

Warszawa, dn. 8 kwietnia 2019 r.

dr hab. inż. Krzysztof Trojanowski, prof. UKSW  
Wydział Matematyczno-Przyrodniczy. Szkoła Nauk Ścisłych  
Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie  
ul. Wóycickiego 1/3, 01-938 Warszawa

Sz. Pan  
dr hab. inż. Zbigniew Ranachowski, prof. IPPT PAN  
Sekretarz Rady Naukowej  
Instytutu Podstawowych Problemów Techniki PAN  
ul. Pawińskiego 5B, 02-106 Warszawa

Szanowny Panie Profesorze,

przesyłam recenzję rozprawy doktorskiej mgr Oskara Wyszyńskiego pt. „Evolutionary Algorithm for Particle Trajectories Reconstruction”, której wykonanie powierzyła mi Rada Naukowa Instytutu Podstawowych Problemów Techniki PAN decyzją podjętą na posiedzeniu w dniu 31 stycznia 2019 r. Oceniam przedłożoną rozprawę doktorską jako spełniającą wymagania stawiane rozprawom doktorskim przez obowiązującą ustawę o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 r. (Dz. U. Nr. 65 poz. 595 z późniejszymi zmianami) z późniejszymi zmianami w brzmieniu z dnia 15 września 2017 r. (Dz. U. 2017 r. poz. 1789.), zgodnie z Art. 175. 1. Ustawy z dnia 3 lipca 2018 r. Przepisy wprowadzające ustawę – Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 r. poz. 1669). Wnioskuje o dopuszczenie autora do dalszych etapów postępowania o nadanie tytułu doktora. Szczegółowa opinia zawarta jest w recenzji.

Łączę wyrazy szacunku,



Krzysztof Trojanowski

Warszawa, 8 kwietnia 2019 r.

dr hab. inż. Krzysztof Trojanowski, prof. UKSW  
Wydział Matematyczno-Przyrodniczy. Szkoła Nauk Ścisłych  
Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie  
ul. Wóycickiego 1/3, 01-938 Warszawa

Recenzja pracy doktorskiej mgr Oskara Wyszyńskiego pt.  
„Evolutionary Algorithm for Particle Trajectories Reconstruction”

Recenzowana praca dotyczy zastosowań heurystycznych metod optymalizacji parametrycznej. Autor został postawiony przed zadaniem rekonstrukcji trajektorii cząstek elementarnych, poruszających się w polu magnetycznym, a dokładniej znalezienia parametrów modelu trajektorii cząstki. Parametry modelu definiują przestrzeń rozwiązań, w której pracuje proponowana metoda optymalizacyjna, tj. strategie ewolucyjne. Ponadto strategie ewolucyjne wykorzystane zostały do identyfikacji produktów reakcji jądrowych, powstałych w reakcjach ciężkich jonów przy pośrednich energiach i zarejestrowanych w wielodetektorowym systemie. Celem pracy było więc opracowanie dwóch algorytmów optymalizacyjnych, w obydwu przypadkach wywodzących się z dziedziny strategii ewolucyjnych, a postawiona hipoteza badawcza stanowi, że proponowane algorytmy są w stanie znajdować właściwe wartości parametrów poszukiwanych modeli oraz w pełni automatycznie identyfikować produkty reakcji jądrowych. Realizacja celu, jaki postawił sobie autor, prowadzi do uzasadnienia prawdziwości hipotezy badawczej.

Rozprawa doktorska mgr Oskara Wyszyńskiego jest napisana w języku angielskim i liczy 179 stron. Na rozprawę składa się wstęp, 9 rozdziałów, podsumowanie i bibliografia. Bibliografia zawiera 136 pozycji, z czego 4 jest współautorstwa doktoranta. Struktura pracy jest dość regularna, właściwa dla tekstów technicznych.

