

Recenzja rozprawy doktorskiej

Mgr Jakuba Kwiatkowskiego

zatytułowanej:

Staged Self-Supervision for Solving Raven's Progressive Matrices

1. Problem badawczy i jego znaczenie

Rozprawa dotyczy bardzo istotnego obszaru badawczego, mianowicie umiejętności abstrakcyjnego rozumowania wizualnego (ang. *Abstract Visual Reasoning* - AVR) współczesnych systemów sztucznej inteligencji (AI). Doktorant rozważa Progresywne Matryce Ravena (RPM), stanowiące najpopularniejszy rodzaj problemów AVR rozważanych w literaturze. W standardowej postaci problem RPM składa się z dwóch części: paneli kontekstowych oraz paneli odpowiedzi. Panele kontekstowe tworzą matrycę 3x3 z pustym prawym-dolnym elementem. Zadanie polega na wybraniu spośród zadanych odpowiedzi (zwykle jest ich osiem) tej, która wstawiona w miejsce brakującego panelu prawidłowo uzupełni matrycę. Wybór właściwej odpowiedzi wymaga odkrycia reguł reprezentowanych w rozważanej matrycy (np. *progresja* (rozmiaru, liczby elementów, itd.), *wybór z trzech* (ustalonych opcji np. kształtu, rozmiaru, itd.), *stała wartość* (np. koloru, kształtu, położenia), *operacja logiczna* (np. OR, XOR, AND, itd.)). Reguły te aplikowane są w wierszach i/lub kolumnach matrycy.

Rozwiązywanie zadań AVR wykracza poza standardową analizę i rozpoznawanie obrazów i wymaga umiejętności dostrzeżenia/odkrycia abstrakcyjnych relacji pomiędzy obiektami oraz ich zastosowania w nowych konfiguracjach. Umiejętności tego typu, w sposób bardzo efektywny wykorzystywane przez ludzi, wciąż stanowią problem dla współczesnych modeli AI, choć głównie w zadaniach innego rodzaju niż RPM (np. ARC-2/3 czy Problemy Bongarda).

Badania prezentowane w rozprawie mają charakter naukowy i odnoszą się do aspektów metodologicznych oraz architektonicznych modeli AI rozwiązujących zadania RPM. W sensie ogólniejszym, badania dotyczące AVR (w tym RPM) mają istotne implikacje praktyczne związane z możliwością autonomicznego funkcjonowania agentów AI w nowych środowiskach oraz ich współpracy z ludźmi. Badania w obszarze AVR są kluczowe z punktu widzenia rozwoju systemów AI zmierzających w kierunku AGI (*Artificial General Intelligence*), zdolnych do samodzielnego rozpoznawania i rozwiązywania nowych problemów wymagających abstrakcyjnego rozumowania i adaptacji do zmieniających się warunków.

2. Wkład autora

Główny wkład Autora rozprawy polega na zaproponowaniu dwuetapowego podejścia do rozwiązywania zadań RPM w drodze uczenia samonadzorowanego (ang. *staged self-supervision*).

Doktorant proponuje oddzielenie procesu predykcji atrybutów brakującego panelu od procesu formalnego wyboru odpowiedzi, co stanowi interesującą koncepcję w kontekście przeważających w literaturze podejść monolitycznych.

Etap predykcji/generowania atrybutów brakującego panelu, opisany w rozdziale czwartym rozprawy, realizowany jest w oparciu o zaproponowaną w dysertacji architekturę ACT (*Abstract Compositional Transformer*). W skład ACT wchodzi 3 moduły (*Image Tokenizer*, *Transformer*, *Property Predictor*). Podział na tokeny (moduł pierwszy) realizowany jest na jeden z trzech sposobów, określonych w pracy jako (*Panel Tokenizer*, *Task Tokenizer* oraz *Raw Tokenizer*) i różniących się zakresem przetwarzanej informacji: pojedynczy panel, cała matryca, wiersz matrycy. Moduł transformatora bazuje na standardowej implementacji odpowiednio dopasowanej do warunków rozpatrywanego problemu. Przewidywanie własności brakującego panelu odbywa się w module trzecim, w drodze uczenia samonadzorowanego polegającego na maskowaniu paneli w matrycach treningowych (ze znanymi odpowiedziami). Kandydat rozważa trzy strategie maskowania indywidualnych paneli macierzy RPM: *random masking* maskuje dowolny z dziewięciu paneli matrycy, *query masking* – maskowany jest wyłącznie brakujący panel, oraz sposób mieszany (*combined masking*) – początkowo maskowany jest dowolny panel, a w fazie drugiej wyłącznie panel brakujący.

