

dr hab. inż. Mariusz Kamola
Instytut Automatyki i Informatyki Stosowanej
Politechnika Warszawska
mariusz.kamola@pw.edu.pl

Recenzja rozprawy doktorskiej
opracowana na wniosek
Rady Dyscypliny Naukowej
Informatyka Techniczna i Telekomunikacja
Wojskowej Akademii Technicznej w Warszawie

Tytuł rozprawy: Modele i metody komputerowego generowania aranżacji muzycznych oraz badania ich podobieństwa

Autor rozprawy: mgr inż. Szymon Muszyński

Przedmiot rozprawy stanowi badanie możliwości wykorzystania grafowego opisu jednogłosowych utworów muzycznych w zadaniu automatycznej generacji aranżacji utworu oraz w zadaniu wyznaczania podobieństwa pary utworów. Zadanie to zostało podzielone przez Autora na część teoretyczną, obejmującą formalizację zapisu problemu i algorytmów prowadzących do rozwiązania, oraz część praktyczną, obejmującą implementację i ocenę prototypów i niezbędnego oprogramowania narzędziowego, jak również badanie ankietowe dot. wybranych wyników. W rozprawie zostały sformułowane następujące cele:

1. *rozwińnięcie grafowo-sieciowych modeli utworu muzycznego,*
2. *opracowanie i formalne zdefiniowanie metod automatycznego generowania aranżacji muzycznych,*
3. *opracowanie i formalne zdefiniowanie metod porównywania utworów muzycznych.*

W pracy nie są stawiane żadne tezy w sposób jawny, jednak można je wywieść z powyższych celów, i dotyczą one użyteczności (trafności, niezawodności i wydajności) metod z pp. 2-3 zaproponowanych przez Autora.

Zgodnie z art. 187 ustawy Prawo o szkolnictwie wyższym, rozprawa doktorska ma

- 1) prezentować ogólną wiedzę teoretyczną kandydata w dyscyplinie,
- 2) prezentować umiejętność samodzielnego prowadzenia pracy naukowej,
- 3) przedstawiać oryginalne rozwiązanie problemu naukowego *lub* rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej lub społecznej.

W odniesieniu do warunku 3., w opinii recenzenta charakter przedstawionych prac jest w przeważającej części badawczy. Dojrzałość technologiczna wykonanego oprogramowania i zakres jego zastosowania nie pozwalają na ocenę dokonań autora pod kątem aplikacyjnym. Opinia o rozprawie została wyrażona poniżej względem ww. kryteriów cząstkowych.

1. Przedstawienie w rozprawie ogólnej wiedzy teoretycznej kandydata w dyscyplinie

Informatyka jest zazwyczaj dyscypliną służebną wobec innych dziedzin, zatem interdyscyplinarność badań jest w niej zjawiskiem powszechnym. Jednakże stopień penetracji dziedzin przez informatykę jest mocno zróżnicowany i zależy zarówno od dostępności materiałów w postaci cyfrowej, jak i gotowości specjalistów z tychże dziedzin. Zastosowanie informatyki w muzykologii ma wciąż charakter nowatorski i eksperymentalny – i taką miarę należy przykładać do prac wykonanych przez Autora. Interdyscyplinarność badań manifestuje się najdobitniej w pracy w postaci dwóch leksykonów: terminów muzycznych, a następnie pojęć z teorii grafów, otwierających odpowiednie części rozprawy. Mimo, że muzyka jest już od wieków ujęta w karby matematyki, nie doczekała się uniwersalnych, ogólnodostępnych i użytecznych modeli matematycznych, analogicznie np. do dużych modeli językowych dla języków naturalnych. Dlatego rozprawa wpisuje się we w sumie prekursorski nurt badań.

