Министерство образования и науки Российской Федерации
Московский физико-технический институт (государственный университет)

Физтех-школа прикладной математики и информатики
Кафедра дискретной математики

Выпускная квалификационная работа магистра

Топологические методы в некоторых задачах анализа данных

Автор:
Студент 115 группы
Снопов Павел Михайлович

Научный руководитель:
д-р физ.-мат. наук, проф.
Мусин Олег Рустумович

Москва 2023

Аннотация

Топологические методы в некоторых задачах анализа данных
Снопов Павел Михайлович

Краткое описание задачи и основных результатов, мотивирующее прочитать весь текст

Abstract

Topological methods in some problems of data analysis

Оглавление

Глава 1

Введение

В этой части надо описать предметную область, задачу из которой вы будете решать, объяснить её актуальность (почему надо что-то делать сейчас?). Здесь же стоит ввести определения понятий, которые вам понадобятся в постановке задачи.

Глава 2

Необходимые теоретические сведения

В этой главе мы напомним некоторые определения и факты, используемые в дальнейшем.

2.1 Симплициальные гомологии

Пусть K – симплициальный комплекс, то есть множество симплексов, такое, что

1. Для каждого симплекса из K его грани тоже лежат в K.
2. Пересечение любых двух симплексов $\sigma, \tau \in K$ либо пусто, либо является гранью и σ, и τ.

Тогда можно рассмотреть *n-мерные цепи* c, т.е. линейные комбинации с целыми коэффициентами (лишь конечное число которых ненулевые) всех ориентированных n-симплексов в K

$$c = \sum_i z_i \sigma_i^n.$$

Множество $C_n(K)$ всех n-цепей называется *n-й группой цепей* и имеет очевидную структуру свободного $\mathbb{Z}$-модуля. Прямая сумма $\bigoplus_{n \geq 0} C_n(K)$ таких групп также является свободным $\mathbb{Z}$-модулем и обозначается $C_*(K)$.

Из n-цепи можно получить $(n-1)$-цепь путем взятия границы. А именно, *границей* n-цепи $c = \sum_i z_i \sigma_i^n$ называется $(n-1)$-цепь

$$\partial_n(c) = \sum_{j=0}^{n} (-1)^j \sum_i \varepsilon_i \partial_j \sigma_i^n,$$

где $\partial_j \sigma_i^n = \partial_j [v_0, ..., v_n] = [v_0, ..., \hat{v}_j, ..., v_n]$ – это $(n-1)$-симплекс, порожденный всеми вершинами, кроме вершины v_j. Гомоморфизм модулей ∂_n называется *граничным оператором.*

Лемма (Пуанкаре)**.** *Для любого $n \geq 2$ справедливо*

$$\partial_{n-1} \circ \partial_n = 0.$$

Последовательность $\mathbb{Z}$-модулей и гомоморфизмов ∂ между ними (на самом деле, вместо $\mathbb{Z}$-модулей и гомоморфизмов могут быть объекты любой аддитивной категории и морфизмы между ними [1]), для которых верно, что $\partial^2 = 0$, называется *цепным комплексом.*

Нетрудно заметить, что в таком случае $\operatorname{im} \partial_{n+1} \leq \ker \partial_n$.

Определение 1 (Группа гомологий). *n-й группой гомологий $H_n(K)$ симплициального комплекса K называется $\mathbb{Z}$-модуль*

$$H_n(K) = {\ker \partial_n} \Big/ {\operatorname{im} \partial_{n+1}}.$$

2.2 Pas de Deux: категории и функторы

Конструкция групп гомологий обладает одним замечательным свойством: она функториальна. Для того, чтобы наиболее полно описать это, придется воспользоваться аппаратом теории категорией.

