

dr hab. inż. Joanna Domańska
Instytut Informatyki
Teoretycznej i Stosowanej PAN
ul. Bałtycka 5, 44-100 Gliwice

Gliwice, 31 stycznia 2025 r.

Recenzja rozprawy doktorskiej

mgr. Jeremiaha Oluwagbemi Abimbola

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

Cel, zakres i charakter rozprawy

Przedstawiona do recenzji rozprawa doktorska została wykonana pod kierunkiem naukowym dr hab. inż. Pawła Kasprowskiego, profesora Politechniki Śląskiej. Promotorem pomocniczym jest dr inż. Daniel Kostrzewa.

Rozprawa dotyczy problematyki klasyfikacji utworów muzycznych poprzez analizę tempa i metrum. Tempo określa prędkość utworu w uderzeniach na minutę (BPM), natomiast metrum definiuje jego strukturę rytmiczną. Autor rozprawy jako cel pracy wskazał zbadanie skuteczności różnych technik ekstrakcji cech, metod przetwarzania sygnału oraz algorytmów uczenia maszynowego w detekcji tempa i metrum.

W ramach pracy autor rozprawy zamierzał opracować nowatorskie metody algorytmiczne do wykrywania tempa i metrum, wykorzystując zarówno klasyczne techniki przetwarzania sygnału, jak i algorytmy uczenia maszynowego. Autor formułuje cztery hipotezy badawcze dotyczące skuteczności różnych metod:

1. Wykorzystania Mel-Frequency Cepstral Coefficient (MFCC) do detekcji metrum;
2. Optymalizacji algorytmów detekcji poprzez zastosowanie algorytmów genetycznych;
3. Skuteczności głębokich sieci neuronowych ResNet w rozpoznawaniu metrum;
4. Lepszej efektywności modeli zespołowych (ensemble) w porównaniu do pojedynczych klasyfikatorów.

Zawartość rozprawy

Forma drukowana recenzowanej rozprawy obejmuje 146 stron. Praca została napisana w języku angielskim i składa się ze wstępu, ośmiu rozdziałów, podsumowania pracy oraz wykazu literatury. Tytuł odzwierciedla zawartość pracy, a przyjęty układ

pracy jest prawidłowy. Autor rozprawy numeruje rozdziały wliczając do numeracji Wstęp oraz Podsumowanie.

Wstęp składa się z 5 sekcji. Pierwsza sekcja zawiera podstawy teoretyczne, które wyjaśniają w jaki sposób atrybuty utworów muzycznych, takie jak tempo i metrum, wpływają na charakter utworu. Autor rozprawy opisuje w tej sekcji w jaki sposób te atrybuty mogą służyć do podziału utworów wg gatunku, nastroju czy też stylu. Sekcja druga porównuje dwa typy reprezentacji dźwięku: MIDI oraz audio pod kątem przydatności do detekcji metrum. Kolejna sekcja zawiera poprawnie sformułowany cel rozprawy. W sekcji czwartej autor opisuje hipotezy badawcze, które dotyczą różnych podejść do detekcji metrum i tempa w utworach muzycznych. Są one oparte na zastosowaniu technik przetwarzania sygnału, klasycznych metod analizy muzycznej oraz algorytmów uczenia maszynowego i głębokiego. Każda z hipotez odpowiada na konkretne aspekty celu pracy i testuje różne podejścia, które mogą przyczynić się do poprawy dokładności detekcji metrum i tempa. Sekcja piąta opisuje metodologię badawczą wykorzystaną w ramach pracy. Obejmuje ona zarówno klasyczne metody analizy muzycznej (ASM, BSM), jak i nowoczesne techniki uczenia maszynowego, a także sieci neuronowe. Opisano również bazy danych, które wykorzystano do trenowania modeli. Przedstawiono również zastosowanie metod wykrywania tempa i metrum, które obejmuje zastosowanie w systemach rekomendacji muzycznych, automatycznej transkrypcji muzyki oraz edukacji muzycznej.

