

Prof. dr hab. Daniel Krzysztof Wójcik

Warszawa, 4.1.2023

Instytut Biologii Doświadczalnej

im. Marcelego Nenckiego PAN

02-093 Warszawa, ul. Pasteura 3

Recenzja rozprawy doktorskiej mgr Judyty Jabłońskiej pt.: „Analiza modeli uczenia się ze wzmocnieniem związanym z działaniem nagród naturalnych i substancji uzależniających w zadaniu z probabilistycznym odwróceniem” wykonanej w Instytucie Farmakologii im. Jerzego Maja Polskiej Akademii Nauk, pod kierunkiem dr. hab. Jana Rodriguezza Parkitny z Instytutu Farmakologii im. Jerzego Maja Polskiej Akademii Nauk, oraz prof. dr. hab. Piotra Zielińskiego z Instytutu Fizyki Jądrowej im. Henryka Niewodniczańskiego Polskiej Akademii Nauk.

Rozprawa mgr Judyty Jabłońskiej dotyczy badania procesów podejmowania decyzji z wykorzystaniem inteligentnych klatek IntelliCage oraz modeli obliczeniowych. Celem badań doktorantki było opracowanie testu behawioralnego pozwalającego na obserwację uczenia ze wzmocnieniem w warunkach zbliżonych do naturalnych oraz sprawdzenie, czy obserwowane zachowanie można opisać przy użyciu modeli matematycznych. Swoją metodę autorka oparła o test probabilistycznego uczenia z odwróceniem, który zaadoptowała do klatki IntelliCage, dzięki czemu przebiegał on w naturalnym cyklu okołodobowym, bez deprywacji pokarmu i przy minimalnej interwencji badacza. Opracowaną metodologię wykorzystano do porównania wzmacniającego działania nagrody naturalnej (sacharyna) oraz substancji uzależniającej (etanol) oraz do zbadania wpływu selektywnych mutacji upośledzających działanie układu serotoninowego na wykonanie testu.

Rozprawa ma klasyczny układ wprowadzenia, metod, wyników i dyskusji. Rozprawa ma ponad 100 stron, jest bogato ilustrowana, wyniki są przedstawione graficznie oraz w czytelnych tabelach. Literatura jest dobrana adekwatnie do potrzeb pracy. Praca jest zredagowana i wydrukowana bardzo estetycznie, co w połączeniu z dobrą ekspozycją przedmiotu badań sprawia, że czyta się ją z przyjemnością.

We wprowadzeniu doktorantka omawia tło swojej pracy: zaczyna od omówienia zjawiska uczenia ze wzmocnieniem na poziomie biologicznym w szerszym kontekście badania mechanizmów uczenia się zaczynając od warunkowania, a następnie dyskutuje mechanizmy nagradzającego działania sacharyny i etanolu oraz zaangażowane szlaki komunikacyjne w mózgu. Autorka wskazuje, że pomimo

angażowania układu dopaminergicznego w brzuszny mózg zarówno przy przetwarzaniu informacji o etanolu jak i sacharozie, istnieją znaczące różnice w przetwarzaniu informacji o tych substancjach, co uzasadnia jej badania zachowania myszy w odpowiedzi na te dwa różne bodźce. W rozdziale 2.3 doktorantka omawia wykorzystanie testu uczenia z odwróceniem w badaniach funkcji wykonawczych i elastyczności poznawczej, następnie obszernie omawia neuronalne podłoże uczenia ze wzmocnieniem, ze szczególnym uwzględnieniem roli dopaminy i serotoniny, oraz odpowiednich komórek i receptorów. Na zakończenie autorka omawia modelowanie obliczeniowe uczenia ze wzmocnieniem. Przedstawia tu zarówno modele znane wcześniej jak i warianty, które sama proponuje (model „podwójnie mieszany”).

