

Prof. dr hab. inż. Bożena Kostek, prof. zw. Pol. Gdańskiej
Katedra Systemów Multimedialnych
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska

Gdańsk, dnia 14 grudnia 2010 r.

Opinia nt. rozprawy doktorskiej mgr Elżbiety Kubery pt.: „Rola atrybutów temporalnych w rozpoznawaniu instrumentów muzycznych w nagraniu wielobrzmieniowym”

Przygotowana rozprawa doktorska pani Elżbiety Kubery dotyczy zaawansowanych technik analizy, parametryzacji i rozpoznawania dźwięków muzycznych i mieści się obszarze badań naukowych, które rozwijają się w bardzo szybkim tempie. Należy podkreślić, że w zakresie badań naukowych dotyczących dziedziny wyszukiwania informacji muzycznej (ang. *Music Information Retrieval*) w ostatnich latach został dokonany ogromny postęp. Uzyskane wyniki badań światowych zaowocowały technicznymi realizacjami systemów i aplikacji uzyskujących skuteczność pozwalającą na praktyczne zastosowania. Obecnie prowadzi się dalsze badania, których celem jest m.in. uzyskanie metod pozwalających na poprawę skuteczności i efektywności rozpoznawania muzyki (np. w sensie wydajnościowym). Jest to zagadnienie niezwykle ważne zwłaszcza w kontekście wyszukiwania rozwoju baz multimedialnych, w których reprezentacje dźwięku stanowią drugi po informacjach wizualnych dominujący składnik przesyłanych treści.

Głównym celem rozprawy jest zaprezentowanie nowych cech opisujących zmiany charakterystyk dźwięku w czasie i ocena przydatności tych cech w budowie systemów klasyfikacji instrumentów muzycznych, a także możliwość zastosowania zaproponowanych wektorów cech w nagraniach wielobrzmieniowych. Dodatkowo badania te prowadzone są w kontekście systemów hierarchicznych opartych na tradycyjnej taksonomii zaproponowanej przez Hornbostela i Sachs'a oraz o podziały tworzone automatycznie w procesie klasteryzacji danych.

Doktorantka postawiła sobie za cel udowodnienie następujących tez rozprawy:

1. Analiza zmian charakterystyk dźwięku w czasie (czyli analiza temporalna) umożliwia dokładniejszy opis barwy dźwięku instrumentów muzycznych niż inne analizy nie uwzględniające informacji o sposobie zmienności cech dźwięku.

2. Lepsze wyniki klasyfikacji instrumentów w wielobrzmieniowych fragmentach muzyki uzyskuje się, gdy do zbioru dźwięków wykorzystywanych w procesie uczenia klasyfikatorów dodamy dźwięki wielobrzmieniowe, np. sztucznie utworzone miksy pojedynczych dźwięków.

3. Klasyfikacja hierarchiczna daje lepsze rezultaty w zadaniu rozpoznawania wielu instrumentów w nagraniach polifonicznych i wielobrzmieniowych niż klasyczna klasyfikacja bez wykorzystania podziału systematycznego instrumentów.

Badania naukowe dotyczące szeroko pojętego rozpoznawania dźwięku są trudne i żmudne i najczęściej mają charakter empiryczny. Dotychczas opracowane techniki są dość skomplikowane, biorą pod uwagę analizę muzyczną, akustyczną oraz psychofizjologiczne właściwości ludzkiego systemu słyszenia, a nowo proponowane techniki z konieczności muszą skierować uwagę na dodatkowe analizy. Dlatego dziedzina ta wymaga przeprowadzenia wielu badań eksperymentalnych, które przy niewielkich zaniedbaniach proceduralnych mogą łatwo prowadzić do fałszywych wniosków. Przeprowadzenie badań i ocena uzyskiwanych rozwiązań wymaga dużego wysiłku ze strony badacza.

