



**POLITECHNIKA
RZESZOWSKA**
im. IGNACEGO ŁUKASIEWICZA

al. Powstańców Warszawy 12, 35-959 Rzeszów
tel./fax: +48 17 854 12 60, tel.: +48 17 865 11 00
www.prz.edu.pl

dr hab. inż. Dominik Strzałka, prof. PRz

Rzeszów, dn. 09.06.2023 r.

Politechnika Rzeszowska

Wydział Elektrotechniki i Informatyki

Zakład Systemów Złożonych

Tytuł rozprawy: Neural Simulation Pipeline for Liquid State Machines

Autor rozprawy: mgr Karol Chlasta

Promotor rozprawy: dr hab. Grzegorz Wójcik, prof. PJAiT

Promotor pomocniczy: dr hab. Izabela Krejtz, prof. USWPS

Dziedzina: nauki inżynieryjno-techniczne

Dyscyplina: informatyka techniczna i telekomunikacja



ZAKŁAD SYSTEMÓW ZŁOŻONYCH

Wydział Elektrotechniki i Informatyki
ul. Wincentego Pola 2, 35-959 Rzeszów, tel. 17 865 1340
zsz.prz.edu.pl



Uwaga główna/General remark

Praca doktorska została (z wyjątkiem rozdziału Polish language version) w całości napisana w języku angielskim. W związku z powyższym recenzja pracy została przygotowana w dwóch językach: polskim i angielskim.

The doctoral thesis (except for the Polish language version chapter) was written entirely in English. Therefore, the review of the thesis was prepared in two languages: Polish and English.





A synthetic presentation of the dissertation:

The main goal of the thesis

In this multidisciplinary dissertation author is taking the reader for a long, interesting journey devoted to computer-based simulations of neural networks in order to overcome some key challenges in neuroscience and also is giving a helping hand for understanding the joint, complex behavior of large groups of different types of neurons. In the dissertation thesis, author states that numerical simulations should integrate a robust model development methodology and appropriate simulation and testing steering workflows to understand the current and future challenges in neuroscience. This should significantly increase research throughput and improve utilization of on-premise and in-cloud current and next-generation computational infrastructure especially lowering the entry barriers.

Scope of the work

Generally the scope of the work is focused on neuroinformatics, as a branch of computer science and engineering, and related to processing of data obtained in computer simulations of large, biologically plausible neural network models using GENESIS simulation engine in order to build appropriate scientific workflows, here called Neural Simulation Pipeline (NSP), with the use of Liquid State Machines. Moreover, there is a significant part related to hardware solutions and containerization based on on-premise and in-cloud models. The scope of the work is located in the information and communication technology discipline.

Description of the work

The whole dissertation consists of six main chapters and four appendixes. In the structure of dissertation the reader can observe two main parts. The first one is devoted to theoretical considerations about presented topic including: introduction to selected issues of neuroscience, development of adequate models, testing and simulation steering workflows methodology, and next-generation computational infrastructure on both on-premise and in-cloud solutions for tasks such as pattern recognition and classification. The second part shows main author's practical achievements related to the scope of work, its topic and results of experimental part.





In chapter 1 that consists of 11 subsections we have a justification why the brain simulations are critical for neuroscience. Next chapter of dissertation, with again 11 subsections, provides a new computing paradigm – the idea of liquid computing with justification of its importance in brain modelling. Chapter 3, with 7 subsections, is focused on the Liquid State Machines models and their simulation setting as a part of biocybernetic modelling. Author proposed 3 models that are the base of currently common used approaches. This finishes the theoretical, introductory part of dissertation. Chapter 4, built of 6 subsections, is focused on experimental work that is computer simulations based on the cybernetic models with setups proposed in Chapter 3. Author tested the performance of twelve machine learning algorithms used in the training of own LSM model. The next 5th Chapter proposes NSP – Neural Simulation Pipeline – own author's solution of processing pipeline. In 6 successive sections of this chapter we can see an approach where the prototype built of single board computer clusters is given. This prototype is given with many technical details and photos showing interesting design results on both hardware and software level. There is also a proposal of pipeline with the use of AWS cloud. Both chapters, 4 and 5 are the core of this thesis. The whole thesis has a summary in Section 6.

This chapters layout is correct and gives impression to the reader that structure of dissertation is well-designed and well thought out.

Review about the presentation of general theoretical knowledge:

- the correctness of the originality of the thesis and the extent to which it has been demonstrated

The thesis stated in the dissertation is:

... numerical simulations must integrate a robust model development methodology, with adequate testing and simulation steering workflows to increase scientific throughput, and improve utilisation of current and next-generation computational infrastructure, available both on-premise and in-cloud.

This thesis is validated by several actions taken by the author. Among them we have:

1. Modelling of different spiking neural networks basing on the concept of Liquid State Machines.
2. Study the impact of different computing resources on model performance, and data generation delivered by system GENESIS.
3. Introduction of automation into the model development to perform testing and deployment process prototypes of the neural simulation pipeline using own neurobiological simulations.





We also have investigations related to 6 research questions stated at the end of Section 1.11 in pages 24 and 25.

The dissertation thesis is original and innovative in relation to the broad context of the entire dissertation – it refers to the real, significant and current scientific and practical problems with reference to the international scientific community activity and existing research teams, trends, research directions and commonly meet problems in this field.

Taking into account the whole content of the work, the considered scope of experimental and research performed work, and the individual author's contribution, it should be stated that the proposed thesis has been proven in relation to specific examples considered in the work, including case studies.

- the analysis of sources

Author used in dissertation references to 247 different sources. They are listed in the alphabetical order. Depending on the needs most of them (>80%) refers the research and scientific achievements that were obtained in the last 10-12 years. Taking into account the wide scope of presented references, the spectrum of considered problems and constant progress in neuroscience and computer engineering the analysis of sources is more than satisfactory.