Po krótkim przedstawieniu celów autorka przedstawia metody doświadczalne i obliczeniowe, które wykorzystała w pracy. Najpierw omówiono szczepy, na których wykonano doświadczenia, poza szczepem C57BL/6J były to myszy transgeniczne $NR1^{Tph2CreERT2}$, które nie posiadają funkcjonalnych receptorów NMDA na komórkach serotoninowych, oraz genetycznie zmodyfikowane myszy $Htr7^{tm1Sut}$, które pozbawione są receptorów serotoninowych 5-HT₇. Następnie omówiono wykonane doświadczenia. Większość doświadczeń (11 kohort po ok. 14 zwierząt każda) stanowiły testy probabilistycznego odwrócenia w klatce IntelliCage, ale autorka przeprowadziła też testy probabilistycznego odwrócenia w klatce do warunkowania instrumentalnego oraz test preferencji sacharyny. W klatkach IntelliCage autorka badała wzmocnienie uczenia w sytuacji gdy nagrodą był etanol lub sacharyna, oddzielnie lub razem, oraz woda jako kontrola. Następnie doktorantka zbadała uczenie ze wzmocnieniem obu szczepów modyfikowanych genetycznie w doświadczeniach, gdzie nagrodą była sacharyna a kontrolą woda. W tych doświadczeniach brały udział grupy mieszane, połowa zwierząt to były mutanty, połowa szczepu dzikiego. W pozostałych doświadczeniach (klatki Skinnera, test preferencji sacharyny) autorka wykorzystała samce myszy transgenicznych $NR1^{Tph2CreERT2}$; w klatkach IntelliCage używano wyłącznie samic. Na zakończenie rozdziału czwartego autorka opisuje wykorzystane techniki statystyczne, zarówno uogólnione modele liniowe jak i dopasowywanie modeli uczenia ze wzmocnieniem poprzez maksymalizację wiarygodności modelu.

W rozdziale 5 autorka przedstawiła otrzymane wyniki. Doświadczenia pogrupowano w trzy naturalne grupy: badania uczenia ze wzmocnieniem w zależności od różnych nagród w klatkach IntelliCage, zbadanie wpływu dysfunkcji wybranych receptorów na uczenie ze wzmocnieniem w klatkach IntelliCage, porównanie z wynikami w klasycznych klatkach do warunkowania instrumentalnego. W rozdziale 5.1 autorka analizowała dane zebrane uprzednio przez kolegów ze swojego zespołu. Doktorantka najpierw wykazała, że użyte substancje (etanol, sacharyna, kombinacja obu) rzeczywiście stanowią nagrodę dla zwierząt typu dzikiego. Następnie pokazała, że liczba prób

uzyskania nagrody utrzymywała się na stałym poziomie (Rycina 5-2) po czym wykorzystwała uogólnione modele liniowe w celu wyjaśnienia obserwowanych częstości wyborów otrzymania dostępu do nagrody w poprzedniej próbie (efekt nagrody), indywidualnej preferencji lokalizacji narożnika (efekt miejsca) i czasu decyzji (efekt czasu). Autorka wykazała znaczenie nagrody i czasu od poprzedniej nagrody, który elegancko zilustrowała na rycinie 5-4. Następnie dopasowała kilka popularnych i mniej znanych modeli obliczeniowych uczenia ze wzmocnieniem, w tym dwa modele, gdzie niektóre parametry zmieniają się z czasem. Otrzymane wyniki autorka zanalizowała na kilka sposobów skupiając się na identyfikacji modeli, które najlepiej reprezentują zebrane dane, a jednocześnie najlepiej różnicują badane grupy, a w konsekwencji odpowiedź na różne nagrody (tabele 5-2 i 5-3).

Rozdział 5.2 realizuje ten sam program analityczny co rozdział 5.1, tym razem w celu zbadania zależności uczenia ze wzmocnieniem od wybranych receptorów. Wykorzystując wyłącznie sacharozę porównano uczenie ze wzmocnieniem myszy transgeniczných NR1^{Tph2CreERT2} oraz myszy Htr7^{tm1Sut}. Wyniki nie odbiegają znacząco od wyników dla typu dzikiego C57BL/6J. W rozdziale 5.3 dla porównania doktorantka przeprowadziła test probabilistycznego odwrócenia w klatkach do warunkowania instrumentalnego tylko dla samców linii NR1^{Tph2CreERT2} z wykorzystaniem sacharyny jako nagrody. Wyniki jakościowo nie odbiegają od wyników z rozdziału 5.2.

