

Recenzja rozprawy doktorskiej

Jakuba Kwiatkowskiego

zatytułowanej:

Staged Self-Supervision for Solving Raven's Progressive Matrices

1. Problem badawczy i jego znaczenie

Rozprawa doktorska mgr. inż. Jakuba Kwiatkowskiego dotyczy tworzenia i wszechstronnej weryfikacji i walidacji metod sztucznej inteligencji do rozwiązywania zadań, które wymagają rozumowania wizualnego, efektywnego uchwycenia istotnych zależności między obiektami w scenie (np. przestrzennych i przyczynowo-skutkowych) oraz zastosowania ich do rozwiązywania nowych instancji analizowanego problemu. W swojej pracy Autor skupił się głównie na rozwiązywaniu Progresywnych Matryc Ravena (*Raven's Progressive Matrices*, RPM), ale metody wnioskowania wizualnego i automatycznej analizy relacji pomiędzy obiektami mają szerokie praktyczne zastosowania w wielu dziedzinach nauki i przemysłu (np. w medycynie, robotyce, obrazowaniu satelitarnym i w wielu innych), daleko wykraczające poza problemy benchmarkowe dotyczące rozwiązywania wizualnych zagadek.

Autor zauważa, że istniejące techniki (głębokiego) uczenia maszynowego do rozwiązywania RPM – zestawu benchmarków, który jest szeroko stosowany w literaturze do oceny metod wnioskowania wizualnego – są w przeważającej większości architekturami monolitycznymi, których zadaniem jest przewidzenie, który z obrazów-kandydatów jest poprawny. Doktorant słusznie podkreśla, że tak sformułowane zadanie może kierować modele uczenia maszynowego w stronę eksploatacji wzorców i statystycznych regularności danych treningowych (w konsekwencji prowadzi do tzw. uczenia na skrótach), a nie wypracowywania reguł, które rzeczywiście odzwierciedlają zależności przyczynowo-skutkowe pomiędzy kolejnymi obrazami (panelami) w obrazach RPM. Mgr inż. Jakub Kwiatkowski podejmuje się zaprojektowania i oceny nowego systemu opartego na etapowym uczeniu samonadzorowanym do rozwiązywania zadań związanych z abstrakcyjnym rozumowaniem, w którym zjawisko uczenia na skrótach zostanie wyeliminowane.

Problem, którym mgr inż. Jakub Kwiatkowski zajmuje się w badaniach omówionych w swojej rozprawie jest precyzyjnie i poprawnie zdefiniowanym problemem badawczym o istotnym znaczeniu naukowym oraz praktycznym. Tematyka podjęta w pracy bezpośrednio wpisuje się w *Informatykę Techniczną i Telekomunikację*.

2. Wkład autora

Najważniejszym wkładem mgr. inż. Jakuba Kwiatkowskiego jest zaprojektowanie, zaimplementowanie oraz wszechstronne zwalidowanie – również w bardzo rozbudowanych badaniach ablacyjnych – metody dwuetapowego uczenia samonadzorowanego do rozwiązywania RPM. W pierwszym etapie wyznaczane są symboliczne opisy paneli RPM (z wykorzystaniem zaproponowanej w pracy architektury transformerowej: *Abstract Compositional Transformer*, ACT), a w drugim wybierana jest odpowiedź ze zbioru paneli kandydujących (za pomocą jednego z trzech opracowanych modułów

wyboru odpowiedzi). Autor zidentyfikował wady istniejących algorytmów (związane z uczeniem na skrótach wspomnianych architektur monolitycznych czy z obciążeniami i statystycznymi regularnościami zbiorów danych) – pozwoliło to na precyzyjne przedstawienie motywacji, które kierowały Doktorantem w trakcie opracowywania przedstawionych algorytmów.