- the position of the dissertation in relation to the state of knowledge and/or state of the art represented by world literature

The scope of work is special since it belongs to not easy domains of science that combine, to put it simply, biology and computers. It is not easy, in general, to focus on such problems without special knowledge, research testing stands, big laboratory teams, different resources and, unfortunately big money. Author, based on several papers [8, 107, 151, 179, 152, 230] that were the main inspiration for his research and having in mind all wide context of the considered scoped, proposed something that is really new. This is not only the proposal of pipeline but also the approach based on quite cheap processing units that are reachable for many. This can be quickly expanded and extended taking into account different problems that come from limited resources and easily be repeated in small labs by small teams – note: author did almost all work alone.

The use of machine learning techniques in general is a not new approach and from the perspective of fourth thesis Chapter (I will expand this lately) is also unfortunately little bit disappointing from scientific point of view, but on the other hand shows author's broad view on different selected techniques and tools that were successfully applied during the research giving some important results. Maybe their presentation is not satisfactory and can be improved but on the other hand this dissertation is not about a machine learning itself.





During author's research activity he was able to prepare 7 scientific papers, but not all of them are related to the scope of thesis (see Table A.1, page 195). All of them are published in good, reviewed scientific papers and conference proceedings.

So comparing dissertation thesis to the current state of knowledge and literature we have quite well-prepared scientific work that convinces the reader that author knows the ropes.

- opinion on the importance of the obtained results for a given scientific discipline

As it was written above, the author basing on own's idea proposed a whole processing pipeline that refers not only to processing of data with the use of ML techniques but also can be obtained thanks to the own pipeline based on hardware solution. Having in mind that during data processing there is a need to have big computing power which means the necessity of big money not only for its purchase but also for maintenance, the alternative proposals related to cluster build of single board computers (here Raspberry Pi) and GENESIS simulation engine together with the use as an alternative cloud computing are really good. It shows that there is a way to create valuable solutions that significantly lower the entry barriers related to hardware in neurocomputing.

In my opinion, Chapter 5 is very interesting since it solves very important problem: how to have or to ensure necessary computing power in order to perform data processing in acceptable time and with available, sometimes small amount of money. It is clear that all similar researches related to this topic are done with big money that is also spend on computers with bit computing power. In Polish conditions this is not allowed to everyone, but there are possible ways to omit this problem. The thesis proposal of on-premise solution or the use of cloud resources to do the processing seems to be natural, but as we can read in the thesis it is not commonly used in neuroinformatics. We can suppose that this could be caused by limited knowledge and practical skills in this field. Author made a significant progress and showed how to do this.

In summary, this dissertation gives an important contribution to the discipline IT and telecommunications.

- opinion about the author's ability to correctly and convincingly present obtained results

The main results of work are given in Chapter 4 and 5. The first one consists of around 50 pages, the second one has around 40 pages.





Chapter 4 is mostly devoted to experiments with Liquid State Machines and the use of 12 machine learning algorithms for data processing. We have experiments related to attempt for: (i) computational complexity of single LSM column and set of experiments related to (ii) recognitions of 3 signs (chars) with the use of ML algorithms.

In the case of (i) the final result (page 80) is expressed as $O(0.000679468 * n^2 - 1.09334 * x + 716.782)$ but there are the following problems that should be explained by author:

- technical problem – in the result we have variable x instead of n . Since we are talking about the size n of task, variable x should be considered here as n and is rather a typo;
- the final result on page 80 has been achieved basing on several (less than 10?) measured points for one pattern of '0' – see also the remark 9 given in the disadvantages of work, but I think that it is not enough to write so general statements. I'm not saying that this is not the truth, but it will be better to provide more convincing proofs. In statistics, having several points its quite easy to draw many functions for curve fit and say that this is the proof;
- the result was achieved using MS Excell, but it will be good to have more details about this fit and show fitted function on Fig. 4.5, for instance the R^2 parameter that measures the trendline reliability.

In the case of results presented on Figs 4.7 – 4.9 some comments to data presented should be given. There are box plots, maybe there are important, but they are left without a special attention.

In the case of (ii) the major problems can be described as follows:

- Tables 4.3 – 4.5 are given without any special comments and even references into the text. In Table 4.5 we have a column cost [USD] – here the reader can suppose for example that this was the cost of access to AWS Cloud, but how it was calculated and is it important in the context of results presented in this Chapter is unclear. In fact, the details are given in Chapter 5.5, page 163.
- on page 90 we have: “... *that are organised in 273519 distinct spike observations, as per Fig. 4.10.*” but I cannot see this in Fig. 4.10.
- on page 90 we have: “... *set using the 80/20 principle.*” Is it a general rule in the case of similar research?
- on page 92 we have Eqs. (4.5) and (4.6) but if may way of understanding this good they can be exchanged by Eqs. (4.2) and (4.3).





- in subsection 4.5.3 we have many tables (4.7 – 4.30) and no references (except 4.21) in the text to any of them. For tables: 4.8, 4.10, 4.12, 4.14, 4.16, 4.18, 4.20, 4.22, 4.24, 4.26, 4.28, 4.30 its quite easy to understood what the show, but for tables 4.9, 4.11, 4.13, 4.15, 4.17, 4.19, 4.23, 4.25, 4.27, 4.29 we have some parameters that are typical for used ML algorithms and they are left without any comments. Some of these tables show different parameters of algorithms and it will be good to know at least are these parameters given with default values or maybe there were some experiments related to fixing these values in order to improve the final results of algorithms' work. This is especially important in the case of these algorithms where the final results were scored as *a not good result*.

