dr hab. Michał Stukow, prof. UG
Instytut Matematyki
Wydział Matematyki, Fizyki i Informatyki
Uniwersytet Gdański

Gdańsk, 6 lutego 2025

# Recenzja rozprawy doktorskiej
# **New Shape Descriptors for Topological Data Analysis**
magistra Davide Gurnariego

Przedstawiona rozprawa doktorska jest poświęcona rozszerzeniu aparatu narzędzi stosowanych w topologicznej analizie danych (ang. Topological Data Analysis).

## Zawartość pracy

Rozprawa liczy 108 stron i składa się z 5 rozdziałów. Poprzedzona jest bardzo starannie napisanym wstępem. Na końcu znajduje się bardzo wyczerpująca bibliografia (116 pozycji). Duża liczba pozycji bibliograficznych znajduje odzwierciedlenie w tekście rozprawy – autor bardzo starannie wskazuje na źródła przytaczanych definicji, twierdzeń, ale też inspiracji. Pod tym względem rozprawa sprawia wrażenie bardzo uczciwie napisanej.

Pierwszy rozdział ma charakter wstępny i stanowi zbiór podstawowych definicji ułatwiających lekturę kolejnych rozdziałów. Zagadnienia omówione przez autora w tym wstępnym rozdziale koncentrują się głównie wokół podstaw topologii algebraicznej i kombinatorycznej. Pobieżnie omówiono klasyczne konstrukcje teorii homologii kompleksów symplicjalnych oraz adaptację tych konstrukcji do przypadku kompleksów stowarzyszonych ze skończonymi zbiorami (chmurami) punktów w przestrzeniach metrycznych – homologie persystentne. Przytoczono też standardowe twierdzenia dotyczące stabilności homologii persystentnych – Twierdzenia 1-4 oraz algorytm obliczania tych homologii – Algorytm 1. Rozdział kończy się podstawowymi informacjami o kompleksach kubicznych, znajdujących zastosowanie na przykład przy analizie obrazów.

Drugi (najdłuższy) rozdział rozprawy jest poświęcony analizie możliwości zastosowania charakterystyki Eulera kompleksów związanych z chmurami punktów do analizy kształtu topologicznego tych zbiorów. Pierwsza część tego rozdziału jest poświęcona dowodowi (pewnego wariantu) stabilności charakterystyki Eulera dla kompleksów komórkowych z filtracją (funkcji ECC zdefiniowanej w Definicji 29) – Proposition 7. Następnie autor rozszerza analizę niezmienników związanych z charakterystyką Eulera na przypadek kompleksów z filtracją w zbiorze częściowo uporządkowanym (funkcja ECP zdefiniowana w Definicji 32) i ponownie dowodzi (pewnego rodzaju) stabilności tak zdefiniowanej funkcji ze względu na perturbacje analizowanego zbioru danych – Proposition 9. Wspomniane wyniki o stabilności należą do głównych wyników teoretycznych recenzowanej rozprawy i zostały opublikowane w artykule [35].

Dalsza część rozdziału drugiego jest poświęcona analizie możliwości efektywnego obliczania zdefiniowanych wcześniej niezmienników (ECC, ECP) oraz ich wektoryzacji na poczet zastosowań w uczeniu maszynowym. W szczególności opisano autorski algorytm obliczania ECC i ECP oraz podano przykłady użycia tych algorytmów na przykładowych danych.

Rozdział 3 jest poświęcony zagadnieniu interpretacji informacji topologicznych (na przykład o nietrywialnych klasach homologii) otrzymanych w wyniku obliczania homologii persystentnych w języku cech analizowanego zbioru danych. Podstawowym narzędziem technicznym umożliwiającym taką interpretację jest rodzaj analizy harmonicznej konstruowanych klas homologii. Rozdział zawiera autorski algorytm oraz opis różnych przykładowych obliczeń takich harmonicznych homologii persystentnych. Przeprowadzono też analizę związku wag sympleksów w cyklach harmonicznych z istnieniem tzw. istotnych sympleksów – Definicja 40. Przeprowadzona analiza skłoniła autora do sformułowania ciekawej Hipotezy 18. Na zakończenie rozdziału 3 opisano przykłady zastosowania obliczania cykli harmonicznych do analizy pewnych zbiorów danych medycznych.