Oprócz samej metody dwuetapowego uczenia samonadzorowanego, Autor opracował metodykę badawczą, która umożliwiła przeprowadzenie dogłębnej analizy eksperymentalnej i porównanie algorytmów Autora z konkurencyjnymi technikami znanymi z literatury. Doktorant zaprojektował i przeprowadził eksperymenty, które pozwoliły na zweryfikowanie postawionych hipotez, np. związanych z obciążeniami zbiorów danych, które mogą powodować (i w praktyce powodują) zmaterializowanie się zjawiska uczenia na skrótach i zapamiętywania powierzchownych zależności statystycznych obecnych w danych treningowych. Mgr inż. Jakub Kwiatkowski zadbał o dokładne zaprezentowanie i omówienie metodyki eksperymentalnej oraz o zapewnienie reprodukowalności swoich badań (implementacje są dostępne na platformie GitHub, a modele i dane na platformie HuggingFace) – uważam, że jest to niezwykle ważne i wskazuje na dojrzałość naukową Autora.

Mgr inż. Jakub Kwiatkowski jest współautorem artykułów opublikowanych w czasopismach i na konferencjach międzynarodowych. Jako publikacje wskazane jako te bezpośrednio związane z tematyką rozprawy, Autor wymienił dwa artykuły czasopismowe (jedna z nich została opublikowana w światnym czasopiśmie *IEEE Transactions on Neural Networks and Learning Systems*) oraz poster zaprezentowany na konferencji PP-RAI 2024 w Warszawie.

3. Poprawność

Badania, których wyniki zostały przedstawione i omówione w rozprawie były prowadzone rzetelnie, metodyka badań jest poprawna, hipotezy zostały precyzyjnie i jasno sformułowane, a otrzymane wyniki zostały przekrojowo omówione i poprawnie zinterpretowane przez Doktoranta.

W czasie lektury nasunęły mi się pewne pytania, które zawarłem w kolejnej sekcji niniejszej recenzji – nie wpływają one negatywnie na mój bardzo pozytywny odbiór pracy.

4. Wiedza kandydata

Rozprawa mgr. inż. Jakuba Kwiatkowskiego została napisana w j. angielskim i składa się z siedmiu numerowanych rozdziałów, bibliografii oraz licznych załączników.

Rozdział pierwszy jest wprowadzeniem do tematyki rozprawy, w którym Autor krótko omawia problemy rozumowania wizualnego i ich istotność, zarysowując jednocześnie wady istniejących metod uczenia głębokiego – są one w dużej mierze bezpośrednio związane z uczeniem na skrótach, zamiast odkrywaniem przez takie modele istotnych reguł i zależności pomiędzy analizowanymi obiektami. Doktorant omawia RPM i przedstawia benchmarki wykorzystywane do oceny algorytmów rozwiązywania RPM – omówione zostały obciążenia takich zbiorów (np. *answer set bias* i *similarity bias*), które powodują, że możliwe jest tworzenie metod osiągających bardzo wysokie wyniki ilościowe, ale jednocześnie praktycznie zupełnie ignorujące reguły, informacje kontekstowe i zależności pomiędzy poszczególnymi panelami w RPM. W Sekcji 1.2 Autor omawia cel i zakres pracy oraz trzy, poprawnie i precyzyjnie zdefiniowane hipotezy badawcze, do których będzie odnosił się w dalszej części rozprawy, a w Sekcji 1.3 omawia swój najistotniejszy wkład. W dalszej części tego rozdziału mgr inż. Jakub Kwiatkowski zarysowuje swoje zainteresowania badawcze i komercyjne, zwłaszcza związane

z opracowywaniem bezpiecznych i wyjaśnialnych systemów sztucznej inteligencji. Uważam, że tematyka, którą zajmuje się Doktorant jest niezwykle ważna i aktualna.

Rozdział 2 wprowadza czytelnika w najistotniejsze aspekty dotyczące tematyki rozprawy oraz związane z automatycznym uczeniem reprezentacji przez modele uczenia głębokiego, wyzwaniami związanymi z tworzeniem „odpornych” i wiarygodnych systemów AI, uczeniem na skrótach oraz z wyzwaniami związanymi z tworzeniem i efektywnym wykorzystaniem dużych modeli językowych (*Large Language Models*, LLM) i modeli podstawowych. Warto podkreślić, że Autor precyzyjnie i przekrojowo (ale nie „rozwlekle”) omawia aspekty, które pomogą czytelnikowi umiejscowić badania Doktoranta w obecnym stanie wiedzy – jednocześnie odnosząc się do postawionych wcześniej hipotez. W kolejnym rozdziale 3, Autor przechodzi do omówienia RPM: dokładnie został przedstawiony zbiór benchmarkowy RAVEN (i inne warianty tego zbioru, a także inne zbiory benchmarkowe), wraz ze swoimi „obciążeniami”. Na uwagę zasługuje fakt, że Doktorant obiektywnie kwantyfikował i oceniał zidentyfikowane obciążenia, a w pracy umieścił bardzo informatywne rysunki i przekonujące przykłady wizualne (mogłoby być ich więcej, np. w dołączonych załącznikach), które istotnie ułatwiają przyswojenie informacji. Autor zaprezentował przegląd metod rozwiązywania RPM, jasno omawiając ich zalety i wady. Nie mam żadnych wątpliwości, że Doktorant świetnie orientuje się w literaturze związanej z tematyką rozprawy.

