

РАЗРАБОТКА МЕТОДА СЖАТИЯ ДАННЫХ МОНИТОРИНГА С СОХРАНЕНИЕМ РЕПРЕЗЕНТАТИВНОСТИ

А. Ф. Михайлов

ФГАОУ ВО «Снежинский физико-технический институт Национального исследовательского
ядерного университета МИФИ», Снежинск, Россия

E-mail: chelaxe@gmail.com

Эффективный мониторинг сложных систем требует обработки значительных объемов данных, что представляет проблему для быстрого анализа и обеспечения отказоустойчивости. В этой работе рассматривается сжатие данных как стратегия для уменьшения генерируемых данных. Предлагается улучшенный алгоритм *Swinging Door* с автономной настройкой погрешности сжатия. Экспериментальные результаты демонстрируют эффективность подхода, подтверждая его практическую применимость.

В эпохе глобальной цифровизации, мониторинг систем с использованием информационных технологий становится критически важным, особенно в контексте обработки и анализа массивных данных, генерируемых множественными устройствами. С увеличением данных возникает проблема их хранения и передачи, что актуализирует потребность в эффективном сжатии данных без потери их репрезентативности.

В ответ на эти вызовы, предлагается разработка алгоритма сжатия данных, реализуемого на аппаратном уровне для ускорения обработки данных и снижения нагрузки на вычислительные системы. Основная особенность алгоритма – его поточный характер, позволяющий обрабатывать данные в реальном времени без необходимости временного хранения.

Цель статьи – представить концепцию, разработку и применение этого поточного алгоритма сжатия в мониторинговых системах, с упором на сохранение точности данных и возможности аппаратной интеграции.

В качестве основы был выбран алгоритм *Swinging Door* («вращающаяся дверь») [2] из-за его эффективности в прореживании данных. Основываясь на механизме «окна» [1] для отслеживания значимых изменений, алгоритм устанавливает диапазон допустимых отклонений («дверные полотна») вокруг последнего зарегистрированного значения. Запись новых данных происходит только при выходе за эти пределы, обеспечивая экономию объема данных без значительной потери информации. Это особенно актуально для данных с незначительными или повторяющимися изменениями.

В алгоритме *Swinging Door* были реализованы ключевые модификации для повышения его эффективности. Во-первых, вместо создания новой точки при закрытии «коридора» теперь сохраняется последнее фактическое значение, гарантируя реальную временную метку и значение прореженных данных. Введен параметр продолжительности «коридора», позволяющий автоматически закрывать его после определенного количества пропущенных точек или временного интервала, что улучшает обработку данных с незначительными изменениями.

Погрешность сжатия теперь разделяется на верхнюю и нижнюю границы, обеспечивая более тонкую настройку вокруг среднего значения. Добавлена функция автономной настройки погрешности сжатия, использующая экспоненциальное скользящее среднее и абсолютное значение наклона между последовательными точками [3].

Эти изменения обеспечили возможность реализации алгоритма на аппаратном уровне с использованием микроконтроллера ATmega328, значительно расширяя его применение в различных системах мониторинга.

Улучшения алгоритма *Swinging Door* для сжатия данных мониторинга, повысили его эффективность и адаптивность. Значительный вклад внесла функция автономной настройки погрешности.

Аппаратная реализация алгоритма расширяет его применение для эффективного сжатия данных в реальном времени в разнообразных сложных системах. Эти усовершенствования способствуют улучшению функциональности систем мониторинга, сохраняя при этом качество и точность обрабатываемых данных, что является ключевым в современном информационном контексте.

В рамках исследования, направленного на оценку эффективности модифицированного алгоритма сжатия данных, использовался открытый набор данных. Основной задачей тестирования было определение степени сжатия данных при одновременном сохранении их репрезентативности, критически важной для аналитической точности и целостности принятия решений.