Chapter 5 shows the purpose of hardware architecture based on single board computers and a processing pipeline that allows to use cloud (public/private) resources for data processing. This chapter is also little bit disappointing, like Chapter 4, since it mainly focuses on convincing the reader why Docker, why Raspberry Pi, why containerization, etc. I expected that inside, in the text there will be given more details on how everything was built and how the prepared Bash scripts were developed and work. In fact this is really valuable contribution of author but reading the main text of dissertation is very hard to have at least general impression about the difficulty of done work. Some of these flaws and doubts are dispelled in Appendix B, but not all of them, especially to details about prepared Bash scripts. Relating to my own experience: here, a lot was done, but the author has some problems how to show these results in a convincing way; the choice that was done is not a good one.

The main disadvantages of the dissertation and its weaknesses.

Some of the dissertation disadvantages were presented above, here it is worth to note that: author presented the Neural Simulation Pipeline (NSP) as a novel setup that allows running neural simulation in the cloud environment. This is important practical contribution since I suppose that many of scientists involved in neuroscience can have big problems how this could be done. Moreover, we have a practical proposal of cheap computing cluster called Neural Simulations Cluster (NSC) built with Raspberry Pi and ROCKPro64 boards that can support different simulations when the accessible budget is limited and some simulations can be done even at home environment. But the presented work is not without flaws.

Weaknesses: some of particular and general weak elements of dissertation were indicated above, but here I would like to point other remarks and questions:





1. In Sect. 1.3, page 3, author wrote: “*It is built of 100 billion interconnected neural cells forming an extremely large and complex network.*”. Is there any paper that explains this complexity? How we can model this complexity taking for instance time and spatial/topological domain? What really means that the brain is a complex structure? How author understands the term complexity? Moreover, on page 58, end of Sect. 3.3 we have a view where author wrote “... *different researchers might develop different parts of that large computational model of a reference brain. (...) If such a reference brain can be simulated in reasonable time, the next steps could be to simulate the whole, thinking brain...*”. In my, but not only, opinion this is a reductionist approach that is not necessary suitable for any complex systems.

2. In Sect. 1.6.2, page 9, author wrote: “*There is a consensus that information processing specific to the neocortex is performed by neuronal microcircuits.*”. How it is similar to our imaginations and existing theories about information processing especially when it is digital versus analogue?

3. On page 4, we have: “*These large computers, often need to also consume a lot of electricity, ...*” – let’s say that they consume electrical energy.

4. On page 17 there is: 9×10^6 neurons with 18×10^9 synapses and 2.2×10^6 neurons with 45×10^9 synapses but I suppose that there should be: 9×10^6 neurons with 18×10^9 synapses and 2.2×10^6 neurons with 45×10^9 synapses as it is given in the abstract of work [44].

5. On page 29 we have a paragraph where there is a description of LSM theory and its computing power equal to Turing machines: LSM can simulate any Turing machine, but what about the reverse statement?

6. In section 2.6 on page 34 we have: “... *the input info function $Y^N(t)$* ”, but in Eq. (2.1) and (2.2) there is $y^N(t)$. I suppose that there is a typo.

7. Fig. 2.2 a) and b) on page 37 shows some results taken from [151], but how they should be interpreted?

8. In Sect. 2.11 in point 2 (c) we have a mention about bifurcations during learning. Did author meet with something like this during preparation of this thesis?

9. What could be the results of simulations if the digit “0” given in Fig. 3.1 won’t be a pattern with straight lines? For example the corners of digit 0 will be rounded.

10. When the results of experiments with machine learning are given in Chapter 4, there is a lack of explanations about different parameters that are used in ML algorithms. For example on page 113 we have: “... *using a number of pre-selected parameters*”. This is too enigmatic and it will be good to discuss more details about these issues.

11. Why the AWS cloud was taken?





Technical and language errors:

1. A quick comparison of photos for hardware solutions with some figures presenting important results of work is not convincing. The quality of many figures with important results is very low, for instance Fig. 4.12 a), 4.15 a), 4.21 a).

2. Many figures and tables given in the dissertation haven't got any references and comments in the text. Some of the also mentioned above, but there are more, for instance Fig. 3.5, Fig. 3.6. If there is no reference in the text, why they are given?

3. In many parts of the text, the author writes in the first person. For example: page 3: *I assess*, page 5: *I analyse, I explore, I run, I design, I evaluate, I attempt*, and so on. This is not a typical for such dissertation and usually it is expected to write in a rather impersonal form.

4. There are also some technical errors related to missing spaces, wrong measurement units, etc. but here I will omit them as not important ones.

Generally the work is well prepared and there aren't many flaws that significantly influences the overall overview of the thesis.

Review about whether the doctoral dissertation demonstrate general theoretical knowledge

Taking into account the so far presented arguments, the content of work and its general structure it should be said that author demonstrated necessary amount of general theoretical knowledge together with many practical skills typical in computer science and engineering. The most important of these statements comes from the fact that thesis gives a specific language, skills, knowledge that belong to more than two disciplines where we have a connection to brain behavior and attempts to its modelling with computers. Many important details and considerations presented in Chapters 1-3 are significantly enough in order to state that author has very good theoretical background for scientific activities related to neuroscience. Most of parts were supported by literature review, detailed explanations and the several doubts presented in this review above can be easy explained by the author.





Review about whether the doctoral dissertation demonstrates the ability to independently conduct scientific work

In the dissertation author presented two Chapters (4 and 5) that are mainly focused on his independent and personal work. Special attention should be paid on Chapter 5 where there are original own proposals related both to the hardware and software level. Despite the fact that in my opinion the author had some problems with better presentation of the work results, I'm convinced that there are no doubts that author has a full ability to conduct scientific work. To be more precise: I'm convinced that the author put a lot of work in order to present the final results, but he could pay more attention to better presentation of results. The above indicated flaws are minor remarks that can be explained giving the answers or during PhD defense.