Rozdział drugi składa się również z 5 sekcji i jest rozdziałem teoretycznym omawiającym różne podejścia do klasyfikacji muzyki. Sekcja pierwsza dotyczy klasyfikacji gatunków muzycznych. Omawia klasyczne podejścia wykorzystujące cechy niskopoziomowe (np. tekstura barwy, rytm) oraz nowoczesne metody bazujące na uczeniu głębokim. Sekcja druga omawia detekcję tempa, które jest kluczowym parametrem rytmicznym wpływającym na klasyfikację gatunkową i percepcję muzyki. Opisane zostały w tej sekcji tradycyjne metody oparte są na analizie periodyczności sygnału oraz nowoczesne algorytmy wykorzystujące uczenie maszynowe i analizę widmową sygnału. Sekcja trzecia dotyczy klasyfikacji nastroju muzycznego. Metody klasyczne wykorzystują cechy akustyczne (np. rytm, barwa), modele ML analizują teksty piosenek i ich sentyment (NLP + MIR), deep learning (CNN, RNN) umożliwia automatyczne wykrywanie wzorców emocjonalnych. Sekcja czwarta omawia klasyfikację instrumentacji czyli rozpoznawanie instrumentów muzycznych w utworze, natomiast sekcja piąta dotyczy detekcji wokalu w utworze muzycznym. W obu tych sekcjach podkreślono wysoką skuteczność detekcji przy wykorzystaniu głębokich sieci neuronowych.

Rozdział trzeci jest również rozdziałem teoretycznym, analizującym różne podejścia do wykrywania metrum w utworach muzycznych. W rozdziale przedstawione zostały tradycyjne metody cyfrowego przetwarzania sygnału (DSP), oparte na matematycznych transformacjach, np. transformaty Fouriera, auto-korelacja, śledzenie bitów (beat tracking). Metody te są zazwyczaj regułowe, bazujące na analizie struktury sygnału. Opisano szerzej dwie metody wykorzystane w niniejszej pracy: model ASM (Audio Similarity Matrix), który porównuje powtarzające się wzorce rytmiczne w utworze, tworząc macierz podobieństwa między fragmentami utworu, celem analizy odległości euklidesowych między ramkami sygnału oraz model BSM (Beat Similarity Matrix), który analizuje rytm na poziomie uderzeń (beats) zamiast pełnych taktów, stosując detekcję onsetów (momentów rozpoczęcia dźwięku lub uderzenia w nagraniu audio) do analizy podobieństwa między sąsiadującymi beatami. Rozdział trzeci przedstawia również wykrywanie metrum przy wykorzystaniu metod ML, które uczą się na dużych zbiorach danych, wykrywając wzorce i zależności rytmiczne.

Rozdział czwarty jest również rozdziałem teoretycznym przedstawiającym podstawy metod ML i DL stosowanych do detekcji metrum. W rozdziale tym przedstawione zostały modele ML takie jak SVM, Random Forest, KNN oraz klasyfikator probabilistyczny bazujący na twierdzeniu Bayesa (Naive Bayes). Modele te charakteryzują się wysoką skutecznością w przypadku dobrze zdefiniowanych cech wejściowych (ustrukturyzowanych danych). W rozdziale tym opisane zostały również głębokie sieci neuronowe, które pozwalają na automatyczne wykrywanie złożonych wzorców w danych audio, eliminując konieczność ręcznej ekstrakcji cech. Doktorant opisał sieci w tym rozdziale CNN, CRNN, ResNet oraz hybrydę ResNet i LSTM.

