ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

УДК 515.1

Отчет об исследовательском проекте

на тему Основы прикладной топологии

Выполнил:
студент группы БПМИ198 Подпись В. О. Копчев
И.О. Фамилия

29.05.2021
Дата

Принял:
руководитель проекта Антон Андреевич Айзенберг
Имя, Отчество, Фамилия

Заведующий лабораторией
Должность

Международная лаборатория алгебраической топологии и ее приложений
Место работы

Дата 1 июня 2021 10
Оценка (по 10-тибалльной шкале) Подпись

Москва 2021

Реферат

Целью данной работы было изучить основные понятия алгебраической топологии и основанные на этих понятиях алгоритмы. Было изучено понятие топологического пространства, гомотопии, гомеоморфизма, симплициального комплекса, симплициальных гомологий и устойчивых гомологий. Были изучены методы топологического анализа и их обоснование, понятия облака точек, фильтрации Чеха, фильтрации Вьеториса–Рипса, альфа-фильтрации, проделан ряд упражнений. Были изучены алгоритмы топологического анализа данных и их имплементация на примере библиотеки gudhi с упором на конструкцию альфа-фильтрации. В качестве примера разбирались искусственно сгенерированные выборки точек с тора и бутылки Клейна, для которых были построены баркоды и сформированы альфа-фильтрации. Также методами топологического анализа данных был проанализирован датасет, содержащий некоторую информацию о 80 странах: для этого датасета были вычислены устойчивые гомологии фильтрации Вьеториса-Рипса и альфа-фильтрации, построен баркод, проведена кластеризация данных и сделаны выводы из проделанной работы. Также была разобрана статья о применении топологических методов в теории динамических систем, сделаны выводы об актуальности этого направления исследований. При проведении работы были прочитаны несколько учебников и статей, проведены вычисления на языке Python, собраны данные по 80 странам. В результате сделанной работы было показано, как из исследований по теории множеств появилась аксиоматика топологических пространств, были разобраны доказательства важнейших теорем из области общей и алгебраической топологии, решен ряд упражнений, разобрана статья и проведен вычислительный эксперимент.

Ключевые слова: топология, топологический анализ данных, gudhi

Содержание

Основные термины и определения

Топология — это раздел математики, изучающий свойства геометрических объектов, которые не изменяются под действием непрерывных деформаций.

Топологическое пространство — множество с определенной на ней топологией.

Общая топология — раздел топологии, изучающий основные теоретико-множественные определения и конструкции, используемые в топологии.

Алгебраическая топология — раздел топологии, изучающий топологические пространства методами общей алгебры.

Прикладная математика — область математики, рассматривающая применение математических методов, алгоритмов в других областях науки и техники.

Анализ данных — область математики и информатики, занимающаяся исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных данных.

Топологический анализ данных — подход к анализу данных, использующий методы топологии.

Введение

Топологический анализ данных — это недавно возникшая область прикладной математики, развивающая новый подход к задачам анализа данных, который основан на использовании аппарата алгебраической топологии для решения этих задач. Топологический анализ данных широко применяется для анализа изображений, многомерных данных, обработки текстов. Во время работы над проектом я буду исследовать методы и задачи именно этой области математики.

Целью моей работы является изучение прикладной топологии и топологического анализа данных. Для этого я планирую выполнить несколько задач. Начать изучение топологического анализа данных я собираюсь с основ общей топологии, поскольку именно этот раздел топологии и разработал самые базовые понятия и методы этой науки. Для этого я планирую прочитать учебник Виро О. Я. «Элементарная топология». После изучения основ, необходимо перейти к алгебраической топологии: узнать, что такое симплициальные комплексы, фильтрации, гомологии. Для этого я воспользуюсь учебниками Хатчера А. «Алгебраическая топология» и Айзенберга А. «Методичка по симплициальным комплексам и гомологиям». Закрепив полученные сведения с помощью учебника Васильева «Введение в топологию», я приступлю к третьей задаче: изучение основ топологического анализа данных. Для этого я планирую прочитать статью Chazal F. «An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists», решить ряд упражнений в «Методичке по симплициальным комплексам и гомологиям», прочитать книгу R. Ghrist, «Elementary Applied Topology». После этого, снова воспользовавшись «Методичкой по симплициальным комплексам и гомологиям», я планирую решить последнюю и основную задачу: изучить альфа-комплексы и альфа-шейпы, области их применения в прикладной математике, выяснить, какие существуют пакеты, способные формировать альфа-фильтрации и вычислять их устойчивые гомологии.

Таким образом, объектом моего исследования является прикладная топология, предметом — методы топологического анализа данных, приложения альфа-шейпов и альфа-комплексов. Целью моего исследования является изучение прикладной топологии и топологического анализа данных, задачами — прочтение ряда учебников, решение упражнений в них, изучение имеющихся пакетов. Актуальность моего проекта заключается в том, что топологический анализ данных — это молодая область прикладной математики, которая позволяет решить большое число задач, поэтому крайне важно разобраться с методами этой области, чтобы в будущем иметь возможность применять их для решения самых разных задач анализа данных.

Обзор и сравнительный анализ источников

Виро О. Я. «Элементарная топология». Книга представляет собой подробное введение в общую и алгебраическую топологию с большим количеством задач и упражнений. Разбираются темы от топологических пространств и непрерывности до гомотопий и прочих основных тем в алгебраической топологии, именно поэтому я планирую прочитать этот учебник самым первым. В отличие от остальных книг, которыми я воспользуюсь во время выполнения проекта, посвященных более узким областям топологии, данная книга является введением в топологию для читателей, не сталкивавшихся раньше с этой областью математики.

Хатчер А. «Алгебраическая топология». Этот учебник посвящен основам алгебраической топологии. В нем рассматриваются такие темы как гомологии, накрытия. При этом он является более специализированным, чем учебник Виро, и требует от читателя некоторой подготовки, включающей в себя знание общей топологии. Поэтому его я прочитаю после предыдущего учебника. Также, помимо большей специализированности на алгебраической топологии и рассмотрения более узких тем в рамках этой области топологии, учебник отличается от предыдущего тем, что в нем существенного меньше упражнений — по всей видимости, он написан не для читателей, начинающих изучать топологию, которым полученные знания необходимо закреплять упражнениями, но для тех, кто собирается подробнее погрузиться в более узкую область этого раздела геометрии.

Васильев В. А. «Введение в топологию». Эта книга является учебником по основам общей и алгебраической топологии, но отличается от предыдущих сжатым и кратким изложением основных разделов этих областей математики без рассмотрения большого количества примеров и упражнений. При этом в книге Виро рассматриваются в первую очередь основы общей топологии, в книге Хатчера — в первую очередь алгебраической. Однако же в данном учебнике можно найти главы, посвященные как определению топологических пространств (рассматривается у Виро, но не у Хатчера), так и гомологиям (рассматривается у Хатчера, но не у Виро). Поэтому его я решил использовать для закрепления изученного материала.

R. Ghrist, «Elementary Applied Topology». Эта книга, также как книга Хатчера, рассмотренная ранее, посвящена основам алгебраической топологии, однако отличается от нее тем, что в первую очередь в ней разбирается большое количество сложных примеров, а уклон сделан скорее на прикладное применение алгебраической топологии читателем, который уже знаком с данной областью математики. Поэтому я прочитаю эту книгу уже после изучения основной теории, когда перейду к рассмотрению приложений изученного материала — кажется разумным сначала изучить основные идеи и понятия алгебраической топологии, а потом уже разбирать углубляющие понимание изученной теории примеры, приложения и частные случаи. При этом данный учебник отличается и от учебников Васильева и Виро — в нем рассматриваются в первую очередь именно прикладные аспекты топологии, а не общая теория.

Айзенберг, А. А. «Методичка по симплициальным комплексам и гомологиям». Эта, также как «Elementary Applied Topology», рассматривает алгебраическую топологию с практической точки зрения. При этом содержащаяся в ней информация дополнена большим числом простых примеров, позволяющих лучше понять топологию и не требующих предварительной подготовки. Это отличает ее от первых трех учебников, рассмотренных мной в этом разделе. При этом простота примеров отличает ее и от книги «Elementary Applied Topology», которая, как мне кажется, требует большей подготовки. Поэтому этим учебником я воспользуюсь при изучении основ алгебраической топологии — в нем можно найти главы, посвященные гомологиям, способам их вычислений, симплициальным комплексам. Наряду с этим данная книга содержит и более углубленные темы (например, в ней рассматриваются нервы покрытий), поэтому, когда во время выполнения проекта я перейду к применению полученных знаний на практике, я также воспользуюсь именно этой книгой.