Review about whether the doctoral dissertation is an original solution to a scientific problem, an original solution in the application of the results of one's own research in the economic or social sphere

As it was written above and taking into account the surprising fact that the neuroscience community did not often use (see page 123) any single board computer clusters (SBCC) focused on simulating cybernetic models; neither in the area of research, nor even in teaching neuroinformatics, biomedical engineering or computational neuroscience at least the proposed in the work original software and hardware pipeline solution is original to existing scientific problems. It is not only presented as a theoretical one but is given in practice with necessary engineering knowledge and practical skills.





Final conclusion

After the detailed analysis of the submitted doctoral dissertation prepared by MSc Karol Chlasta, I would like to express that it makes a valuable contribution to the development of the information and communication technology discipline in the engineering and technology field (domain). It is an original solution to a scientific problem and original solution in the application of the results of PhD candidate's scientific research, confirming that the doctoral dissertation presents general theoretical knowledge, very good substantive preparation, practical skills and the ability to independently conduct scientific research by a PhD student – Karol Chlasta. The subject matter is up-to-date and has a very high application potential. The reviewed doctoral dissertation meets the statutory requirements for candidates for the doctoral degree, specified in the Act on Academic Degrees and Titles - included in Art. 13 sec. 1 of the Act of March 14, 2003 on academic degrees and titles and degrees and titles in the field of art in connection with art. 179 of the Act of July 3, 2018. Provisions introducing the Act - Law on Higher Education and Science (Journal of Laws of 2018, item 1669).

Basing on this conclusion, I am applying to the IT and Telecommunications Discipline Council of the Polish-Japanese Academy of Information Technology for admission to its public defense.

Dominik Stępień





Syntetyczna prezentacja rozprawy

Główny cel pracy

W recenzowanej multidyscyplinarnej rozprawie doktorskiej, autor zabiera czytelnika w długą, interesującą podróż poświęconą komputerowym symulacjom sieci neuronowych w celu przezwyciężenia niektórych kluczowych wyzwań neuroinformatyki, a także podaje pomocną dłoń w zrozumieniu wspólnego, złożonego zachowania dużych grup różnych typów neuronów. W rozprawie doktorskiej autor stwierdza, że symulacje numeryczne powinny integrować solidną metodologię opracowywania modeli oraz odpowiednie symulacje i testy sterujące przepływami strumieni danych, aby zrozumieć obecne i przyszłe wyzwania w neuroinformatyce. Powinno to znacząco zwiększyć przepustowość badań naukowych i poprawić wykorzystanie obecnej i nowej generacji infrastruktury obliczeniowej w modelach on-premise i in-cloud, zwłaszcza obniżając tzw. bariery wejścia.

Zakres prac

W dużym uogólnieniu praca koncentruje się na neuroinformatyce, jako gałęzi informatyki, i jest związana z przetwarzaniem danych uzyskanych w symulacjach komputerowych dużych, biologicznie wiarygodnych modeli sieci neuronowych z użyciem silnika symulacyjnego GENESIS w celu zbudowania odpowiednich przepływów strumieni danych i przetwarzania, tutaj zwanych jako Neural Simulation Pipeline (NSP), oraz z użyciem Liquid State Machines. Ponadto, znaczna część pracy dotyczy rozwiązań sprzętowych i konteneryzacji w oparciu o modele on-premise i in-cloud. Zakres prac mieści się w dyscyplinie informatyka techniczna i telekomunikacja.

Opis pracy

Całość rozprawy składa się z sześciu głównych rozdziałów oraz czterech dodatków. W strukturze rozprawy czytelnik może zaobserwować dwie główne części. Pierwsza poświęcona jest rozważaniom teoretycznym na prezentowany temat, w tym: wprowadzeniu do wybranych zagadnień neuroinformatyki, opracowaniu odpowiednich modeli, metodologii testowania i sterowania przepływami strumieni danych oraz infrastrukturze obliczeniowej nowej generacji w rozwiązaniach on-premise i in-cloud do zadań takich jak rozpoznawanie wzorców i klasyfikacja. W drugiej części przedstawiono autorskie rozwiązanie praktyczne związane z zakresem pracy, jej tematyką oraz wynikami części eksperymentalnej.





W rozdziale 1, który składa się z 11-u podrozdziałów, mamy uzasadnienie, dlaczego symulacje mózgu są kluczowe dla neuroinformatyki. Kolejny rozdział rozprawy, ponownie liczący 11 podrozdziałów, przedstawia nowy paradygmat obliczeniowy – ideę Liquid State Machines wraz z uzasadnieniem jej znaczenia w modelowaniu mózgu. Rozdział 3, składający się z 7-u podrozdziałów, koncentruje się na modelach Liquid State Machines i ich symulacji jako części modelowania biocybernetycznego. Autor zaproponował 3 modele, które stanowią podstawę obecnie powszechnie stosowanych podejść. Na tym kończy się teoretyczna, wstępna część rozprawy. Rozdział 4, składający się z 6-u podrozdziałów, koncentruje się na pracy eksperymentalnej, czyli symulacjach komputerowych opartych na modelach cybernetycznych z rozwiązaniami zaproponowanymi w rozdziale 3. Autor przetestował działanie dwunastu algorytmów uczenia maszynowego używanych w uczeniu własnego modelu LSM. W kolejnym, piątym rozdziale zaproponowano NSP – Neural Simulation Pipeline – autorskie rozwiązanie potoku przetwarzania. W 6-u kolejnych podrozdziałach tego rozdziału możemy zobaczyć podejście, w którym podany jest prototyp zbudowany z jednopłytkowych platform komputerowych. Ten prototyp jest opatrzony wieloma szczegółami technicznymi i zdjęciami przedstawiającymi interesujące wyniki projektowe zarówno na poziomie sprzętu, jak i oprogramowania. Jest też propozycja potoku z wykorzystaniem chmury AWS. Obydwa rozdziały, 4 i 5, stanowią praktyczne trzon rozprawy. Całość pracy jest podsumowana w rozdziale 6.