#### **Konstrukcja i treść rozprawy**

We wstępie (numerowanym jako rozdział pierwszy) autor zaprezentował problem badawczy, cel, tezę i zakres prac, oraz omówienie układu całej pracy. W pierwszej części wstępu znajdujemy krótki opis instrumentów badawczych fizyki wysokich energii oraz podstawowych zasad prowadzenia obserwacji i badań z wykorzystaniem tych instrumentów. Podrozdział pierwszy prezentuje skrótowo aktualny stan wiedzy dotyczącej metod rekonstrukcji trajektorii cząstek stosowanych w fizyce cząstek elementarnych oraz metod identyfikacji jąder atomowych stosowanych w fizyce jądrowej. Podrozdział drugi prezentuje tezy badawcze pracy, a trzeci – metodologie rekonstrukcji trajektorii cząstek oraz identyfikacji jąder atomowych, w których zastosowanie znalazły algorytmy strategii ewolucyjnych. Ostatni podrozdział zawiera omówienie układu pracy.

Następne rozdziały prezentują infrastrukturę obiektu NA61/SHINE Experiment, co pozwala czytelnikowi lepiej zrozumieć charakterystykę i rozmiary gromadzonych danych oraz wyzwania, jakim trzeba sprostać podczas ich analizy. W rozdziale drugim naszkicowana została budowa detektora, w tym komór projekcji czasu, oraz ogólna zasada dokonywania pomiarów i dalszego przetwarzania zebranych danych. Rozdział trzeci krótko prezentuje stosowane obecnie różne standardy elektroniki modularnej, na której oparta jest budowa części pomiarowej detektora: NIM, CAMAC, FASTBUS,





VMEBus i ATCA. Rozdział czwarty zawiera opis systemu akwizycji danych, którego zadaniem jest odczyt sygnałów analogowych, ich konwersja do postaci cyfrowej i zapis na nośnikach danych pod kontrolą serwera DAQ (Data AcQuisition). W piątym rozdziale omówione zostało działanie systemu filtrowania rejestrowanych danych (tzw. trigger system). Filtrowanie ma znaczenie, ponieważ rejestracja wszystkich danych pomiarowych wymaga wysokiej przepustowości systemu rejestracji, a co ważniejsze, nie wszystkie rejestrowane zdarzenia są jednakowo istotne dla analizy wykonanych pomiarów. Ogólny zarys nowej wersji oprogramowania Shine Offline Framework znajdujemy w rozdziale szóstym. Poprzednia wersja bazująca na oprogramowaniu dostarczonej przez Pierre Auger Observatory zawierała ograniczenia, które utrudniały jej dalszy rozwój. Nowe oprogramowanie, napisane w całości w C++, zachowuje funkcjonalność poprzedniej wersji, dając możliwości rozwoju, który został opisany w następnych rozdziałach pracy. Warto podkreślić, że autor pracował w zespole, którego celem było stworzenie tego oprogramowania. Rozdział szósty jest jednocześnie ostatnim z rozdziałów charakteryzujących środowisko, w którym muszą działać metody zaproponowane przez autora pracy.

W rozdziale siódmym autor szkicuje ogólne zasady działania strategii ewolucyjnych ze szczególnym uwzględnieniem CMA-ES. Po dość pobieżnym opisie strategii ewolucyjnych, kolejne podrozdziały prezentują podstawowe kroki pętli głównej algorytmu: próbkowanie (podrozdział 7.1), oraz aktualizację podstawowych parametrów: średniej  $m$  (podrozdział 7.2), macierzy kowariancji  $C$  (podrozdział 7.3) oraz adaptacji długości kroku  $\sigma$  (podrozdział 7.4). Trzeba jednak powiedzieć, że nie jest to szeroki przegląd istniejących podejść, ale omówienie jednego, wybranego z podstawowych. Podrozdział 7.5 traktuje o złożoności obliczeniowej poszczególnych kroków pętli głównej CMA-ES.