Chazal, F. «An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists». Эта статья посвящена исключительно топологическому анализу данных, его методам и задачам. В ней отсутствуют объяснения основных терминов общей и алгебраической топологии, и это отличает ее от рассмотренных мной раннее в этом разделе учебников. При этом она требует подготовки в алгебраической топологии, понимания прикладных аспектов этой области математики, и поэтому данную статью я буду изучать во время выполнения прикладной части проекта.

1 Основная часть

1.1 Введение в топологию.

Основной целью моего проекта является изучение топологии. Топология — это раздел математики о непрерывности и свойствах пространств, инвариантных относительно гомеоморфизма. У понятий пространства, непрерывности, гомеоморфизма существует строгое аксиоматическое определение, однако перед тем, как изложить это общее определение, стоит рассмотреть различные частные случаи, чтобы понять, как подойти к этим определениям интуитивно.

Определение 1.1. Граф G гомеоморфен графу G', если существуют их изоморфные подразбиения (H — подразбиение G, если H получен из G добавлением вершин на ребрах).

Определение 1.2. Подмножество $Z \subseteq M$ метрического пространства называется открытым, если $Z = \cup_\alpha B_{\varepsilon_\alpha}(x_\alpha)$ и $\forall z \in Z : B_\varepsilon(z) \subseteq Z$

Определение 1.3. Пусть $\{z_i\}$ — последовательность точек в метрическом пространстве (M,d). Говорят, что $\{z_i\}$ сходится к z, если $\lim_{i\to\infty} d(z_i,z) = 0$

Определение 1.4. Пусть $(M_1,d_1),(M_2,d_2)$ — метрические пространства. Отображение $f : M_! \to M_2$ называется непрерывным, если:

1. $z_i \to z \Rightarrow f(z_i) \to f(z)$

2. $\forall z \in M_1, \varepsilon > 0 \exists \delta > 0 : d_1(x,z) < \delta \Rightarrow d_2(f(x),f(z)) < \varepsilon$

3. Прообраз открытого множества открыт.

Определение 1.5. Непрерывная замкнутая кривая — непрерывное отображение $f : [a,b] \to \mathbb{R}^2 : f(a) = f(b)$. Непрерывное отображение $F : [a,b] \times [0,1] \to \mathbb{R}^2$ называется гомотопией кривой $f_0(t) = F(t,0)$ в кривую $f_1(t) = F(t,1)$, если для произвольного $t \in [0,1] f_t(t) = F(x,t)$ задает замкнутую кривую: $f_t(a) = F(a,t) = F(b,t) = f_t(b)$

Понятие гомотопии кривой настолько сильно, что позволяет доказать основную теорему алгебры:

Теорема 1.6. $P(z) = z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0$ степени $n \geq 1$ с $a_i \in \mathbb{C}$ имеет $c \in \mathbb{C} P(c) = 0$.

Итак, мы видим, что было осуществлено обобщение понятия непрерывности на произвольные метрические пространства, разработано понятие гомеоморфизма для графов (и его требовалось обобщить на более широкий класс геометрических объектов). Необходимо было создать аксиоматику для науки о математических объектах, инвариантных относительно гомеоморфизмов. Такое обобщение, к примеру, мы можем увидеть у Хаусдорфа в книге «Теория множеств», где операции над метрическими пространствами или графы были обобщены до множеств с заданной на ней операцией замыкания. Опишу эту аксиоматику.

Определение 1.9. Говорят, что на множестве R задана операция замыкания, если $\forall M \subseteq R$ поставлено в соответствие подмножество $\overline{M} \subseteq R$, называемое замыканием множества M. Множество R, на котором задана операция замыкания, называется общим топологическим пространством, его элементы — точками пространства R; подмножества — точечные множества. Точки $\overline{M}$ — точки прикосновения M.

Пример 1.10. Пусть $R = \mathbb{R}$, x — точка прикосновения $m \Leftrightarrow \forall (a,b) : x \in (a,b) : \exists m \in M : m \in (a,b)$. Этим устанавливается операция замыкания, а такое общее топологическое пространство R называется числовой прямой.

Определение 1.11. Общее топологическое пространство R называется топологическим, если:

$\overline{A \cup B} = \overline{A} \cup \overline{B}$

$A \subseteq \overline{A}$

$\overline{\overline{A}} \subseteq \overline{A}$

$\overline{0} = 0$

Определение 1.12. Множество $F \subseteq R$ называется замкнутым, если $\overline{F} \subseteq F$.

Утверждение 1.13. F замкнуто $\Leftrightarrow \overline{F} = F$ по аксиоме 2.

Определение 1.14. Множество G открыто, если $R \backslash G$ замкнуто.

Дальнейшие улучшения данной аксиоматики привели к современному пониманию топологических пространств. Рассмотрю ее.

1.2 Теоретико-множественная топология

Определение 2.1 . [8] Пусть X — множество, Ω — набор его подмножеств, для которого:

1. Объединение любого семейства множеств, принадлежащих Ω, также принадлежит Ω

2. Пересечение любого конечного семейства множеств, принадлежащих Ω, также принадлежит Ω

3. $\emptyset, X \in \Omega$

В таком случае Ω называется топологией (или топологической структурой) в X. Пара (X, Ω) называется топологическим пространствам. Элементы X называются точками пространства. Элементы Ω называются открытыми множествами пространства (X, Ω). Условия 1–3, описанные выше, называются аксиомами топологической структуры

Зачастую топологическое пространство (X, Ω) обозначают как X.

Пример 2.2. Пусть X — произвольное множество, $\Omega = 2^X$. Такая топология называется дискретной топологией.

Пример 2.3. Пусть X — произвольное множество, $\Omega = \{\emptyset, X\}$. Такая топология называется антидискретной, а также топологией слипшихся точек.

Пример 2.4. $X = [0, +\infty), \Omega = \{\emptyset\} \cup \{(a, +\infty) : a \geq 0)\} \Rightarrow (X, \Omega)$ — топологическая структура, называемая стрелкой.

Пример 2.5. $X = \mathbb{R}, \Omega$ — совокупность объединений всевозможных семейств открытых интервалов. Тогда (X, Ω) — это топологическая структура, называемая вещественной прямой. (лучше сформулировать + доказать примеры)

Определение 2.6. Множество $F \subseteq X$ в пространстве (X, Ω) называется замкнутым, если $X \backslash F$ открыто.

Упражнение 2.7. Любой отрезок $[a,b]$ замкнут в $\mathbb{R}$.

Решение. Очевидно, что $(-\infty,a)\cup(b,+\infty)=(\bigcup\limits_{\alpha<a}(\alpha,a))\cup(\bigcup\limits_{\beta>b}(b,\beta))\Rightarrow\overline{(-\infty,a)\cup(b,+\infty)}=[a,b]$ — дополнение открытого множества.

Упражнение 2.8. $[0,1)$ не открыт и не замкнут в $\mathbb{R}$, но представим и как объединение замкнутых множеств, и как пересечение открытых.

Решение. Не замкнут, поскольку дополнение $[0,1)$ равно $(-\infty,0)\cup(0,1)\cup(1,+\infty)\cup\{1\}$ не является объединением интервалов из-за одноточечного множества $\{1\}\Rightarrow$ дополнение не открыто. Не открыт, поскольку равен $\{0\}\cup(0,1)\Rightarrow$ не является объединением интервалов из-за одноточечного множества $\{1\}$. При этом $[0,1]=\bigcup\limits_{0<a<1}[0,a)=\bigcap\limits_{1>a>0}(a,1)$

Упражнение 2.9. $A=\{0\}\cup\{\frac{1}{n}\}_{n=1}^{+\infty}$ замкнуто в $\mathbb{R}$.

Решение. $A=\overline{(\cup_n(0,\frac{1}{n}))\cup(-\infty,0)}\Rightarrow$ дополнение открыто.

Рассмотрим теперь некоторые понятия общей топологии, позволяющие более удобно работать с пространствами.

Удобно задавать топологическую структуру с помощью некоторой ее части, восстанавливая всю топологию с помощью операций над этой частью.