Taki układ rozdziałów jest poprawny i daje czytelnikowi wrażenie, że struktura rozprawy jest dobrze zaprojektowana i przemyślana.

Przegląd dotyczący prezentacji ogólnej wiedzy teoretycznej

- poprawność oryginalności pracy i stopień jej wykazania

Teza postawiona w rozprawie to:

... numerical simulations must integrate a robust model development methodology, with adequate testing and simulation steering workflows to increase scientific throughput, and improve utilisation of current and next-generation computational infrastructure, available both on-premise and in-cloud.

Teza jest weryfikowana przez szereg działań podjętych przez autora. Wśród nich mamy:

1. Modelowanie różnych sieci neuronowych w oparciu o koncepcję Liquid State Machines.

2. Zbadanie wpływu różnych zasobów obliczeniowych na wydajność modelu i generowanie danych dostarczanych przez system GENESIS.





3. Wprowadzenie automatyzacji do budowy modelu w celu wykonania testów i wdrożenia prototypów procesów neuronowego potoku symulacji z wykorzystaniem własnych symulacji neurobiologicznych.

W pracy pokazano również analizy związane z 6-oma pytaniami badawczymi podanymi na końcu sekcji 1.11 na stronach 24 i 25.

Praca doktorska jest oryginalna i nowatorska a w stosunku do szerokiego kontekstu całej rozprawy odnosi się do rzeczywistych, znaczących i aktualnych problemów naukowych i praktycznych w odniesieniu do działalności międzynarodowego środowiska naukowego i istniejących zespołów badawczych, trendów, kierunków badań i powszechnie spotykanych problemów w tym zakresie.

Biorąc pod uwagę całokształt treści pracy, rozpatrywany zakres wykonanych prac eksperymentalnych i badawczych oraz indywidualny wkład autora należy stwierdzić, że postawiona teza została udowodniona w odniesieniu do konkretnych przykładów uwzględnionych w pracy, w tym studiów przypadku.

- analiza źródeł

Autor użył w rozprawie odniesień do 247 różnych źródeł literaturowych. Są one wymienione w porządku alfabetycznym. W zależności od potrzeb większość z nich (>80%) odnosi się do badań i osiągnięć naukowych uzyskanych w ciągu ostatnich 10-12 lat. Biorąc pod uwagę szeroki zakres prezentowanej literatury, spektrum rozważanych problemów oraz stały postęp neuroinformatyki i inżynierii komputerowej analiza źródeł jest więcej niż zadowalająca.

- pozycja rozprawy w stosunku do stanu wiedzy i/lub stanu sztuki reprezentowanej przez literaturę światową

Zakres pracy jest poniekąd szczególny, gdyż należy jej interdyscyplinarność wynika z połączenia nietrywialnych obszarów nauki, które łączą, najprościej mówiąc, biologię i informatykę. Generalnie nie jest łatwo skupić się na takich problemach bez specjalistycznej wiedzy, stanowisk badawczych, dużych zespołów laboratoryjnych, różnych zasobów i niestety dużych pieniędzy. Autor, opierając się na kilku pracach [8, 107, 151, 179, 152, 230], które były główną inspiracją dla jego badań i mając na uwadze cały szeroki kontekst rozważanego zakresu, zaproponował nowe podejście. To nie tylko propozycja potoku, ale również podejście oparte na dość tanich jednostkach przetwarzających, które są finansowo osiągalne niemal dla każdego. Można tę propozycję szybko rozszerzyć biorąc pod uwagę różne problemy, które wynikają z ograniczonych zasobów i łatwo powtórzyć w małych laboratoriach przez małe zespoły – uwaga: tutaj autor prawie wszystko opracował sam.





Użycie technik uczenia maszynowego zasadniczo nie jest nowym podejściem i z punktu widzenia czwartego rozdziału pracy doktorskiej (będzie to rozwinięte dalej) jest niestety trochę rozczarowujące zwłaszcza z naukowego punktu widzenia, ale z drugiej strony pokazuje szerokie spojrzenie autora na różne wybrane techniki i narzędzia, które z powodzeniem zastosowano podczas badań, dając kilka ważnych wyników. Być może ich prezentacja nie jest zadowalająca i można ją poprawić, ale z drugiej strony ta rozprawa nie jest o samym uczeniu maszynowym.

W trakcie działalności badawczej autora udało mu się przygotować 7 prac naukowych, jednak nie wszystkie z nich są związane z zakresem pracy (patrz tab. A.1, str. 195). Wszystkie są publikowane w dobrych, recenzowanych czasopismach naukowych i materiałach konferencyjnych.

Porównując więc tezę pracy do obecnego stanu wiedzy i literatury mamy całkiem dobrze przygotowaną pracę naukową, która przekonuje czytelnika, że autor rozumie istotę problemów.

- opinia o znaczeniu uzyskanych wyników dla danej dyscypliny naukowej

Jak już napisano wyżej, autor rozprawy opierając się na własnym pomysłe zaproponował cały potok przetwarzania, który odnosi się nie tylko do przetwarzania danych z wykorzystaniem technik ML, ale również może być realizowany dzięki własnemu podejściu opartemu na rozwiązaniu sprzętowym. Mając na uwadze, że podczas przetwarzania danych potrzebna jest duża moc obliczeniowa, co oznacza konieczność ponoszenia dużych nakładów finansowych nie tylko na jej zakup, ale również na utrzymanie, alternatywne propozycje dotyczące budowy klastrów z komputerów jednopłytkowych (tutaj Raspberry Pi) oraz silnik symulacyjny GENESIS wraz z ewentualnym użyciem jako alternatywy cloud computingu są naprawdę bardzo wartościowe. Pokazuje to, że istnieje sposób na tworzenie wartościowych rozwiązań, które znacznie pokonują bariery wejścia związane ze sprzętem w neuroinformatyce.