Autor dobitnie prezentuje swoją szeroką wiedzę z zakresu informatyki, poczynając od aspektów teoretycznych, takich jak formalizacja pojęć i algorytmów, poprzez zagadnienia implementacyjne z nieustanną uwagą skierowaną na kwestie wydajności i wymagań zasobowych, wraz z wynikającymi zeń działaniami, takimi jak stosowanie heurystyk. Dobitym przykładem troski o praktyczną stosowalność opracowanych metod jest skrupulatna analiza ich złożoności obliczeniowej wraz z eksperymentalną weryfikacją (zał. A). Zaproponowane w rozprawie podejścia wymagają biegłego połączenia rozmaitych technik i zagadnień, m.in. optymalizacji wielokryterialnej i kombinatorycznej, strategii ewolucyjnych, procesów stochastycznych, teorii grafów. Nie wszystkie te połączenia wykonano bezbłędnie, niemniej rozpiętość tematyczna prac i wiedza Autora pozostają bezsprzecznie udokumentowane.

Bardzo interesującym, choć formalnie pobocznym zagadnieniem, jest geneza i umotywowanie zadania automatycznej generacji utworów muzycznych. Autor, podając jego rys historyczny, jednocześnie dowodzi swojej wiedzy w zakresie historii i teorii muzyki. Niestety, osadzenie własnych prac w kontekście współczesnych wyzwań i potrzeb nie prezentuje się równie przekonująco. O ile generowanie utworów to rzeczywiście teren pionierski, o tyle ich porównywanie jest już gałęzią rozwiniętą komercyjnie i znajdującą zastosowanie w mechanizmach rekomendacji stosowanych przez serwisy muzyczne oraz w sporach o prawa autorskie. Ponieważ

obecnie powszechną praktyką w działaniach komercyjnych jest stosowanie modeli o strukturze sieci neuronowej, dotkliwy jest w rozprawie brak dyskusji nt. potencjału wykorzystania grafowych sieci neuronowych w porównywaniu utworów. Sieci tego typu umożliwiają reprezentację wierzchołków albo całych grafów w abstrakcyjnej przestrzeni cech, i jako takie z powodzeniem znalazły zastosowanie np. do klasyfikacji związków chemicznych czy wykrywania kampanii dezinformacyjnych. Będąc same w sobie modelami przystosowanymi do uczenia maszynowego, z pewnością nie rozwiązywałyby problemu same z siebie z uwagi na ilość wymaganych przykładów treningowych – jednak mogłyby dostarczać cennych reprezentacji wierzchołków, uzupełniając zastosowane przez Autora proste ich atrybuty. Przy czym podnoszony przez Autora (przypis 59) problem braku odpowiednich zbiorów danych jest co najmniej dyskusyjny; wystarczy wskazać choćby projekt music21^a, udostępniający kilkaset zapisów nutowych.

W tym miejscu warto również zauważyć, że sama liczność cytowanych publikacji nie może stanowić w tej pracy miary oceny zakresu wiedzy z uwagi na manierę posługiwania się bibliografią, polegającą na cytowaniu wielu pozycji ze stosunkowo lakonicznym komentarzem (s. 79). Autor w pojedynczym cytowaniu odwołuje się aż do 26 publikacji dotyczących generowania utworów muzycznych poprzez *deep learning* po to, by stwierdzić, że nie są one przydatne w porównywaniu utworów. Czy są zatem przydatne w generacji utworów i jak metoda generacji przedstawiona w rozprawie ma się do cytowanych prac?

Ostatnia poważniejsza uwaga dotyczy traktowania odległości edycyjnych (s. 99): należy pamiętać, że odległość Levenshteina (jak i wiele pokrewnych) nie jest *miarą* sensu stricto gdyż, bez wdrożenia dodatkowych zabiegów normalizujących, nie spełnia nierówności trójkąta^b.

W opinii recenzenta, rozprawa należycie prezentuje ogólną wiedzę kandydata w dziedzinie.

2. Przedstawienie w rozprawie umiejętności samodzielnego prowadzenia pracy naukowej

Ocenę indywidualnego warsztatu badawczego można przeprowadzić w wielu aspektach, wśród których należy wyróżnić: a) dobór metod badawczych lub rozwiązania problemu, b) wybór metodyki oceny uzyskiwanych rozwiązań oraz c) sposób interpretacji i prezentacji wyników i wniosków.