Определение 2.10. Базой топологии называется некоторый набор открытых множеств такой, что всякое непустое открытое множество представимо в виде объединения множеств из некоторого набора. Рассмотрим один из таких способов.

Пример 2.11. Базой $\mathbb{R}$ является $\{(a,b):a<b\}$.

Для алгебраических структур есть понятия подпространства, подгруппы и т. д. Интересно было бы задать аналогичное понятия и для топологических структур.

Определение 2.12. Пусть (X,Ω) — топологическое пространство, $A\subseteq X$, $\Omega_A=\{A\cap V:V\in\Omega\}$. Пара (A,Ω_A) называется подпространством пространства (X,Ω), а Ω_A — топологией, индуцированной в A топологией Ω.

Упражнение 2.13. Определение 2.12 корректно: Ω_A является топологической структурой в A

Решение. Доказывается простой проверкой аксиом.

Упражнение 2.14. Стандартная топология в $\mathbb{R}$ и топология, индуцированная в $\mathbb{R}$ как в подмножестве $\mathbb{R}^2$, совпадают.

Решение. Следует из того, что пересечение интервала и круга — это интервал $\Rightarrow$ пересечение интервала и объединения кругов — это объединение интервалов. При этом $\forall(a_i,b_i):\exists B_{\frac{b_i-a_i}{2}}(\frac{b_i-a_i}{2})=:B_i:B_i\cap\mathbb{R}=(a_i,b_i)$.

Упражнение 2.15. F замкнуто в $A\subseteq X\Leftrightarrow F=A\cap E$

Решение. Следует из законов Моргана.

В пункте 1.1 было рассмотрено понятие замыкание. Ему есть место и в современной аксиоматике топологических пространств. Рассмотрим термины, позволяющие говорить о расположении точек относительно множеств.

Определение 2.16. Пусть $x\in X$ — точка топологического пространства, $x\in U\subseteq X$ открыто. Тогда U называется окрестностью точки x пространства X.

Определение 2.17. Пусть X — топологическое пространство, $a \in A \subseteq X$. Тогда a называется внутренней точкой множества A, если некоторая окрестность точки a лежит в A, внешней точкой A, если существует окрестность точки a, не пересекающаяся с A, граничной точкой A, если любая окрестность a имеет непустое пересечение и с A, и с $X \backslash A$.

Определение 2.18. Замыкание множества A $\mathrm{Cl}\, A$ — это наименьшее замкнутое множество, в котором лежит A.

Вспомним теорему из курса матанализа о том, что множество $C \subseteq \mathbb{R}$ замкнуто $\Leftrightarrow C = \emptyset \vee \forall \{a_n\}_{n=1}^{+\infty} : a_i \in C : a_n \underset{n\to\infty}{\to} a : a \in C$. В данном случае замкнутость (следовательно, и открытость) множества определяется через последовательности. Эта идея может быть использована для построения в некотором роде альтернативной аксиоматики топологического пространства, которая называется секвенциальным подходом к топологии.

Определение 2.19. Пусть $A \subseteq X, X$ — топологическое пространство. Секвенциальным замыканием A $\mathrm{Scl}\, A$ называется множество пределов последовательностей точек из A. Отображение $f : X \to Y$ секвенциально непрерывно, если $\forall x \in X, \{a_n\}_{n=1}^{+\infty} : a_i \in X : a_n \underset{n\to\infty}{\to} a : f(a_n) \underset{n\to\infty}{\to} f(a)$.

1.3 Отображения топологических пространств

Для начала рассмотрим в качестве напоминания несколько важных понятий из теории множеств, связанных с отображениями.

Определение 3.1. Отображение $g : Y \to X$ называется обратным к $f : X \to Y$, если $g \circ f = \mathrm{id}_X$, $f \circ g = \mathrm{id}_Y$.

Определение 3.2. Пусть $A \subseteq X$, где X — некоторое множество. Тогда отображение $\mathrm{in} : A \to X$ называется включением, если $\mathrm{in} : x \mapsto x$.

Уже было отмечено, что непрерывность была обобщена на произвольные метрические пространства. Однако современное понятие непрерывности является еще более общим. Рассмотрим его.

Определение 3.3 . [8] Пусть X, Y — топологические пространства. Отображение $f : X \to Y$ называется непрерывным, если прообраз любого открытого подмножества пространства Y является открытым подмножеством пространства X.

Упражнение 3.4. Непрерывно ли в топологии $[0,2]$, индуцированной прямой, $f : x \mapsto \begin{cases} x, & x \in [0,1) \\ 3-x, & x \in [1,2] \end{cases}$?

Решение. $U = (1,2] = [0,2] \cap (1,3)$ открыто, но $f^{-1}(1,2] = [1,2)$ не открыто $\Rightarrow$ не непрерывно.

Упражнение 3.5. Непрерывно ли $f(x) = \begin{cases} x, & x \in [0,1] \\ x+1, & x \in (1,2] \end{cases}$ из $[0,2]$ в стрелку?

Решение. В стрелке открыто $(a, +\infty)$. Очевидно, что $f^{-1}(a, +\infty)$ будет открытым.

Аксиоматизируем теперь понятие гомеоморфизма, которое интуитивно должно пониматься как отображение, деформирующее пространство, не разрывая его.

Определение 3.6. [8] Отображение $f : X \to Y$ называется гомеоморфизмом, если оно непрерывно, обратимо, и обратное к нему непрерывно.

Определение 3.7. Говорят, что пространство X гомеоморфно пространству Y, если существует гомеоморфизм из X в Y.

Пример 3.8. $(-1,1) \simeq \mathbb{R}$, гомеоморфизм задает функция $f(x) = \mathrm{tg}(\frac{\pi}{2}x), f^{-1}(y) = \frac{2}{\pi}\mathrm{arctg}(y)$ непрерывны.

Стоит отметить, что гомеоморфизм не будет задавать отношение эквивалентности но множестве топологических пространств, если убрать хоть одно из условий в определении. Например, рассмотрим $f : [0,2\pi) \to S^1 := \{(x,y) : x^2+y^2 = 1\}$. Пусть $f(t) = (\cos t, \sin t)$. Тогда f непрерывно, взаимоοднозначно и имеет $f^{-1} : S^1 \to [0,2\pi)$. Однако обратное отображение уже не является непрерывным $\Rightarrow f$ — не гомеоморфизм в определенном выше смысле $\Rightarrow$ гомеоморфизм, если бы в его определении не требовалась непрерывность обратного отображения, не задавал бы отношение эквивалентности.

Рассмотрим еще некоторые важные классы непрерывных отображений.

Определение 3.9 Пусть X,Y — произвольные множества $A \subseteq X, B \subseteq Y$. Тогда всякому отображению $f : X \to Y : f(A) \subseteq B$ ставится в соответствие отображение $\mathrm{ab}(f) : A \to B, x \mapsto f(x)$, называемое сокращением отображения f. Если $B = Y$, то $\mathrm{ab}(f) : A \to Y$ обозначается символом $f|_A$ и называется сужением f на A.

Определение 3.10. Непрерывное отображение $f : X \to Y$ называют вложением, если $\mathrm{ab}(f) : X \to f(X)$ — гомеоморфизм.

Определение 3.11. Вложения $f_1, f_2 : X \to Y$ называются эквивалентными, если $\exists$ гомеоморфизмы $h_X : X \to X, h_Y : Y \to Y : f_2 \circ h_X = h_Y \circ f_1$

Рассмотрим еще один вид непрерывных деформаций, на этот раз это деформации отображений, а не пространств.

Определение 3.12. [8] Пусть f,g — непрерывные отображения топологического пространства X в топологическое пространство Y, $H : X \times I \to Y$ — непрерывное отображение такое, что $H(x,0) = f(x), H(x,1) = g(x)$ $\forall x \in X$. Тогда отображения между f и g называют гомотопными, а H называется гомотопией между f и g.

Пусть для $(x,t) \in X \times I$ $h_t(x) := H(x,t)$. Тогда H можно рассматривать как семейство отображений h_t, занумерованных числами $t \in I$. Условия $H(x,0) = f(x)$ и $H(x,1) = g(x)$ можно заменить в такой нотации на $h_0 = f, h_1 = g$. Тем самым гомотопию между f и g можно рассматривать и как непрерывное семейство непрерывных отображений, соединяющие f и g.