Выбранный для тестирования набор данных отличался значительным объемом и разнообразием, что позволило всесторонне оценить производительность алгоритма в различных сценариях. После применения модифицированного алгоритма к этим данным, был проведен анализ полученных результатов.

Согласно результатам, представленным на рис. 1, алгоритм продемонстрировал высокую эффективность, сократив исходный объем данных на 87%. Это подчеркивает его способность к значительному уменьшению объема данных, что особенно важно в контексте ограниченных ресурсов хранения и обработки больших данных. Ключевым моментом является сохранение репрезентативности данных алгоритмом, что подтверждает его пригодность для практического использования в сферах мониторинга и анализа данных.

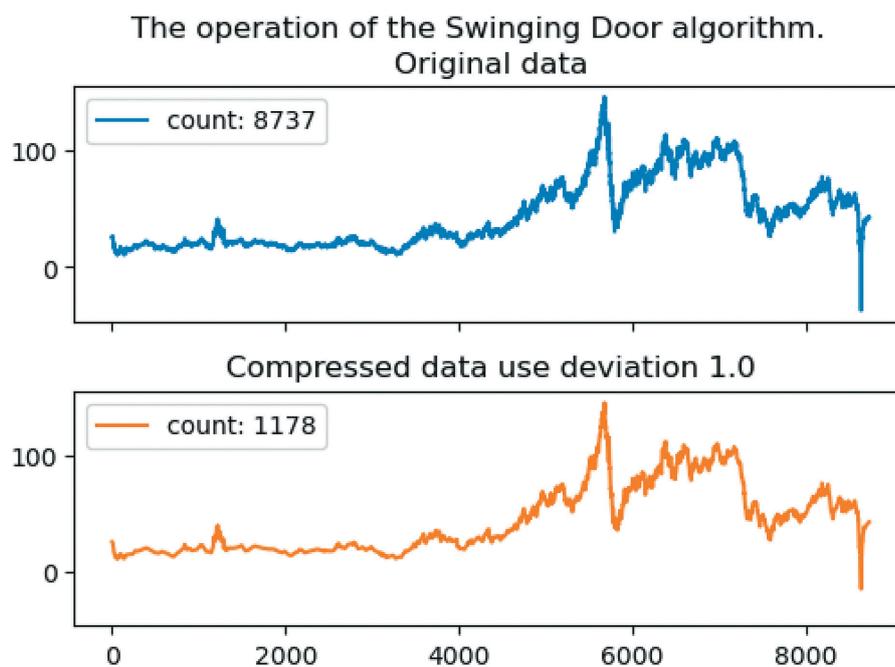


Рис. 1. Демонстрация работы алгоритма на открытом наборе данных

Таким образом, результаты тестирования на открытом наборе данных эффективно демонстрируют высокую производительность модифицированного алгоритма сжатия данных, подчеркивая его значимость в области обработки и анализа больших объемов данных.

Литература

1. М. В. Демешко, И. А. Канаева, И. А. Ботыгин. Визуализация метеорологических данных; [Электронный ресурс] – Режим доступа: URL: https://earchive.tpu.ru/bitstream/11683/46433/1/conference_tpu-2017-C04_p249-250.pdf (дата обращения: 25.12.2023).
 2. Edgar H. Bristol. Data compression for display and storage // Патент США № US4669097A; [Электронный ресурс] – Режим доступа: URL: <https://patentimages.storage.googleapis.com/f6/ca/83/c34a2602ede2ac/US4669097.pdf> (дата обращения: 25.12.2023).
 3. Juan David Arias Correa, Alex Sandro Roschildt Pinto, Carlos Montez, Erico Meneses Leão. Swinging Door Trending Compression Algorithm for IoT Environments; [Электронный ресурс] – Режим доступа: URL: <https://pdfs.semanticscholar.org/fe00/bc1bba2147c338898a8fa6c31a84afce467.pdf> (дата обращения: 25.12.2023).
-