

prof. dr hab. Piotr Porwik
Instytut Informatyki
Uniwersytet Śląski, Katowice
Wydział Nauk Ścisłych i Technicznych
ul. Będzińska 39, 41-200 Sosnowiec

Katowice 10.02.2025

OCENA ROZPRAWY DOKTORSKIEJ

*"Algorithmic methods for detecting the tempo
and time signatures of music tracks"*

Pana mgra inż. Jeremiaha Oluwagbemi Ambibola

W piśmie Przewodniczącego Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej dnia 27.11.2024 zostałem poinformowany, że zgodnie z decyzją Rady, zostałem powołany na recenzenta rozprawy doktorskiej Pana Jeremiaha Oluwagbemi Ambibola.

Promotorem pracy doktorskiej jest dr hab. inż. Paweł Kasprowski z Katedry Informatyki Stosowanej na Wydziale Informatyki i Telekomunikacji Politechniki Śląskiej przy współudziale konsultanta dr inż. Daniela Kostrzewy.

Ocena tematyki rozprawy

Rozprawa dotyczy wykrywania tempa i metrum w utworach muzycznych przy użyciu metod algorytmicznych, uczenia maszynowego i głębokiego uczenia. Tematyka jest więc aktualna, interdyscyplinarna i łączy informatykę, inżynierię dźwięku i sztuczną inteligencję, co podkreśla jej potencjał aplikacyjny.

W rozprawie doktorskiej Autor propozuje nowe sposoby analizy utworów muzycznych oparte na połączeniu metod klasycznych i głębokiego uczenia, co może prowadzić do ulepszeń w dziedzinie analizy dzieła muzycznego, ponieważ wykrywanie metrum i tempa ma szerokie zastosowania w:

1. Systemach rekomendacji muzycznej (np. Spotify, Apple Music).
2. Automatycznej transkrypcji muzycznej, wspomagającej muzykologów i producentów muzycznych.
3. Algorytmach analizy emocji i stylu muzycznego.
4. Rozwoju systemów sztucznej inteligencji do generowania muzyki.

Ocena rozprawy będzie przeprowadzona z uwzględnieniem zaproponowanych czterech hipotez badawczych, które w ostatnim rozdziale rozprawy są przez Autora walidowane.

Hipoteza 1 (H1) „Wykorzystanie macierzy podobieństwa MFCC (MFCC Similarity Matrix - MFCCSM) jako narzędzia do detekcji metrum może znacząco poprawić dokładność identyfikacji sygnatur czasowych w muzyce.”

Hipoteza 2 (H2) „Algorytm genetyczny może być użyty do optymalizacji modelu opartego na MFCC w celu poprawy detekcji metrum.”

Hipoteza 3 (H3) „Zastosowanie sieci neuronowej ResNet pozwoli na skuteczną detekcję metrum w utworach muzycznych.”

Hipoteza 4 (H4) „Model zespołowy (ensemble) oparty na głębokim uczeniu zapewnia lepsze wyniki niż pojedyncze klasyfikatory.”

Charakterystyka rozprawy i jej zakres

Lektura rozprawy sprawia wrażenie, że Autor przedstawia do oceny przewodnik i zbiór podstawowych, znanych faktów związanych z tematem rozprawy. Doktorant robi to w sposób ogólny, rozwlekły i bezreflexyjny. Mamy np. porozdział 3.4 "Machine Learning Methods", a za chwilę rozdział 4 o tym samym tytule, gdzie informacje się powtarzają. Trudno wyłuskać z tekstu rozprawy indywidualny wkład Autora w rozwój dziedziny. Praca składa się z 9 rozdziałów. Rozdziały 1-4 to obszernie wprowadzenie w tematykę zagadnienia na bazie dostępnej wiedzy literaturowej. Rozdział 5 zawiera opis własnej bazy danych. Rozdziały 6 oraz 7 to 2, 3 stronicowe krótkie teksty na tematy bardzo ważne, z punktu widzenia tematyki rozprawy. Te rozdziały zostały jednak opracowane bez należytej staranności i precyzji wywodów. Rozdziały 8 i 9 przedstawiają ilościowo-jakościowe osiągnięcia Autora rozprawy oraz wnioski.