Утверждение 3.13. Гомотопность — это отношение эквивалентности.

Гомотопность, будучи отношением эквивалентности, определяет множество всех непрерывных отображений пространства X в пространство Y $C(X,Y)$ на классы эквивалентности, называемые гомотопическими классами.

Определение 3.14. Топологические пространства X и Y называются гомотопически эквивалентными, если существует гомотопическая эквивалентность $X \to Y$.

Пример 3.15. Пусть $f,g : [-2,2] \to [-2,2]$ такое, что $g(x) = x, f(x) = 0$. Тогда $f_t(x) = tx$ задает гомотопию между ними.

Полезно знать также следующее обобщение гомеоморфизма пространств:

Определение 3.16. $X,Y \subseteq \mathbb{R}^d$ изотопны, если $\exists$ непрерывное семейство гомеоморфизмов $H : [0,1] \times$

$\mathbb{R}^d \to \mathbb{R}^d$ (H непрерывно) такое, что $H_t := H(t,\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ является гомеоморфизмом при любом t.

1.4 Топологические инварианты и конструкции

Топологические инварианты — это свойства, сохраняющиеся при гомеоморфизме. Соответственно, зная, какие бывают топологические инварианты, мы можем доказать, что два пространства не гомеоморфны, просто посчитав количество компонент связности или проверив, компактно ли пространство.

Один из важнейших топологических инвариантов — это связность, обобщение понятия связности для графов.

Определение 4.1. Топологическое пространство X называется связным, если любое его подмножество, открытое и замкнутое одновременно, либо пусто, либо совпадает со всем пространством X.

Упражнение 4.2. Топологическое пространство связно $\Leftrightarrow$ его нельзя разбить на два непустых открытых множества $\Leftrightarrow$ его нельзя разбить на два непустых замкнутых.

Решение. Пусть $X = A \cup (X \backslash A)$, где $\emptyset \neq A \neq X$ открыто (в случае замкнутости решение аналогично). Очевидно, что множество A открыто и замкнуто $\Leftrightarrow A, X\backslash A$ открыты $\Leftrightarrow A, X\backslash A$ замкнуты. Тогда $A, X\backslash A$ открыты и замкнуты одновременно $\Rightarrow$ одно из них пусто — противоречие.

Упражнение 4.3. $[0,1]$ с топологией $\mathbb{R}$ связен.

Решение. Пусть $[0,1] = U \cup V$ — разбиение на открытые подмножества. Пусть $1 \in V, a = \sup U$. Заметим: $a = 1 \Rightarrow a \in V, a < 1 \wedge a \in U \Rightarrow \exists \varepsilon : U_\varepsilon(a) \subseteq U \subseteq [0,1] \Rightarrow b := a = \varepsilon/2 \in U, b > a$ — противоречие с определением sup. Итак, мы доказали, что $a \in V$. Тогда $a \neq 0$. Действительно: $a \in V$ — открыто $\Rightarrow \exists \varepsilon : U_\varepsilon(a) \subseteq V \Rightarrow (a-\varepsilon, a) \cap U = \emptyset$ — снова противоречие с определением sup. Итак, от противного связность отрезка доказана.

Упражнение 4.4. Образ связного пространства при непрерывном отображении связен

Решение. Пусть X связно, $f : X \to Y$ непрерывно, $f(X) = U \cup V \Rightarrow f^{-1}(U), f^{-1}(V)$ — разбиение X. Противоречие.

Теперь перейдем к понятию линейной связности, которое обобщает понятие выпуклости.

Определение 4.3. Путем в топологическом пространстве X называется непрерывное отображение отрезка $I = [0,1]$ в X. Началом пути $s : I \to X$ называется точка $s(0)$, концом — точка $s(1)$. При этом говорят, что пусть s соединяет $s(0)$ и $s(1)$.

Определение 4.4. Пусть $u : I \to X, v : I \to X : u(1) = v(0)$. Тогда можно определить произведение путей

$$uv(t) = \begin{cases} u(2t), & t \in [0, \frac{1}{2}] \\ v(2t-1), & t \in [\frac{1}{2}, 1] \end{cases}$$

Определение 4.5. Топологическое пространство называется линейно связным, если в нем любые две точки можно соединить путем.

Упражнение 4.6. I связен.

Решение. $a, b \in I \Rightarrow f(t) = at + (1-t)b$ — пусть, соединяющий a и b.

Упражнение 4.7. Линейное пространство связно.

Решение. Пусть X линейно связно, но $X = a \in U \cup V \ni b, \gamma$ соединяет a, b. Тогда $\gamma^{-1}(U), \gamma^{-1}(V)$ — разбиение отрезка. Противоречие.

Для обсуждения следующего инварианта необходимо напомнить один термин из теории множеств.

Определение 4.8. Множество $\Gamma \subseteq 2^X$ называется покрытием X, если $X = \bigcup\limits_{A \in \Gamma} A$.

Теперь рассмотрим компактность. Компактность — это топологическое обобщение свойства множества быть замкнутым и ограниченным.

Определение 4.9. Топологическое пространство называется компактным, если любое его покрытие открытыми множествами содержит конечную часть, также являющуюся покрытием.

Определение 4.10. Топологическое пространство называется секвенциально компактным, если любая последовательность его точек содержит сходящуюся последовательность.

Естественно стандартные теоретико-множественные конструкции (фактормножества, прямые произведения) распространить и на топологические пространства.

Определение 4.11. Пусть X, Y — топологические пространства. Множества $U \times V$, где U, V открыты в X, Y соответственно, называются элементарными.

Определение 4.12. Произведением пространств $X < Y$ называется множестве $X \times Y$ с топологией, базой которой является совокупность всех элементарных множеств.

Определение 4.13 Фактормножество $X \backslash S$ топологического пространства X по любому его разбиению S на непустые подмножества наделяется естественной топологией: $U \subseteq X \backslash S$ открыто в $X \backslash S$, если открыт его прообраз $\mathrm{pr}^{-1}(U)$ при отображении факторизации $\mathrm{pr} : X \to X \backslash S$.

Для удобного описания факторпространств используются термины склеивание, стягивание и отождествление. Дадим им формальное определение.

Определение 4.14. Факторизация пространства X по разбиению, состоящему из множества A и одноточечных подмножеств дополнения $X \backslash A$, называется стягиванием множества A в точку, а соответствующее факторпространство обозначается как $X \backslash A$.

Если A, B — непересекающиеся подпространства пространства X и $f : A \to B$ — гомеоморфизм, то факторизация пространства X по разбиению на одноточечные множества $X \backslash (A \cup B)$ и двухточечные множества $\{x, f(x)\}, x \in A$ называется склеиванием или отождествлением A и B с помощью гомоморфизма f.

Разбиения также удобно описывать с помощью соответствующего отношения эквивалентности. Например, факторпространство пространства X, полученное с помощью отождествления A и B с помощью гомеоморфизма $f : A \to B$ и обозначается через $X \backslash_{[a \sim f(a) \forall a \in A]} =: X \backslash_{[a \sim f(a)]}$.

Пример 4.15. $I_{[0 \sim 1]}$ гомеоморфно S^1

1.5 Симплициальные комплексы

Симплициальный комплекс — одна из важнейших комбинаторных топологических структур. Над ними крайне удобно производить вычисления, и они часто встречаются при решении задач топологического анализа данных. Именно поэтому мы их здесь рассмотрим.

Определение 5.1. [4] Симплициальным комплексом на конечном множестве вершин M называется совокупность $R \subseteq 2^M$, удовлетворяющая следующим двум условиям:

1. $I \in K, J \in K \Rightarrow J \in k$

2. $\emptyset \in K$.

Элементы множества M называются вершинами симплициального комплекса K, элементы $I \in K$ — его симплексами, а если $i \in I$, то говорят, что i есть вершина симплекса I. Множество вершин M обозначается $V(K)$. Число $|I|-1$ называется размерностью симплекса I и обозначается $\dim I$. Размерность симплициального комплекса K определяется как максимальная размерность его симплексов.

Рассмотрим некоторые примеры симплициальных комплексов.

Пример 5.2. Симплициальный комплекс размерности не больше 1 является простым графом.

Пример 5.3. n pt = $\{\emptyset, \{1\}, \{2\}, \ldots, \{n\}\}$ — симплициальный комплекс из n точек.

