

wpłynęło dnia 28.12.2023...

nr zał.

Bytom 23.12.2023

Prof. dr hab. inż. Konrad Wojciechowski
Polsko-Japońska Akademia Technik Komputerowych
Wydział Informatyki

Recenzji: rozprawy doktorskiej Pani mgr inż. Eryki Probierz

Modelling Social and Emotional Components in Social Robotics Using Robot Artificial Intelligence

Charakterystyka rozprawy i uwagi ogólne.

Recenzowana rozprawa liczy 186 stron, napisana jest w języku angielskim i ma formę rozprawy opartej na cyklu publikacji. Główną część rozprawy stanowią 4 rozdziały zatytułowane kolejno : *Modelling emotional response based on implementations of artificial intelligence and machine learning methods in robotics*, *Modelling emotional response based on action planning in PDDL*, *Modelling social behaviour of robots based on the application of neural networks in robotics* *Modelling selected social components using the PID controller in social robotics*. Dodatkowo w skład rozprawy wchodzi następujące wydzielone części merytoryczne: *Discussion*, *Practical application*, *Summary and conclusions*, *Appendix 1*, *Appendix 2*, *Summary in Polish*, *Summary in English*. Każdy z wymienionych rozdziałów posiada własną taką samą dla wszystkich rozdziałów strukturę w postaci: *Introduction*, *Literature review*, *Introduction to the publication*, *Summary of Chapter*, *Publications*. Dopiero w ostatnim punkcie Doktorantka wstawia kopie publikacji przyporządkowanych do omawianego rozdziału. Należy podkreślić że objętość tekstów poprzedzających publikacje czyli fragmenty: *Introduction*, *Literature review*, *Introduction to the publication*, *Summary of Chapter* wielokrotnie przewyższają objętość samej publikacji. W szczególności dotyczy to wprowadzenia do tematyki rozdziału oraz dodatkowego przeglądu biorąc pod uwagę fakt, że publikacja zawiera swój własny przegląd literatury. Z tego względu rozprawa jest hybrydą pomiędzy formą klasyczną a formą rozprawy opartej na publikacji. W mojej opinii jest to konsekwencja łatwości pisania raczej rzadka wśród informatyków.

Pierwsza uwaga ogólna dotyczy tytułu rozprawy. Brakuje w nim zaznaczenia, że Doktorantka przedstawia w rozprawie zbiór koncepcji z obszaru określanego jako *Social Robotics* i związanych z nimi problemów badawczych dla których przedstawia „ziarna” rozwiązań. Dlatego uważam, że dodanie na początku tytułu „*Perspective of ...*” albo „*Future of ...*” stanowiłby sygnał dla czytelnika, że jest to dopiero początkowa faza badań prac w tym obszarze. Rozprawa w mojej ocenie jest rodzajem predykcji co nie umniejsza jej wartości. Szkoda, że nie odniesiono się do dostępnych w Internecie informacji popularno- naukowych o

roli humanoidalnych robotów i ich relacjach z ludźmi. Stosunkowo niedawno częste były odwołania do testu Turinga, obecnie Chat GPT pozwala na prowadzenie rozmowy która będzie sklasyfikowana jako rozmowa z człowiekiem. Możliwe jest też strojenie robota do interakcji z daną osobą.

Druga uwaga również ogólna dotyczy tezy: *The proposed solutions for recognizing and modeling social and emotional components based on robot artificial intelligence allow for implementation and application in social robotics*. Teza to formułowanie które może być prawdziwe lub nieprawdziwe. W mojej ocenie wersja przyjęta w rozprawie ma charakter tautologii.