Rozdział ósmy zawiera opis metody rekonstrukcji trajektorii cząstek. W pierwszym podrozdziale znajdujemy ogólne wprowadzenie do metody, która składa się z dwóch części: ustalenia trajektorii cząstki i określenia jej punktu początkowego. Drugi podrozdział wprowadza pojęcie operatora ekstrapolacji, który na podstawie początkowej reprezentacji trajektorii i kolejnych wartości obserwowanego położenia z detektora aktualizuje reprezentację trajektorii. W drugim podrozdziale opisany jest ewolucyjny tracker – tropiciel trajektorii cząstek. Składa się on z trzech komponentów: pierwszy selekcjonuje zaobserwowane pomiary odrzucając szum, drugi tworzy początkowe postaci trajektorii, a trzeci optymalizuje parametry trajektorii z wykorzystaniem algorytmu CMA-ES. Komponenty te zostały opisane kolejno w podrozdziałach 8.3.1, 8.3.2 i 8.3.3. Ostatni podrozdział 8.3.4 omawia złożoność obliczeniową metody rekonstrukcji trajektorii cząstek.

Weryfikacja skuteczności proponowanej metody rekonstrukcji trajektorii cząstek została zaprezentowana w rozdziale dziewiątym. Testy zostały przeprowadzone na zbiorze danych testowych spreparowanych za pomocą zestawu symulacyjnego GEANT oraz wzbogaconych o szum pochodzący z prawdziwych detektorów. Ocena polegała na porównaniu zidentyfikowanych trajektorii z trajektoriami wykorzystanymi przez zestaw symulacyjny. Rozdział kończy sugestie zmian mogących prowadzić do dalszego poprawienia skuteczności metody.

W rozdziale dziesiątym zostało zaprezentowane zastosowanie algorytmu CMA-ES do identyfikacji produktów reakcji jądrowych. Pomiary zmiany energii przedstawiane są w postaci charakterystycznych wykresów składających z linii, które mogą być mo-



delowane za pomocą funkcji Tassana-Gota. Na podstawie tych wykresów dokonuje się klasyfikacji, co w dotychczas stosowanych podejściach wymagało interakcji z użytkownikiem. Proponowana przez autora metoda jest natomiast w pełni automatyczna, jej zadaniem jest optymalizacja parametrów modelu w celu dopasowania go do danych. Metodę również przetestowano na zbiorze danych testowych, w których dane pochodziły z modelu i były dodatkowo wzbogacone o szum Gaussowski, a celem było poprawne zidentyfikowanie parametrów tego modelu. Metoda rekonstruowała parametry z wysoką dokładnością, co przekładało się na wysoką trafność klasyfikacji.

Podsumowanie zawiera listę zrealizowanych celów pracy, wnioski końcowe oraz kierunki dalszych badań.

### **Ocena merytoryczna**

Przedłożona rozprawa prezentuje rozwiązanie problemów praktycznych dotyczących badań trajektorii cząstek prowadzonych w Europejskim Ośrodku Badań Jądrowych w Genewie (CERN). Ze względu na bardzo szczególne i niezbyt szerokie grono adresatów rozwiązania, zakres pracy jest zdefiniowany w pierwszym rzędzie jej funkcjonalną przydatnością dla badań prowadzonych w CERN. Autor wykazuje się bardzo dużą wiedzą dotyczącą natury i pochodzenia danych, które będą analizowane. Sam jednak zaznacza, że rozdziały, w których jest to opisane, nie zawierają informacji niezbędnych dla potwierdzenia prawdziwości tez postawionych w pracy. Rozdział opisujący sposób działania CMA-ES jest krótki, pobieżnie przedstawia ogólne zasady działania strategii ewolucyjnych i szybko przechodzi do opisu metody opracowanej przez Hansena i jego kolegów. W całym rozdziale znajdujemy odwołania do dziewięciu pozycji bibliograficznych, prawie wszystkich współautorstwa Nikolausa Hansena i dotyczących CMA-ES. Inne publikacje, dotyczące szeroko pojętej optymalizacji heurystycznej, występują również w bibliografii. Odwołania do nich znajdują się głównie w pierwszym rozdziale pracy, tj. we Wstępie.