Rozdział piąty opisuje autorską pracę doktoranta, polegającą na stworzeniu zbioru danych METER2800, dedykowanego do detekcji metrum. Autor dokonał przeglądu popularnych zbiorów danych stosowanych w MIR (Music Information Retrieval), podkreślając ich ograniczenia w kontekście detekcji metrum. W rozdziale tym opisano szczegółowo proces tworzenia autorskiej bazy danych METER2800 oraz proces ekstrakcji cech akustycznych z nagrań, który obejmował wydobycie kluczowych niskopoziomowych cech dźwiękowych, często wykorzystywanych w analizie muzyki i mowy.

Rozdział szósty przedstawia metodę MFCCSM (Mel-Frequency Cepstral Coefficient Similarity Matrix), której celem jest poprawa dokładności detekcji metrum. Metoda ta rozwija wcześniejsze podejścia (ASM i BSM), wykorzystując MFCC do analizy rytmicznych i akcentowych właściwości utworów muzycznych. Główna różnica polega na lepszym uchwyceniu barwy dźwięku, akcentowania i niuansów rytmicznych, co przekłada się na wyższą dokładność klasyfikacji metrum.

Rozdział siódmy przedstawia optymalizację modelu MFCCSM (Mel-Frequency Cepstral Coefficient Similarity Matrix) w celu poprawy jego skuteczności w detekcji metrum. Doktorant wykorzystał dwie kluczowe techniki: optymalizację Bayesowską i algorytmy genetyczne (GA), które pozwoliły na znaczącą poprawę dokładności klasyfikacji metrum w porównaniu do modelu bazowego. Rozdział opisuje szczegółowo obie techniki optymalizacji. Przedstawiono również wyniki eksperymentów - Doktorant porównał skuteczność modeli przed i po optymalizacji, uzyskując znaczącą poprawę dokładności.

Rozdział ósmy przedstawia porównanie klasycznych i głębokich modeli uczenia maszynowego w detekcji metrum przy użyciu zbioru METER2800. Autor testuje cztery klasyczne modele ML (SVM, Random Forest, Naive Bayes, KNN) oraz cztery modele głębokiego uczenia (CNN, CRNN, ResNet18, ResNet18-LSTM). Wyniki eksperymentów pokazują, że metody głębokiego uczenia przewyższają klasyczne algorytmy, szczególnie w zadaniu detekcji metrum na danych wieloklasowych.

Rozdział dziewiąty przedstawia zastosowanie technik Ensemble Learning w celu poprawy dokładności detekcji metrum. Doktorant testował kilka metod łączenia modeli, aby znaleźć najlepszą strategię klasyfikacji metrum. Doktorant testował takie metody łączenia modeli, jak Soft Voting, Weighted Voting, K-Ranked Voting, Stacked Voting i Bayesian Model Averaging (BMA), a także optymalizował ważone głosowanie za pomocą algorytmów genetycznych (GA). Wyniki eksperymentów pokazują, że modele zespołowe przewyższają pojedyncze modele, szczególnie w zadaniach wieloklasowej klasyfikacji metrum.

Podsumowanie pracy (oznaczone jako rozdział dziesiąty) zawiera skrócony opis głównych celów pracy, uzyskanych w ramach pracy wyników oraz wniosków płynących z przeprowadzonych badań.

Bibliografia zawarta w rozprawie składa się z 167 pozycji. Cytowane w tekście rozprawy pozycje oraz analiza ich zawartości potwierdzają dostateczną znajomość stanu wiedzy doktoranta - związanego z tematyką rozprawy.

Ocena rozprawy

Rozprawa doktorska dotyczy zagadnień związanych z metodami detekcji metrum w utworach muzycznych. Praca ma na celu analizę skuteczności różnych podejść do klasyfikacji metrum, w tym zarówno klasycznych metod przetwarzania sygnału (ASM, BSM), jak i zaawansowanych modeli opartych na sieciach neuronowych.