W rozdziale 6 autorka podsumowała swoje wyniki i przedyskutowała je na tle literatury, na zakończenie podając główne wnioski. Główne wyniki tej rozprawy to

1. Opracowanie nowego testu behawioralnego uczenia się ze wzmocnieniem opartego o klatki IntelliCage;
2. Opisanie różnic w uczeniu ze wzmocnieniem myszy dla różnych nagród;
3. Zbadanie roli wybranych receptorów związanych z transmisją serotonergiczną w badanych procesach, oraz
4. Systematyczne wykorzystanie modelowania obliczeniowego otrzymanych danych w celu głębszego zrozumienia i interpretacji wyników.

Rozprawę oceniam bardzo wysoko. Uważam, że projekt był ciekawy i ambitny a autorka poradziła sobie z nim bardzo dobrze. Moje szczególne uznanie budzi połączenie pracy doświadczalnej z nietrywialnym modelowaniem danych.

Zgodnie z aktualną ustawą ocena rozprawy doktorskiej powinna zawierać trzy elementy:

- 1) ocenę, czy rozprawa doktorska prezentuje ogólną wiedzę teoretyczną kandydata;
- 2) ocenę, czy rozprawa doktorska wykazuje umiejętność samodzielnego prowadzenia pracy naukowej przez kandydata;
- 3) ocenę, czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego.

Nie mam wątpliwości, że przedstawiona rozprawa mgr Judyty Jabłońskiej spełnia wszystkie te warunki. Wprowadzenie do rozprawy i dyskusja prowadzone są na tle dobrze rozpoznanej i dobranej literatury i jasno wskazują na zrozumienie bieżących zagadnień dotyczących badania zachowania i uczenia ze wzmocnieniem, a także wieloznacznych interpretacyjnych, przez które autorka musiała przebrnąć (jak np. potencjalne efekty społeczne w klatkach IntelliCage). Autorka przedstawiła konkretny problem – wpływu różnych nagród na uczenie ze wzmocnieniem myszy oraz zbadanie roli wybranych receptorów związanych z transmisją serotonergiczną w tych procesach – a następnie rozwiązała go w oryginalny sposób samodzielnie przeprowadzając obszerne doświadczenia i wykonując złożone analizy.

Mojej wysokiej oceny rozprawy nie umniejszają pewne uwagi krytyczne, które z obowiązku recenzenta przedstawiam. Część z nich dotyczy fragmentów, których nie zrozumiałem w rozprawie. Chciałbym, żeby autorka na obronie je wyjaśniła. Pozostałe uwagi mają charakter głównie polemiczny.

Uwagi krytyczne

W rozdziale 4.5 autorka pisze „Testowane modele uczenia się zostały napisane w języku C++”, ale nie pisze przez kogo. **Proszę o wyjaśnienie.** Nie zrozumiałem też **w jaki sposób autorka wykonała bootstrap opisany na końcu rozdziału 4. Jakie dane były mieszane i jak wykonano mieszanie?** Przy uczeniu ze wzmocnieniem sekwencja czasowa danych jest istotna, więc nie można dowolnie danych mieszać.

Nie jestem entuzjastą miary „ułamek poprawnych przewidywań modelu” używanej przez autorkę, która nie jest nigdzie zdefiniowana ani omówiona. Użyte przez nią modele są probabilistyczne. Podejrzewam, że autorka patrzyła dla każdej próby uzyskania nagrody, czy model przewidywał większe od 0.5 prawdopodobieństwo wyboru tego rogu. **Proszę doktorantkę o wyjaśnienie tego na obronie.** Jeżeli dobrze zrozumiałem definicję, to odpowiada ona losowaniu kul o dwóch kolorach (większa i mniejsza szansa nagrody) z różnymi prawdopodobieństwami przed dwie osoby (prawdziwe prawdopodobieństwa i odgadnięte). Prawdopodobieństwo > 0.5 dostaniemy wtedy, kiedy obie osoby preferują ten sam róg, chociaż preferencje mogą być zupełnie inne (np. 0.6 vs 0.9), natomiast jeżeli preferują przeciwne rogi, prawdopodobieństwo będzie mniejsze niż 0.5. Zatem jest to bardzo zgrubna miara jakości dopasowania. Ponadto ignoruje ona wszystkie aspekty czasowe dopasowania modelu. Nie mam przekonania do jej użyteczności. Wydaje mi się, że NLL, czy AIC są bardziej adekwatnymi miarami.