Można więc powiedzieć, że autor opiera się na aktualnym stanie wiedzy z dziedziny, której dotyczą tezy jego pracy. Wnioski wyciągnięte z analizy źródeł są wykorzystane z pożytkiem dla celów, które autor sobie postawił.

Autor w swojej pracy jasno przedstawił tezę oraz wymienił cele, których osiągnięcie zaprowadziło go do wykazania prawdziwości postawionej tezy. Na tej podstawie jako samodzielny i oryginalny dorobek autora należy uznać:

1. metodę rekonstrukcji trajektorii cząstek, w której identyfikacja i optymalizacja parametrów trajektorii odbywa się z wykorzystaniem algorytmu CMA-ES,
2. metodę identyfikacji produktów reakcji jądrowych, poprzez określenie na podstawie danych pomiarowych parametrów modelu pozwalającego na identyfikację,
3. implementację i wdrożenie metody rekonstrukcji trajektorii cząstek w nowej wersji oprogramowania zarządzającego danymi zebranymi w ramach eksperymentu NA61/SHINE,
4. badania symulacyjne weryfikujące poprawność metod.

Struktura rozprawy jest poprawna, choć widać wyraźnie jak podporządkowana jest raczej opisowi eksperymentu NA61/SHINE, niż zakresowi zadeklarowanemu w tytule pracy. Cel pracy został dobrze sformułowany, a prawdziwość tezy została potwierdzona wynikami symulacji. Trzeba jednak powiedzieć, że fragment pracy opisujący metody



strategii ewolucyjnych i ich adaptacji w celu rozwiązania postawionego zadania jest niezbyt obszerny objętościowo, a informacje dotyczące natury problemu postawionego przed metodą ewolucyjną i struktury danych podawane są dość ogólnie. Znacznie więcej miejsca autor poświęca opisowi eksperymentu NA61/SHINE podając wiele szczegółów technicznych, chociaż we wstępie jasno deklaruje, że praca dotyczy algorytmów ewolucyjnych i w tym zakresie chciałby być oceniany.

Skupiając się na zakresie określonym w tytule pracy można wskazać kilka miejsc, gdzie tekst należałoby rozwinąć lub uszczegółwić.

