

Кирилук Б.И. — рецензент *Петренко А.И.*
УНК «ИПСА» НТУУ «КПИ»

Графовая кластеризация геологических данных

Исследование землетрясений – одна из самых важных и сложных задач современной геофизики и компьютерного моделирования. За последние 50 лет много сил было потрачено на создание модели происхождения землетрясения и попытки предсказать их происшествия. На данном этапе развития науки это ещё невозможно, что очень ярко видно на примере недавнего землетрясения в Гаити. А найти некоторые зависимости, выделить проблемные регионы уже под силу современной науке [1].

Целью моей работы являлось создание программы, которая убирает шум мелких и редких землетрясений и разбивает данные на отдельные кластеры по координатам, визуализируя их в двухмерном пространстве. В качестве инструментария я выбрал пакет прикладных программ Matlab. Основной его особенностью является его широкие возможности по работе с матрицами и простая визуализация массивов данных.

Данные взяты из службы геологии, геодезии и картографии США. В них содержится информация о 30000 землетрясений с 1973 по 2010 год. Данные характеризуются координатами центра, временем происшествия, магнитудой и глубиной землетрясения.

Для кластеризации геологических данных наиболее подходят иерархические и графовые алгоритмы. В данной работе реализован графовый алгоритм на отфильтрованных данных. Земля разбита на квадраты и рассчитано количество землетрясений в каждом квадрате (плотность распределения).

Квадраты, не преодолевающие определённый порог, отсеиваются. В результате мы имеем вместо 30000 событий около 1000 точек для кластеризации. Создав матрицу расстояний между каждой парой узлов графа и рассчитав среднее расстояние между узлами, мы удаляем связи, превышающие порог. В результате – выделяем из матрицы отдельные кластеры. Программа позволяет проработать данные из разных временных промежутков. Контролируя параметры кластеров и исследовав данные с периодом 5 лет, можно получить информацию о стабилизации некоторых участков земной коры и также о проявлении сейсмической активности в ранее стабильных регионах. Наиболее опасными оказались территории Японии, Индии, Малайзии и западное побережье Южной Америки.

Кластеризация – является одним из перспективных направлений в интеллектуальном анализе данных. Алгоритм позволил выделить независимые тектонические области земной коры и проследить изменения в сейсмологической активности.

Литература

1. Fox G, Sung-Hoon Ko, Pierce M, (2002) Grid services for earthquake science. Concurrency Computation Practice Experience 14(6–7):371–393.
2. Andenberg MR (1973) Clusters analysis for applications. Academic Press, New York.