Do głównych osiągnięć opisanych w rozprawie należy zaliczyć opracowanie oryginalnej metodyki ekstrakcji cech temporalnych dźwięku muzycznego oraz zastosowanie zaproponowanych wektorów parametrów do poprawy klasyfikacji dźwięku. Konstrukcja zaproponowanej metody parametryzacji wynika z potrzeby uwzględnienia cech dźwięku zmiennych w czasie. Można stwierdzić, że parametry zawarte w standardzie MPEG-7, choć dobrze opisujące dźwięk czy sygnał muzyczny nie koncentrują się na zmianach tych cech w czasie. Zaproponowane przez Doktorantkę metody wynikają z pogłębionej analizy zalet i wad różnych technik parametryzacji, które są znane z literatury i stosowane w praktyce.

Należy podkreślić, że w swojej pracy pani Kubera zaprezentowała rzetelną i wieloaspektową ocenę zaproponowanych rozwiązań. Ta ocena została dokonana poprzez przyjęcie pewnego algorytmu działania, który pokrótce można streścić jako kolejne kroki badań. W pierwszej kolejności podjęta jest klasyfikacja pojedynczych dźwięków instrumentów muzycznych w oparciu o typowe wektory parametrów. Eksperyment ten służy do wyboru klasyfikatora. Zaproponowane i skonstruowane wektory cech, które uwzględniają cechy temporalne, są wykorzystywane w kolejnym kroku badań. Doktorantka stosuje też okna analizy o różnej rozdzielczości i dodatkowo zmiennym nakładkowaniu, co pozwala na bardziej wydajną analizę dźwięków o różnych wysokościach. Następny eksperyment ma na celu sprawdzenie wyników klasyfikacji realizowanych przez klasyfikatory wieloetykiętowe. Kontynuacja eksperymentów idzie dalej w kierunku konstrukcji klasyfikatorów wieloetykiętowych, ale dotyczących klasyfikacji wielodźwięków. Ponieważ uzyskane wyniki nie są w pełni zadowalające, dlatego Doktorantka wykorzystuje również wektory ekstrahowane w oparciu o przykłady dźwięków zmiksowanych. Ostatnim krokiem eksperymentu badawczego jest zastosowanie klasyfikacji hierarchicznej w rozpoznawaniu dźwięków w nagraniu wielobrzmieniowym.

W szczególności w rozprawie zostały opisane i przebadane dwa warianty zastosowania zaproponowanych wektorów cech w rozpoznawaniu przy zastosowaniu klasyfikacji hierarchicznej. Pierwszy z nich opiera się na ogólnie przyjętej taksonomii instrumentów muzycznych, zaś w drugim wariantcie Doktorantka dokonuje automatycznego podziału instrumentów na grupy i podgrupy za pomocą mechanizmu klasteryzacji, uzyskując poprawę skuteczności. Na uwagę zasługuje fakt pogłębionej analizy uzyskanych wyników.

Zaproponowane rozwiązania parametryzacji uwzględniają aktualny stan nauki w zakresie analizy, parametryzacji i rozpoznawania dźwięków muzycznych. Zgromadzone źródła są adekwatne, choć budzi może niedosyt liczba pozycji z dziedziny separacji sygnałów fonicznych, a w szczególności muzycznych. Dlatego opis zawartych zagadnień w rozdziale 2 nie jest pełny, w szczególności rozdział 2.3 (Separacja źródeł i analiza jednobrzmieniowa) został przedstawiony w sposób wrywkowy, należałoby odnieść się szerzej do tych zagadnień, przywołując źródła literaturowe.

Stwierdzam, że opisane badania zostały przeprowadzone w sposób prawidłowy, choć pewien niedosyt budzi brak szerszej informacji na temat wykorzystywanych klasyfikatorów. Dla przykładu wykorzystywane metody zostały przedstawione w sposób opisowy bez podania formalnego zapisu tych metod, dodatkowo brak jest odniesienia do aspektów wydajnościowych zastosowanych algorytmów. Doktorantka uzyskała w pierwszym proponowanym eksperymencie porównywalne wyniki jakości klasyfikacji dla metody SVM i lasów losowych (ale przy dokładności rzędu 20%). Wybór algorytmu lasów losowych uzasadnia, odwołując się do prac swoich i współautorów, a przecież można było odnieść się właśnie do wydajności, skalowalności systemów czy specyfiki związanej z właściwościami klasyfikatorów (jak np. brak zjawiska „przeuczenia” w przypadku algorytmu lasów losowych). W tym kontekście odniesienia do wykorzystywanych algorytmów są mało precyzyjne, na przykład nie znalazłam w pracy opisu zastosowanych funkcji jądra w przypadku maszyny wektorów wspierających (SVM).

