

КАК ОПИСАТЬ ОБЛАКО ТОЧЕК: ТОПОЛОГИЧЕСКИЙ АНАЛИЗ ДАННЫХ

Макаренко Н.Г.

ГАО РАН, г. Санкт-Петербург, Россия

ng-makar@mail.ru

Поэтому, если тебе сначала все покажется пустым бредом, знай что причиной твоя слабость!

(Н.Кузанский.Берилл.)

Топологическим анализом данных (ТДА) называют инструментарий для анализа абстрактных данных - облака точек. Примером могут быть точки, случайно выбранные из окружности. Преобразуем этот набор в «непрерывный» объект, топологически похожий на окружность, используя фильтрацию - последовательное объединение окрестностей соседних точек. Увеличивая размер окрестности, мы получим *флаг* форм, все более близких к кольцу. Слияние окрестностей уменьшает число компонент связности, до единицы - окружности. При этом рождается новые топологические свойства – дыры. Далее, фильтрация замостит дыры и даст диск. Удобно кодировать фильтрацию *комплексом* с вершинами (точками), ребрами, соединяющими центры пересекающихся пар окрестностей и гранями (пересечением трех и более окрестностей). Такой скелет называют комплексом, который можно снабдить наборами топологических инвариантов, маркирующих упомянутые свойства – компоненты и дыры. Фильтрация сложных данных позволяет получить универсальный язык для описания и диагностики множества процессов и структур из разных областей знания. Интересно, что идеи фильтрации можно вывести из некоторых удивительных свойств нашего зрения. Так, ее аналогом в живописи может служить пуантилизм Жоржа Сёра, т.е. картины сотавленные мозаикой отдельных точек. Каналы Марса, обнаруженные Скиапарелли в 1877г., оказались зрительной иллюзией: глаз соединяет близкие точки на пределе разрешения в линейные структуры. Математическим аналогом этой иллюзии является ε - связанная цепь Кантора, последовательность случайных точек с ребрами, образованными парами, с расстоянием $\leq \varepsilon$. Эту идею использовала Робинс в 1998 г., для извлечения топологии многообразия из комплексов, построенных на конечной выборке точек [1]. Комплекс включает циклы и грани. Леонард Эйлер, еще в 1752 году предложил для комплекса целочисленный инвариант, выраженный знакопеременной суммой числа входящих в него клеток разной размерности, или число классов, которые образованы циклами разной размерности. Их называют числами Бетти, важными инвариантами ТДА. В 1993г. их использовали М.Малдун и его коллеги [2], для реконструкции топологии аттрактора из наблюдений. Эта интересная работа является обобщением эмбедологии Такенса на гомологии. Настоящая Лекция является эвристическим введением в методы ТДА, с примерами некоторых их приложений. Это прогулка по проблеме, цель которой привлечь интерес молодых физиков к вычислительной топологии и облегчить их знакомство с многочисленной специальной литературой (см., например, [3–5]).

Список литературы

- 1.Robins V., et al. Computing connectedness: An exercise in computational topology // Nonlinearity. 1998. V. 11, 913.
- 2.Muldoon M. R., et al. Topology from time series // Physica D. 1993. V. 65, 1.
- 3.Kaczynski T., et al. Computational homology. New York : Springer, 2004. 496 с.
- 4.Edelsbrunner H., Harer J. L. Computational topology: an introduction. Amer. Mathem.Soc. 2010.
- 5.Dey T. K., Wang Y. Computational topology for data analysis. Cambridge University Press, 2022.