Problem naukowy rozprawy

Podstawowym problemem naukowym rozprawy jest opracowanie skutecznych metod algorytmicznych do wykrywania tempa i metrum utworów muzycznych na podstawie sygnałów audio. Tradycyjne podejścia, oparte na analizie MIDI, okazują się niewystarczające, ponieważ w rzeczywistych zastosowaniach dominują cyfrowe nagrania audio, które są znacznie bardziej złożone pod względem struktury rytmicznej i zawierają dodatkowe zniekształcenia, takie jak szumy, dynamika wykonania i zmiany tempa. Rozwój metod analizy audio wymaga połączenia wiedzy z zakresu przetwarzania sygnałów, uczenia maszynowego i teorii muzyki. Autor przedstawia metody, które mogą znaleźć zastosowanie w automatycznym wykrywaniu metrum, które uwzględniają zarówno strukturalne aspekty rytmu, jak i możliwości technologiczne sztucznej inteligencji. Przeprowadzone badania mają na celu stworzenie uniwersalnego modelu, który będzie mógł znaleźć zastosowanie w różnych domenach – od muzykologii po inżynierię dźwięku.

Analiza utworu muzycznego jest zadaniem trudnym, gdyż wymaga uwzględnienia kilku aspektów:

- obserwacji zmienności rytmicznej – muzyka obejmuje szeroką gamę me-

trum, zarówno regularnych (np. 3/4, 4/4), jak i nieregularnych (np. 5/4, 7/4),

- ograniczeń klasycznych metod DSP, gdyż techniki oparte na prostych filtrach czy transformacjach Fouriera nie są wystarczająco precyzyjne w warunkach rzeczywistych nagrań muzycznych. Trzeba więc poszukiwać nowych skuteczniejszych metod,
- brak odpowiednich zbiorów danych – dotychczasowe bazy danych nie zawierają wystarczająco zróżnicowanych przykładów metrum dla uczenia modeli głębokiego uczenia. Trzeba więc budować nowe bazy referencyjne o dużej liczbie zróżnicowanych przykładów metrum, szczególnie dla modeli głębokiego uczenia.

Główne pytanie badawcze można więc sformułować następująco: *„Jakie algorytmy i techniki pozwolą na skuteczną detekcję tempa i metrum utworów muzycznych bezpośrednio z sygnału audio?”*

W realizacji tego celu zastosowano podejścia wykorzystujące zarówno klasyczne metody przetwarzania sygnałów, jak i techniki uczenia maszynowego oraz uczenia głębokiego. Szczególny nacisk położono na optymalizację procesu ekstrakcji cech oraz integrację różnych modeli klasyfikatorów w celu zwiększenia odporności modelu klasyfikatora na zmienność sygnału muzycznego. Przeprowadzona analiza ma także na celu poprawę dokładności klasyfikacji i dostosowanie modeli klasyfikatorów do różnorodnych gatunków muzycznych.

Celem Autora jest opracowanie i wdrożenie algorytmów umożliwiających automatyczne wykrywanie tempa i metrum utworów muzycznych w sposób, który może znaleźć zastosowanie w rzeczywistych systemach analizy muzycznej, takich jak systemy rekomendacji, oprogramowanie do transkrypcji muzycznej czy inteligentne syntezatory dźwięku. Opracowane rozwiązania opierają się na połączeniu technik przetwarzania sygnałów z metodami sztucznej inteligencji, co poprawia jakość analizy nagrań audio.

Autor proponuje zastosowanie własnych rozwiązań, ale brak bezpośredniego porównania z istniejącymi systemami MIR (np. Essentia, Sonic Visualiser) osłabia jednoznaczność wniosków. Autor proponuje istotne ulepszenia w stosunku do metod klasycznych, ale przewaga ulepszeń nad narzędziami komercyjnymi nie została dostatecznie wykazana.

Ocena wyników przedstawionych w rozprawie

Rozprawa wnosi nowe rozwiązania w zakresie detekcji metrum i tempa, jednak nie jest przełomowa, ponieważ bazuje na istniejących metodach lub ich modyfikacjach. Do największych osiągnięć Autora rozprawy zaliczam:

- a) Utworzenie nowej bazy danych Meter2800 składającej się z połączenia bazy własnej oraz baz FMA, GTZAN i MAG. Baza zawiera 2800 rekordów. Jest to jednak relatywnie mała baza na potrzeby uczenia głębokiego.

