

УДК 004.932

**АЛГОРИТМЫ УМЕНЬШЕНИЯ КОЛИЧЕСТВА ТОЧЕК, НЕОБХОДИМЫХ ДЛЯ ВИЗУАЛЬНОГО ПРЕДСТАВЛЕНИЯ ТРЕНДОВ ИСТОРИЧЕСКИХ ДАННЫХ В СВУ АСУ ТП АЭС**

ГАРКАВЧЕНКО Д.А., ЗАХАРИН И.В., КУЛИКОВ А.М.

*Национальный исследовательский ядерный университет «МИФИ»  
(Москва, Российская Федерация)*

**Аннотация.** В работе рассматриваются различные методы понижающей дискретизации исторических данных автоматизированных систем управления для их визуального представления в виде линейных диаграмм. Предложенный алгоритм на основе картографической техники значительно сокращает количество точек данных, но сохраняет важные визуальные характеристики трендов.

**Ключевые слова:** понижающая дискретизация, аппроксимация, визуальный анализ, тренды, архив, SCADA, АСУ ТП, АЭС.

**ALGORITHMS FOR THE REDUCTION OF THE NUMBER OF POINTS REQUIRED TO REPRESENT HISTORICAL DATA TRENDS IN UPPER LEVEL SYSTEM OF I&C NPP**

DMITRIY.A. GARKAVCHENKO, ALEXANDER.M.KULIKOV, IVAN.V. ZAHARIN

*National Research Nuclear University MEPHI  
(Moscow, Russian Federation)*

**Abstract.** The article discusses several methods for downsampling historical data of automated control system, to enable visualization by line charts. The algorithm based on cartographical technic significantly reduces the number of points, but retains important visual characteristics of trends.

**Keywords:** downsampling, approximation, visual analysis, trends, archive, SCADA, I&C, NPP.

Представление значений технологических параметров во времени в виде линейных диаграмм, называемых трендами, является неотъемлемой функцией систем верхнего уровня АСУ ТП на базе любой SCADA-системы, так как способствует лучшему пониманию динамики параметров технологического процесса.

Обычно выделяют 2 вида трендов: тренды реального времени – для отслеживания динамики технологического процесса, и исторические – для оценки, анализа технологического процесса на большом интервале.

Как правило, построение линейной диаграммы на компьютере – простая задача. Однако, если визуализируемые данные содержат большое количество точек, значительно превышающее количество пикселей для изображения, то возникают проблемы: итоговый тренд получается сжатым (на одной вертикальной линии отображается множество значений), что затрудняет визуальный анализ. Также требуется большее время на перерисовку тренда при изменении его визуальных параметров (типа линии, масштаба, добавления маркеров). И если исходные данные хранятся централизованно, то требуется большее время для передачи и накладывается большая нагрузка на сеть передачи данных.

Данная проблема актуальна и для СВУ АСУ ТП АЭС на базе ПО ПОРТАЛ, разрабатываемого АО «РАСУ». В архиве СВУ для многих технологических параметров хранится огромное количество значений, так как зоны нечувствительности и апертуры архивирования настроены так, чтобы, в первую очередь, не пропустить изменение параметра при оперативном контроле и управлении. Но при визуальном анализе процесса на длительном интервале многие из этих изменений не нужны, а только усложняют процесс получения, отображения и анализа данных. Поэтому ранее при достижении установленного предела по количеству точек на тренде для уменьшения их количества применялось «тренды с фиксированием шагом», полученные в результате стробирования. При таком подходе просто выбираются значения параметра через равные промежутки времени. Это является уместным только для относительно «гладких» линий, которые не имеют отклонений. Но локальные экстремумы, а также значительные колебания значений параметров являются очень важными для анализа технологического процесса, поэтому необходим другой метод, уменьшающий количество точек, но сохраняющий визуальные характеристики тренда.

Задача состоит в том, чтобы показать только хранящиеся в архиве СВУ значения параметров с их реальными метками времени, исключив некоторое количество точек, наименее

важных для визуального анализа. То есть алгоритм должен возвращать подмножество точек, существующих в исходных данных. Поэтому в данной работе не рассматриваются общеизвестные методы регрессионного анализа, которые по сути создают новые точки данных на основе исходных.