W fazie doboru metod rozwiązania problemu, tj. reprezentacji utworu, a następnie metod generacji i porównywania, budzi wątpliwości wybrana, dość powierzchowna grafowa reprezentacja utworu. Jeżeli Autor zdecydował się na reprezentację grafową (de facto, ograniczył się do niej), to przypuszczalnie miał świadomość jej kluczowego niedomagania, którym jest zatarcie się linii melodycznych wykorzystujących tę samą nutę, czyli wierzchołek. Nuta taka jest wykorzystywana w utworze w wielu kontekstach, w wielu akordach, w fazie budowania czy rozładowania napięcia, zaskoczenia, zmiany konwencji czy wreszcie zmiany tonacji w obrębie utworu. Żadna z zaproponowanych w rozprawie reprezentacji grafowych nie jest w stanie ww. zjawisk kodować. Wydaje się, że ten niedostatek można by usunąć np. inspirować się głębiej metodami analizy tekstu (*n*-gramy), stosując grafy dwudzielne czy wręcz sieci semantyczne.

Faza oceny otrzymanych rozwiązań jest rozbudowana; Autor dokonuje systematycznego przeglądu wyników porównań utworów oraz wyników generacji utworów dla rozmaitych kombinacji parametrów i metod. W obu przypadkach zastosowana metodyka budzi pewne obawy. Porównywanie utworów wyjściowych z ich automatycznie wygenerowanymi aranżacjami (rozdz. 8.5) jest niejako z góry skazane na sukces, gdyż narzędzie pomiaru deformacji (tu: reprezentacja uproszczona) jest o wiele bardziej zgrubne niż narzędzie deformujące (reprezentacja pełna). Rzeczywista, choćby nawet wrywkowa, ocena przydatności opracowanej metody powinna zostać wykonana na parach utworów będących przedmiotem sporów sądowych o plagiat. Sprawy te są ogólnie znane, a utwory – dostępne. Niezależnie od doboru utworów, skalaryzacja kryteriów powinna uwzględniać znikomą rozpiętość wartości podobieństwa strukturalnego w porównaniu z podobieństwem ilościowym (por. np. utwór *Yellow Submarine*, tab. 20). Wstępne wzajemne unormowanie kryteriów prawdopodobnie uprościłoby ich późniejsze dokładne zestrojenie metodą ekspercką. Przechodząc do metodyki oceny wygenerowanych utworów należy zauważyć, że natura ich generacji jest stochastyczna, podczas gdy zastosowana ocena, również ilościowa, opiera się na pojedynczych próbkach losowych (rozdz. 7.5). Czym Autor uzasadnia takie podejście i jakie są jego skutki? Wreszcie, pytanie nr 1 w ankiecie zostało sformułowane zbyt ogólnie, gdyż odbiorca, nie znając charakteru utworu wyjściowego, mógł ocenić utwór wygenerowany nie w kontekście aranżacji, lecz posługując się wyłącznie własną intuicją i gustem muzycznym, co znacząco obciąża wyniki, zasadniczo udanej, ankiety.

Faza interpretacji i prezentacji wyników eksperymentów wyraża się w sposób najbardziej powierzchowny w strukturze i redakcji samej rozprawy. Struktura jest jasna, przemyślana i dobrze przedstawiona we wstępie. Strona edycyjna jest zasadniczo poprawna, użyty język – klarowny, wypowiedzi – spójne lecz miejscami (rozdz. 3) rozwlekłe. Autor zamieszcza w tekście głównym rozbudowane tabele charakterystyk grafowo-sięciowych (rozdz. 7.4), z których sam korzysta w niewielkim stopniu, odwołując się do zapisu nutowego, którego akurat brakuje. Podnoszone walory intuicyjnej i syntetycznej wizualizacji grafowej utworów są wątpliwe: w przypadkach większych grafów (rys. 38) wierzchołki, które „można zauważyć” są zwyczajnie niewidoczne i w gruncie rzeczy analiza podobieństwa utworów sprowadza się do porównania ambitusu, tonacji, tempa i częstości nut, opierając się w dużej mierze na podstawowych terminach i statystykach. Pozostałe ważniejsze uwagi redakcyjne dotyczą definicji utworu muzycznego jako *zbioru* nut *N* (zamiast *ciągu*), błędnie zapisanej nuty *C* (rys. 12) oraz braku żywej paginy z numerem rozdziału, co utrudnia orientację w rozbudowanym dokumencie.