Категорией **C** называют коллекцию объектов и морфизмов между парой объектов. То есть, категория состоит из

1. (классов) объектов X,Y,Z,
2. для любых двух объектов X и Y существует класс морфизмов $\mathrm{Hom}_{\mathbf{C}}(X,Y)$ (если категория понятна из контекста, то индекс будем опускать),
3. для любого объекта X существует единичиный морфизм $1_X \in \mathrm{Hom}(X,X)$,
4. на морфизмах задана операция композиции: для любого пары морфизмов $f \in \mathrm{Hom}(X,Y)$ и $g \in \mathrm{Hom}(Y,Z)$ существует морфизм $gf \in \mathrm{Hom}(X,Z)$,
5. для любого морфизма $f \in \mathrm{Hom}(X,Y)$ верно, что $1_Y f = f = f 1_X$,
6. операция композиции ассоциативна.

......

Если имеется два симплициальных комплекса K и L, и симплициальное отображение f между ними, то оно индуцирует гомоморфизм модулей

$$f_{*,n} : H_n(K) \to H_n(L)$$

2.3 Персистентные гомологии

Под облаком точек D здесь и далее мы будем подразумевать конечное множество в $\mathbb{R}^n$. О персистентных (устойчивых) гомологиях можно думать как об адаптации понятия гомологии к облаку точек.

А именно, имея облако точек, можно построить семейство симплициальных комплексов, параметризованное некоторым частично упорядоченным множеством (например, на практике зачастую это $\mathbb{R}$ или его подмножества). Такая параметризация естественным образом задает отображения между получаемыми симплициальными комплексами. Тогда устойчивыми гомологиями будут в точности гомологии полученных комплексов вместе с индуцированными гомоморфизмами между ними.

Определим теперь устойчивые гомологии формально.

Зафиксируем некоторое частично упорядоченное множество $T \subseteq \mathbb{R}$ и соответствующую ему категорию **T**. Тогда функтор

$$F : \mathbf{T} \to \mathbf{Simp}$$

называется *фильтрацией*.

Наиболее важным для нас примером фильтрации на облаке точек D будет являться *фильтрация Вьеториса-Рипса*

$$\mathrm{Rips}(D) : [\mathbf{0},\infty) \to \mathbf{Simp} : r \mapsto X(\mathcal{N}(D)_r),$$

где $\mathcal{N}(D)_r$ – это граф соседей радиуса r облака точек D, а $X(G)$ – это кликовый комплекс графа G, т.е. наибольший (по включению) симплициальный комплекс, чей 1-скелет – это G.

Алгоритм 1: Алгоритм построения комплекса Вьеториса—Рипса

Исходные параметры: Облако точек X, вещественное число $\alpha > 0$.
Результат: Симплициальный комплекс Вьеториса—Рипса
Для каждой точки x строим её α-окрестность $B_\alpha(x)$;
$i = 1$;
до тех пор, пока $i + 1$ *окрестностей попарно имеют непустое пересечение* **выполнять**
    строим i-ый симплекс на соответствующих вершинах;
    $i \leftarrow i + 1$;
конец

Глава 3

Обзор существующих решений

Здесь надо рассмотреть все существующие решения поставленной задачи, но не просто пересказать, в чем там дело, а оценить степень их соответствия тем ограничениям, которые были сформулированы в постановке задачи.

Глава 4

Исследование и построение решения задачи

Здесь надо декомпозировать большую задачу из постановки на подзадачи и продолжать этот процесс, пока подзадачи не станут достаточно простыми, чтобы их можно было бы решить напрямую (например, поставив какой-то эксперимент или доказав теорему) или найти готовое решение.

Глава 5

Описание практической части

Если в рамках работы писался какой-то код, здесь должно быть его описание: выбранный язык и библиотеки и мотивы выбора, архитектура, схема функционирования, теоретическая сложность алгоритма, характеристики функционирования (скорость/память).

Глава 6

Заключение

Здесь надо перечислить все результаты, полученные в ходе работы. Из текста должно быть понятно, в какой мере решена поставленная задача.

Список использованных источников

1. Weibel C. An Introduction to Homological Algebra. — Cambridge University Press, 1994. — (Cambridge Studies in Advanced Mathematics). — URL: https://books.google.ru/books?id=flm-dBXfZ%5C_gC.