Przechodząc do omówienia zawartości rozprawy w rozdziale pierwszym nagrania głosowe konwertowane są na tekst, a następnie dzielone na krótkie i długie wypowiedzi te ostatnie z intencją uchwycenia kontekstu. Wypowiedzi w języku polskim są analizowane z wykorzystaniem odpowiedniego (z adnotacjami emocji) leksykonu polskiego. W kolejnym kroku teksty tłumaczone są na język angielski i analizowane z wykorzystaniem angielskiego leksykonu emocjonalnego. Analiza tekstu pod względem nastroju jest aktywnym obszarem badawczym o czym świadczą zmieszane publikacje. Uzasadnione wydaje się też badanie wpływu translacji na analizę nastroju. Można pogłębić ten aspekt stosując translacje krzyżowe lub wielokrotne. Podstawowym jest założenie, że dysponujemy tylko tekstem. Koncepcja zamiany nagrania głosowego na tekst po to aby analizować nastrój na podstawie tekstu nie wydaje się racjonalna a przynajmniej wymaga dyskusji. W rozdziale przedstawiono również poprawnie zastosowaną koncepcję modeli Markowa do wyznaczania tranzycji pomiędzy emocjami. Problem jest pozyskanie odpowiednich danych uczących reprezentujących przejścia pomiędzy emocjami.

Drugi rozdział przedstawia koncepcję wykorzystania języka Planning Domain Definition Language (PDDL) do automatycznego planowania zachowań robota w sytuacji niepewności co do stanu emocjonalnego człowieka z którym robot jest w interakcji. Wynikiem pracy są dwa przykłady takiego planowania przy założeniu, że robot swoimi zachowaniami powinien redukować negatywne emocje człowieka. Niezależnie w rozdziale przedstawione są modyfikacje heurystycznych metod przekształcających zadania planowania do zadania programowania liniowego. Podejście oceniam jako ciekawe a w odniesieniu do zarządzanie emocją innowacyjne.

Rozdział trzeci dotyczy rozpoznawanie emocji w trybie online na podstawie obrazu twarzy z wykorzystaniem wstępnie wytrenowanej sieci neuronowej. Potok przetwarzania strumienia

wideo jest opisany pobieżnie. Brak szczegółów odnośnie wyznaczania punktów kluczowych, ich podziale na punkty obrazu twarzy i sylwetki. Widać brak normalizacji niezbędnej ze względu na dowolną odległość, orientację i oświetlenie. Czy rzeczywiście wystarczy YOLO5 i Raspberry Pi.

Czwarty rozdział dotyczy wykorzystania kontrolera PID do modelowania wybranych zachowań robota, w rozpatrywanym przypadku „śledzenia” wzrokiem człowieka. W mojej ocenie trudność polega na wymaganym przetwarzaniu informacji wizyjnej z kamery robota. Przykładowo nie chodzi tylko o segmentację obrazu człowieka ale również wyznaczenie orientacji jego twarzy i kierunku patrzenia. Chodzi o to że aby nawiązać kontakt wzrokowy, co jest niemożliwe jeśli człowiek jest obrócony tyłem albo patrzy w inną stronę. W zakresie PID badania są poprawne.

Problem naukowy badany w rozprawie i jasność jego sformułowania.

Problem naukowy sformułowany i badany przez Doktorantkę przesłonięty jest wieloma wątkami ubocznymi związanymi z perspektywą robotyki socjalnej w różnych obszarach. Dlatego poniżej przedstawiam bardziej skondensowane ujęcie naukowego problemu rozprawy.

Interakcję człowiek robot humanoidalny można sformułować w kategoriach sprzężenia zwrotnego i sterowania a bardziej aktualnie uczenia ze wzmocnieniem. Robot na podstawie swoich systemów sensorycznych, głównie kamery dokonuje „pomiaru” stanu środowiska, zawężając, stanu emocjonalnego człowieka. Z dostępnego zbioru akcji robot wybiera jedną, przykładowo, może to być stan emocjonalny twarzy robota w konsekwencji stan środowiska, i znowu zawężając stan emocjonalny człowieka zmienia się. Model można wzbogacić zakładając, że człowiek ma możliwość podejmowania akcji mających wpływ na środowisko rozumiane jako stan emocjonalny robota. W takim ujęciu stan środowiska jest łącznym stanem emocjonalnym człowieka i robota. Powracając do modelu prostszego można formułować różne problemy takie jak tranzycie z dowolnego stanu emocjonalnego do pożądanego w zadanej lub minimalnej liczbie akcji których koszt mierzony jest przez przyjęty wskaźnik.

