НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Основы прикладной топологии

Руководитель: Айзенберг Антон Андреевич, кандидат физ.-мат. наук, заведующий лабораторией в международной лаборатории алгебраической топологии и ее приложений

Автор работы: Копчев Владислав Олегович, БПМИ198

Национальный исследовательский университет
«Высшая школа экономики» (Москва)
Факультет компьютерных наук
Прикладная математика и информатика

Описание предметной области

Топологический анализ данных и прикладная топология — новейшая область исследований, раздел анализа данных, использующий топологические методы.

Основные термины, сокращения, определения

Определение 1

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

Определение 2

Топологическое пространство — множество с определенной на ней топологией.

Определение 3

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

Основные термины, сокращения, определения

Определение 4

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

Определение 5

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

Определение 6

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

Основные термины, сокращения, определения

Определение 7

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

Актуальность работы

1. Топологический анализ данных — недавно возникшая область исследований
2. Новый подход к анализу данных
3. Применение в анализе динамических систем, для анализа данных, связанных с экономикой
4. Не затронутые в работе приложения: биоинформатика, анализ текстов

Цель и задачи работы

Целью моей работы является изучение прикладной топологии и топологического анализа данных
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.

Цель и задачи работы

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.

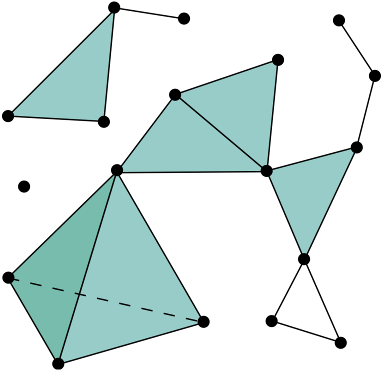
Цель и задачи работы

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.

Анализ существующих подходов / методов/ моделей / алгоритмов / решений

Существующие подходы основаны на терминах и идеях из алгебраической топологии.
Гомотопии и гомеоморфизмы как обобщения понятия деформации без разрывов. Аксиомы топологических пространств как продолжение исследований множеств с заданной на них операцией замыкания. Симплициальный комплекс как мощная комбинаторная топологическая конструкция. Нерв стягиваемого покрытия гомотопен самому пространству. Предположим, что облако точек лежит на топологическом пространстве. Тогда можно исследовать его топологическую структуру с помощью теоремы о нерве.

Simplicial complex example



Анализ существующих подходов / методов/ моделей / алгоритмов / решений

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

Анализ существующих подходов / методов/ моделей / алгоритмов / решений

Нервом покрытия $U = \{U_i\}_{i \in [m]}$ пространства X называется симплициальный комлекс K_U на множестве вершин $[m] = \{1, \ldots, m\}$, 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 $\emptyset \neq I = \{i_1, \ldots, i_k\} \subseteq [m]$ рассмотрим $\bigcap_{i \in I} U_i =: U_I \subseteq X$. Отдельно рассмотрим случай $I = \emptyset : U_\emptyset := X$.

Анализ существующих подходов / методов/ моделей / алгоритмов / решений

Фильтрация: комплекс I появляется во время t. С помощью теории R-модулей это офрмализуется.
Пусть $X = \{x_1, \dots, x_m\} \subseteq^n$. Фильтрация Чеха для X — это фильтрация $\{K_t^{\hat{C}}\}_{t \in R_{\geq 0}}$ на множестве вершин $[m]$ такая, что

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

Анализ существующих подходов / методов/ моделей / алгоритмов / решений

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

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

.

Анализ существующих подходов / методов/ моделей / алгоритмов / решений

Пусть $X \subseteq^n$ — не более, чем счетное, множество. Ячейкой Вороного точки x называется множество $V_x = \{y \in^n | \ d_2(x, y) \leq d_2(x', y) \forall x \ \in X : x \ \neq x\}$.
Пусть $X = \{x_1, \dots, x_m\} \subseteq^n$ — конечное множество точек в общем положении (например, не лежащих на одной прямой). Альфа-фильтрация X — это фильтрация $\{K_t^\alpha\}_{t \in_{\geq 0}}$, где $I \in K_t^\alpha$ — нерв покрытия множествами $V_{x_i} \cap B_t(x_i)$.

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

Выбор используемых в работе методов / алгоритмов / моделей и т.п.

В работе были выбраны методы алгебраической топологии, для визуализации метод построения баркода, имплементация алгоритмов в библиотеке gudhi.

Планирование эксперимента

1. Были объединены данные об экономических показателях 80 стран из разных источников: название госудпрства, площадь, население, ВВП, количество публикаций, индекс уровня начального образования, траты на образование, средняя зарплата.
2. Планировалось провести топологический анализ этих данных, а также искусственно сгенерированных облаков точек с тора и бутылки Клейна.

Результаты экспериментов

1. Были построены альфа-комплекс и комплекс Вьеториса–Рипса.
2. Были посчитаны баркоды для них
3. Был сделан вывод о том, что данные можно кластеризовать, который был подтвержден с помощью алгоритма k-means
4. Были найдены баркоды для бутылки Клейна и Тора.

Основные результаты и выводы. Новизна и/или практическая значимость работы.

1. В работе были проаеализированы области применения топологического анализа данных
2. Было продемонстрировано, как исследовать его в теории динамических систем, для анализ экономических показателей.

Направления дальнейшей работы.

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 использованных 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)

Список использованных источников

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)

Список использованных источников

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.

Список использованных источников

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.

Спасибо за внимание!