Nancy (France), November 19, 2023

Report on the PhD thesis of W. Bołt,
prepared in the framework of the doctoral procedure **:**
"Identification of partially
observed deterministic and stochastic cellular automata"

On the problem studied in this PhD thesis

Cellular automata (CA) are a class of spatially-extended discrete dynamical systems ; they can be studied from three complementary perspectives :

- they provide a simple way to explore how complex global rules can emerge from simple local interactions, thus providing a clear mathematical description of how complexity can be observed in various systems with a low number of states and elementary interactions rules ;
- they allow researchers to invent new computing models, with an emphasis on using distributed anonymous elements which obey simple interaction patterns ;
- they provide a framework for modelling a wide range of phenomena observed in various sciences, whether natural sciences as Physics, Chemistry, Biology, etc., or "human sciences" such as Geography, Linguistics, Economics, Sociology, etc.

The topic studied by W. Bołt is the *identification problem*, that is, to know how one can identify the local rules which generate a particular behaviour represented by a given set of observations. In the case studied here, the observations take the form of space-time diagrams with incomplete information: this means that the initial condition of the system is supposed to be known, but the following states of the system are only partially known : either that some time steps are not observed, or some cells states are not known, or both. The difficulty of the problem is that there is a huge set of rules available to "reconstruct" the observations, even if one considers only binary states for the cells. There is thus a need to develop specific techniques to select the right local rules that totally or partially match a given set of observations. Note that the work of W. Bołt relies a fundamental hypothesis : unlike the missing time steps, which are never known, the missing states are marked with a particular symbol and the observed does know precisely where these symbols are found. As mentioned by the author himself, this is a strong assumption, which is nevertheless necessary a first step to lay the foundations of algorithmic techniques which can be later adapted to the case where the altered states cannot be distinguished from the "regular" states…

The importance of the topic studied W. Bołt is mainly related to the first and third aspects of cellular automata mentioned above. Indeed, it is intended to serve as a "bridge" between the world of simple binary systems and the field of modelling real-world systems, where the information would be incompletely known. The second aspect can also be related to the work of the candidate since it can also be quite interesting to perform some identification on systems which compute in a distributed way, as in the case of the density classification problem for instance. In all cases, relating the universe of simple dynamical systems and the world of real-world modelling is a key topic in the field of cellular automata, which has not been much studied to this date, given its tremendous difficulty.

Scientific content

The thesis of W. Bołt is presented in the form of a summary of his work of 37 pages, followed by five articles related to the CA identification problem. The comments that follow thus only concern the scientific aspects of this dissertation and not on the form of the presentation, which can be fully understood given the numerous responsibilities that he holds outside the academic world.

The instantiation of the identification problem concerns deterministic cellular automata with time gaps and partial spatial observations. According to our understanding, the main difficulty relies in the presence of *time gaps:* indeed, if there were only spatial observations missing, then it would be easy to fill the unknown look-up table of the CA by observing the places where complete transitions are present. The candidate chooses genetic algorithms in order to search for the original rule that generated a set of observations. He works on the space of radius-2 rules, which means that the identification has to select one rule among approximately 4.10^{10} rules. A preliminary work is presented where the candidate and his co-authors select two rules (ECA 150 and ECA 180, considered as radius-2 rules), and apply these searching techniques. They obtain encouraging results and notice a difference of how the algorithm performs on these two rules. This study is deepened in a another series of experiments, where the candidate presents a series of numerical simulations which are much more developed, and where a wider range of rules is studied (radius 2 and beyond) with a more diverse type of missing observations (in particular, time gaps).

These first results are quite encouraging and clearly demonstrate that W. Bołt has good mastery of his subject and knows how to make it interesting and how to develop the adequate settings of genetic algorithms to reach a satisfying solution. This also raises the question to know to which extent these techniques would apply on a wider range of rules, e.g., by extending the number of states, the dimension, etc. One would be also interested in changing the value of various parameters, in particular **Γ**, the bound for time gaps : for example, does the identification become impossible as **Γ** increases? Does it become trivial when **Γ** is too small ?

In a second step, the candidate considers alpha-asynchronous rules, that is, a stochastic case where the local rule is deterministic but where each cell can randomly either update its state or keep it unchanged. The probability to update a cell, alpha, is called the *synchrony rate*. This case is particularly interesting as the different degrees of synchrony may reveal to which extent the global behaviour of a given automaton depends on the hypothesis that all transitions occur at the same time. In the context of modelling natural phenomena, this is particular important to detect spurious behaviours that are only observed with a perfect simultaneous updating.

