

Recenzja rozprawy doktorskiej Pana mgr inż. Konrada Kluwaka
Modele i algorytmy detekcji znaczników w problemach analizy ruchu obiektów

Recenzowana rozprawa liczy 160 stron i napisana jest w języku polskim. Główna część rozprawy podzielona jest na trzy elementy zatytułowane kolejno: Część I, *Stan wiedzy*, Część II, *Badania własne*, Część III, *Bibliografia i publikacje*. Rozprawę rozpoczynają streszczenia w języku polskim i angielskim oraz *Wprowadzenie* zawierające: Wstęp, Tezy pracy, Zakres pracy. Rozprawę kończą: *Wykaz literatury*, *Spis rysunków*, *Spis tabel*. Podsumowując w ujęciu formalnym rozprawa jest poprawnie ustrukturyzowana a nawet innowacyjnie ze względu na szczegółowe opisy baz danych w Części III.

Na początku chciałbym przypomnieć, że rozprawa doktorska, jako element kariery naukowej ma pokazać, że osoba ubiegająca się o uzyskanie stopnia doktora umie poprawnie sformułować nowy problem naukowy, postawić istotne dla tego problemu hipotezy, sformułować tezy a następnie stosując metody naukowe rozwiązać problem to jest zweryfikować hipotezy i dowieść prawdziwości postawionych tez.

Korzystając z przypomnianego rozumienia celu i struktury rozprawy doktorskiej stwierdzam, że wypełniają je wyniki przedstawione w Części II, *Badania własne*. W każdym z trzech opisanych w tej części „eksperymentach” oznaczonych, jako E1, E2, E3 prawidłowo określono: i) nowy problem badawczy (hipotezy, tezy), ii) dane, które podlegają tagowaniu, iii) model/cechy/trybuty fragmentów danych, które powinny zostać otagowane, iv) algorytmy wyszukiwania fragmentów danych najbardziej podobnych do zdefiniowanego modelu danych, v) kryteria oceny poprawności tagowania, Dodatkowo przeprowadzono analizę wrażliwości procesu tagowania na: i) rodzaj algorytmu, ii) reprezentację danych w bazie, iii) redukcję danych.

W E1 „Tagowanie pozy w sporcie” jako problem do rozwiązania przyjęto detekcję i na tej podstawie tagowanie postaw dla 5, mistrzowskich uderzeń długim mieczem należących do dawnych europejskich sztuk walki. Ze względu na atrakcyjność i unikalność problemu cytuję nazwy jak poniżej:

1. Zornhau (Strike of Wrath) — potężne ukośne uderzenie z góry. Cios wyprowadzony z prawego ramienia w lewe ucho przeciwnika.
2. Schielhau (Squinting Strike) — uderzenie fałszywą krawędzią. Odsuwa broń przeciwnika trafiając go w ramię.
3. Zwerchhau (Cross Strike) — poziome, wysokie uderzenie celujące w głowę odbijające atak przeciwnika.
4. Krumphau (Crooked Strike) — pionowe uderzenie z góry w nadgarstki lub miecz przeciwnika.
5. Scheitelhau (Crown Strike) — wysoki, pionowy atak długą krawędzią, skierowany na czubek głowy lub twarz z jednoczesnym wycofaniem się.

Wybrany problem oceniam jako bardzo ciekawy i unikalny, ze względu na fakt, że opiera się na materiałach źródłowych z XVI wieku. Pewne zastrzeżenia mam co do koncepcji wyboru klatki na której poza jest analizowana. Stwierdzam jednocześnie, że Autor jest świadomy tego problemu. Pierwszym oryginalnym wynikiem rozprawy było utworzenie obszernej bazy danych surowych a następnie wydzielenie dla potrzeb badawczych trzech podzbiorów danych zawierających następujące atrybuty: B.M1 - pozycje markerów na ciele aktora, 117 atrybutów, B.M2 - kątów Eulera ugięcia stawów aktora, 78 atrybutów, B.M2 - kątów Eulera ugięcia stawów aktora, 78 atrybutów, B.M3 - pozycje markerów na