Пример 5.4. Пусть $l \in \mathbb{Z}_{\geq 0}, K$ — симплициальный комплекс на M. Тогда $K_l = \{I \in K : \dim I \leq l\}$ — симплициальный комплекс на M, называемый l-мерным остовом комплекса K.

Мы рассмотрели комбинаторное определение симплициального комплекса. Рассмотрим его геометрические интерпретации.

Определение 5.5. n-мерный симплекс — это выпуклая оболочка набора из $n+1$ точек $v_0, \ldots, v_n$, не лежащих в одной $(n-1)$-мерном аффинном подпространстве. Тогда v_i — вершины симплекса, а сам симплекс обозначается $[v_0, \ldots, v_n]$. Правильный n-мерный симплекс $\Delta^n = \{(t_0, \ldots, t_n) \in \mathbb{R}^{n+1} : t_i \geq 0, \sum\limits_i t_i = 1\}$.

Определение 5.6. Пусть K — симплициальный комплекс на множестве $[m] = \{1, \ldots, m\}$, $e_1, \ldots, e_m$ — базис пространства $\mathbb{R}^m$, и пусть $\Delta_I = \text{conv}(e_i : i \in I)$ — симплекс, натянутый на базисные векторы, соответствующие индексам из подмножества $I \subseteq [m]$. Тогда $|K| = \bigcup\limits_{I \in K} \Delta_I \subseteq \mathbb{R}^m$ называется геометрической реализацией симплициального комплекса K.

Аналогично обозначениям топологических пространств, зачастую геометрическая реализация симплициального комплекса K также обозначается K.

Определение 5.7. Симплициальные комплексы K и L называются изоморфными, если между множествами их вершин существует биекция, индуцирующая биекцию между K и L.

Утверждение 5.8. Граница любого трехмерного многогранника гомеоморфна S^2.

Рассмотрим теперь понятие триангуляции, одно из важнейших для топологического анализа данных.

Определение 5.9. Пусть X — топологическое пространство, K — симплициальный комплекс такой, что $|K| \sim X$. Тогда K называется триангуляцией X.

Рассмотрим некоторые классы симплициальных комплексов, особенно полезных для топологического анализа данных. Подробнее и с более общей точки зрения они будут рассмотрены в главе 7.

Определение 5.10. Облако точек — это любое конечное подмножество $X \subseteq \mathbb{R}^d$.

Определение 5.11. Комплекс Чеха на облаке точек X — это симплициальный комплекс $\hat{C}_\varepsilon(X)$: $[x_0,\ldots,x_n] \in \hat{C}_\varepsilon(X) \Leftrightarrow B_\varepsilon(x_0) \cap \ldots \cap B_\cap(x_n) \neq \emptyset$.

Определение 5.12. Альфа-комплекс на облаке точек X — это симплициальный комплекс $V_\alpha(X)$ такой, что $[x_0,\ldots,x_n] \in V_\alpha(X) \Leftrightarrow V_\alpha(x_0) \cap \ldots \cap V_\alpha(x_n) \neq \emptyset$, где $V_\alpha(x) = V(X,x) \cap B_\alpha(x)$, где $V(X,x)$ — ячейка Воронова точки $x \in X$: $V(X,x) = \{v \in \mathbb{R}^d : \forall y \in X : ||x-v|| \leq ||y-v||\}$.

1.6 Симплициальные гомологии

Гомологии — классическое понятие алгебраической топологии, которое позволяет алгебраическими методами исследовать топологические свойства пространств (путем сопоставления алгебраического объекта топологическому). Рассмотрю его. Интуитивно симплициальные гомологии подсчитывают дырки в симплициальном комплексе. Однако необходимо это интуитивное представление формализовать с помощью алгебры.

Определение 6.1. Пусть R — фиксированное кольцо. R-модулем называется абелева группа M, снабженная операцией $R \times M \to M$ со свойствами:

1. $\forall r_1, r_2 \in R, m \in M : (r_1 r_2)m = r_1(r_2 m)$
2. $1 \cdot m = m$
3. $\forall r_1, r_2 \in R, m \in M : (r_1 + r_2)m = (r_1 m) + (r_2 m)$
4. $\forall r \in R, m_1, m_2 \in M : r(m_1 + m_2) = rm_1 + rm_2$

Определение 6.2. Пусть S — конечное множество. Свободным R-модулем на множестве S называется модуль $R < S >$, состоящий из всевозможных выражений вида $\sum\limits_{s \in S} a_s \cdot s, a_s \in R$.

Определение 6.3 . [4] Пусть K — симплициальный комплекс на множестве вершин $[m]$. Будем считать, что $1 < 2 < \ldots < m$. Фиксируем основное кольцо R — поле или кольцо $\mathbb{Z}$. При $j \geq 0$ определим свободный R-модуль, порожденный множеством симплексом размерности j комплекса K: $C_j(K;R) = \{ \sum\limits_{I \in K, \dim I = j} a_I \cdot I : a_I \in R\}$. Этот модуль называется модулем j-мерных цепей комплекса K с коэффициентами в R.

Определение 6.4. Пусть M,N — R-модули. Отображение $f : M \to N$ называется гомоморфизмом R-модулей, если:

1. f — это гомоморфизм абелевых групп
2. $\forall r \in R : \forall m \in M : f(rm) = rf(m)$

Итак, мною были разобраны основные понятия, связанные с R-модулями в алгебре. Теперь введем более специальные термины, нужные для корректного определения симплициальных гомологий.

Определение 6.5. Пусть $I = \{i_0,\ldots,i_j\}$. Положим $\partial I = \sum\limits_{s=0}^{j} (-1)^s \{i_0 \ldots \hat{i_s} \ldots i_j\} \in C_{j-1}(K;R)$. Получим гомоморфизм модулей $\partial : C_j(K;R) \to C_{j-1}(K;R)$, называемый симплициальным дифференциалом.

В простейшем случае $R = \mathbb{Z}_2$. тогда дифференциал будет равен просто сумме по всем k симплексов, порожденных всеми вершинами, кроме i_k, поскольку в $\mathbb{Z}_2$ $(-1)^s = 1$.

Итак, мы получили послежовательность гомоморфизмов модулей $\ldots \xrightarrow{\partial} C_j(K;R) \xrightarrow{\partial} C_{j-1}(K;R) \to \ldots \to 0$.

Определение 6.6. Модуль j-мерных симплициальных циклов определяется как $Z_j(K;R) = \mathrm{Ker}(\partial : C_j(K;R) \to C_{j-1}(K;R))$. Модуль j-мерных симплициальных границ определяется как $B_j(K;R) = \mathrm{Im}(\partial : C_{j+1}(K;R) \to C_j(K;R))$.

Итак, основные определения6 связанные с симплициальными гомологиями, сформулированы. Теперь перейдем к самому важному.

Утверждение 6.7. Легко заметить, что $B_j(K;R) \subseteq Z_j(K;R) \subseteq C_j(K;R)$.

Определение 6.8. Модулем j-мерных гомологий с коэффициентами в R называется фактор-модуль $H_j(K;R) := Z_j(K;R) \backslash B_j(K;R)$. Притом $\beta_j(K) = \dim_R H_j(K;R)$ называется j-м числом Бетти комплекса K.

Определим теперь гомологии топологического пространства так, чтобы спообс определения не зависел от триангуляции.

Определение 6.9. Пусть X — топологическое пространство. Сингулярным j-мерным симплексом в X называется непрерывное отображение $\sigma : \Delta^j \to X$, где Δ^j — j-мерный симплекс. Рассмотрим $C_j(X;R)$ — свободный R-модуль, порожденный всеми j-мерными сингулярными симплексами пространства X. Пусть Δ^j имеет вершины $p_0, \ldots, p_j$. Введем сингулярный дифференциал $\partial : C_j(X;R) \to C_{j-1}(X;R)$, действующий на сингулярном симплексе как $\partial\sigma = \sum\limits_{s=0}^{j} (-1)^s \sigma|_{\{p_0,\ldots,\hat{p}_s,\ldots,p_j\}}$. Имеем последовательность R-гомоморфизмов $\ldots \xrightarrow{\partial} C)j(X;R) \xrightarrow{\partial} \ldots \xrightarrow{\partial} 0$. Тогда группой j-сингулярных гомологий пространства X с коэффициентами в R называется R-модуль

$$H_j(X;R) = \frac{Ker(\partial : C_j(X;R) \to C_{j-1}(X;R))}{Im(\partial : C_{j+1}(X;R) \to C_j(X;R))}$$

Рассмотрим теперь идею устойчивых гомологий. Она берется из следующего мысленного эксперимента: представим, что с течением времени симплициальный комплекс меняется, а не является статичным объектом. Множество вершин не изменяется, а вот симплексы рождаются и умирают. Чем дольше живет гомологический цикл, тем он более устойчив — устойчивые гомологии формализуют эту идею. В частности, понятие фильтрации симплициального комплекса формализует частный случай такой ситуации: когда симплексы рождаются, но не умирают.