W rozdziałach 4 i 5, Doktorant omawia poszczególne elementy opracowanej architektury dwuetapowej dla RPM. Rozdział 4 został poświęcony ACT – architekturze transformerowej, której celem jest predykcja opisów paneli RPM. Predykcja takich symbolicznych opisów paneli RPM realizowana jest poprzez rekonstrukcję zamaskowanych paneli, z wykorzystaniem informacji kontekstowych. Autor dokładnie omawia składowe architektury ACT, tj. *Image Tokenizer*, *Transformer* oraz *Property Predictor*. Zaprezentowanych zostało kilka wariantów tokenizera (*panel tokenizer*, *task tokenizer*, *row tokenizer*) – Autor przedstawił motywację stojącą za wprowadzeniem kilku wariantów tego mechanizmu, a wizualizacje pozwalają na łatwe ich zrozumienie. W kolejnych podrozdziałach rozdziału 4 Doktorant omawia poszczególne elementy architektury ACT, a w podrozdziale 4.3 przedstawia metodykę strategii treningu samonadzorowanego w ACT. Kolejny podrozdział 4.4 omawia badania eksperymentalne Doktoranta. Autor zaczyna od przedstawienia zbiorów danych i podziałów na podzbiory treningowe, walidacyjne i testowe w wykorzystanych benchmarkach (RAVEN, I-RAVEN, RAVEN-FAIR), prezentuje wykorzystane strategie maskowania paneli w czasie treningu (które zostały dokładnie omówione w poprzednich sekcjach – ponownie, wraz z dokładną motywacją stojącą za ich wprowadzeniem i wykorzystaniem) i metrykami, które zostaną użyte do oceny analizowanych podejść.

Podrozdział 4.6 jest opisem i omówieniem uzyskanych wyników eksperymentalnych – został podzielony na kilka podsekcji. Każda z tych podsekcji jest poświęcona bardzo dobrze zdefiniowanym eksperymentom, mającym na celu odpowiedzenie na poprawnie zdefiniowane przez Autora pytania badawcze (m.in. analizowany jest wpływ doboru strategii maskowania, tokenizera czy technik augmentacji danych treningowych). Warto podkreślić, że przestrzeń możliwych (i jednocześnie sensownych) eksperymentów i konfiguracji, które można byłoby przebadać jest niezwykle duża – Autor świadomie ją zawężył w kolejnych badaniach, poprawnie projektując kolejne eksperymenty. Oprócz samego przedstawienia wyników eksperymentalnych, mgr inż. Jakub Kwiatkowski nie tylko je szczegółowo omawia i interpretuje, ale też odnosi się bezpośrednio do wcześniej postawionych trzech hipotez badawczych (H1 – H3). W podrozdziale 4.5.6 Autor przedstawia wyniki swoich badań dotyczących oceny możliwości generalizacyjnych modeli pomiędzy zbiorami benchmarkowymi (modele były trenowane na zbiorze RAVEN, a testowane na zbiorach I-RAVEN i RAVEN-FAIR). Wyniki pokazują, że analizowane modele pozwalają na uzyskanie praktycznie takich samych wyników

na wszystkich zbiorach (rankingi tokenizerów i metod maskowania są również niezwykle spójne pomiędzy zbiorami) – to w istocie wskazuje na to, że opracowana metoda ma silne możliwości generalizacyjne nawet w przypadku zbiorów RPM, dla których strategie tworzenia zbiorów odpowiedzi się różnią. W podrozdziale 4.8 Doktorant analizuje wypracowane reprezentacje, a w podrozdziale 4.9 podsumowuje wyniki badań dotyczących ACT w odniesieniu do postawionych hipotez.