1. Brakuje jawnego opisu struktury danych rozwiązania  $\Theta$  oraz dziedziny potencjalnych rozwiązań (choć "nie wprost" dowiadujemy się o tym ze zdań w górnej połowie strony 119 oraz wcześniejszych opisów w rozdziale 8).
2. Z opisu funkcji oceny rozwiązania podanego na stronie 127 nie widać, w jaki sposób otrzymywana jest wartość skalarna określająca jakość rozwiązania.
3. Ze względu na brak w podrozdziale 8.3.3 szczegółów działania kolejnych kroków wykonywanych w pętli głównej CMA-ES należy się domyślać, że opis podstawowej wersji CMA-ES z rozdziału 7 to jednocześnie wariant wybrany do późniejszej implementacji.
4. W podrozdziale 8.3.3 dowiadujemy się, że do obliczeń zaimplementowany został Active-CMA-ES (aCMA-ES), który jest rozszerzeniem standardowego podejścia opisanego w rozdziale 7, jednak w tekście pracy nie ma szczegółowego opisu podejścia aCMA-ES.
5. Działanie operatora  $\mathcal{Q}(\cdot)$  występującego we wzorze (8.18) na str. 127 mają objaśniać wzory (8.3) i (8.4). Z opisu nad wzorem (8.4) dowiadujemy się, że wynikiem  $\mathcal{Q}$  dla danych parametrów trajektorii  $\Theta$  oraz położenia  $z$  jest nowy wektor parametrów trajektorii. Brakuje klarownej informacji, na czym polega działanie  $\mathcal{Q}$ . Uwzględniając informację ze strony 119 o tym, że wektor parametrów trajektorii  $\Theta = p \oplus o$ , można przypuszczać, że działanie  $\mathcal{Q}$  polega na konkatenacji kolejnego pomiaru położenia  $z$  na końcu wektora.
6. Na stronie 127 pod wzorem (8.18) znajdujemy stwierdzenie, że operator  $\mathcal{S}(\cdot)$  transformuje klastę (*cluster*) lub wektor parametrów trajektorii na wektor w przestrzeni Euklidesowej (co brzmi dość enigmatycznie) wraz z odwołaniem do wzoru (8.6). Wzór (8.6) wygląda następująco:  $S(\xi) = (\xi_x, \xi_y, \xi_z)$  i jest objaśniony wyrażeniem  $\mathcal{S} : T \cup C \rightarrow \mathbb{R}^3$ , gdzie wg opisu pod wzorem (8.6)  $C$  to wektor parametrów klastra, a  $T$  to wektor parametrów trajektorii. Operator  $\mathcal{S}(\cdot)$  występuje już wcześniej, na stronie 118 we wzorze (8.5) gdzie wg opisu pod wzorem (8.5) jego argumentem jest wektor  $\mathbf{c} = (x, y, z, s_1, \dots, s_n)$  będący konkatenacją wektora reprezentującego środek ciężkości klastra wraz z sygnałami depozycji ładunku w klastrze. Zbierając razem te informacje czytelnik pozostaje z poczuciem, że nie dostał klarownej odpowiedzi na pytanie, co może być argumentem operatora  $\mathcal{S}(\cdot)$ , na czym polega jego działanie i co reprezentuje wynik jego działania.
7. W zdaniu na str. 117 "Predykcja polega na ekstrapolacji wektora parametrów trajektorii za pomocą równania (8.3) na pozycję  $Z$  następnego pomiaru" autor powołuje się w rzeczywistości nie na równanie, ale na wyrażenie definiujące dziedzinę i przeciwdziedzinę funkcji  $\mathcal{Q}$  co nie objaśnia działania ekstrapolacji.
8. Autorowi zdarza się złamanie zasady, aby symbol w całym tekście reprezentował jedno i tylko jedno znaczenie, np.  $\Theta$  w podrozdziale 7.5 oznacza złożoność oblicze-

niową, a w rozdziale 8 wektor parametrów trajektorii cząstki, przy czym jest on czasem reprezentowany przez  $\Theta$ , a czasem przez  $\Theta$  (pogrubiona czcionka).

9. Na str. 22 autor wyraża opinię, że z twierdzenia "No Free Lunch" może wynikać niższa wydajność [algorytmów] przy rozwiązywaniu problemów gdzie można wykorzystać wiedzę *a priori*. Zdanie nie jest jasne (niższa w porównaniu z czym?).
10. Zdania wprowadzające i objaśniające pojęcia klastra i trajektorii znajdują się w podrozdziale 6.5 (trzeci akapit) zatytułowanym "Walidacja numeryczna" rozdziału "Shine Offline Framework". Sądząc po tytułach rozdziałów, nie jest to miejsce, gdzie należałoby ich szukać. Tym bardziej, że w tekście wprowadzającym do tego rozdziału znajdujemy zdanie: "Comprehension of this dissertation does not depend on this content".