- b) Eksperymentalne porównanie klasycznych i głębokich modeli uczenia maszynowego w kontekście rozpoznawania metrum i tempa w plikach muzycznych.
- c) Optymalizację metod detekcji metrum, choć nie jest jasne, jak bardzo przewyższa dotychczasowe podejścia.
- d) Wprowadzenie do obliczeń własnej macierzy podobieństwa MFCCSM współczynników cepstralnych uzyskanych na podstawie cepstralnych częstotliwości mel (MFCC).

Przedstawione wyniki badawcze w rozprawie doktorskiej wskazują na różne aspekty skuteczności proponowanych metod wykrywania metrum w utworach muzycznych. Poniżej znajduje się ich szczegółowa analiza, uwzględniająca wartość naukową, wiarygodność, mocne strony oraz ograniczenia.

A) Wyniki oparte na metodach DSP.

Autor porównał skuteczność klasycznych metod przetwarzania sygnałów dźwiękowych (ASM, BSM) oraz nowej macierzy podobieństwa MFCC (MFCCSM). Eksperymenty przeprowadzono na zbiorze Meter2800, a ocena modeli obejmowała klasyfikację binarnej (2 klasy) i wieloklasowej (4 klasy).

Dla klasyfikacji 2 klas (3/4 i 4/4):

ASM: 53% , BSM: 50%, MFCCSM: 55% (najlepszy wynik)

Dla klasyfikacji 4 klas (3/4, 4/4, 5/4, 7/4):

ASM: 51%, BSM: 49%, MFCCSM: 53% (najlepszy wynik)

Wyniki pokazują nieznaczną przewagę dla MFCCSM w stosunku do metod klasycznych, ale osiągnięta dokładność (55% w najlepszym przypadku) jest zbyt niska, aby uznać ją za znaczący postęp. Nie przeprowadzono również testów istotności statystycznej, aby zweryfikować czy różnice są znaczące.

B) Wyniki uczenia maszynowego.

Autor ocenił klasyczne algorytmy ML: SVM, KNN, Naïve Bayes i Random Forest.

Dla klasyfikacji 2 klas:

SVM: 86.67%, KNN: 84.83%, Naïve Bayes: 84.50%, Random Forest: 86.00%.

Dla klasyfikacji 4 klas:

SVM: 74.29%, KNN: 70.43%, Naïve Bayes: 66.00%, Random Forest: 72.57%.

Wyniki klasycznych algorytmów ML są znacznie lepsze niż metody DSP.

C) Wyniki głębokiego uczenia.

Testowano sieci neuronowe: CNN, CRNN, ResNet18, ResNet18-LSTM na cechach MFCC i spektrogramach.

Dla klasyfikacji 2 klas:

(MFCC) - CNN: 89.00%, CRNN: 0.88%, ResNet18: 88.50%, ResNet18-LSTM: 89.00%

(Spektrogramy) - CNN: 98.00%, CRNN: 0.99%, ResNet18: 99.67%, ResNet18-LSTM: 98.50%.

Dla klasyfikacji 4 klas:

(MFCC) - CNN: 81.29%, CRNN: 0.80%, ResNet18: 85.14%, ResNet18-LSTM: 86.00%

(Spektrogramy) - CNN: 82.86%, CRNN: 0.86%, ResNet18: 90.71%, ResNet18-LSTM: 85.57%

Wyniki pokazują wyraźną przewagę głębokiego uczenia nad innymi metodami, szczególnie gdy używane są spektrogramy. ResNet18 osiągnął najlepsze wyniki (do 99.67%), co potwierdza skuteczność głębokiego uczenia w analizie metrum muzycznego.

D) Wyniki metod zespołowych.

Autor testował różne techniki łączenia modeli:

Weighted Voting (głosowanie ważone)

k-Ranked Voting

Bayesian Model Averaging (BMA)

Stacking (XGB, GB, CatB, RF)

Najlepsze wyniki osiągnięto przy zastosowaniu Weighted Voting, BMA lub Soft Voting:

(MFCC) - F-Score = 0.8543, Accuracy = 0.8571

(Spektrogramy): F-Score = 0.9967, Accuracy = 0.9967.