Mam problem z wnioskowaniem na temat zachowania myszy kiedy nagrodą był alkohol, ponieważ moim zdaniem badanych grup nie można traktować jako różnych próbek tego samego zbioru, a więc i wyników nie należy integrować. Po pierwsze, na rycinie 5-2 mamy wyraźną systematyczną różnicę pomiędzy liczbą prób uzyskania nagrody w obu grupach. Podobne różnice

widzimy na rycinie 5-3, nieco mniej wyraźne na 5-4. Zachowanie modeli pokazane na rycinach 5-7 i 5-8 również wskazuje na bardzo duże różnice pomiędzy obiema grupami, widoczne także w dopasowanych parametrach modeli z tabeli 5-1. W dyskusji autorka wskazuje, że przyczyną tych różnic mogły być różne schematy początkowej adaptacji. Niestety, mogło być wiele innych przyczyn, np. wada klatki czy ogólniej systemu pomiarowego, czy mniej oczywiste różnice między dwiema kohortami. Podobne różnice widzimy między kohortami sacharynowymi, ale są one mniejsze. Moim zdaniem widoczne podobieństwa w strukturze wyników dla sacharyny na rycinach 5-7 i 5-8, większe wewnątrz grupy niż pomiędzy grupami, są pośrednim argumentem za znaczeniem interakcji społecznej w grupach, czego autorka jest świadoma, o czym świadczy jej dyskusja. Podobnie wyniki pokazano na rycinie 5-13 czy 5-14B. Dlatego uważam, że do pełnego zrozumienia danych z klatek IntelliCage musimy nauczyć się uwzględniać te aspekty społeczne w analizie czy modelowaniu. Oczywiście na razie nie ma dobrych metod, które by to pozwoliły zrobić, a autorka wykonała ważny pierwszy krok w dobrą stronę. Jednocześnie chciałbym podkreślić moja wysoką ocenę prezentacji danych przez autorkę, która świadoma tych problemów interpretacyjnych pokazuje swoje wyniki w taki sposób, który nie ukrywa, ale pozwala zidentyfikować potencjalne problemy.

Moim zdaniem ryciny 5-4 i 5-15 zyskałyby na czytelności gdyby na osi y odłożyć logarytm prawdopodobieństwa, a nawet może narysować wykresy podwójnie logarytmiczne – widać bardzo długie ogony obserwowanych rozkładów. W tym drugim przypadku użyłbym przedziałów rozłożonych równomiernie w skali logarytmicznej, np. o granicach dla 2^n minut (czyli $(0,5, 1]$, $(1,2]$, $(2,4]$ min itd. lub gęściej, np. co $2^{n/2}$ minut itd.). Chociaż bardzo mi się podobają te wyniki, także w zakresie badania zależności od czasu uważam, że to dopiero pierwszy krok. Docelowo powinniśmy brać pod uwagę lokalną aktywność populacji w czasie. Ten sam odstęp od ostatniej nagrody może mieć dla zwierzęcia inne znaczenie kiedy jest ono w fazie aktywnej niż kiedy podsypia, co może wpływać na wyniki analizy i ich interpretację. Dla celów porównawczych przyjęta metodologia jest adekwatna.

Myślę, że dla spójności wyniki pokazane na rycinie 5-2 lepiej byłoby reprezentować jako funkcję fazy doświadczenia, a nie okresu, tak jak to zrobiono na pozostałych rycinach tego typu (5-3, 5-10, 5-11, 5-13, 5-14).