The candidate and his co-authors develop analytical techniques to identify the rule from the observations and to obtain an estimation of the synchrony rate (this parameter being continuous, its perfect knowledge is out of reach). Their analysis is mainly based on statistical estimation of the proportion of the number of occurrences of a given neighbourhood configuration on the number of cases where the application of the local rules sets the cell to a 1. The estimation of this proportion allows one to reconstruct the original rule with a good accuracy. They demonstrate the relevance of their analysis by presenting results which show that in all the cases examined (the 256 elementary cellular automata); in all cases, they succeed to identify the rule examined and to obtain a good estimate of the synchrony rate.

In this case too, one would be curious to know more about the conditions that can make this problem difficult: in particular, I may have missed a discussion on the values of the synchrony rate (e.g. : Is the identification more difficult in the case of a low value of the synchrony rate?)

In a third step, the candidate generalises the problem to diploid cellular automata, that is, the case of stochastic CA where two deterministic rules are “stochastically mixed”: in effect, this means that at each time step, each cell independently decides to apply either rule *f1* with probability *p* or rule *f2* with probability *1-p*. This step is a natural stage before reaching the world of “purely probabilistic rules” where each probability would be set independently for each corresponding transition.

Here too, each of the eight probabilities of the transition table of a stochastic ECA is estimated with statistical methods, and the original couple of rules is retrieved with this estimation, together with the mixing rates (the analogue of the synchrony rate in the previous case). The techniques used in this part of the work are quite similar to those of the previous step, except that the space to explore is now much bigger (256*255 rules, as the candidate does not use the symmetries of the problem to reduce the number of rules to examine). Consequently, this set of experiments now require a great amount of computations, which calls for specific HPC techniques to be put in place. The candidate also examine how the evolution of the CA can be reconstructed cell by cell in spite of the missing observation, what the authors call the *gap-filling procedure*. The results show that the success of this gap-filling procedure is dependent on the type of couples of rules that make up the diploid CA : in general, the more difference between the two, the more efficient the gap-filling procedure is, which is somehow logical as the differences between rules create more change of states, and thus allow one to gain more information on the transitions at play.

These results obtained are solid and convincing. They nevertheless raise the question to know which are the most challenging cases and *where are the limits* of the methods that were chosen…

In the last chapter of his dissertation, the candidate evokes various possibilities for future work: the first possibility is a key extension of this work, as it concerns the case where noise affects the observations (one does not know *which* observations are missing). The other suggestions regard an extension to other families of CA (continuous state, general probabilistic CA, etc.) and the use of machine-learning techniques ; all these suggestions form are quite stimulating and we do hope that the candidate will continue to explore these directions.

It also worth mentioning that W. Bołt has contributed to other aspects of the research on cellular automata. In particular, he has some interesting contributions on the density classification problem with various numerical simulations that explore different settings of this important problem.

To sum up, the dissertation presented testifies that the candidate has a deep comprehension of his research topic and that he knows how to present his results of his research in a synthetic form. The strong quality of his contributions are attested by publications in highly-recognised refereed international conference proceedings and journals. The originality of his research is obvious and calls for praise, as it is never easy to propose a new research path and to persevere on this path to demonstrate its validity and relevance.

Contribution of the doctoral candidate

The role of the candidate and his co-authors is described precisely in the first part of the summary (Sec. 2). According to this information, and according to my personal knowledge of the research team where the candidate was involved, and given that I could assist to presentations he made in international conferences, it clearly appears that the candidate always played a central role in the development of this research, as well as in the writing of the articles.

Suggested improvements

The summary is written with care and is pleasant to read. It is however a pity that is does not contain a single space-time diagram (nor any figure). Maybe this is due to the space limit that was imposed. In all cases, if this is possible, we strongly recommend to add a few visual elements that would help understanding the mathematical definitions of the problem.

The bibliography could slightly be more developed. In particular, some more references on the identification problem would be welcome (e.g., see the article “Identification of cellular automata” in the Encyclopedia of Complexity and Systems Science DOI: 10.1007/978-0-387-30440-3_280). It would be a good thing to learn more about the other approaches that have been used to tackle the identification problem and what were the limits of these approaches (in a few words).

Concluding comments

To sum up, this dissertation presents a rich palette of solutions to the identification problem in cellular automata. The candidate demonstrated a mastery of various techniques, both at the mathematical analytical level and at the computer science level (numerical simulations, in a classical framework or with HPC tools). He proved his capacity of conducting effective research work, to integrate a research team and to raise his level up to the international standards.

I thus strongly recommend to accept Mr Witold Bołt as a Doctor of Philosophy in Information And Communication Technology for the Polish Academy of Sciences. Given the solidity of his scientific quality of his work, given the originality of his research theme, given his personal qualities of perseverance and clarity in the expression, I also support the attribution of honours to him if a consensus on this point emerges in the jury.

Nazim Fatès,

full-time researcher at the Inria National Institute for Research in Digital Science and Technology,
chair of the IFIP WG 1.5 working group on cellular automata and discrete complex systems.

N. Fatès