Rozdział 4 rozprawy jest poświęcony nowym algorytmom wizualizacji chmur punktów umieszczonych w przestrzeniach euklidesowych i opiera się na wynikach opublikowanych w pracy [38]. Autor opisuje trzy warianty algorytmów typu Mapper: Equivariant Ball Mapper, Mapper on Ball Mapper i MappingMappers. Jako ilustrację zastosowania tych algorytmów przeprowadzona jest porównawcza analiza danych otrzymanych w wyniku obliczania podstawowych niezmienników dla węzłów (wielomiany Jonesa, Alexandera i HOMFLYPT) o niewielkiej liczbie skrzyżowań (do 17 dla wielomianów Jonesa i Alexandera i do 15 dla HOMFLYPT). Autor opisuje przykłady użycia tego typu wizualizacji do ilustracji teoretycznych zależności między np. sygnaturą węzła a współczynnikami odpowiednich wielomianów. Rozdział kończy się jeszcze jednym przykładem zastosowania zdefiniowanych algorytmów wizualizacji – do możliwych przebiegów gry w kółko i krzyżyk.

Rozdział 5 rozprawy jest poświęcony zagadnieniu połączenia informacji uzyskiwanych przez zastosowanie algorytmów klasyfikacyjnych (klastrujących) oraz algorytmów redukujących wymiar zanurzenia dyskretnych zbiorów punktów w przestrzeniach euklidesowych. Autor definiuje graf – ClusterGraph, który zawiera informacje zarówno o klastrach (otrzymanych w wyniku zastosowania standardowych algorytmów klastrujących) jak i o odległościach pomiędzy puntami tych klastrów. Następnie przeprowadzona jest analiza wierności informacji reprezentowanych przez tego typu grafy oraz opisane są algorytmy przerzedzania tych grafów (pruning). Podobnie jak w poprzednich rozdziałach, na koniec autor ilustruje użycie zdefiniowanych pojęć analizą wybranych ogólnie dostępnych zbiorów danych.

## Ocena rozprawy

Ocenę rozprawy rozpocznę od rzeczy bezdyskusyjnie pozytywnych. Rozprawa jest napisana bardzo dobrze – zarówno jej struktura jak i język nie budzą zastrzeżeń, czyta się ją z prawdziwą przyjemnością. Autor w wielu miejscach zamieszcza rozbudowane komentarze często ilustrowane przykładami. Jest też sporo wysokiej jakości rysunków, diagramów i wykresów – stanowią one poniekąd dowód możliwości praktycznych zastosowań rozwijanych przez autora metod.

Na zdecydowanie pozytywną ocenę zasługuje też fakt udostępnienia (na github-ie) przez autora repozytoriów z przykładowymi implementacjami omawianych algorytmów.

W tekście jest trochę literówek, ale nie utrudniają one znacząco lektury rozprawy. Niestety autor nie stosuje spójnej numeracji twierdzeń, definicji, uwag itd. Z tego powodu normalną jest sytuacja gdy bezpośrednio po Uwadze 10 mamy Proposition 20, a po nim Definicję 43. Przy dość znacznej objętości rozprawy, ten mankament utrudnia poruszanie się po jej tekście. Również sposób odwoływania się do numerowanych wzorów, definicji i twierdzeń pozostawia sporo do życzenia – w wielu miejscach mamy bardzo lakoniczne odsyłacze w stylu *... as a function like in 32 ...* (dół strony 10), czy *... left-hand side in 2.2...* (góra strony 24). Ten aspekt redakcyjny mógłby być zrobiony lepiej.

Jeżeli chodzi o ocenę merytoryczną rozprawy, tematyka pracy jest dość obszerna i dotyka różnych aspektów metod stosowanych w Topologicznej Analizie Danych (TDA). Ponieważ jest to dziedzina stosunkowo młoda (choć oczywiście bazuje na klasycznych już narzędziach topologii algebraicznej i kombinatorycznej), to widać że wciąż ciężar badań w znacznej mierze koncentruje się na definiowaniu nowych struktur kombinatorycznych i opisie algorytmów służących ich efektywnemu obliczaniu. Z tego punktu widzenia oceniania rozprawa zawiera wiele tego przykładów - o czym dokładniej pisałem omawiając zawartość pracy. W mojej ocenie definiowane przez autora deskryptory cech topologicznych (kształtu) analizowanych dyskretnych zbiorów punktów mają sens i są ciekawe. Również przykłady podawane przez autora nie są całkowicie trywialne i na ogół dobrze ilustrują wprowadzane algorytmy.

