.1016/j.biosystems.2019.103976.
5. W. Bołt, J. M. Baetens, and B. De Baets, “Identification of cellular automata based on incomplete observations with bounded time gaps,” IEEE Transactions on Cybernetics, vol. 50, no. 3, pp. 971–984, 2020. doi: 10.1109/TCYB.2018.2875266.

In all works, the first author is Witold Bołt. All the algorithms and methods were created by Witold Bołt in papers 1 and 5, and together (with equal contribution) with Barbara Wolnik in papers 2 , 3 and 4. Extensive computational experiments were caried out by Witold Bołt, and with support of Aleksander Bołt in papers 3 and 4. Jan M. Baetens and Bernard De Baets contributed to the research through in-depth discussions. All texts were written by Witold Bołt, the others contributed as correctors. In conclusion, the contribution of the PhD candidate is substantial, crucial and well distinguished.

The topic of this doctoral dissertation is the problem of automatic (algorithmic) identification of cellular automata based on incomplete observations. The scientific problem of research is related to important practical aspects of modeling natural phenomena, namely dealing with the incompleteness of real data. The methods for dealing with missing information depend fundamentally on the type of cellular automata under consideration, so the problem is solved separately for deterministic 1-dimensional CA (as presented in papers 1 and 5) and probabilistic 1-dimensional diploid CA (as presented in papers 2, 3 and 4). In the case of deterministic CAs, an evolutionary algorithm is used, while in the case of probabilistic ones, the direct solution is based on statistical inference and parameter estimation. In the latter case, the overall inference about missing time steps is limited (the amount of missing information is unknown), hence only spatial incompleteness is taken into account. In both

cases, a significant set of computational experiments were performed, and the presented results demonstrate the effectiveness of the algorithms.

Although the title of the dissertation is general, solutions to the identification problem have been proposed for two classes of cellular automata: binary, one-dimensional, deterministic cellular automata with variable neighborhood sizes and binary, one-dimensional, diploid stochastic cellular automata with fixed neighborhoods (including a special subclass of α-Asynchronous CAs). This should not be considered a limited scope of work, because even in the case of binary, one-dimensional systems, the problem is difficult and requires the invention of appropriate methods, which constitutes an important contribution to the discipline.

On the other hand, I am more cautious than the candidate in terms of generalizations and I believe that his optimistic claim that " all the presented methods can be generalized to multi-state and multi-dimensional cases in a straightforward manner" requires more detailed justification. I am interested in the perspectives and possible limitations of the presented approach. Even though the full answer may require further research, which is not part of the PhD thesis, I would like the candidate to share his current experiences in this aspect in general and also especially about triploid (and also consisting of more components) stochastic cellular automata.

Going back to deterministic case, it is assumed that the first row of space-time diagram is completely observed, and not contain of “?” symbol, i.e. that $I[1] \in \{0, 1\}^M$. This assumption is “transferred” to the subsequent steps of the algorithm – each subsequent line treated as initial configuration is “filled” only with the values 0 and 1 values only; namely the positions occupied with “?” are filled with values obtained from the action of global rule F currently considered. I believe that this assumption can be relaxed, especially for M being (somehow) big enough. It would be good to clarify whether there is any substantial obstacle to relaxing this assumption, other than runtime performance degradation. Why a solution of discarding entries containing symbol “?” cannot be used as in the case of stochastic cellular automata?

It is also intriguing, however not of significant importance, why the rule 150 is compared to the rule 180 in paper 1, while in paper 5 the rule 184 is used as a reference to an orderly behaviour for comparison to the rule 150. Is there a specific reason for the change?

Assessment of the candidate's general theoretical knowledge in the discipline and the ability to independently carry out scientific research

The dissertation demonstrates that the candidate has the deep knowledge of cellular automata both in theoretical and computational aspects. He designed the solution, performed calculations introducing improvements on efficiency, and analysed obtained results on the level required for high-impact journal publications (including *IEEE Transactions on Cybernetics* with current Impact Factor 11.8 and CiteScore 22.3).

It should be emphasized that the candidate consequently worked on advancing the obtained results, which is visible in the progress of work in both series of articles (1,5 and 2,3,4), as well as in his future plans (including taking into account noise - erroneous observations – additionally to the previously recognized incompleteness of "observational" data). This means that the candidate formulated a scientific research program and, by implementing this plan, was able to obtain results that advanced the subject.

The candidate proved, that he can successfully share obtained results both by writing and publishing scientific articles as well as by giving oral presentations (on p.12 there is internet address to a recording of live presentation given at Summer Solstice 2022). He also is experienced in international collaboration, taking part in research projects and organizing necessary computational resources.

The justification that the solution of the problem in the doctoral dissertation is original

To my knowledge, the solution to the problem in the presented doctoral dissertation is original. In addition, the results were originally published in 5 scientific articles in monographs and scientific journals by the group of authors as indicated earlier, in all of which the candidate is the first author.

Opinion regarding the candidate's admission to the public defence of the doctoral dissertation

I find the dissertation as satisfying the requirements for compilation-of-papers type PhD thesis and on this basis the candidate deserves to be admitted to the public defence of a doctoral dissertation.

Distinction

I consider the subject of the dissertation to be very interesting and potentially of practical importance, promoting the use of cellular automata to model natural phenomena. In my opinion, the approach is very promising and may, in the long run, contribute to the wider use of CA as a modern, practical and efficient modeling tool. The developed approach is direct and elegant, provides methods that are effective for selected classes of cellular automata and allows for generalizations. To sum up, the presented results are scientifically justified, which is why I propose this dissertation for distinction.