Moim zdaniem rozdział 5 jest bardzo interesujący, ponieważ rozwiązuje bardzo ważny problem: jak mieć lub zapewnić niezbędną moc obliczeniową, aby przetwarzanie danych odbywało się w akceptowalnym czasie i przy dostępnych, czasem niewielkich nakładach finansowych. Oczywiście jest, że wszystkie podobne badania związane z tym tematem są wykonywane za duże pieniądze, które są również wydawane na komputery o bardzo dużej mocy obliczeniowej. W polskich warunkach nie zawsze jest to możliwe, ale istnieją jak widać sposoby na ominięcie tego problemu. Propozycja rozwiązania zawarta w rozprawie, aby przy użyciu podejścia on-premise lub używania zasobów chmury do przetwarzania danych, wydaje się naturalna, ale jak czytamy w pracy nie jest powszechnie stosowana w neuroinformatyce. Można przypuszczać, że jest to być może spowodowane ograniczoną wiedzą i umiejętnościami praktycznymi naukowców w tym zakresie. Autor swoją ciekawą propozycją pokazał, jak to zrobić.

Podsumowując, niniejsza rozprawa stanowi ważny wkład w dyscyplinę informatyka techniczna i telekomunikacja.

- opinia o umiejętności autora do poprawnego i przekonującego przedstawienia uzyskanych wyników

Główne wyniki pracy podano w rozdziałach 4 i 5. Pierwszy składa się z około 50 stron, drugi zaś z około 40 stron.

Rozdział 4 jest w większości poświęcony eksperymentom z Liquid State Machines i użyciu 12 algorytmów uczenia maszynowego do przetwarzania danych. Mamy eksperymenty związane z: (i) analizą złożoności obliczeniowej pojedynczej kolumny LSM oraz zestaw eksperymentów związanych z (ii) rozpoznawaniem 3 znaków (liter) za pomocą algorytmów ML.

W przypadku (i) wynik końcowy (str. 80) jest wyrażony jako $O(0,000679468 * n^2 - 1,09334 * x + 716,782)$, ale autor powinien wyjaśnić następujące kwestie:

- problem techniczny – w powyższym wyniku mamy zmienną x zamiast n . Ponieważ mówimy o rozmiarze n zadania, zmienną x należy tutaj traktować jako n i jest to chyba raczej literówka;

- wynik końcowy na str. 80 uzyskano na podstawie kilku (mniej niż 10?) punktów pomiarowych dla jednego wzorca znaku '0' – patrz też uwaga 9 podana w słabych stronach pracy, ale myślę, że jest to niewystarczające, aby podać tak ogólne stwierdzenie. Nie twierdę, że to jest nieprawda, ale lepiej byłoby przedstawić bardziej przekonujące dowody. W statystyce, mając na wykresie kilka punktów, dość łatwo jest narysować wiele różnych funkcji dopasowania krzywych i powiedzieć, że to jest dowód;





- wynik uzyskano za pomocą MS Excell, ale dobrze byłoby mieć więcej szczegółów na temat tego dopasowania i pokazać dopasowaną funkcję na rys. 4.5, na przykład parametr R^2 mierzący rzetelność wyniku uzyskanej linii trendu.

W przypadku wyników przedstawionych na rysunkach 4.7 – 4.9 należy dodać komentarz do przedstawionych danych. Mamy wykresy typu box-plot i być może są one ważne, ale pozostawiono je bez szczególnej uwagi.

W przypadku (ii) główne problemy można opisać następująco:

- Tabele 4.3 – 4.5 podano bez specjalnych komentarzy, a nawet odniesień do tekstu. W tabeli 4.5 mamy kolumnę koszt [USD] – tutaj czytelnik może np. założyć, że był to koszt dostępu do AWS Cloud, ale nie jest jasne, w jaki sposób został on obliczony i czy ma to znaczenie w kontekście wyników przedstawionych w tym rozdziale. W rzeczywistości szczegóły podano w rozdziale 5.5 na stronie 163.

- na stronie 90 mamy: “... *that are organised in 273519 distinct spike observations, as per Fig. 4.10.*”, ale nie widzę tego na Rys. 4.10.

- na stronie 90 mamy: “... *set using the 80/20 principle.*”. Czy jest to jakaś ogólna zasada w przypadku podobnych badań?

- na stronie 92 mamy równania (4.5) i (4.6), ale jeśli dobrze to rozumiem, można je zamienić przez równania (4.2) i (4.3).

- w podrozdziale 4.5.3 mamy wiele tabel (4.7 – 4.30) i brak odniesienia (poza tabelą 4.21) w tekście do którejkolwiek z nich. Dla tabel: 4.8, 4.10, 4.12, 4.14, 4.16, 4.18, 4.20, 4.22, 4.24, 4.26, 4.28, 4.30 dość łatwo zauważyć, o co chodzi, ale dla tabel 4.9, 4.11, 4.13, 4.15, 4.17, 4.19, 4.23, 4.25, 4.27, 4.29 mamy kilka parametrów typowych dla zastosowanych algorytmów ML i pozostawiono je bez komentarza. Niektóre z tych tabel pokazują różne parametry algorytmów i dobrze byłoby przynajmniej pokazać, czy te podane parametry są z wartościami domyślnymi, czy może były jakieś eksperymenty związane z ustalaniem tych wartości, aby poprawić końcowe wyniki pracy algorytmów. Jest to szczególnie ważne w przypadku tych algorytmów, gdzie końcowe wyniki zostały ocenione jako *a not good result*.





