MAPCLUST - СЕРВИС ДЛЯ ЭПИДЕМИОЛОГИЧЕСКОГО И ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА

Смирнов Антон Сергеевич,Захаров Олег Станиславович,Пятницкий Алексей Михайлович, Гукасов Вадим Михайлович

Mar 23, 2023

Оглавление

# Аннотация

Эта статья посвящена описанию разработанного авторами веб-сервиса для кластеризации пространственных и пространственно-временных данных MapClust. MapClust является практической реализацией исследований, проведенных нами в предыдущих работах. В этой работе мы демонстрируем применение сервиса для решения задач эпидемиологического и экологического мониторинга.

Ключевые слова: MapClust; эпидемиологический мониторинг; экологический мониторинг; алгоритм кластеризации.

# Введение

Поиск кластеров в пространственных данных является нетривиальной задачей, решение которой важно для мониторинга ситуации в некоторой изучаемой области. В текущий момент, в связи с неблагоприятной эпидемиологической и экологической ситуацией в мире остро стоит задача обнаружения областей, на которые в первую очередь необходимо обратить внимание для принятия мер. Один из наиболее популярных видов методов поиска кластеров, например, в анализе эпидемиологии туберкулеза – сканирующие статистики[2]. Самый известный метод сканирующей статистики – метод Куллдорфа[1]. Однако, метод зависит от формы и размера сканирующую окна, выдает один наиболее вероятный кластер и обладает довольно большой вероятностью ложноположительного результата, то есть в кластеры часто попадают области, которые не имеют никакого отношения к ним[3]. Наш критерий не зависит от формы и размеров кластеров, позволяет регулировать свою чувствительность, осуществляет поиск нескольких кластеров при заданной ошибке первого рода.

# Принцип работы

Данный сервис получает на вход данные о взаиморасположении областей исследования и о значении исследуемой величины в этой области. Исследуемая величина зависит от цели исследования: летальность заболевания, количество случаев, площадь лесных пожаров и многое другое. Данные должны быть представлены в специальном общеиспользуемом геоинформационном формате – ESRI Shapefile. Принцип поиска кластеров описан в статьях[4–7]. Для каждой области, для которой p-value меньше или выше некоторого порога, рассчитывается статистики ([Equation 1](#eq-clust-stat)) для кластеров и ([Equation 2](#eq-discharges-stat)) для разряжений.

Программа проверяет вхождение значения статистик в критическую область S(n), где n – число регионов соседей, для которых p-value меньше или выше определенного порога. Программа строит эту область методом Монте-Карло. Кластеры или разрежения, выходящие за критическую область, являются значимыми.

# Примеры

# Заключение

# MAPCLUST IS SERVICE FOR EPIDEMIOLOGICAL AND ECOLOGICAL MONITORING

**Smirnov Anton Sergeevich1, Zakharov Oleg Stanislavovich1, Pyatnitskiy Alexey Mikhailovich1, Gukasov Vadim Mikhailovich2,**

1 - PRNRMU, Moscow,

2 - FRCEC, Moscow

# Abstract

Keywords: MapClust; epidemiological monitoring; ecological monitoring; clustering algorithm.

Ключевые слова

1. Kulldorff M. [A spatial scan statistic](https://doi.org/10.1080/03610929708831995) / M. Kulldorff // Communications in Statistics - Theory and Methods. – 1997. – Vol. 26. – № 6. – P. 1481-1496.

2. Shaweno D. [Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review](https://doi.org/10.1186/s12916-018-1178-4) / D. Shaweno [et al.] // BMC Medicine. – 2018. – Vol. 16. – Methods used in the spatial analysis of tuberculosis epidemiology. – № 1. – P. 193.

3. Tango T. [Spatial scan statistics can be dangerous](https://doi.org/10.1177/0962280220930562) / T. Tango // Statistical Methods in Medical Research. – 2021. – Vol. 30. – № 1. – P. 75-86.