Autor rozprawy dokonał prawidłowej analizy stanu wiedzy na temat detekcji metrum, klasyfikacji muzyki oraz metod ML/DL stosowanych w analizie sygnału audio. Przedstawił analizę istniejących metod detekcji metrum (ASM, BSM) oraz w prawidłowy sposób wskazał ich ograniczenia. Zaprojektował i opracował nowy zbiór danych – METER2800, który zawiera ręcznie oznaczone nagrania muzyczne. Doktorant dokonał porównania skuteczności klasycznych metod ML i głębokiego uczenia w analizie metrum, w tym algorytmów takich jak SVM, KNN, Random Forest, CNN, CRNN, ResNet18, ResNet18-LSTM. W ramach przeprowadzonych badań przedstawił optymalizację detekcji metrum poprzez zastosowanie algorytmów genetycznych (GA) i optymalizacji Bayesowskiej (BO). W ramach pracy doktorskiej autor rozprawy przeprowadził również badania wpływu technik Ensemble Learning na poprawę dokładności detekcji metrum, w tym metod Stacking, Weighted Voting i Bayesian Model Averaging (BMA).

Hipotezy postawione w pracy są oryginalne, a uzyskane i opisane w pracy wyniki potwierdzają osiągnięcie celu pracy.

Praca została napisana w sposób przyjazny dla czytelnika, bez zauważalnych dużych niedociągnięć edycyjnych. Można zauważyć drobne błędy formatowania w bibliografii (niespójności w zapisach nazwisk - czasem pełne imiona, czasem tylko inicjały). Czytelność i przejrzystość struktury pracy można by było poprawić rezygnując z niektórych zbyt krótkich sekcji, m.in. sekcji 1.3, która zawiera tylko 5 linii tekstu. Rozdział 6 jest również nieproporcjonalnie krótki - w stosunku do pozostałych rozdziałów pracy.

Przedstawione poniżej uwagi i spostrzeżenia nie wpływają na pozytywną ocenę merytorycznej części pracy oraz nie podważają dobrej oceny umiejętności autora do poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników. Oczekuję jednakże od doktoranta ustosunkowania się do poniższych uwag podczas publicznej obrony pracy doktorskiej:

- niektóre modele mogą działać lepiej na czystych nagraniach niż na utworach o słabej jakości audio. Niedosyt budzi brak opisu jakości dźwięku w bazie METER2800 – uwzględniającego jakość plików w GTZAN, FMA-medium i Magna-TagATune. Brakuje w pracy dyskusji dotyczącej odporności modeli na szum i zniekształcenia.
- autor opisuje w pracy regularne oraz nieregularne metrum podkreślając wpływ rodzaju metrum na strukturę rytmiczną utworu. Nie wspomina natomiast o polimetrii, która oznacza występowanie różnych metrum w tym samym czasie lub naprzemiennie w utworze. Interesująca byłaby dyskusja, w jaki sposób można by wykorzystać metody opisane w pracy do klasyfikacji tego typu utworów muzycznych.

Wnioski końcowe

Mgr Jeremiaś Oluwagbemi Abimbola przedstawił rozprawę doktorską zawierającą oryginalne rozwiązania w zakresie klasyfikacji utworów muzycznych poprzez analizę tempa i metrum. Rozprawa doktorska dotyczy ważnego problemu oraz stanowi wartościowe osiągnięcie naukowe Autora. Dysertacja doktorska powinna również potwierdzać ogólną teoretyczną wiedzę Doktoranta w danej dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej. Wg mojej opinii przedstawiona do recenzji praca spełnia wymienione wymagania.

Biorąc pod uwagę powyższe stwierdzam, że niniejsza rozprawa spełnia wymogi ustawowe stawiane pracom doktorskim w *dziedzinie nauk inżynieryjno-technicznych*, w *dyscyplinie informatyka techniczna i telekomunikacja*, określone w Ustawie z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce (t.j.: Dz.U. z 2023 poz. 742, z późn.zm.) i wnioskuję o dopuszczenie Autora rozprawy do publicznej obrony.