Для описания устойчивых гомологий используют баркод и диаграмму устойчивости. Баркод наглядно показывает время жизни каждой компоненты связности симплициального комплекса, отмечая для каждой компоненты связности интервал $[t_{birth}, t_{death})$.

Если гомологий слишком много, то удобнее использовать диаграмму устойчивости. Для этого на плоскости $\mathbb{R}^2$ отмечают векторы (t_{birth}, t_{death}) для каждой компоненты.

Более формально определить понятья рождения и смерти компоненты связности можно с помощью понятия модулей устойчивости.

Определение 3.11. Последовательность R-модулей и гомоморфизмов между ними $A_0 \xrightarrow{x} A_1 \xrightarrow{x} A_2 \xrightarrow{x} \ldots$ называется модулем устойчивости.

Построения, связанные с этим понятием, не будут применяться в данной работе при анализе данных и поэтому не будут здесь вопроизведены.

1.7 Нервы покрытий

Наконец, мы рассмотрим понятие нерва покрытия и теорему Александрова о нерве. Именно результаты, полученные в этой области, позволяют теоретически обосновать применение топологических методов для анализа данных.

Для начала введем основные определения.

Определение 7.1. [4] : Нервом покрытия $U = \{U_i\}_{i\in[m]}$ пространства X называется симплициальный комлекс K_U на множестве вершин $[m] = \{1,\dots,m\}$, определяемый следующим образом:

$$K_U = \{I \subseteq [m] \mid \bigcap_{i\in I} U_i \neq \emptyset\}$$

Таким образом, для построения нерва мы каждому элементу U_i покрытия сопоставляем вершину i, после чего для $\emptyset \neq I = \{i_1,\dots,i_k\} \subseteq [m]$ рассмотрим $\bigcap\limits_{i\in I} U_i =: U_I \subseteq X$. Отдельно рассмотрим случай $I = \emptyset : U_\emptyset := X$.

До сих пор мы рассмотрели только один способ перейти от облака точек к симплициальному комплексу — триангуляции. Однако этот способ неудобен в силу того, что во многих случаях требуются очень сложные вычисления. Как раз в таких случаях гораздо лучше использовать нервы покрытий, чтобы построить покрытие облака точек, по нему построить нерв, и дальше для исследования топологических свойств облака точек исследовать топологические свойства симплициальных комплексов методами, описанными в предыдущих главах.

Теперь сформулируем ключевую для топологчиеского анализа данных теорему.

Определение 7.2. [4] : Покрытие U называется стягиваемым, если $\forall I \in K : I \neq \emptyset\ U_I$ — стягиваемое пространство.

Теорема 7.3 (Теорема Александрова о нерве покрытия). Если U — стягиваемое покрытие пространства X, то $X \simeq K_U$.

Таким образом, нервы стягиваемых покрытий стягиваемо. Заметим, что шар и сфера являются стягиваемыми. На основе этого наблюдения возникает идея метода, с помощью которого можно использовать топологию для анализа данных. Рассмотрим эти вещи.

Пусть $X \subseteq \mathbb{R}^n$ — облако точек (набор данных). Предположим, что ёти данные лежат на некотором подпространстве $A \subseteq \mathbb{R}^n$. Необходимо исследовать топологию пространства A по выборке X. Например, гомологии A.

Для решения этой задачи воспользуемся теоремой о нерве и стягиваемостью шаров.

Итак, мы имеем конечное множество $X \subseteq \mathbb{R}^n$, топология которого имеет $|X|$ компонент связности (каждая компонента — это точка). Теперь зафиксируем число $t > 0$ и по каждой точке $x_i \in X$ посмотрим шар $B_t(x_i)$. Рассмотрим $U_t = \bigcup\limits_{x_i\in X} B_t(x_i)$. Всегда можно подобрать радиус t так, что U_t будет являться покрытием A, следовательно, $U_t \simeq A$. Можно сделать задачу еще проще: взять нерв K_t покрытия $U : \{i_1,\dots,i_k\} = I \in K_t \Leftrightarrow \bigcap\limits_{i\in I} B_t(x_i) \neq \emptyset$. Тогда мы имеем $U_t \simeq K_t \Rightarrow A \simeq K_t$. Таким образом, задачу, например, вычисления гомологий A мы свели к задаче вычисления гомологий симплициального комплекса K_t — а решение этой задачи уже было описано раньше в моей работе.

Осталась всего одна идея: подобрать t так, что U_t будет покрытием, можно не одним способом. При этом при выборе радиуса t возникают проблемы: если t будет достаточно маленьким, то объединение шаров все еще даст $|X|$ компонент связности и не исправит ситуацию. При слишком большом t все шары могут слипнуться в один кластер. Эту проблему также можно решить с помощью введенных нами раннее понятий. Для этого мы рассмотрим устойчивые гомологии K_t при увеличении параметра t.

Рассмотренную идею можно использовать для построения различных симплициальных комплексов. Рассмотрим некоторые из них.

Определение 7.4 [4] : Пусть $X = \{x_1, \ldots, x_m\} \subseteq \mathbb{R}^n$. Фильтрация Чеха для X — это фильтрация $\{K_t^{\hat{C}}\}_{t \in R_{\geq 0}}$ на множестве вершин $[m]$ такая, что

$$I \in K_t^{\hat{C}} \Leftrightarrow \bigcap_{i \in I} B_t(x_i) \neq \emptyset$$

Комплекс $K_t^{\hat{C}}$ при каком-то фиксированном значении tß$\mathbb{R}_{\geq 0}$ называется комплексом Чеха, соответствующим параметру t.

Проблема с параметром t была решена с помощью понятия фильтрации. Однако в связи с таким методом анализа данных возникает и другой вопрос: как подобрать метрику для шара? Везде в данном пункте под шарами имелись в виду шары в евклидовой метрике $\mathrm{d}_2(x,y) = \sqrt{\sum\limits_{i=1}^{n} (x_i - y_i)^2} : B_t(x) = \{y \in \mathbb{R}^n \mid d_2(x,y) \leq r\}$. Однако для некоторых данных лучше использовать манхеттенскую метрику $d_1(x,y) = \sum\limits_{i=1}^{n} |x_i - y_i|$ или чебышевскую $d_\infty(x,y) = \max\limits_{1 \leq i \leq n} |x_i = y_i|$. В результате при исслежовании данных может понадобиться даже придумать собственную метрику, которая позволила бы описать топологчиеские свойства облака точек.

Зачастую в топологическом анализе данных используется также комплекс Вьеториса–Рипса.

Определение 7.5. Пусть $X = \{x_1, \ldots, x_m\} \subseteq \mathbb{R}^n$. Фильтрацией Вьеториса–Рипса для X называется фильтрация $\{K_t^{VR}\}_{t \in \mathbb{R}_{t \geq 0}}$ на множестве вершин $[m]$ такая, что

$$I \in K_t^{VR} \Leftrightarrow \forall i, j \in I : B_{t/w}(x_i) \cap B_{t/w}(x_j) \neq \emptyset$$

.

Интересна фильтрация Вьеториса–Рипса, что верно также и следующее:

$$I \in K_t^{VR} \Leftrightarrow \forall i, j \in I : d_2(x_i, x_j) \leq t$$

Таким образом, данную фильтрауию можно задать не только для облаков точек из $\mathbb{R}^n$, но и для произвольного конечного метрического пространства $X = ([m], d)$.

При фиксированном параметре $t \in \mathbb{R}_{\geq 0}$ комплекс K_t^{VR} называется комплексом Вьеториса–Рипса.