Первый рассмотренный алгоритм основан на идее статистического представления данных. Исходные данные разбиваются на примерно равные по количеству точек наборы. Далее для каждого набора выбирается либо точка с самым часто повторяющимся значением ординаты в наборе (мода), либо точка с медианным значением ординаты, либо точка глобального экстремума, если он достигается на интервале. Основным недостатком такого подхода – практически всегда исключаются локальные пики и впадины в каждом наборе, так как учитывается не значение ординаты, а частота ее появления в наборе. При этом часто практически все значения в наборе уникальны, поэтому если в наборе попадает хотя бы две точки с одинаковой ординатой, то выбирается одна из них.

Во втором алгоритме для выбора точек используется понятие стандартной ошибки оценки. Обычно этот показатель используется, чтобы оценить насколько сильно точки исходных данных отклоняются от прямой регрессии [2]. В данном случае прямая регрессии не строится, а заменяется всевозможными отрезками между реальными точками. В начале фиксируются первая и последняя точки исходных данных, остальные точки разбиваются на примерно равные по количеству точек наборы. Для каждого набора вычисляются прямые, соединяющие каждую точку этого набора со всеми точками следующего набора, затем для каждой такой прямой вычисляется стандартная ошибка оценки относительно точек, мимо которых она проходит. После этого нужно выбрать из каждого набора одну точку так, чтобы сумма стандартных ошибок оценки для отрезков прямых между всеми выбранными точками была минимальной. Данную задачу, можно решить разными способами, например, представить ее в виде ориентированного ациклического графа, в котором вершины – это точки исходных данных, ребра – отрезки прямых, а вес ребер – значение стандартной ошибки оценки. Тогда остается решить задачу о кратчайшем пути, которая имеет множество стандартных решений, например, алгоритм Дейкстры [3]. После нахождения кратчайшего пути вершины этого пути возвращаются в качестве искомого подмножества точек исходных данных. Несмотря на то, что в результате работы алгоритма получается статистически наиболее близкая линия к исходной, ее нельзя назвать хорошим визуальным представлением данных, потому что многие локальные пики сглаживаются, так как обычно получается, что сумма стандартных ошибок оценки оказывается меньше, если исключить значительные выбросы. Также полученный алгоритм является довольно медленным, но есть и другие способы решения задачи динамической оптимизации, которые, вероятно, могут оказаться быстрее и предпочтительнее, так как для графического представления не обязательно нахождение абсолютной минимальной суммы, возможно, будет достаточно небольшой суммы.

Основной проблемой первых двух алгоритмов является сглаживание локальных пиков, поэтому третий алгоритм основан на идее максимизации длины итоговой линии тренда. Действительно, чем больше локальных пиков охватит итоговая линия, тем больше окажется ее длина. Для этого, как и во втором алгоритме, фиксируются первая и последние точки, остальные точки разбиваются на наборы, вычисляются все отрезки прямых, соединяющие точки одного набора с точками следующего, после этого строится ориентированный ациклический граф, но в данном случае весом ребер являются длины отрезков. Тогда остается решить задачу поиска самого длинного пути. Такая задача решается за линейное время на ориентированных ациклических графах, например, если заменить на противоположный знак весов всех ребер и найти кратчайший путь. Визуально результаты работы алгоритма получаются значительно лучше предыдущих, большинство локальных пиков и впадин попадают в итоговую выборку, но сам алгоритм получается слишком сложным и неэффективным, поэтому в таком виде не может быть применен в системах верхнего уровня АСУ ТП, так как может оказывать негативное влияние на выполнение других функций СВУ.

Во всех описанных выше алгоритмах используется разбиение исходных данных на наборы с примерно одинаковым количеством точек, что является тривиальным только при равномерном распределении точек по оси абсцисс. Но в архиве СВУ АСУ ТП АЭС данные могут располагаться крайне неравномерно, если рассматривать длительный интервал времени, так как запись значений в архив, осуществляется по изменению с учетом апертуры, а не

циклически, и в разных режимах работы энергоблока параметры технологического процесса могут меняться с разной скоростью и в разных диапазонах. Поэтому вместо попыток упрощения и оптимизации описанных выше алгоритмов был применен другой подход.