Rozdział 5. został poświęcony omówieniu opracowanych modułów wyboru odpowiedzi w RPM. W kolejnych podrozdziałach zaprezentowano techniki DCM (*Direct Choice Maker*), LCM (*Learnable Choice Maker*) i CCM (*Contrastive Choice Maker*) – Autor prowadzi czytelnika przez analizę ilościową i jakościową otrzymanych wyników, jasno motywując kolejne kroki badawcze i potrzebę tworzenia bardziej zaawansowanych modułów wyboru odpowiedzi. Również w rozdziale piątym Doktorant porównuje wyniki otrzymane dla benchmarków RAVEN i I-RAVEN z wykorzystaniem opracowanego frameworka (ACT w połączeniu z modułami wyboru odpowiedzi DCM, LCM i CCM) z metodami znanymi z literatury – zarówno opartymi na uczeniu głębokim, jak i z metodami neurosymbolicznymi. Otrzymane wyniki wskazują na to, że zaproponowane przez mgr. inż. Jakuba Kwiatkowskiego metody są konkurencyjne względem analizowanych algorytmów rozwiązywania RPM.

W rozdziale szóstym, Doktorant przedstawia bardzo ciekawe, dodatkowe eksperymenty, które pozwoliły na dalszą walidację opracowanych metod (z wykorzystaniem innych zbiorów danych) oraz zarysowanie najciekawszych, według Autora, kierunków dalszych możliwych badań. Badania eksperymentalne wykroczyły poza analizę ACT i objęły także LLMy, w kontekście analizy wpływu wnioskowania pośredniego na ich działanie oraz w kontekście ich dalszego rozwoju i możliwej diagnostyki. Rozprawa została zwieńczona podsumowaniem przedstawionym w rozdziale 7, w którym Autor odnosi się i dokładnie omawia wszystkie hipotezy badawcze.

Opracowane metody oraz otrzymane wyniki eksperymentalne przedstawione w rozprawie pozwalają mi z pełnym przekonaniem stwierdzić, że Doktorant jest świetnie przygotowany do tego, żeby prowadzić dalsze badania związane z rozumowaniem wizualnym, analizą obrazów oraz projektowaniem, implementacją i wszechstronną oceną systemów sztucznej inteligencji w praktycznych zastosowaniach. Warto także podkreślić, że Doktorant świetnie orientuje się w obecnym stanie wiedzy w obszarach związanych z tematyką rozprawy – bibliografia została bardzo dobrze skomponowana i przeanalizowana.

Rozprawa doktorska mgr. inż. Jakuba Kwiatkowskiego, a także artykuły, których jest współautorem prezentują dojrzałość badawczą oraz wiedzę teoretyczną i praktyczną Doktoranta w dyscyplinie *Informatyka Techniczna i Telekomunikacja*. Stronę merytoryczną rozprawy oceniam bardzo wysoko.

W czasie lektury pracy nasunęły mi się następujące uwagi i pytania:

1. Na str. 59 Autor wspomina, że „*The original panel size in the RAVEN dataset is 160×160 , which we downsample to 84×84 pixels. Preliminary experiments showed that this resolution strikes an optimal balance between model performance and computational cost (...)*” – jak dokładnie były zaprojektowane te eksperymenty? Podobne pytanie można byłoby sformułować w przypadku innych badań wstępnych, o których wspomina Autor (np. na str. 76).
2. Jak wyglądał proces projektowania architektury LCM (i oceny potencjalnie innych architektur, jeśli były analizowane)? Czy, według Doktoranta, miałoby sens wykorzystać tutaj techniki automatycznego uczenia maszynowego do wypracowania/optimalizacji architektury