^a <https://github.com/cuthbertLab/music21>

^b Li Yujian, Liu Bo; A Normalized Levenshtein distance metric; IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007

Ostatecznie, pośrednim ale miarodajnym probierzem jakości warsztatu badawczego jest działalność publikacyjna Autora, która owocowała 2 samodzielnymi referatami konferencyjnymi oraz jedną publikacją w czasopiśmie, we współautorstwie. Różnorodność tych form publikacyjnych i ich chronologia znamionują należyty rozwój własnych umiejętności badawczych.

Mimo uwag i pytań dotyczących metodyki, których liczba jest też pochodną samego zakresu badań, w opinii recenzenta rozprawa dowodzi umiejętności samodzielnej pracy naukowej Autora.

3. Przedstawienie w rozprawie oryginalnego rozwiązania problemu naukowego

Jeśli przyjąć, iż problemem *naukowym* stawianym w pracy jest zbadanie możliwości zastosowania teorii grafów w kodowaniu utworów muzycznych, w celu ich porównywania oraz tworzenia utworów pochodnych – to rozprawa takie kryterium spełnia. Autor przedstawia 3 metody kodowania grafów, kilka wariantów ich generacji oraz kilka wariantów wyznaczania podobieństwa ilościowo-strukturalnego. Zaproponowane podejścia są rozwinięciem wcześniejszych prac [12, 73, 74 itp.], co samo w sobie jest jak najbardziej pożądane. Problem polega na tym, iż w pracy nie pojawia się zestawienie własnego, *oryginalnego* wkładu Autora. Dość wyraźnie zarysowują się dwie takie innowacje: pierwsza polega na wprowadzeniu bardziej rozbudowanych reprezentacji nut w postaci węzłów, druga – na wprowadzeniu kolejnych wariantów generowania ścieżki w grafie i mutacji samego grafu. Autor powinien odnieść się do tej oceny podczas obrony.

Jeśli przyjąć, iż problemem jest *użyteczne* porównywanie i generowanie utworów, wówczas wyniki badań raczej wskazują, iż zaproponowane metody nie prowadzą do jego osiągnięcia w sposób satysfakcjonujący. Ocena podobieństwa pod kątem wykrywania plagiatu została wykonana w warunkach odległych od rzeczywistych, a ocena ankietowa jakości aranżacji dała wyniki gorsze od oczekiwanych. Nie oznacza to, że praca nie jest wartościowa; napotkane problemy stanowią inspirację i punkt wyjścia do dalszych udoskonaleń, w szczególności takich, które wykorzystywałyby muzykologiczną wiedzę dziedzinową. Można i należy ją wdrażać do algorytmów w różnych punktach, np. modyfikując algorytm odległości edycyjnej, adaptując metodę generacji nut do metrum utworu, dbając o tonalność i inne detale kompozycyjne – co zresztą Autor sam postuluje. Ostatecznie należy pamiętać, co również Autor sam stwierdza (s. 206), że *samo* podobieństwo strukturalne nie sprawdza się w porównywaniu utworów – ale też stwierdzenie to domaga się uzupełnienia o wytypowanie innych, perspektywicznych cech uzupełniających w takim porównywaniu.

W opinii recenzenta, problem naukowy w pracy należy postrzegać jako ograniczony do zbadania możliwości zastosowania grafów w ww. celach, z wyłączeniem wymagania osiągnięcia w pełni zadowalających wyników praktycznych. (Praca prezentuje już zmysł praktyczny Autora w wielu miejscach.) Oryginalny wkład własny w metody badawcze, choć jawi się skromnie, jest dostateczny.

Uważam, że rozprawa mgra inż. Szymona Muszyńskiego spełnia wymagania Ustawy z dn. 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. 2023, poz. 742) w odniesieniu do rozpraw doktorskich. W związku z tym wnoszę o przyjęcie rozprawy i dopuszczenie Autora do dalszych, przewidzianych przepisami, etapów postępowania awansowego.

Mariusz Kamuda