Были рассмотрены алгоритмы из области картографической генерализации, где выполняется научно обоснованный отбор и обобщение географических объектов для их отображения на карте. Задача упрощения отображения, например, береговой линии на карте во многом похожа, на задачу понижающей дискретизации исторического тренда, только точки тренда расположены строго по возрастанию абсциссы в отличие от точек на карте. Одним из наиболее известных алгоритмов из этой области являются алгоритм Рамера-Дугласа-Пекера [4].

Оригинальный алгоритм Рамера-Дугласа-Пекера получает на вход множество исходных точек и заданное значение  $\epsilon$ , которое определяет максимальное расстояние между исходной и итоговой линиями. Первая и последняя точки фиксируются, после этого среди оставшихся точек находится наиболее удаленная точка от отрезка, соединяющего первую и последнюю точки. Если расстояние от найденной точки до отрезка, меньше заданного  $\epsilon$ , то алгоритм завершается, и все невыбранные ранее точки отбрасываются. Если это расстояние больше, то алгоритм рекурсивно вызывает себя для подмножеств точек от начальной до текущей и от текущей до конечной точки до тех пор, пока все точки не будут выбраны или отброшены. По окончании всех рекурсивных вызовов итоговая линия строится только из тех точек, что были выбраны. В результате итоговая линия охватывает большинство локальных экстремумов, при этом реализация алгоритма достаточно простая и эффективная, и она не требует предварительного разбиения исходных данных на наборы с примерно одинаковым количеством точек.

Алгоритм был доработан для практического применения для трендов технологических параметров. Во-первых, для упрощения вычислений вместо расстояния от точки до прямой используется отклонение по вертикали. Во-вторых, нельзя использовать абсолютное значение  $\epsilon$ , так как значения разных технологических параметров изменяются в разных диапазонах, поэтому используется относительное значение  $\epsilon$  от диапазона изменений. В-третьих, в различных режимах работы энергоблока один и тот же технологический параметр может изменяться в разных диапазонах, и для лучшего наблюдения таких параметров на трендах используется логарифмическая шкала. Поэтому, чтобы реализовать возможность понижающей дискретизации в логарифмическом масштабе, исходные данные предварительно разбиваются на наборы, так чтобы в пределах одного набора были точки, отличающиеся по ординате незначительно (не более чем в 100 раз – определено эмпирически) и алгоритм применяется к каждому из наборов отдельно. Также при предварительном разбиении исходных данных учитываются признак качества (достоверности) данных. В-четвертых, в отличие от предыдущих алгоритмов оригинальный алгоритм Рамера-Дугласа-Пекера возвращает заранее неизвестное количество точек, поэтому если итоговое количество точек оказывается больше установленного предела, то алгоритм выполняется заново с увеличенным значением максимального расстояния  $\epsilon$ , количество таких итераций ограничено. Благодаря простоте и эффективности алгоритма, его повторение остается незаметным для конечного пользователя системы. Понижающая дискретизация применяется только при достижении установленного предела по количеству точек на тренде и подтверждается пользователем.

Таким образом, на основе алгоритма Рамера-Дугласа-Пекера разработан и реализован в ПО ПОРТАЛ алгоритм понижающей дискретизации, сохраняющий общий вид тренда с его локальными экстремумами при значительном сокращении числа точек. Данный алгоритм проверен для различных наборов данных и уже применяется в системах верхнего уровня АСУ ТП АЭС.

#### Список литературы

1. Зверков В.В. Автоматизированная система управления технологическими процессами АЭС / В.В. Зверков. – М.: Национальный исследовательский ядерный университет «МИФИ», 2013. – 558 с.
2. Драйпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Статистика, 1973. – 210 с.
3. Алексеев В.Е. Графы и алгоритмы. Структуры данных. Модели вычислений: учебник / В.Е. Алексеев, В.А. Таланов. – М.: БИНОМ, 2012. – 319 с.
4. Douglas D., Peucker T. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature». The Canadian Cartographer, 1973, 10(2), 112–122