Modele zespołowe, zwłaszcza przy użyciu spektrogramów, niemal perfekcyjnie klasyfikują metrum, co stanowi najsilniejszy dowód na skuteczność metod zespołowych.

Analizując wszystkie zastosowane przez Autora techniki analizy utworów muzycznych można zauważyć, że wprowadzenie nowej macierzy podobieństwa MFCCSM, co miało być innowacyjnym rozwiązaniem, nie podniosło skutecz-

ności rozpoznawania metrum, co wydaje się być rozczarowujące.

Pan mgr inż. Jeremiaś Oluwagbemi Ambibola posiada niewątpliwie szczegółową wiedzę na temat metod analizy plików muzycznych przy pomocy metod uczenia maszynowego. Doktorant potrafi również projektować eksperymenty badawcze oraz formułować konkluzje na temat osiągniętych rezultatów badań.

Krytyczna ocena wyników

Uwagi na temat wyników uzyskanych na podstawie uzyskanych badań skategoryzować można w postaci mocnych i słabych stron rozprawy.

Pozytywne (mocne) strony dysertacji:

Autor wskazał, że modele uczenia głębokiego, takie jak ResNet18 i modele zespołowe, znacznie przewyższają klasyczne podejścia oparte na klasycznych klasyfikatorach uczenia maszynowego i technikach DSP. Wśród wymienionych metod największą skuteczność wykazują techniki oparte na spektrogramach. Doktorant przeprowadził też wartościowe porównanie różnych technik analizy plików muzycznych, porównując przydatność metod DSP, ML, DL i modeli zespołowych, co wzmacnia wartość tych badań.

Autor, Jeremiaś Oluwagbemi Abimbola, ma na swoim koncie kilka publikacji związanych z tematyką rozprawy doktorskiej. Jego dorobek koncentruje się głównie na detekcji metrum i tempa w muzyce oraz zastosowaniu metod sztucznej inteligencji w analizie dźwięku. Publikacje autora znalazły się w renomowanych czasopismach i konferencjach, takich jak EURASIP Journal on Audio, Speech, and Music Processing oraz Sensors, co świadczy o pewnym stopniu uznania w środowisku naukowym. Dorobek autora jest tematycznie spójny i wskazuje na konsekwentne ukierunkowanie badań autora.

Praca wykazuje również pewne ograniczenia i wady. Autor nie przeprowadził nawet najprostrzych testów statystycznych w badaniach porównawczych różnych metod analizy danych audio. Wyniki różnych strategii dawały niekiedy bardzo zbliżone rezultaty liczbowe, to np. obliczenie odchylenia standardowego dawałoby już jakąś wiedzę na temat istotności tych wyników. Wyraźnie brakuje także testów istotności statystycznej np. testu Wilcoxona, co by jednoznacznie rozstrzygało czy różnice między metodami są statystycznie znaczące.

Brakuje również porównań z istniejącymi narzędziami MIR – nie wiadomo więc jak modele wypadają w stosunku do popularnych narzędzi, takich jak Essentia czy Sonic Visualiser.

W pracy znaleźć można również usterki we wszystkich wzorach rozdziału 9, gdzie zmienna N jest błędnie opisywana w tekście jako zmienna M .

Uwagi krytyczne, nie wpływają na moją pozytywną ocenę rozprawy doktorskiej mgr inż. Jeremiaha Oluwagbemi Ambibola.

Wnioski końcowe

Podsumowując przedstawioną powyżej szczegółową charakterystykę rozprawy doktorskiej Pana mgr inż. Jeremiaha Oluwagbemi Ambibola, stwierdzam, że praca spełnia wymagania stawiane pracom doktorskim z obszaru nauk technicznych w myśl odpowiednich przepisów ustawowych, tj. Ustawy Prawo o szkolnictwie wyższym i nauce, Dz. U. 2023, poz. 742 z późn. zm.

Wnoszę o dopuszczenie opiniowanej rozprawy doktorskiej do dalszych etapów przewodu doktorskiego.