- i hiperparametrów LCM (jeśli tak, to jakie?). Czy ich wykorzystanie mogłoby znacząco polepszyć jakość działania modułu LCM?
3. Czy podzbiory próbki treningowej zbioru RAVEN (5%, 25%, 50% przykładów treningowych) wykorzystane do eksperymentów omówionych na str. 105-106 zostały upublicznione (nie znalazłem tych podzbiorów na platformie HuggingFace)? Jeśli nie zostały, warto byłoby rozważyć ich upublicznienie, żeby zapewnić reprodukowalność również tych eksperymentów.
 4. Warto byłoby popracować nad opisem modeli i zbiorów danych umieszczonych na platformie HuggingFace – odpowiednie *readme* znacznie ułatwiłoby ich eksplorację i późniejsze wykorzystanie.
 5. Wyniki badań eksperymentalnych, np. dla różnych analizowanych konfiguracji frameworka (np. przedstawione w Tabeli 5.13) są czasami bardzo zbliżone – czy Autor rozważał przeprowadzenie testów statystycznych, aby zweryfikować czy różnice pomiędzy badanymi metodami są statystycznie istotne? Czy takie testy miałyby sens – jeśli tak, to jakie testy statystyczne Doktorant sugerowałby, żeby zastosować?
 6. W Tabeli 5.15, w której Doktorant porównuje opracowany framework z metodami opartymi na uczeniu głębokim i technikami neurosymbolicznymi na zbiorze RAVEN widzimy, że najnowsza zaprezentowana w tym porównaniu metoda uczenia głębokiego dla zbioru RAVEN została opublikowana w 2021 r. (DCNet). Czy taki kierunek prac nie jest obecnie eksplorowany w literaturze (np. przez wzgląd na zidentyfikowane przez Autora wyzwania związane z uczeniem na skrótach), czy odchodzi się od wykorzystania tego zbioru (ponownie, mając na względzie jego ograniczenia) i obecnie wykorzystuje się głównie (tylko?) nowe warianty zbiorów RAVEN? Jak, według Doktoranta, powinny wyglądać procedury weryfikacji i walidacji nowych metod rozwiązywania RPM – czy dałoby się zdefiniować „ustandaryzowany” proces takiej oceny (i czy miałoby to sens)?
 7. Dlaczego w tokenizerze Doktorant zdecydował się na wykorzystanie akurat architektury EfficientNetV2B0? Czy były rozważane inne architektury? Jeśli tak, to jakie?
 8. Na str. 160 Autor krótko omówił eksperymenty związane z automatyczną optymalizacją hiperparametrów, w której wykorzystał metodę TPE (*Tree-structured Parzen Estimator*). Dlaczego akurat na tę metodę zdecydował się Autor? Czy były rozważane inne algorytmy optymalizacji hiperparametrów?

Tak jak wspomniałem wyżej, powyższe pytania i uwagi w żaden sposób nie wpływają negatywnie na mój bardzo pozytywny odbiór rozprawy.

5. Inne uwagi

Rozprawa jest napisana niezwykle starannie pod względem językowym i redakcyjnym. Zauważyłem tylko kilka drobnych uchybień:

1. Na Rys. 2.1 Autor użył brytyjskiego angielskiego („colour”); podobnie na str. 47 („summarise”) – w całej pracy stosowany jest amerykański angielski (powinno to zostać uspojnione),
2. Zauważyłem, że w niektórych miejscach rozprawy tekst niepoprawnie „wypływa” na margines (np. na str. 37, 115 i in.),
3. W niektórych wpisach w bibliografii brakuje istotnych informacji (np. stron, wydawcy, itd.; np. w Alam et al., 2024, Camposampiero et al., 2024); zauważyłem też niespójność zapisu wielkich/małych liter (np. w Marcus 2020).
4. Na str. 21 Autor odwołuje się do Listingu 2.3.2 – chodzi najpewniej o Listing 2.2.

5. W rozdziale 3.1.2, wartości orientacji (-135° , -90°) powinny być umieszczone w otoczeniu matematycznym (\$..\$).

6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami)¹ moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

- A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? (wybierz jedną opcję stawiając znak X)

☒	☐	☐	☐	☐
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

- B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja?

☒	☐	☐	☐	☐
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

- C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

☒	☐	☐	☐	☐
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

Ponadto, biorąc pod uwagę jakość prowadzonych badań i otrzymanych wyników, ich praktyczne i teoretyczne implikacje, dojrzałość naukową Doktoranta, która emanuje z przedstawionej rozprawy oraz jakość publikacji, których mgr inż. Jakub Kwiatkowski jest współautorem, **rekomenduję wyróżnienie rozprawy doktorskiej.**

Podpis

¹ <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276>