Głównym problemem w interpretacji danych w rozdziale 5.2 jest dla mnie fakt, że mamy tu mieszane grupy, mutanty + kontrola, gdzie znowu efekty społeczne mogą odgrywać pewną rolę. Widać istotne korelacje między aktywnością grupy badanej a kontrolnej na rycinach 5-13 i 5-14B, większe, niż pomiędzy odpowiadającymi sobie grupami z różnych kohort. To nie musi być problemem, na rycinach 5-17 i 5-18 nie widać tak silnych różnic między kohortami jak w rozdziale 5.1, ale dla pełnego oglądu sytuacji brakuje mi zbadania grup jednorodnych (nie mieszanych; i oczywiście analizy uwzględniającej



efekty społeczne, patrz dyskusja wyżej). Pod koniec rozdziału 5.2 autorka pisze: „Testowane grupy charakteryzowały się zróżnicowanym składem osobników, ponieważ w jednej kohorcie testowano zarówno osobniki z mutacją jak i kontrolne. W porównaniu do wskazanych grup jednolita kohorta myszy C57BL/6J charakteryzowała się lepszą zdolnością do rozpoznania zróżnicowanego prawdopodobieństwa uzyskania nagrody w obu narożnikach [], większym efektem nagrody [] oraz silniejszą tendencją do powtórzenia działania zależnie od czasu mijającego od poprzedniej próby [].” Możemy więc mieć tu do czynienia z takim samym efektem, co w pracy Kiryk i wsp. (2011), gdzie porównując znacząco różne wyniki grup jednolitych i mieszanych autorzy wskazują na możliwe efekty społeczne. Powodowały one znaczącą poprawę skuteczności mutantów (tam: myszy model choroby Alzheimera) przy jednoczesnym pogorszeniu się wyników myszy typu dzikiego. Tutaj autorka obserwuje drugi efekt bo ma jednolitą grupę myszy typu dzikiego. Nie może zaobserwować pierwszego efektu, bo nie ma jednolitych grup mutantów. W tym kontekście bezpieczniejsze do interpretacji są wyniki z klatek Skinnera, gdzie efektu społecznego na pewno nie ma, ale tracimy inne aspekty doświadczenia w klatkach IntelliCage (np. efekt przestrzeni, itp.).

Zamiast badać odsetek zwierząt z najlepszym dopasowaniem do danego modelu (rycina 5-9) wolałbym patrzeć na łączną ujemną wiarygodność logarytmiczną (NLL) dopasowania danego modelu do całej grupy i to bym wykorzystał do globalnego uszeregowania modeli pod kątem wyjaśniania przez nie zebranych danych.

Dla kompletności recenzji wymieniam również drobne uwagi, głównie redakcyjne, częściowo polemiczne. Nie oczekuję, że doktorantka się do nich ustosunkuje.

- Zrezygnowałbym z numeracji streszczenia, piśmiennictwa, spisów rycin i tabel. Wnioski bym włączył do dyskusji. Spisy rycin i tabel umieściłbym po spisie treści. Rozdziały powinny się zaczynać od nowej strony nieparzystej, tak jak teraz rozdział 4 i 6.

- Spis skrótów jest przydatny i w zasadzie adekwatny, zabrakło mi definicji *nll* (*negative log-likelihood*), także w tekście.

- W podrozdziale 2.2 zabrakło mi diagramów omawianych szlaków. Nie są kluczowe, ale moim zdaniem ułatwiłyby czytanie. W podrozdziale 2.4 warto było zilustrować opisywane wyniki z doświadczeń Hollermana i Schultza (1998), Fiorillo i in. (2003).

- p. 11: „Szlak mezolimbiczny (ang. mesolimbic pathway) prowadzi od (ang. ventral tegmental area, VTA) do jądra półleżącego (ang. nucleus accumbens, NAc).” Zabrakło polskiej nazwy VTA, chociaż jest w spisie skrótów.