W rozdziale 5 przedstawiono realizację architektury sprzętowej opartej na komputerach jednopłytowych i potoku przetwarzania, który umożliwia wykorzystanie zasobów chmury (publicznej/prywatnej) do przetwarzania danych. Ten rozdział również trochę rozczarowuje, podobnie jak 4, ponieważ skupia się głównie na przekonaniu czytelnika, dlaczego Docker, dlaczego Raspberry Pi, dlaczego konteneryzacja, itp. Spodziewałem się, że w tekście, będzie podane więcej szczegółów, jak to wszystko było zbudowane oraz jak powstały i działają przygotowane skrypty Bash. W rzeczywistości jest to moim zdaniem naprawdę cenny wkład autora, ale czytając główny tekst rozprawy bardzo trudno jest mieć choćby ogólne wrażenie o trudności wykonanej pracy. Niektóre z tych słabych stron i moich wątpliwości zostały rozwiązane w Dodatku B, ale nie wszystkie, zwłaszcza jeśli chodzi o szczegóły dotyczące przygotowanych skryptów Bash. Odnosząc się do mojego własnego doświadczenia: w rozprawie sporo zostało zrobione, ale autor ma pewne problemy z przedstawieniem tych wyników w przekonujący sposób; sposób, w który tego dokonano, nie jest za dobry.

Główne wady rozprawy i jej słabe strony.

Niektóre z wad rozprawy zostały już przedstawione powyżej, w tym miejscu warto zauważyć, że: autor przedstawił Neural Simulation Pipeline (NSP) – potok przetwarzania, jako nowatorską konfigurację umożliwiającą prowadzenie symulacji neuronowej w środowisku chmurowym. Jest to ważny wkład praktyczny, ponieważ przypuszczam, że wielu naukowców zajmujących się neuroinformatyką może mieć duże problemy, jak to zrobić. Ponadto mamy praktyczną propozycję taniego klastra obliczeniowego o nazwie Neural Simulations Cluster (NSC) zbudowanego z komputerów Raspberry Pi i ROCKPro64, który może obsługiwać różne symulacje, gdy dostępny budżet na sprzęt jest ograniczony, a niektóre z tych symulacji można wykonać nawet w warunkach domowych. Ale prezentowana praca nie jest pozbawiona wad.

Słabe strony: niektóre szczególne i ogólne słabe strony rozprawy zostały już wskazane powyżej, ale w tym miejscu chciałbym zwrócić uwagę na inne kwestie:

1. W rozdziale 1.3, strona 3, autor napisał: *“It is built of 100 billion interconnected neural cells forming an extremely large and complex network.”* Czy jest jakiś dokument wyjaśniający tę złożoność? Jak możemy modelować tę złożoność, biorąc na przykład dziedzinę czasu i kontekst przestrzenno/topologiczny? Co tak naprawdę oznacza, że mózg jest złożoną strukturą? Jak autor rozumie pojęcie złożoności? Ponadto na stronie 58, koniec rozdz. 3.3 mamy widok, w którym autor napisał: *“... different researchers might develop different parts of that large computational model of a reference brain. (...) If such a reference brain can be simulated in reasonable time, the next steps could be to simulate the whole, thinking brain...”*. Moim zdaniem, ale nie tylko, jest to podejście redukcjonistyczne, które niekoniecznie jest odpowiednie dla jakichkolwiek złożonych systemów.





2. W rozdziale 1.6.2, strona 9, autor napisał: *There is a consensus that information processing specific to the neocortex is performed by neuronal microcircuits.* W jakim stopniu jest to podobne do naszych wyobrażeń i istniejących teorii dotyczących przetwarzania informacji, zwłaszcza gdy jest ono cyfrowe, a nie analogowe?

3. Na stronie 4 mamy: *“These large computers, often need to also consume a lot of electricity, ...”* – powiedzmy, że zużywają energię elektryczną.

4. Na stronie 17 jest: *9 × 106 neurons with 18 × 109 synapses and 2.2 × 106 neurons with 45 × 109 synapses*, ale przypuszczam, że powinno być: *9 × 106 neurons with 18 × 109 synapses and 2.2 × 106 neurons with 45 × 109 synapses*, jak podano w streszczeniu pracy [44].

5. Na stronie 29 mamy akapit, w którym znajduje się opis teorii LSM i jej mocy obliczeniowej równej maszynom Turinga: LSM może symulować dowolną maszynę Turinga, ale co ze stwierdzeniem odwrotnym?

6. W podrozdziale 2.6 na stronie 34 mamy: *“... the input info function $Y^N(t)$ ”*, ale w równaniu. (2.1) i (2.2) istnieje $y^N(t)$. Domyślam się, że jest literówka.

7. Rysunek 2.2 a) i b) na stronie 37 pokazuje niektóre wyniki zaczerpnięte z [151], ale jak należy je interpretować?

8. W rozdz. 2.11 w punkcie 2 (c) mamy wzmiankę o bifurkacjach podczas użycia algorytmów ML. Czy autor spotkał się z czymś takim podczas przygotowywania swojej rozprawy?

9. Jakie mogłyby być wyniki symulacji, gdyby cyfra „0” podana na rys. 3.1 nie była wyrażona liniami prostymi? Na przykład rogi cyfry 0 zostałyby zaokrąglone.

10. Gdy w rozdziale 4 podawane są wyniki eksperymentów z uczeniem maszynowym, brakuje wyjaśnień na temat różnych parametrów, które są wykorzystywane w algorytmach ML. Na przykład na stronie 113 mamy: *“... using a number of pre-selected parameters”*. To jest zbyt enigmatyczne i dobrze byłoby omówić więcej szczegółów na ten temat.

