

## ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ ФОРМИРОВАНИЯ КОНТЕНТА ГЕОГРАФИЧЕСКИ РАСПРЕДЕЛЕННОГО КЛАСТЕРА

Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь

Гузов Д. С.

Отвагин А. В. – к-т. техн. наук, доцент

В настоящее время повсеместное использование сетей для дистрибуции контента приводит к существенному увеличению нагрузки на сети передачи данных. С целью минимизировать временные задержки, а также снизить нагрузку на опорные сети Интернета активно используется реплицирование данных, основанное на географическом признаке (см. рисунок 1).

Очевидно, что полное реплицирование данных на всех кластерах приводит к избыточности. При этом можно выделить дополнительные подклассы контента: например, существует контент, специфичный для конкретного региона, либо данные, потребность в которых циклически меняется в течение определенного промежутка времени (например, дня).

Использование адаптивного алгоритма позволяет не выполнять полное копирование данных на кластерах всех регионов, а поддерживать некоторую заданную степень реплицирования данных и перемещать циклически используемые данные из одного кластера в другой. Таким образом, в момент пиковой активности для конкретного региона данные будут находиться в непосредственной близости от клиентов, что положительно скажется на времени отклика, и значительно снизит нагрузку на сети передачи данных.

Целью исследования было определение возможности использования генетического алгоритма (см. рисунок 2) для определения данных, обращение к которым происходит циклически из различных регионов.

В результате был получен генотип из пяти геномов, которые с высокой степенью вероятности (более 70%) предсказывали запросы к определенным данным. На основании близости генома конкретных данных к эталонному геному стало возможно заблаговременное перемещение данных, которые будут востребованы в другом регионе. Условием успешной работы алгоритма была принята близость к оптимальному использованию сети передачи данных (75% от рассчитанного эталона).

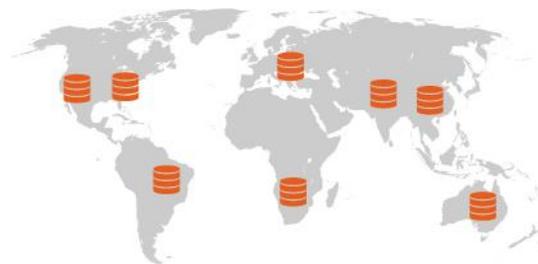


Рис. 1 – Географически распределенный кластер

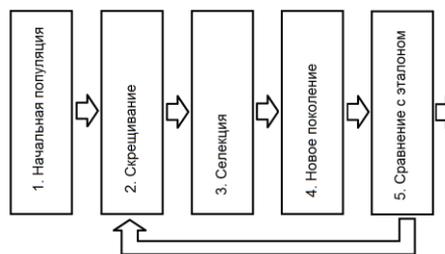


Рис. 2 – Генетический алгоритм

Для создания моделей был использован язык Python, а для генерации выборки запросов к данным использовался набор скриптов на языке JavaScript.

Основной сложностью при создании модели является поиск базы с релевантными запросами пользователей к данным с предоставлением множества характеристик, которые могли бы использоваться в качестве геномов. Важным фактором являлась возможность соотнесения запросов по регионам и дате запроса с точностью в 2-4 часа. В результате анализа общедоступных данных, предоставляемых различными сайтами, было решено использовать запросы пользователей к видео-контенту сайта YouTube.

В результате проведенного исследования было получено несколько результатов: а) стали известны основные метрики видео-контента сайта YouTube, на основании которых можно предположить циклическость обращения к данным, а также целесообразность их перемещения в другой географический регион с целью уменьшения нагрузки на сеть; б) анализ различных генотипов показал, что точность определения циклических данных не пропорциональна количеству геномов; в) генетический алгоритм целесообразно использовать лишь для данных, которые обладают множеством тэгов, на основании которых можно построить определяющие генотипы.

Таким образом, использование генетических алгоритмов позволяет сократить трафик, передаваемый по опорным сетям и уменьшить время отклика для конечного пользователя.

Список использованных источников:

1. Панченко, Т. В. Генетические алгоритмы / Т. В. Панченко // Уч. метод. пособие под ред. Ю. Ю. Тарасевича. – Астрахань, 2007. – 87 с.
2. Гладков, Л. А. Генетические алгоритмы / Л. А. Гладков, В. В. Курейчик // М.: ФИЗМАТЛИТ, 2006. – 320 с.