Badania eksperymentalne zostały wykonane w zakresie wystarczającym do prawidłowej

oceny zaproponowanych rozwiązań, choć problem został w jakimś stopniu ograniczony ze względu na brak potrzeby przygotowania algorytmu automatycznej segmentacji przykładów dźwiękowych, jak również znaną informację dotyczącą wysokości dźwięku. Jeśli wysokość dźwięku nie jest znana, to duża część wysiłku badawczego koncentruje się wokół detekcji częstotliwości podstawowej. Dokładność tej detekcji rzutuje z kolei w sposób zasadniczy na dokładność wielu parametrów wektora cech.

Uwagi krytyczne, dyskusyjne, pytania:

1. Na stronie 76 pojawia się wynik eksperymentu wstępnego (tab. 5.1), tj. klasyfikacja pojedynczych dźwięków na podstawie wektorów parametrów zawierających średnie wartości cech energetycznych, widmowych i harmoniczných. Dlaczego tylko takie cechy zostały wzięte pod uwagę w tym eksperymencie? Ponadto w opisie wykresu widnieje słowo: dokładność, która zawiera się max w 20%. Czy na pewno w tym przypadku jest mowa o jakości klasyfikacji? Jeśli tak, to uzyskany wynik praktycznie dyskwalifikuje przygotowane wektory cech. Ponadto eksperyment ten służy do wyboru klasyfikatora, więc tym bardziej pojawia się pytanie o zasadność tego testu.
2. Str. 91 – opis dotyczący mechanizmu klasteryzacji nie jest pełny, przydałby się opis formalny wykorzystywanej metody.
3. Rozdział Podsumowanie i Wnioski zawiera trzy akapity, z których jeden odnosi się w sposób bardzo ogólny do wyników rozprawy, drugi w sposób bardzo szczegółowy, w trzecim pojawiają się plany przyszłych badań. Wyraźnie w Posumowaniu brakuje szerszego określenia, co jest oryginalnym dorobkiem Autorki rozprawy i dodatkowo, że zaproponowane tezy rozprawy zostały udowodnione. Przy tej okazji należałoby zwrócić uwagę, że uwagi dotyczące udowodnienia poszczególnych tez rozprawy powinny pojawić się w podsumowaniu rozdziałów, w których Doktorantka udowadnia te tezy.
4. Sadzę, że rozprawa zyskałaby jeszcze na wartości, gdyby Doktorantka przeprowadziła testy subiektywne (dla wybranego eksperymentu) i pokazała, jak dla określonych przykładów radzi sobie z rozpoznawaniem dźwięków ekspert. Porównanie takie dałoby odniesienie czy uzyskane wyniki klasyfikacji są zbieżne z odczuciami subiektywnymi ekspertów czy też odbiegają w znaczny sposób.

Szczegółowe uwagi krytyczne:

1. Tekst rozprawy został napisany poprawnie, a stosowana nomenklatura jest w zasadzie prawidłowa, choć Doktorantka wprowadza również własne tłumaczenie wyrażen anglojęzycznych, co w niektórych przypadkach nie jest konieczne lub nieuzasadnione ze względu na powszechnie przyjęte i stosowane w literaturze wyrażenia (np. współczynniki mel-cepstralne, tłumaczenie własne podane przez Doktorantkę - współczynniki cepstralne widma w skali melowej, macierz konfuzji – wyrażenie to funkcjonuje formalnie jako macierz błędów, itd.).
2. W tekście rozprawy zauważyłam dość liczne usterki interpunkcyjne (w szczególności brak przecinków oddzielających zdania składowe w zdaniach złożonych oraz nadużywanie zdań równoważnikowych, w których brak jest orzeczenia). Dodatkowo przyjęta I os. liczby mnogiej powoduje, że zastosowany tzw. „styl książkowy” w rozprawie doktorskiej brzmi pompatycznie. Ze względu na drugorzędny charakter usterek, nie mają one większego znaczenia dla oceny rozprawy, ale jednak sprawiają, że Czytelnik je zauważa.
3. Układ rozdziałów oraz poszczególnych treści wewnątrz rozdziałów jest logiczny i czytelny,