W mojej ocenie słusznie autor środek ciężkości rozprawy zawarł w rozdziale 2, gdyż temat zrozumienia możliwości stosowania charakterystki Eulera jako narzędzia do analizy chmur punktów jest najpełniej opracowaną częścią rozprawy. Znajdujemy tam zarówno wyniki teoretyczne (np. twierdzenia o stabilności), definicje nowych niezmienników (ECP), algorytmy obliczania odpowiednich niezmienników (ECC i ECP) oraz przykładowe zastosowania tych pojęć. Myślę, że w wariancie minimum sama zawartość rozdziału drugiego mogłaby być podstawą do pozytywnej oceny pracy. Dodam też, że wyniki te zostały opublikowane w artykule [39].

Jeżeli chodzi o ocenę pozostałych rozdziałów, a więc innych zagadnień poruszanych w rozprawie, to sytuacja jest mniej jednoznaczna. Na pewno tematyka tych rozdziałów jest interesująca i ma potencjał do ciekawych wyników, jednak odnoszę wrażenie, że pełne wyeksponowanie znaczenia wprowadzanych narzędzi będzie wymagało jeszcze dalszych badań.

Na przykład wydaje mi się, że język i tematyka rozdziału 3, czyli dodanie elementów analizy harmonicznej w teorii homologii persystentnych jest niezwykle interesujące i osobiście (pewnie ze względu na specjalizację zawodową) uważam ten rozdział za najciekawszy w całej rozprawie. Autor czyni szereg ciekawych obserwacji, niemniej wydaje mi się, że potencjał tych narzędzi wykracza poza aktualny stan ich zrozumienia i będzie wymagał dalszych badań.

Rozdział 4 w zasadzie w większym zakresie jest związany z wizualizacją danych, a mniej z aspektami czysto teoretycznymi. Nie będę ukrywał, że pomysł zastosowania algorytmów typu mapper do analizy niezmienników węzłów bardzo mi się spodobał. Otrzymane diagramy są interesujące i mogą być inspirujące. Jednocześnie bardzo ostrożnie bym podchodził do oczekiwań, że pozwolą one odkryć nowe relacje między tymi niezmiennikami. Na dzień dzisiejszy traktowałbym je jako bardzo interesującą ciekawostkę. Z drugiej strony przyznam, że pomysł wizualizacji danych pochodzących z niezmienników czysto matematycznych (współczynników wielomianów w tym przypadku) jest ciekawy i ma potencjał do stosowania w innych sytuacjach.

Rozdział 5, pomimo tego, że dość krótki zawiera szereg ciekawych pomysłów rozszerzających standardowe algorytmy klastrujące. Na ile te rozszerzenia okażą się istotnie w zastosowaniach trudno ocenić na tym etapie.

## Konkluzja

Tematyka rozprawy jest w istocie poświęcona zastosowaniom metod topologicznych w analizie danych i w swojej naturze jest interdyscyplinarna. Autor prowadząc badania i przygotowując rozprawę udowodnił, że dość sprawnie operuje współczesnym językiem matematycznym, prawidłowo forumułuje i analizuje algorytmy oraz potrafi znaleźć przykłady użycia wprowadzonych pojęć np. do analizy danych pochodzenia medycznego. Uzyskane rezultaty są ciekawe i moim zdaniem w istotny sposób przyczyniają się do rozwoju Topologicznej Analizy Danych. W mojej ocenie rozprawa magistra Davide Gurnariego spełnnia zwyczajowe i ustawowe wymogi stawiane rozprawom doktorskim. W związku z tym wnoszę o przyjęcie rozprawy i dopuszczenie magistra Davide Gurnariego do dalszych etapów przewodu doktorskiego.

Michał Stukow