WPLYNEŁO
29-09-2021

RADN-IT / 158/2021

mieczu, 15 atrybutów. Istotnym i pracochłonnym było ręczne tagowanie danych po to aby w dalszych badaniach mogły być podstawą oceny jakości klasyfikacji. Tak przygotowane dane były podstawą do następnego etapu badań polegającego na weryfikacji jaki wpływ ma przyjęta reprezentacja (markery, kąty, markery na mieczu) na wynik klasyfikacji pozy związanej z uderzeniami mieczem. Wybrano odpowiednio reprezentatywny zbiór klasyfikatorów oraz kryteria klasyfikacji. Podstawy do wyboru kryteriów klasyfikacji wyjaśniono szerzej dopiero w E2 jednak sam wybór jest poprawny. Zrealizowano wystarczającą liczbę starannie opracowanych eksperymentów. W podsumowaniu brakuje prób wyjaśnień słabych wyników klasyfikatora Naiwny Bayes i KNN na danych reprezentujących kąty w stawach. Taka reprezentacja jest praktycznie jedyną w przypadku animacji tworzonych dla gier i gdyby była niewystarczająca informacyjnie jak sugerują wyniki badań to gracz widziałby percepcyjne usterki animacji a tak nie jest.

W E2 Tagowanie akcji (jednostki chorobowej pacjenta) na podstawie ścieżek chodu. W tym „eksperymentcie” sformułowano problem tagowania chodu pacjenta z wykorzystaniem sieci LSTM jako klasyfikatora ze względu na czasowy aspekt danych i spodziewaną zależność wyniku klasyfikacji od całego horyzontu ruchu, którym jest pojedynczy krok wydzielony z danych chodu. Do klasyfikacji wykorzystano dane wielomodalne (mocap, GRF, sEMG) Rozpatrzono dwa warianty pierwszy z tagami „chód poprawny”, „chód patologiczny” oraz drugi z tagami nazw schorzeń. Pod względem metodologii realizacja eksperymentu jest poprawna. W odniesieniu do E1 nie przeprowadzono analizy wrażliwości klasyfikatora, zamieszczono za to fragment doboru i strojenia sieci LSTM. Poprawnie analizowano problem oceny jakości klasyfikacji z uwzględnieniem istotnie różnych liczebności poszczególnych klas.

W E3 Tag interakcji - ergonomia ruchu w podnoszeniu pacjenta. W eksperymencie analizowano i w konsekwencji opracowano algorytm tagowania multimodalnych danych zawierających: wideo, mocap, GRF, sEMG, zebranych w procesie podnoszenia pacjenta przez pielęgniarkę. Założono dwa tagi: „nie energomiczne podnoszenie”, „energomiczne podnoszenie”. Pomimo jedynie dwóch tagów proces automatycznego odnajdowania właściwych (nie energomiczne podnoszenie, energomiczne podnoszenie) fragmentów danych jest trudny ze względu na interakcję pomiędzy osobami. Z tego względu „eksperyment” E3 oceniam najwyżej pod względem metodologii badawczej uwzględniając również wykorzystanie w konstrukcji algorytmu odnalezionych w literaturze a zaproponowanych przez M.Mullera 39 cech opartych na relacjach pomiędzy stawami, oraz implementacji tych cech powiązanej z doбором parametrów i autorską wizualizacją.

W rozprawie przedstawiono koncepcję EXTag jako wysokopoziomowego tagu utworzonego na podstawie tagów poziomu niższego. Jest to przykład wielopoziomowej hierarchicznej struktury tagów, w której poziom tagu w hierarchii wynika z uporządkowania względem modalności lub czasu. Koncepcję oceniam, jako ciekawą ponieważ hierarchiczna struktura tagów czyni otagowane w ten sposób dane bardziej elastycznymi na wyszukiwania po tagach. Użytkownik może używać tagów na różnych poziomach od najniższego np. „uderzenie pięty o podłoże” przez coraz bardziej złożone jak „krok poprawny” lub „krok patologiczny” Autor zadeklarował uszczegółowienie i rozwój zarysowanej koncepcji jako element przyszłych prac badawczych.

Szczegółowa analiza pozostałej treści rozprawy nasuwa krytyczne uwagi odnoszące się do tytułu rozprawy, jej merytorycznego układu (układ formalny jest poprawny) precyzji sformułowań a w niektórych przypadkach gramatyki języka polskiego.