Jednym z istotnych problemów częściowych jest pomiar stanu emocjonalnego człowieka, lub również robota aktualny w bardziej rozbudowanej wersji całego problemu. Z tego powodu prowadzone są intensywne badania dotyczące rozpoznawania emocji na podstawie: obrazu twarzy, postawy, głosu. W przypadku obrazu twarzy i postawy rozpatrywane są wersje statyczna w której mamy pojedynczą klatkę z obrazem twarzy i obrazem postawy lub bardziej uzasadniona wersja dynamiczna w której dysponujemy sekwencją klatek biorąc pod uwagę że emocja twarzy jest zjawiskiem czasowym podobnie jak chód. W przypadku głosu zawsze

konieczny jest segment czasowy. Sekwencje obrazów twarzy wyekstrahowane z sekwencji klitek, podobnie jak wyekstrahowane sekwencje postaw z sekwencji klitek, oraz nagranie głosowe przekształcone do reprezentacji *cepstralnej*. Poszczególne strumienie danych (sekwencja twarzy, sekwencja postaw, głos) przewarzone są niezależnie i łączone (wczesne łączenie, późne łączenie) Niezależnie od szczegółowości nie uwzględniono w nim problemu „in wild” to znaczy niekontrolowanych warunków w których pozyskiwane są dane.

Opisany jak powyżej problem rozpoznawania stanu emocjonalnego człowieka jest tylko jednym z wybranych którego rozwiązanie jest konieczne. Pozostaje problemy „sterowania emocją” którego ciekawe rozwiązanie zostało zaproponowane z użyciem języka PDDL bez użycia funkcji kosztu/wypłaty. Kolejnym problemem do rozwiązania jest sterowanie emocjami twarzy, postawy i głosu robota. Elementem takiego rozwiązania jest sterowanie kierunkiem patrzenia robota przez odpowiednią orientację gałek ocznych co zostało pokazane w jednej z cyklu publikacji.

Ocena wiedzy w zakresie obszaru rozprawy.

Doktorantka wykazała się bardzo szeroką i jednocześnie wystarczająco szczegółową wiedzą o czym świadczy analiza literatury do całej rozprawy, analizy literatury poprzedzające każdy z rozdziałów rozprawy i na koniec analiza literatury w każdej z publikacji przyporządkowanej do rozdziału. Nie umie ocenić w jakim stopniu Doktorantka przyswoiła sobie treści cytowanych publikacji dlatego proponuję aby w okresie do obrony zapoznała się z rozprawą doktorską: Xin Chang. Human Emotion Recognition from Image and Speech using Deep Neural Networks w którym promotorem był prof. dr hab. inż. Władysław Skarbek z Politechniki Warszawskiej a w którym byłem recenzentem stąd znam dość dobrze rozprawę i proponuję dyskusję podczas obrony. Ze względu na charakter rozprawy istotnym elementem jest znajomość baz danych niezbędnych w uczeniu i równocześnie stanowiących odniesienie dla uzyskiwanych wyników. Doktorantka wymieniła przykładowe bazy co oceniam jako wystarczające.

Ocena przedstawionego rozwiązania problemu oraz metodologii prowadzenia badań.

Forma rozprawy odbiega od klasycznego schematu obejmującego: zdefiniowanie naukowego problemu badawczego oraz charakteryzujących ten problem hipotez/tez, wybór metody badawczej mającej na celu potwierdzenie hipotez lub wykazaniem prawdziwości tez. Podjęty przez Doktorantkę problem interakcja człowieka z robotem humanoidalnym został zdekomponowany na trzy podproblemy. Pierwszy to rozpoznawanie emocji człowieka na podstawie i) wypowiedzi głosowych przekształczanych do tekstu, ii) obrazu twarzy w ujęciu podobnym do biometrii twarzy, iii) sterowania procesem tranzycji pomiędzy stanami emocjonalnymi, iv) podążanie robota (wzrokiem) za człowiekiem. Każdy z wyróżnionych

problemów częściowych został poprawnie opisany, odniesiony do aktualnego stanu wiedzy w obszarze, oraz rozwiązań γ na poziomie PoC. Niestety za względu na liczbę problemów oraz ich potencjalną głębokość nie było możliwe szczegółowe i wyczerpujące rozwiązanie każdego z nich. Doktorantka jest tego świadoma o czym świadczą konkluzje w podsumowaniach poszczególnych publikacji.