choć pojawiają się nadmiarowe opisy zagadnień podstawowych, które dodatkowo nie zawsze są przypisane do właściwej kategorii, np. podrozdział 1.1.1 (str. 8, 9) nosi tytuł „Subiektywne cechy dźwięku”, a zawarte w nim zostały takie mierzalne charakterystyki dźwięku, jak ciśnienie akustyczne, prędkość rozchodzenia się dźwięku, częstotliwość, itd. Z kolei, jak wcześniej wspomniano brak jest bardziej formalnego opisu stosowanych algorytmów.

4. Odniesienie do klasyfikatora *randomForest* i oprogramowania R znajduje się na stronie 41, ale bez odniesienia do źródła. Odwołanie do literatury pojawia się na stronie 79, ale tylko w postaci cytowania. Warto było przywołać choć podstawowe cechy tego oprogramowania, jego funkcjonalności, itp.
5. Podobnie wykorzystywana baza muzyczna RWC pojawia się na str. 75, dopiero później jest przywołane źródło.
6. Część wyników zostało podane w tabelach w procentach. Ponieważ jednak Doktorantka przytacza miary bardziej szczegółowe, to warto byłoby pokazać liczbę przykładów rozpoznanych odpowiednio jako prawdziwe i fałszywe wskazania klasyfikatora i liczbę przykładów, zwiększyłyby to czytelność wyników.
7. Str. 78 i 81 – Doktorantka w opisie mówi o porównaniu wyników z wykorzystaniem miary jakości klasyfikacji i miary F (średnia harmoniczna), mówiąc, że w przybliżeniu te dwie miary są równe. Ponieważ Doktorantka stosuje obie te miary (ale nie równocześnie), to wprowadza to pewien chaos w prezentacji wyników. Ponadto ze względu na fakt, że większość badaczy stosuje typową metodę prezentacji wyników, czyli jakość klasyfikacji, więc nie ma możliwości bezpośredniego porównania z wynikami uzyskanymi w literaturze.
8. Sposób odniesienia do Bibliografii, np. str. 18: „[92] i [10]”, dlaczego pozycje te zostały przywołane w odwrotnej kolejności?, usterka ta pojawia się w rozprawie w kilku miejscach.
9. Większość rysunków i tabel pojawia się w tekście zanim zostanie umieszczone do nich odwołanie. W tej sytuacji bezpośrednio pod tabelą (czy rysunkiem) powinny się pojawić odniesienia do zawartych symboli (tj. powinno być wyjaśnione ich znaczenie).

Podsumowując ocenę rozprawy, pragnę zaznaczyć, że nie stwierdzam żadnych istotnych jej braków. Oczywiście, jak zawsze w pracach o charakterze konstrukcyjnym, pewne rozwiązania i parametry zostały przyjęte arbitralnie, ale w omawianej rozprawie zgodnie z doświadczeniem z zakresu analizy dźwięków muzycznych. Natomiast na szczególną uwagę zasługuje próba klasyfikacji z wykorzystaniem analizy hierarchicznej, a w szczególności przy zastosowaniu podziału uzyskanego w wyniku automatycznej klasteryzacji. Szkoda tylko, że zastosowana metoda została podana bez szczegółów i pełnego uzasadnienia

W konkluzji recenzji stwierdzam, że rozprawa mgr Elżbietę Kubereę przedstawia oryginalne rozwiązania ważnych problemów naukowych i niewątpliwie spełnia warunki stawiane rozprawom doktorskim przez odpowiednie przepisy. W związku z tym proszę o dopuszczenie jej autorki do publicznej obrony.

prof. dr hab. inż. Bożena Kostek, prof. zw. Pol. Gdańskiej