- p. 13, 23: procesowanie → przetwarzanie. Co prawda według najnowszego słownika języka polskiego użycie słowa „procesowanie” w kontekście informacji jest legalne, ale lepiej pisać o przetwarzaniu informacji, a procesowanie zostawić na salę sądową.
- p. 17: Autorka zwraca uwagę, że Rutledge i wsp. (2019) otrzymali wartości wielkości kroku (alpha) większe niż 1, co prowadzi do trudności w ich interpretacji. W modelu Q-learning można pokazać, że jest on zbieżny dla wartości alpha od 0 do 2. Wartości $2 > \alpha > 1$ dają taką samą zbieżność jak wartości $\alpha - 1$, tylko wyniki oscylują wokół wartości asymptotycznej.
- p. 20: Autorka pisze: „Tutaj chciałabym skomentować...” Zastąpiłbym tą frazę słowem „Zauważmy...” a cały ten fragment komentarza osobistego dał jako przypis, bo obecnie przerywa myśl autorki. Wówczas fragment „Wracając do pracy Sary Matias i wsp.,” można by opuścić.
- p. 25: we wzorze (2) $r_{n+1} \rightarrow r_n$
- p. 28: Opis modeli stosowanych przez autorkę byłby dla mnie czytelniejszy, gdyby poszczególne modele były ponumerowane i ich numery były używane obok nazw przy prezentacji wyników i w dyskusji, albo gdyby modele były przynajmniej omówione w tej samej kolejności, co w tabeli. Nie znalazłem omówienia modelu incydentalnego.
- p. 29: Zakładam, że model z podwójnym tempem uczenia (podpis ryciny 2-4) to model wartościowo zależny.
- Rozdział 3. Cele badawcze mogłyby być z powodzeniem końcowym podrozdziałem Wprowadzenia. Sam ten rozdział napisany jest mniej zgrabnie niż większość pracy.
- p. 46: dla modelu z trzema... $5^3=125 \dots \rightarrow 3^3 = 27$
- p. 46 – dobór początkowych parametrów. Dla modelu Q-learning na pewno wartości alpha od 0 do 2 dają zbieżne wyniki. Pozostałych modeli nie sprawdzałem. Beta wydaje mi się naturalnie wziąć takie, żeby środkową wartością było 1, które jest jedyną wyróżnioną wartością, zresztą doktorantce regularnie wychodziły wartości $\beta < 1$.
- p. 56 i dalsze w podpisach rycin: rząd → lepiej pisać wiersz.
- W Literaturze umieszczanie przecinka na końcu cytowania wydaje mi się nienaturalne. Sugeruję użycie kropki. Nie wszystkie referencje zostały sprawdzone. Patrz np. Cieślak i wsp. („eneuro”), Conway i White (, ,). W szczególności cytowanie klasycznej książki Suttona i Barto zostało zastąpione jej recenzją. Notabene, warto byłoby zastąpić pierwsze wydanie z 1998 roku wydaniem drugim z 2018 roku, które nie tylko jest aktualniejsze ale też dostępne w domenie publicznej <http://incompleteideas.net/book/the-book-2nd.html>.
- literówki:

- p.5: ze odwróceniem → z odwróceniem; w wzmacniających → we wzmacniających
p. 15: Elastyczności → Elastyczność; stosuje w badaniach → stosuje się w badaniach
p. 17: włącznie → włączenie; w układu → w układ
p. 19: z zwiększoną → ze zwiększoną
p. 22: odwrócił deficyt → odwracały deficyt
p.36: powstaje → powstanie; obu → obu.
p. 46: model „wzmocnienie” → wzmocnienia
p. 89: do grupy → od grupy

Podsumowując, uważam, że badania, które autorka przedstawiła w swojej rozprawie są ciekawe, analizy przeprowadzone poprawnie i systematycznie, dyskusja jest obszerna i wskazuje świadomość autorki zarówno co do znaczenia jej wyników i potencjalnych problemów metodycznych i interpretacyjnych. Uważam, że recenzowana rozprawa doktorska pani mgr Judyty Jabłońskiej **spełnia warunki określone w art. 187 Ustawy z dnia 20 lipca 2018 r. o szkolnictwie wyższym i nauce (Dz. U. 2022, poz. 574 z późn. zm.)**, dlatego zwracam się do Rady Naukowej Instytutu Farmakologii im. Jerzego Maja PAN o przyjęcie rozprawy i dopuszczenie mgr Judyty Jabłońskiej do dalszych etapów postępowania w sprawie nadania stopnia doktora w dziedzinie nauk medycznych i nauk o zdrowiu, **dyscyplinie nauki medyczne**. Ponadto, ze względu na nowatorskie połączenie obszernych doświadczeń ze złożoną analizą danych opartą o modele uczenia ze wzmocnieniem, która pozwoliła na dogłębne porównanie badanych przypadków i ciekawą interpretację wyników, z których część została już opublikowana w dobrym czasopiśmie, **wnioskuję o wyróżnienie tej rozprawy**.

Prof. dr hab. Daniel K. Wójcik