**ПЕРСИСТЕНТНЫЕ ГОМОЛОГИИ И АНАЛИЗ МЕДИЦИНСКИХ ДАННЫХ**

Каданцев Г.В., Синицын А.А., Иванов С.О.

*Лаборатория Непрерывного Математического Образования (ЛНМО)*

*Россия, Санкт-Петербург*

Топологический анализ данных — область прикладной математики, в которой топологические идеи применяются для изучения различных видов данных ([1], [2], [3]). В нашей работе мы применяем методы топологического анализа данных для изучения гистологических WSI-изображений (histological whole-slide images). Основной характеристикой изображения для нас является персистентная энтропия изображения, которая извлекается из нулевых персистентных гомологий фильтрованного симплициального комплекса изображения. Эта характеристика уже показала себя как удобный инструмент при изучении идиотипической иммунной сети (idiotypic immune network) (см. [6]). Грубо говоря, персистентная энтропия изображения —это мера хаотичности изображения, выраженная на языке топологического анализа данных. Наша цель — показать, что использование персистентной энтропии может помочь в компьютерной диагностике различных видов рака, в том числе колоректального рака и рака прямой кишки.

Наша работа идейно близка к работам [4] и [5]. В них так же исследуются гистологические WSI-изображения при помощи топологического анализа данных и строится алгоритм для компьютерной диагностики колоректального рака. Однако в этой работе используется только самые простейшие топологические идеи (только числа Бетти, на основе которых они составляют так называемые persistent homology profiles) в совокупности с машинным обучением. В нашей работе мы избегаем использования машинного обучения, но используем более глубокие математические идеи (персистентные гомологии, интервальное разложение, бар-коды, персистентная энтропия).

При исследовании в начале WSI-изображение делится на патчи (patches, маленькие кусочки исходного WSI-изображения), и каждый патч исследуется отдельно. Можно заметить, что левый патч представленный на картинке, выглядит более хаотично, а на правом видно больше структуры. Персистентная энтропия численно учитывает эту разницу.



Слева — раковые клетки; справа — здоровые

Нами был разработан алгоритм, вычисляющий интервальное разложение для нулевых персистентных гомологий и соответствующую персистентную энтропию. Этот алгоритм был реализован на языках Java и Scala. Его можно найти по ссылке: <https://github.com/Sannitsa/Persistent-homology-and-histological-data>

В результате, у нас получилось показать, что средняя персистентная энтропия изображений здоровой ткани значительно ниже средней персистентной энтропии изображений раковой опухоли.

Таким образом, мы даем ответ на концептуальный вопрос: “В чем же разница между структурой здоровой ткани и структурой раковой опухоли?” И мы даем ответ “хаотичностью”, выраженной на языке топологического анализа данных.

Также можно ознакомиться с полной версией нашей статьи, которую можно найти здесь: <https://clck.ru/N5AX3>

[1] G. Carlsson: Topology and data. Bullentin of the

American Mathematical Society 46, 2 (2009), 255–308.

[2] S. Y. Oudot: Persistence Theory: From Quiver Representations to Data Analysis, vol. 209 of Mathematical Surveys and Monographs. American Mathematical Society (2015).

[3] H. Edelsbrunner, D. Morozov: Persistent homology: theory and practice. In Proceedings of the European Congress of Mathematics, pages 31–50, (2012).

[4] Talha Qaiser, Korsuk Sirinukunwattana, Kazuaki Nakane, Yee-Wah Tsang, David Epstein, and Nasir Rajpoot: Persistent homology for fast tumor segmentation in whole slide histology images. Procedia Computer Science, 90:119–124, (2016).

[5] T. Qaiser, Y.-W. Tsang, and N. Rajpoot et al.: Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. M`ed Image Anal., 55:1-14, (2019).

[6] M. Rucco, F. Castiglione, E. Merelli, and M. Pettini.: Characterisation of the idiotypic immune network through persistent entropy. In Proceedings of ECCS 2014, 117–128. Springer, (2016).