Drugi etap proponowanego w rozprawie podejścia, mianowicie wybór właściwej odpowiedzi (spośród ośmiu dostępnych propozycji) bazując na atrybutach brakującego panelu przewidzianych w etapie pierwszym, opisany jest w rozdziale piątym. Pan Jakub Kwiatkowski proponuje trzy metody wyboru, określone jako *Direct*, *Learnable* oraz *Contrastive*. Wszystkie trzy metody odnoszą się do stopnia podobieństwa pomiędzy przewidywaną postacią brakującego panelu w matrycy niepełnej (p), a jego przewidywaną postacią w przypadku uzupełnienia matrycy kolejnymi odpowiedziami: $p_1, \dots, p_8$. Pary (p, p_i) , $i=1, \dots, 8$ stanowią podstawę wyboru jednej z możliwych odpowiedzi. W metodzie bezpośredniej wybierana jest odpowiedź i , dla której odległość $d(p, p_i)$, mierzona jako entropia krzyżowa, jest najmniejsza. Metoda *Learnable* opiera się na wykorzystaniu sieci neuronowej do aproksymacji optymalnej miary odległości $d(\cdot, \cdot)$, w miejsce predefiniowanej miary w strategii *Direct*. Trzecia metoda, kontrastująca, stosuje dodatkową fazę treningową (*fine-tuning*) w oparciu o zmodyfikowaną funkcję straty, która zarówno nagradza bliską odległość do właściwej odpowiedzi w przestrzeni reprezentacji, jak i możliwie duże odseparowanie od odpowiedzi niepoprawnych.

Poza autorskimi rozwiązaniami opisanymi w obu powyższych rozdziałach, Doktorant wykonał także pogłębiony przegląd literatury dotyczącej tematyki rozprawy (rozdziały drugi i trzeci) oraz zaproponował możliwe kierunki dalszego rozwoju przedstawionych w rozprawie badań (rozdział szósty). Oś opisanych w rozprawie badań stanowią trzy hipotezy badawcze H1-H3, sformułowane w rozdziale pierwszym, podsumowane w rozdziale siódmym, stanowiące motywację badawczą zaproponowanego w rozprawie rozwiązania.

3. Poprawność

Oba etapy zaproponowanego przez Kandydata podejścia do rozwiązywania problemów RPM, opisane w rozdziałach czwartym i piątym, stanowią oryginalne i innowacyjne propozycje, szczegółowo zweryfikowane eksperymentalnie. Zaproponowane rozwiązanie cechuje się pomysłowością badawczą i posiada dobre ugruntowanie intuicyjne, sytuując się niejako pomiędzy podejściami koneksonistycznymi a podejściami z zakresu przetwarzania neuro-symbolicznego. Uzyskane wyniki są porównywalne z najlepszymi wynikami prezentowanymi w literaturze światowej.

Ponadto, Doktorant przeprowadził szczegółowe badania dodatkowe (ablace) nakierowane na wykazania zarówno silnych stron jak i pewnych słabości proponowanych wariantów (trzy metody maskowania połączone z trzema metody wyboru rozwiązania). W moim przekonaniu, badania dodatkowe stanowią niezwykle istotny i bardzo profesjonalnie zrealizowany element badawczy rozprawy. Pan Jakub Kwiatkowski wykonał te badania w sposób bezstronny, nie unikając analiz, które wypadają mniej korzystnie dla zaproponowanych metod. Takie podejście, a także szeroki zakres i wnikliwość tych badań potwierdzają wysoki poziom warsztatu badawczego Doktoranta.