11. Dlaczego wybrano chmurę AWS?

Błędy techniczne i językowe:

1. Nawet szybkie porównanie zdjęć rozwiązań sprzętowych z kilkoma rysunkami przedstawiającymi ważne efekty pracy nie jest przekonujące. Jakość wielu rysunków z ważnymi wynikami jest bardzo niska, np. Rys. 4.12 a), 4.15 a), 4.21 a).

2. Wiele rysunków i tabel zamieszczonych w rozprawie nie ma żadnych odnośników i komentarzy w tekście. Część już wymieniono powyżej, ale jest ich więcej, na przykład Rys. 3.5, Rys. 3.6. Jeśli w tekście nie ma odnośników, to dlaczego rysunki są zamieszczone?





3. W wielu miejscach tekstu autor pisze w pierwszej osobie. Na przykład: strona 3: *I assess*, strona 5: *I analyse, I explore, I run, I design, I evaluate, I attempt*, i tak dalej. Nie jest to typowe dla tego typu rozpraw i zazwyczaj oczekuje się ich pisania w bezosobowej formie.

4. Są też błędy techniczne związane z brakującymi spacją, błędnymi jednostkami miary itp., ale tutaj pominię je jako nieistotne.

Ogólnie praca jest dobrze przygotowana i nie ma wielu błędów, które znacząco wpływają na ogólny obraz pracy.

Opinia, czy rozprawa doktorska wykazuje ogólną wiedzę teoretyczną

Biorąc pod uwagę dotychczas przedstawioną argumentację, treść pracy oraz jej ogólną strukturę należy stwierdzić, że autor wykazał się niezbędnym zasobem ogólnej wiedzy teoretycznej wraz z wieloma umiejętnościami praktycznymi typowymi dla informatyki. Najważniejsze z tych stwierdzeń wynika z faktu, że praca prezentuje określony język, umiejętności, wiedzę, które należą do więcej niż dwóch dyscyplin, w których mamy związek z analizą zachowania mózgu i próbami jego komputerowego modelowania. Wiele istotnych szczegółów i rozważań przedstawionych w rozdziałach 1-3 wystarczy, aby stwierdzić, że autor ma bardzo dobre przygotowanie teoretyczne do prac naukowych związanych z neuroinformatyką. Większość z fragmentów została poparta przeglądem literatury, szczegółowymi wyjaśnieniami, a kilka wątpliwości przedstawionych w powyższej recenzji może być łatwo wyjaśnionych przez autora.

Opinia o tym, czy rozprawa doktorska wykazuje zdolność do samodzielnego prowadzenia pracy naukowej

W rozprawie autor przedstawił dwa rozdziały (4 i 5), które koncentrują się głównie na jego samodzielnej i osobistej pracy. Na szczególną uwagę zasługuje rozdział 5, w którym znajdują się autorskie propozycje dotyczące zarówno sprzętowego, jak i programowego rozwiązywania. Pomimo tego, że moim zdaniem autor miał pewne problemy z lepszą prezentacją wyników pracy, jestem przekonany, że nie ma wątpliwości, iż autor posiada pełne zdolności do prowadzenia pracy naukowej. A dokładniej: jestem przekonany, że autor włożył dużo pracy w przedstawienie końcowych wyników, ale mógłby poświęcić więcej uwagi lepszej prezentacji wyników. Wskazane powyżej wady są drobnymi uwagami, które można wyjaśnić udzielając odpowiedzi lub w trakcie obrony doktoratu.





Opinia o tym, czy rozprawa doktorska jest oryginalnym rozwiązaniem problemu naukowego, oryginalnym rozwiązaniem w zastosowaniu wyników własnych badań w sferze gospodarczej lub społecznej

Jak napisano powyżej i biorąc pod uwagę zaskakujący fakt, że społeczność neurobiologów nie używała często (patrz strona 123) żadnych jednopłytkowych klastrów komputerowych (SBCC) skupionych na symulowaniu modeli cybernetycznych; ani w obszarze badań, ani nawet w dydaktyce neuroinformatyki, inżynierii biomedycznej czy neuroinformatyki obliczeniowej, conajmniej zaproponowane w pracy oryginalne rozwiązanie potoku programowo-sprzętowego jest oryginalne dla istniejących problemów naukowych. Jest nie tylko przedstawione jako koncept teoretyczny, ale jest zrealizowane w praktyce z niezbędną wiedzą inżynierską i umiejętnościami praktycznymi.

Wniosek końcowy

Po analizie przedłożonej rozprawy doktorskiej Pana mgra Karola Chlasty stwierdzam, że wnosi ona wkład w rozwój dyscypliny informatyka techniczna i telekomunikacja. Stanowi ona oryginalne rozwiązanie problemu naukowego, potwierdzając wiedzę, bardzo dobre przygotowanie merytoryczne, umiejętności praktyczne oraz umiejętność samodzielnego prowadzenia badań naukowych przez Doktoranta. Podjęta tematyka jest aktualna i ma bardzo duży potencjał aplikacyjny. Recenzowana rozprawa doktorska spełnia ustawowe wymagania stawiane kandydatom do stopnia doktora, określone ustawą o stopniach i tytułach naukowych – uwzględnione w art. 13 ust. 1 Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki w związku z art. 179 Ustawy z dnia 3 lipca 2018 r. Przepisy wprowadzające ustawę - Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r., poz. 1669).

Na tej podstawie wnioskuję do Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Polsko-Japońskiej Akademii Technik Komputerowych w Warszawie o dopuszczenie jej do publicznej obrony.

Dominiła Stręcha