Утверждение 7.6. $\forall X \subseteq \mathbb{R}^n : K_t^{VR} \subseteq K_t^{\hat{C}} \subseteq K_{2t}^{VR}$

Таким образом, комплексы Вьеториса–Рипса легче построить, чем комплексы Чеха. Однако у этих двух комплексов есть общие проблемы: большая размерность и большое количество вершин. Эти проблемы решают альфа-комплексы, для определения которых необходимо сперва определить триангуляции делоне и Разбиения Вороного.

Определение 7.7. Пусть $X \subseteq \mathbb{R}^n$ — не более, чем счетное, множество. Ячейкой Вороного точки x называется множество $V_x = \{y \in \mathbb{R}^n \mid d_2(x,y) \le d_2(x',y) \forall x \in X : x \neq x\}$.

Замечание 7.8. $\mathbb{R}^n = \bigcup\limits_{x \in X} V_x$.

Разбиение пространства на ячейки Вороного называется разбиением Вороного.

Определение 7.9. Комплекс Делоне K^D не более, чем счетного, множества $X = \{x_1, \dots\} \subseteq \mathbb{R}^n$ — это нерв покрытия пространства $\mathbb{R}^n$ ячейками Вороного.

$$I \in K^d \Leftrightarrow \bigcap_{i \in I} V_{x_i} \neq \emptyset$$

Определение 7.10. Пусть $X = \{x_1, \dots, x_m\} \subseteq \mathbb{R}^n$ — конечное множество точек в общем положении (например, не лежащих на одной прямой). Альфа-фильтрация X — это фильтрация $\{K_t^\alpha\}_{t \in \mathbb{R}_{\ge 0}}$, где $I \in K_t^\alpha$ — нерв покрытия множествами $V_{x_i} \cap B_t(x_i)$.

$$I \in K_t^\alpha \Leftrightarrow \bigcap_{i \in I} (V_{x_i} \cap B_t(x_i)) \neq \emptyset$$

КОмплекс K_t^α называют альфа-комплексам.

К множествам $V_{x_i} \cap B_t(x_i)$ применима теорема о нерве $\Rightarrow$ их можно использовать для построения устойчивых гомологий вместо обычных шаров $B_t(x_i)$.

Рассмотрим теперь, почему же альфа-комплекс лучше комплексов Чеха и Вьеториса–Рипса.

Предложение 7.11. $K_t^\alpha \subseteq K_t^{\hat{C}}, K_t^\alpha \subseteq K^D$.

Из этого следует, что $\dim K_t^\alpha \le \dim K^d = n$, что позволяет уменьшить размерность альфа-комплекса по сравнению с другими двумя комплексами.

1.8 Применение топологического анализа данных в теории динамических систем

В главе 7 были сформулированы и обоснованы методы топологического анализа данных. Эти методы могут быть использованы в самых разных областях математики и давать результаты, которые невозможно было бы получить при использовании каикх-либо других методов. К примеру, рассмотрим, как топологчиеский анализ данных может применяться для анализа динамических систем. В статье «Topological Data Analysis of Biological Aggregation Models» этими методами была проанализирована vicsek model — математическая модель, позволяющая описывать стаи птиц, насекомых, косяки рыб. Эта модель — это динамическая система, описывающая движение агентов, собирающихся в группы и социально взаимодействующих между собой с помощью привлечения к себе или отталкивания от себя других агентов. Статья, в которой впервые была введена эта модель, стала крайне популярной и имеет несколько тысяч цитирований. Авторы получили данные с помощью симуляции модели на компьютере. По каждой симуляции было построено облако точек в фазовом пространстве, задающим положение в пространстве и скорость перемещения.

Для начала рассмотрим саму модель. Она задается следующим образом:

$\theta_i(t+\Delta t) = \frac{1}{N}(\sum\limits_{|x_i - x_j| \leq R} \theta_j(t)) + U(-\alpha/2, \alpha/2)$

$v_i(t+\Delta t) = v_0(\cos\theta_i(t+\Delta t), \sin\theta_i(t+\Delta t)$

$x_i(t+\Delta t) = x_i(t) + v_i(t+\Delta t)\Delta t$, где:

$x_i(t) \in \mathbb{R}^2$ — позиция i-й частицы ($i = 1, \ldots, N$) во время t

$v_i(t)$ — скорость во время t i-й частицы

θ_i — направление движения, $v_0 = const, U$ — равномерная случайная величина, заданная на интервале $(-\alpha/2, \alpha/2)$, Δt — шаг времени.

Таким образом, чтобы задать параметры i-й частицы во время $t + \Delta t$, необходимо задать новое направление движения (как среднее по всем предыдущим направлениям частиц в радиусе R) и добавить случайный шум, параметризованный α. Итак, во время $t + \Delta t$ i-я частица проходит расстояние $v_0\Delta t$ с направлением θ. Все частицы располагаются в квадрате с длиной l.

В статье будут рассматриваться симуляции этой модели с параметрами $R = 1, \Delta t = 1$. Также будет полезно ввести плотность частиц $\rho = N/l^2$.

Как утверждают авторы, обычно для исследования поведения частиц используется порядковый параметр $\varphi(t) = \frac{1}{v_0 N}|\sum\limits_{i=1}^{N} v_i(t)|, \varphi(t) \in [0, 1]$. Если группа частиц движется примерно в одну и ту же сторону, то $\varphi(t) \approx 1$. Если же частицы распространяются по квадрату в $\mathbb{R}^2$ случайным и практически независимым друг от друга образом, то $\varphi(t) \approx 0$. Однако при использовании этого параметра проблемы возникают в ситуации, когда части двигаются по одной прямой, но в разных направлениях. Несмотря на то, что частицы явно движутся не независимым друг от друга образом, параметр имеет значение такое, будто бы частицы движутся в случайных направлениях. Авторы статьи решили попробовать исправить этот недостаток, характеризуя систему не с помощью $\varphi(t)$, а с помощью топологических инвариантов, которые авторы решили вычислять с помощью методов топологчиеского анализа данных.

В рамках исследования, проведенного авторами статьи, вначале были посчитаны нулевые и первые числа Бетти $b_0(t, \varepsilon), b_1(t, \varepsilon)$ построенных по данным комплексов Вьеториса–Рипса для каждого момента времени t и радиуса шаров, участвующих в построении комплекса, ε, после чего стали рассматривать получившуюся фильтрацию. Выяснилось, что, хотя при движении частиц по одной прямой в одном направлении и в противоположных направлениях, $\varphi(t) \approx 1$, топологчиеские свойства комплексов Вьеториса–Рипса, построенных по симцяциям vicsek model в первом и во втором свойстве, различаются. Более того, топологические свойства случая почти полностью случайного движения и движения в одном направлении похожи.

Таким образом, топологический анализ данных позволяет описывать свойства динамических систем и даже получать из них информацию, которую неспособны дать привычные методы исследования динамических систем.

Топологический анализ данных применяется также в биологии, в социальных науках. У данной области исследований есть большой потенциал как для теоретических, так и прикладных исследований.

1.9 Библиотеки для топологического анализа данных

В предыдущем пункте была разобрана статья, где теоретические построения из пунктов 1–7 были применены для исследования данных. Пришло время разобарться, как такие исследования проводить. Для этого существуют gudhi, dionysus (Python), phom (R).

Для целей моей работы наиболее подходящей является библиотека gudhi. Разберусь с ее документацией и опишу основные функции и принципы работы с этой библиотекой.

Под фильтрацией здесь понимается любая функция $f : K \to \mathbb{R}$, где K — симплициальный комплекс, такая, что $\tau \subseteq \sigma \Rightarrow f(\tau) \leq f(\sigma)$.

Для представления симплициальных комплексов в работе используется структура данных simplex tree. Рассмотрим, какие операции можно выполнять над этой структурой.

Для названия библиотеки gudhi принято сокращение gd:

```
import gudhi as gd
```

Инициализация структуры данных simplex tree:

```
st = gudhi.SimplexTre()
```

Вернуть числа Бетти симплициального комплекса, представленного в виде simplex tree st, как список целых чисел:

```
st.betti_numbers()
```

Вернуть размерность комплекса как целое число:

```
st.dimension()
```

Вернуть True, если симплекс, заданный списком simplex, был найден в симплициальном комлексе, и False иначе:

```
st.find(simplex)
```

Вернуть значение фильтрации для симплекса simplex, если он существует в комплексе, и $+\infty$ в ином случае:

```
st.filtration(simplex)
```

Вставить в комплекс симплекс simplex с значением фильтрации value:

```
st.insert(simplex, filtration)
```

Эта функция должна быть выполнена перед построением баркода или диаграмм устойчивости, поскольку производит необходимые вычисления:

```
st.persistence()
```

Библиотека позволяет работать с комплексами Вьеториса–Рипса и альфа-комплексами, а также строить баркоды и диаграммы устойчивости. рассмотрим функции, которые позволяют это делать.