Nie mam istotnych uwag do zaproponowanej w rozprawie metody. Mam natomiast kilka pytań dotyczących konkretnych kwestii szczegółowych, o których wyjaśnienie prosiłbym Doktoranta:

- 1) Z czego wynika rozmiar reprezentacji ukrytej (64) w przypadku metody *Task Tokenizer*?
- 2) Na ile istotnym problemem jest pominięcie ostatniego elementu tej reprezentacji, w momencie podziału na siedem porcji dziewięcioelementowych?
- 3) Na stronie 64 jest mowa o wagach poszczególnych cech, które zostały dobrane eksperymentalnie. Na ile metoda jest wrażliwa na ich dobór i jak duży wysiłek obliczeniowy jest niezbędny do ich prawidłowego określenia?
- 4) W kontekście tokenizacji typu *Row*, czy nie warto byłoby rozważyć wariantu *2xRow*, łączącego dwa wiersze (1,2), (1,3) oraz (2,3)? Czy warto byłoby także uwzględnić tokenizację typu *Column*?
- 5) Czym różnią się odpowiedzi p_1 i p_7 w panelu odpowiedzi modelu DCM w przykładzie d) na stronie 104?
- 6) Na stronie 134 Doktorant wskazuje, że uzyskane wyniki są lepsze niż wyniki bazujące na *ground-truth*, co jest dość zaskakującą konkluzją. Jakie jest wytłumaczenie tej (nietypowej) sytuacji?

4. Wiedza kandydata

Rozdziały drugi i trzeci rozprawy poświęcone są analizie literatury dotyczącej obszarów powiązanych z tematyką pracy. Rozdział drugi omawia fundamentalne zagadnienia AI w kontekście aktualnego, dynamicznego rozwoju sztucznej inteligencji, w szczególności w obszarze AVR. Kolejny rozdział poświęcony jest analizie literatury dotyczącej rozwiązywania problemów RPM. Obie części pracy napisane są prawidłowo i stanowią bardzo dobre usytuowanie prezentowanych badań na tle bieżącej literatury światowej. Nie mam wątpliwości, że Autor posiada głęboką wiedzę zarówno w obszarze fundamentów AI, jak i szczegółowych rozwiązań w domenie AVR – w szczególności dotyczących zadań AVR.

Bibliografia, umieszczona na końcu rozprawy, liczy 18 stron i jest bez wątpienia wystarczającym tłem i motywacją przedstawionych prac badawczych oraz uzyskanych wyników. Oczywiście, można wskazać istotne prace, które mogłyby się na tej liście pojawić, ale w żaden sposób nie czyniłbym z tego zarzutu, biorąc pod uwagę obecny bardzo dynamiczny rozwój literatury w obszarze AVR.

Wysoki poziom metodologiczny rozprawy utwierdza mnie w przekonaniu, że Pan Jakub Kwiatkowski posiada szeroką wiedzę w dyscyplinie Informatyka Techniczna i Telekomunikacja, wykraczającą poza zakres zagadnień bezpośrednio poruszanych w rozprawie.

5. Inne uwagi¹

Rozprawa zredagowana jest bardzo starannie, a jej poziom językowy jest bardzo wysoki. Jedyny drobiazg, który zauważyłem to odniesienie się na stronie pierwszej do prawego-dolnego panelu macierzy RPM jako *lower-right*. Wydaje mi się, że *bottom-right* brzmi bardziej naturalnie w tym kontekście – tego określenia używa zresztą Pan Kwiatkowski w dalszej części rozprawy.

Wyniki badań zaprezentowane w rozprawie zostały częściowo opublikowane w prestiżowym czasopiśmie *IEEE Transactions on Neural Networks and Learning Systems*, co dodatkowo potwierdza wysoki poziom naukowy przedstawionych badań.

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)² 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

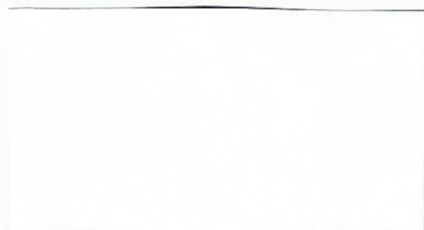
B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja?

☒	☐	☐	☐	☐
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 wysoki poziom metodologiczny i oryginalność przeprowadzonych badań, uzyskane wyniki, porównywalne z najlepszymi wynikami literaturowymi, opublikowanie części wyników w prestiżowym czasopiśmie *IEEE Transactions on Neural Networks and Learning Systems*, a także wysoki poziom merytoryczny i redakcyjny rozprawy, rekomenduję wyróżnienie rozprawy doktorskiej³.



Podpis

¹ Opcjonalnie

² <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276>

³ Oczywiście to zdanie jest opcjonalne.