Słabą stroną rozprawy i jednym z podstawowych utrudnień w czytaniu tekstu było podporządkowanie jej struktury opisowi eksperymentu NA61/SHINE, skutkujące rozproszeniem informacji istotnych dla zrozumienia meritum, czyli sposobu wykorzystania CMA-ES, co jest widoczne w uwagach powyżej. Drugą, może ważniejszą wadą rozprawy jest brak porównań uzyskanych wyników metody identyfikowania trajektorii cząstek z wynikami innych metod (podrozdział 9.2.1 "Reconstruction performance"). Jestem przekonany, że mimo niszowego zastosowania takie alternatywne metody musiały wcześniej istnieć. W podrozdziale 9.2.1 brakuje również opisu metodologii prowadzenia pomiarów wartości zaprezentowanych w tabeli 9.1, czy choćby krótkiego objaśnienia, jak zostały uzyskane te wartości. Podobnie jest w przypadku wyników metody identyfikacji produktów reakcji jądrowych (podrozdział 10.3), gdzie również brak porównań z innymi metodami.

W tekście można też znaleźć różne drobne usterki oraz niespójności na poziomie edytorskim:

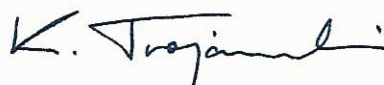
1. na str. 22, niespójność narracji "during **my** participation in Doctoral Student Program at CERN **the author of this dissertation** contributed",
2. str. 36, w pierwszym zdaniu na górze strony słowo „chapters” jest powtórzone dwa razy,
3. str. 83, literówka w „spectific” – powinno być „specific”,
4. str. 114, 2 linijka od góry: "Clustering *simply put*, is a process" – czy słowa "simply put," są zamierzone?
5. str. 113, 7 linijka od dołu: zamiast „form” chyba raczej „from”,
6. str. 116, 3 linijka podrozdziału 8.3: podmiot zdania powinien być w liczbie mnogiej jeżeli zdanie mówi o porządkowaniu elementów ("... are ordered according to an elapsing parameter").

Poziom językowy tekst miejscami pozostawia do życzenia. Zdarzają się dziwne zdania, np. str. 118: "The aim of the competition is an efficient method for distributing valid clusters between tracks and noise" (uważam, że celem rywalizacji raczej nie jest metoda, ale wysoki wynik poprawnej identyfikacji klastrów, które należą do trajektorii, oraz pozostałych, które stanowią szum). W mojej ocenie dobrze by pracy zrobiło, gdyby została napisana po polsku. Być może autor szybciej zobaczyłby niekorzystny — w kontekście tytułu pracy — sposób podania informacji i przemyślał jej organizację jeszcze raz. Zwłaszcza, że czasu było dość, praca bowiem jest datowana na 2016 rok (informacja na dole strony tytułowej), czyli tekst w swojej obecnej postaci był gotów od ponad dwóch lat.



### **Konkluzja końcowa**

Podsumowując stwierdzam, że tematyka rozprawy jest aktualna i ważna, rozprawa prezentuje oryginalne rozwiązanie problemu identyfikowania trajektorii cząstek stworzone w oparciu o metodę strategii ewolucyjnych CMA-ES i wdrożone w eksperymencie NA61/SHINE na akceleratorze SPS w Europejskim Ośrodku Badań Jądrowych w Genewie (CERN) oraz problemu identyfikacji produktów reakcji jądrowych stworzone również w oparciu o metodę strategii ewolucyjnych CMA-ES. Uważam, że rozprawa doktorska mgr Oskara Wyszyńskiego spełnia wymagania stawiane rozprawom doktorskim przez obowiązującą ustawę o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 r. (Dz. U. Nr. 65 poz. 595 z późniejszymi zmianami) z późniejszymi zmianami w brzmieniu z dnia 15 września 2017 r. (Dz. U. 2017 r. poz. 1789.), zgodnie z Art. 175. 1. Ustawy z dnia 3 lipca 2018 r. Przepisy wprowadzające ustawę — Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 r. poz. 1669). Wnioskuje o dopuszczenie autora do dalszych etapów postępowania o nadanie tytułu doktora.



Krzysztof Trojanowski