Для построения баркодов и диаграмм устойчивости необходимо импортировать pyplot:

```
import matplotlib.pyplot as plt.
```

Пусть задано облако точек в виде списка my_points. Тогда можно построить альфа-комплекс или комплекс Вьеториса–Рипса и представить его в виду simplex tree simplex_tree с помощью следующих функций (параметр max_dimension задает размерность комплекса):

```
simplex_tree = rips_complex.create_simplex_tree(points = my_points, max_dimension=1)
alpha_complex = gd.AlphaComplex(points = my_points)
simplex_tree = rips_complex.create_simplex_tree(max_dimension=1)
```

Чтобы построить баркод, необходимо испольщовать следующие функции:

```
diag = simplex_tree.persistence(min_persistence=0.4)
gd.plot_persistence_barcode(diag)
plt.show()
```

1.10 Описание вычислительного эксперимента

Я подготовил данные с 7 непрерывными переменными по 80 странам в формате csv. К этим данным были применены функции библиотеки gudhi для вычисления альфа-фильтрации ac и фильтрации Вьеториса–Рипса rips_complex по этим данным. Были построены баркоды по каждой из фильтраций. После этого были написаны функции для нахождения точек на Торе torus_val в $\mathbb{R}^3$ и бутылке Клейна klein в $\mathbb{R}^4$. С помощью этих функций были построены облака точек, по ним построены альфа-комплексы и найдены их гомологии. Данные расположены в таблице по ссылке

https://github.com/aefrt/project-topology/blob/main/data.csv

Здесь же я приведу их лишь частично:

state	*area*	*population*	*GDP*	*publications_count*
Argentina	2780400	44694198	23283	8811
Australia	7686850	23470145	51036	53610
Austria	83858	8793370	56871	12362
Bangladesh	144000	159453001	3998	3135
Belgium	30528	11570762	52250	15688
Botswana	581730	2249104	18050	281
Brasilia	8515767	208846892	14941	60148
Brunei	5770	450565	61814	294
Canada	9984670	35881659	50078	59968
Chile	756950	17925262	24765	7122
China	9598962	1384688986	15603	528263

Таблица 1. Частично приведенные данные

Опишу переменные (они все взяты за 2018 год):

Название переменной	Описание переменной	Источник
state	Название государства	Wikipedia
area	Площадь государства	Wikipedia
population	Численность населения государства	nonews.co
GDP	ВВП государства на душу населения	Wikipedia
publications_count	Число публикаций, сделанных в стране	gtmarket.ru
primary_ed_idx	Индекс уровня начального образования в стране	nonews.co
education_spending	Государственные расходы на образование в процентах от ВВП	gtmarket.ru
avg_salary	Средняя зарплата в стране	fincan.ru

Таблица 2. Описание переменных

Данные были собраны из различных источников и объединены в одну таблицу.

Вычисления были произведены в среде JupyterLab, код доступен по ссылке

`https://github.com/aefrt/project-topology/blob/main/task.ipynb`

Из проведенного вычислительного экмперимента можно сделать вывод о том, что топологический анализ данных крайне эффективен. Для данных по странам были вычислены 0-гомологии H_0 их альфа-фильтрации, после чего сделан вывод о том, что альфа-комплекс разбивается на несколько компонент связности, из чего был сделан вывод о том, что данные разбиваются на несколько кластеров. И, действительно, после кластеризации методом k-means этот вывод был подтвержден. Таким образом, топологический анализ данных позволяет делать важные выводы о структуре данных.

2 Заключение

Из проделанной работы можно сделать вывод о том, что аксиоматика топологических пространств появилась из работ по изучению классов множеств с заданной на них операцией замыкания. Это дало возможность обобщить топологические методы, уже применявшиеся в метрической геометрии и теории графов, на многие другие геометрические объекты (см. главу 1 основной части). В результате был создан особый язык теоретико-множественной топологии (см. главу 2 основной части), были исследованы различные топологические инварианты, исследование которых было вдохновлено различными другими областями математиик, вроде теории графов (см. главу 4 основной части). В 5–6 главах я перешел к исследованию методов топологического анализа данных и прикладной топологии в целом, узнал о такой комбинаторной топологической структуре, как симплициальные комплексы, узнал, как их применять и какие существуют полезные на практике виды комплексов. Изучил понятие гомологии как способа сопоставить пространству алгебраический объект, узнал, какие гомологии бывают у симплициальных комплексов и как с ними работать на практике. Также я изучил понятие фильтрации, и уже в 7 главе рассмотрел полезные на практике виды фильтраций, привел обоснование методов топологического анализа данных. В 8 главе разобрал статью, посвященную применению прикладной топологии в анализе динамических систем, показал актуальность данного направления исследований. В данной статье было показано, что топологический анализ данных позволяет узнать информацию об облаке точек, которую с помощью каких-либо других методов получить было невозможно. В 9 главе рассмотрел основные функции библиотеки для топологического анализа данных gudhi, после чего перешел к вычислительному эксперименту. В ходе него я построил баркоды для классических топологических конструкций (тора и бутылки Клейна), а также баркод для альфа-фильтрации и фильтрации Вьеториса–Рипса, построенных по моим данным. По баркоду был сделан вывод о том, что данные разбиваются на несколько кластеров, который был подтвержден после кластеризации данных методов k-means. Был сделан вывод о том, что методы прикладной топологии позволяют делать важные выводы о структуре данных. В дальнейшем можно продолжить прочтение статей о применении прикладной топологии в различных областях, например, биоинформатике или в анализе текстов. Можно использовать библиотеку gudhi для анализа любых других датасетов, чтобы проверить, какие выводы можно будет сделать из баркода фильтрации, построенной по данным.

Список источников

1. Chazal, F. An introduction to Topological Data Analysis [Электронный ресурс]: fundamental and practical aspects for data scientists / F. Chazal, M. Bertrandm. — Электрон. текст. дан.— Режим доступа: https://arxiv.org/abs/1710.0 свободный. (дата обращения: 08.02.21)

2. Ghrist, R. Elementary Applied Topology [Электронный ресурс] / R. Ghrist. — Электрон. тест. дан. — Createspace, 2014. — 269 с.— Режим доступа: robert ghrist homepage https://www2.math.upenn.edu/ ghrist/notes.html, свободный. (дата обращения: 08.02.21)

3. Topaz, C. M. Topological Data Analysis of Biological Aggregation Models / C. M. Topaz, L. Ziegelmeier, T. Halverson. — Электрон. тест. дан. – Режим доступа: https://arxiv.org/abs/1412.6430, свободный. (дата обращения: 23.05.21)

4. Айзенберг, А. А. Методичка по симплициальным комплексам и гомологиям [Электронный ресурс] / А. А. Айзенберг. — Электрон. текст. дан. — Режим доступа: Google Drive https://drive.google.com/file/d/1u4TND54HQmLna0XSH-FZPopj6cp5pmCL/view?usp=drive_open, свободный. (дата обращения: 08.02.21)

5. Васильев, В. А. Введение в топологию / В. А. Васильев. — М.: ФАЗИС, 1997. — XII + 132 с. — Библиотека студента-математика. Вып. 37.

6. Хатчер, А. Алгебраическая топология / А. Хатчер; пер. с англ. В. В. Прасолова под ред. Т. Е. Панова — М.: МЦНМО, 2011. — 688 с.

7. Хаусдорф, Ф. Теория множеств / Ф. Хаусдорф; пер. с нем. Н. Б. Беденисова под ред. П. С. Александрова и А. Н. Колмогорова — Ленинград: НКТП СССР, 1937. — 305 стр.

8. Элементарная топология / О. Я. Виро, О. А. Иванов, Н. Ю. Нецветаев, В. М. Харламов. — 2-е изд., исправл. — М.: МЦНМО, 2010. — 358 + x с.