

ПРИМЕНЕНИЕ МЕТОДОВ КЛАСТЕРИЗАЦИИ ДЛЯ ГИС МОДУЛЯ МЕДИЦИНСКОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Балтабаев С.Ш. (ИТМО)

Научный руководитель - старший преподаватель факультета ИКТ, Береснев А. Д.
(ИТМО)

Введение. Развитие здравоохранения неразрывно связано с постоянным ростом объема данных, требующих сбора и анализа. Эти данные включают в себя информацию о заболеваниях в различных регионах, числе пациентов в медицинских учреждениях и перемещениях больных между ними. По данным министерства здравоохранения Российской Федерации на 2023 год около 4 миллиона людей сталкиваются с онкологическими заболеваниями. Каждый год выявляются примерно 600 тысяч новых случаев [1]. В этом контексте применение методов кластеризации в геоинформационных системах становится ключевым, обеспечивая эффективное выявление закономерностей в распределении заболеваний и разработку персонализированных подходов к пациентам.

Основная часть. На первом этапе данного исследования был проведен обзор методов кластеризации данных в ГИС системах. В результате данного обзора были выделены некоторые методы и определены их преимущества и недостатки. В геоинформационных системах часто используется сразу несколько методов кластеризации. Выбор определенного метода зависит от типа данных и конкретной задачи. Для пространственной индексации был выбран метод геохеширования, при котором каждая географическая координата кодируется в строку символов. Данный метод будет реализовываться на языке Python и геохеши будут храниться в базе данных PostgreSQL. Реализация данного метода позволит оптимизировать визуализацию объектов на карте. Для определения областей заболеваний или группировки объектов можно использовать методы на основе плотности. В данную группу входят такие методы как DBSCAN, OPTICS и HDBSCAN. Недостатками последних двух методов является вычислительная сложность, но они требуют минимальных входных параметров и могут определить кластеры различной формы [2-3]. Реализация данных методов позволит снять нагрузку с клиентской части, а также упростит и автоматизирует процессы анализа, что обеспечит более оперативное принятие решений для профилактики или лечения заболеваний.

Выводы. Проведен обзор методов кластеризации данных в ГИС системах. Описаны их преимущества, недостатки и области применения. Определены необходимые для реализации методы, а именно методы геохеширования и DBSCAN. ГИС модуль позволит ускорить и частично автоматизировать анализ медицинских данных, что позволит быстрее принимать важнейшие решения в области здравоохранения.

Список использованных источников:

1. Минздрав сообщил о 4 млн онкологических больных в России [Электронный ресурс] // Официальный сайт РБК. URL: <https://www.rbc.ru/rbcfreenews/63dce6bf9a794740a56c9a43>
2. С. К. Дулин, И. Н. Розенберг, В. И. Уманский, Методы кластеризации в исследовании массивов геоданных, Системы и средства информ., 2009,дополнительный выпуск, 86–113
3. All you need to know about the DBSCAN Algorithm [Электронный ресурс] // Medium. URL: <https://medium.com/analytics-vidhya/all-you-need-to-know-about-the-dbscan-algorithm-f1a35ed8e712>