Pierwsze zastrzeżenie budzi tytuł rozprawy w sformułowaniu „Modele i algorytmy detekcji znaczników w problemach analizy ruchu obiektów” Wynika z niego wprost, że wyszukiwane mają być znaczniki/tagi których przykłady są wymienione w tekście rozprawy w eksperymentach E1, E2, E3. Zadanie takie jest w oczywisty sposób trywialne. W rzeczywistości Autorowi chodzi o algorytmy automatycznego wyszukiwania fragmentów wielomodalnych danych, ogólnie przestrzenno-czasowych, które spełniają określone warunki względem wektora cech lub są „podobne” do określonego wzorca. Odnalezione



fragmenty zostaną opatrzone unikalnym tagiem/znacznikiem. Istotne jest podkreślenie, że proces tagowania powinien być realizowany automatycznie. Ręczne tagowanie danych jest bardzo pracochłonne, o czym pisze Autor w E1. Dane otagowane ręcznie są niezbędne jako referencja dla algorytmów tagowania automatycznego. Z treści rozprawy wynika, że Autor poprawnie rozumie proces tagowania jednak zamienne używanie w tekście terminów „odnajdywanie tagów” i „tagowanie” stanowi dyskomfort dla czytającego. Zastrzeżenia budzi również końcowy fragment tytułu „w problemach analizy ruchu obiektów”. Rzeczywiście Autor w jednym ze zdań wspomina, że opracowane algorytmy tagowania mogą być zastosowane w przypadku innych ruchów niż ruch człowieka ale nie stanowi to podstawy do tak szerokiego ujęcia w tytule.

Część I rozprawy sugeruje, że Autor miał zamiar napisać książkę stąd nieuzasadnione z punktu widzenia rozprawy doktorskiej jest bardzo szerokie potraktowanie tematu. Dla problemu algorytmów tagowania wystarczające jest rozumienie danych i ich formatów a nie ma znaczenia, w jaki sposób zostały pozyskane oraz jakie są alternatywy. Dodatkowo Autor nie ma zdolności literackich, stąd w rozprawie można znaleźć szereg usterek stylistycznych i gramatycznych. Nie widzę celowości ich wymieniania, rozprawę z zaznaczonymi usterkami przekazałem Autorowi. Jednak jako przestrożę na przyszłość, zalecam skrupulatne czytanie tego co się napisało i zacytuję dwa przykładowe wadliwe sformułowania: i) *Systemy bezwładnościowe opierają się na zasadzie oporu wobec zmiany ruchu. Systemy inercyjne najczęściej składają się z akceleratorów i żyroskopów. Służą one do pomiaru bezwładności i orientacji obiektu w czasie* ii) *Akcelerometr wyznacza przyspieszenia modułu w danych chwilach czasu. Żyroskop pozwala na określenie przyspieszenia w lokalnym układzie współrzędnych. Magnetometr określa położenie w układzie współrzędnych Ziemi.* Oczekuję, że Autor podczas obrony odniesie się do cytowanych jak powyżej sformułowań i wskaże, które ich fragmenty są niepoprawne i dlaczego. Na obronę Autora stwierdzam, że w innych fragmentach rozprawy tekst dotyczący IMU jest poprawny.

Uwaga, co do merytorycznego układu treści rozprawy może być uwzględniona w reedycji rozprawy do postaci książkowej. W przypadku podjęcia takiej decyzji zalecam szczegółowe zapoznanie się z naniesionymi uwagami. Zauważone usterki w zakresie precyzji sformułowań obniżają, jakość edycji rozprawy i dlatego powinny być usunięte w wersji książkowej jednak nie podważają one wartości rozprawy doktorskiej.

Część III rozprawy zatytułowana „Bibliografia i publikacje” pomimo mylącego tytułu, jeśli odnieść go do pierwszego fragmentu tej części jest bardzo wartościowa ze względu na szczegółowe i precyzyjne opisy wielomodalnych baz danych oraz procedur, w jakich zostały utworzone. Dorobek publikacyjny Autora jest znaczący, dotyczy obszaru rozprawy i obejmuje: 3 publikacje w czasopiśmie, 3 rozdziały w książkach, 12 publikacji konferencyjnych, 6 raportów. Literatura rozprawy obejmuje 194 pozycje co świadczy o sumienności Autora i dobrej znajomości obszaru. Dzięki tej sumienności Autorowi udało się odnaleźć cechy M.Mullera i konstruktywnie je zastosować. Literatura zawiera też własne publikacje Autora.

Podsumowując stwierdzam, że recenzowana rozprawa pomimo zauważonych usterek stylistycznych i ustawicznego mylenia obiektu (fragment danych) z nazwą obiektu (tag) spełnia warunki stawiane rozprawom doktorskim w sformułowaniu *Ustawy o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki* z dn. 14.03.2003 (Dz. U. nr 65, poz. 595 z późn. zm.) co uzasadniłem w analizie Części II, Badania własne. Wnoszę o dopuszczenie rozprawy do publicznej obrony.

K. Węcierski