Ocena oryginalności rozprawy, w tym samodzielnego i oryginalnego dorobku.

Rozprawa jest oryginalna w aspekcie podjętego problemu. Doktorantka posiada duży dorobek naukowy o szerokim spektrum mieszczącym się w obszarze robotyki społecznej. Doktorantka jest również Autorką samodzielnych publikacji. Ze względu na szerokie spektrum oryginalnego i obszernego dorobku, wybór forma „rozprawa z cyklu publikacji” w mojej ocenie nie była najlepsza. Wystarczyło zawęzić i pogłębić jeden z podproblemów a publikacje zamieścić w literaturze.

Ocena umiejętności poprawnego i przekonującego przedstawienia uzyskanych wyników.

Rozprawa świadczy, że Doktorantka posiada umiejętności poprawnego i przekonującego przedstawiania i analizy problemów. W konsekwencji „dopowiedzenia” do poszczególnych publikacji wielokrotnie przekraczają objętość publikacji co oceniam jako cechę pozytywną rozprawy z badawczego punktu widzenia i negatywną z punktu widzenia czytelnika.

Słabe strony rozprawy i uwagi dyskusyjne.

Rozprawa, ze względu na przyjętą formę „rozprawa z cyklu publikacji” jest trudna w czytaniu nie ze względu na niepoprawność merytoryczną lub narracyjną a ze względu na szerokość omawianego problemu oraz używanie terminologii w rodzaju: „*The third chapter covers research focused on a variety of methods for modeling social behavior, including the use of neural networks and machine learning*” W początkowym fragmencie rozprawy znalazłem sformułowanie „*Adaptive emotional responses are modeled using various machine learning techniques, such as reinforcement learning*” Sformułowanie powtarza się w przeglądach literatury ale brak szczegółów. Proszę o wyjaśnienie na czym miałyby polegać. Z tekstu rozprawy wynika, że Doktorantka zna dobrze dwie podstawowe konwencje opisu emocji. W tym kontekście nie rozumie podpisów pod rysunkami w których występuje słowo microexpression. Zjawisko mikroekspresji zostało wprowadzone przez Ekmana i do chwili obecnej ma porównywalne liczby zwolenników i przeciwników. Rejestracja wymaga kamery rzędu 200 fps i specjalnego oświetlenia. Być może chodziło o obrazy twarzy odpowiadające klasom emocji.

Przydatność rozprawy.

Rozwój AI spowodował, że robot potrafi chodzić jak człowiek, rozmawiać jak człowiek i generować ludzkie emocje i twarze. Wszystko to razem otwiera nowe problemy w tym robotyki społecznej. Doktorantka opisała w rozprawie problemy częściowe oraz przedstawiła zarysy ich rozwiązań co niewątpliwie czyni rozprawę przydatną w dalszym rozwoju obszaru robotyki społecznej. Rozwiązania przedstawione w rozprawie z wykorzystaniem robota OhBot i przykłady interakcji oceniam tylko jako przykłady ze względu na jakość rozmowy i jakość emocji twarzy robota. Zdaję sobie jednak sprawę, że do rozwiązania jest problem natury psychologicznej czy człowiek woli być w interakcji z robotem jak najbardziej humanoidalnym czy robotem z komiksu jak OhBot.

Wnioski końcowe.

Stwierdzam, że przedstawiony w rozprawie problem badawczy został rozwiązany. Doktorantka wykazała się opanowaniem warsztatu naukowego zarówno w zakresie bardzo dobrej znajomości stanu aktualnego jak i dobrej metodologii prowadzenia badań naukowych. Zdefiniowano proces interakcji pomiędzy człowiekiem a robotem humanoidalny, opracowano i przetestowano podstawowe, składowe algorytmy, całego procesu. Rozprawa stanowi wkład Doktorantki w rozwój robotyki społecznej jako ważnego (w przyszłości) elementu dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne. Recenzowaną rozprawę oceniam jako spełniającą wymagania stawiane rozprawom doktorskim. Wnioskuje o jej przyjęcie i dopuszczenie do publicznej obrony.

K. Ojardowski