

«Алгоритмы поиска последовательностей во временных рядах по паттерну, задаваемому отрезком времени»

Автор: Томилов Н. А., Университет ИТМО, Санкт-Петербург

Научный руководитель: Ключев А. О., Университет ИТМО, Санкт-Петербург

Одним из способов анализа временных рядов является поиск последовательностей, удовлетворяющих определенному паттерну или шаблону. Частным случаем такого анализа является поиск временных интервалов, в которых значения попадают под некоторый паттерн, задаваемый определенным временным интервалом, то есть поиск похожих шаблонов данных по временной оси. Необходимость такого поиска возникает, например, при поиске аномалий в работе некоторых датчиков с помощью анализа измеренных этим датчиком данных. Однако в общем случае точность и эффективность поиска зависит от характеристик временного ряда. Поэтому возникает проблема выбора наиболее подходящего алгоритма для поиска по данным, относящимся к конкретной предметной области.

Целью работы стало исследование существующих алгоритмов поиска по временным рядам для выбора алгоритма, наиболее эффективного для поиска по данным датчиков газовой турбины. Данная предметная область имеет свою специфику – размер временного ряда может достигать нескольких лет, а сами данные могут быть неравномерны по времени. Кроме того, в определенный момент времени данные определенного датчика могут вообще отсутствовать, или иметь значения, свидетельствующие о неисправности самого датчика. В связи с этим, была выявлена задача, требующая решения - реализация алгоритма поиска по временному ряду, нечувствительного к неравномерности ряда и к случайным «выбросам» в данных.

В работе рассматриваются несколько реализаций поискового алгоритма, в частности, «наивный» поиск по совпадающим метрикам. Для задаваемого пользователем отрезка временного ряда вычисляется набор метрик, который затем сравнивается с метриками скользящего временного окна на всем диапазоне доступных для определенного датчика – «тэга» – данных. Данный алгоритм, несмотря на тривиальность идеи, показывает хорошую точность поиска, позволяя найти, например, все временные отрезки, соответствующие моментам запуска турбины, при выборе одного из них как шаблона для поиска. Также в работе были затронуты алгоритмы, основанные на DTW – Dynamic Time Warping, и LCS – Longest Common Subsequence. Использованы как готовые реализации этих алгоритмов, так и написанные самостоятельно. Каждый из реализованных алгоритмов поиска предполагает, помимо указания временного интервала, соответствующего шаблону для поиска, ряд специфичных для конкретного алгоритма параметров.

В процессе этого исследования, помимо реализации вышеуказанных алгоритмов, было разработано приложение-микросервис, позволяющий производить поиск данных по шаблону, задаваемому определенным временным отрезком. Этот микросервис, наряду с также разработанными микросервисом-источником данных и приложением для конечного пользователя, определил программный комплекс, предоставляющий пользователю возможность выбрать на графике значений определенного тэга участок времени и найти все временные интервалы, в которых значения этого тэга будут похожи на значения внутри выбранного участка времени.

В результате исследования было установлено, что для каждого из реализованных алгоритмов можно подобрать такие параметры, при которых будут найдены все необходимые временные отрезки. Однако, каждый из реализованных алгоритмов по-разному ведет себя во временных рядах, которые были искусственно прорежены, и по-прежнему нестабилен на данных, содержащих аномальные выбросы по значениям – для каждого алгоритма требуется подбор параметров, который позволяет проводить поиск без влияния этих аномалий на результат. В связи с этим, в дальнейшем планируется поиск путей повышения устойчивости алгоритмов к подобным аномалиям, а также планируется поиск путей оптимизации алгоритмов поиска на больших временных отрезках.

Автор _____ / _____ (Фамилия И.О.)
(подпись)

Научный руководитель _____ / _____ (Фамилия И.О.)
(подпись)

« ____ » _____ марта 2019 года