



# Разработка метода сжатия данных мониторинга с сохранением репрезентативности

МИХАЙЛОВ А.Ф.

ФГАОУ ВО «СНЕЖИНСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ НАЦИОНАЛЬНОГО ИССЛЕДОВАТЕЛЬСКОГО ЯДЕРНОГО УНИВЕРСИТЕТА  
МИФИ», ЧЕЛЯБИНСКАЯ ОБЛ.

CHELAXE@GMAIL.COM

# Различные методы сжатия данных мониторинга

## ▶ Прореживание данных с постоянным шагом.

**Принцип:** выборка каждого  $n$  значения, где  $n$  шаг прореживания.

**Преимущество:** простота реализации.

**Недостаток:** возможная потеря информационных выбросов.

## ▶ Аппроксимация данных

**Принцип:** поиск функции позволяющей близко заменить последовательность данных.

**Преимущество:** значительное прореживание данных.

**Недостаток:** сложность реализации.

## ▶ Прореживание данных с помощью окна.

**Принцип:** сокращение мощности множества исходных данных за счет замены группы из  $k$  множеств одним значением.

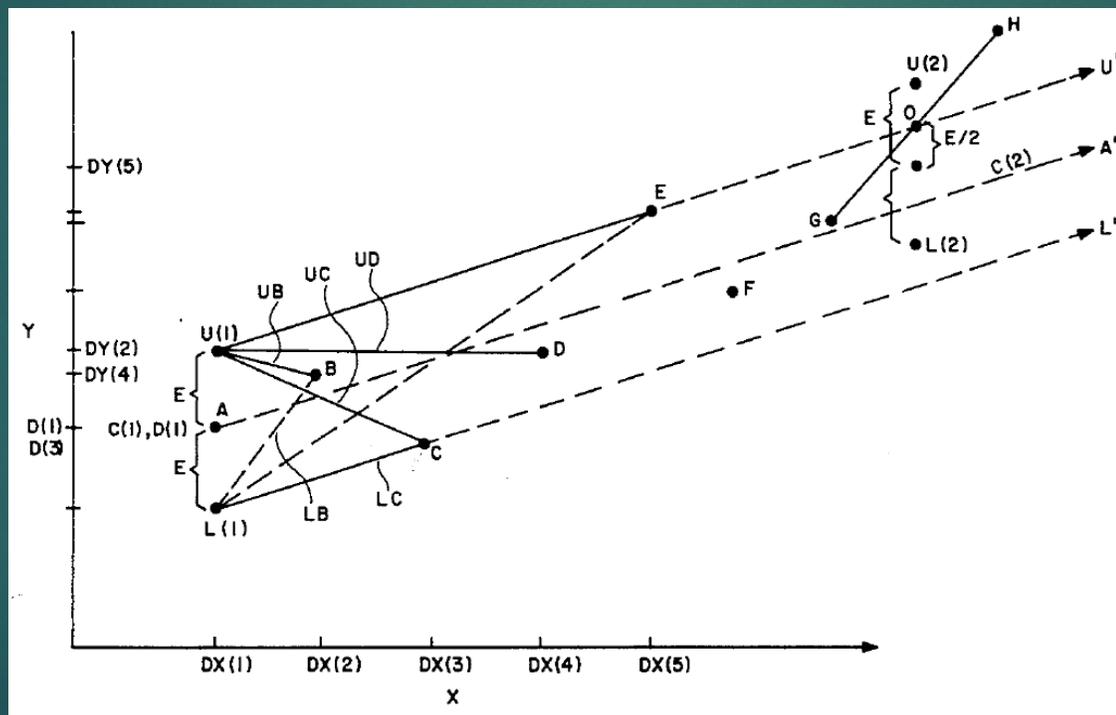
**Преимущество:** сохраняет информативные пики и провалы.

**Недостаток:** зависимость результата от способа выбора значения.

# Алгоритм Swinging Door

Алгоритм сжатия данных мониторинга методом прореживания данных с помощью окна - Swinging Door («вращающаяся дверь»)

Алгоритм был запатентован в США в 1987 году. Принцип работы данного алгоритма отражён в его названии.



# Модификации алгоритма Swinging Door

- ▶ Вместо создания новой точки при закрытии "коридора" теперь сохраняется последнее фактическое значение, гарантируя реальную временную метку и значение.
- ▶ Введен параметр продолжительности "коридора", позволяющий автоматически закрывать его после определенного количества пропущенных точек или временного интервала.
- ▶ Погрешность сжатия теперь разделяется на верхнюю и нижнюю границы.
- ▶ Добавлена функция автономной настройки погрешности сжатия.

# Автономной настройки погрешности сжатия в алгоритме Swinging Door

Экспоненциальное скользящее среднее (EMA - exponential moving average)

$$f_{ema} = a \times s^i(d, p) + (1 - a) \times f_{ema-1},$$

где  $a[0, 1] = 0.3$  сглаживающая константа (smoothing constant) характеризующая скорость уменьшения весов,

$s^i(d, p) = \left| \frac{d_y - p_y}{d_x - p_x} \right|$  абсолютное значение наклона между последовательными точками.

# Показатели эффективности алгоритма сжатия данных

- ▶ ошибка сжатия ( $CE$ ) измеряет относительное количество ошибок, наблюдаемых после сжатия

$$CE = \frac{\sum_{i=1}^n |T_i - T'_i|}{\sum_{i=1}^n |T_i|},$$

где  $T_i$  несжатые данные,  $T'_i$  результаты сжатия данных после процесса декомпрессии

- ▶ степень сжатия ( $CR$ ) направлен на оценку эффективности процесса сжатия и представляет собой уменьшение количества образцов, достигнутое с помощью алгоритма сжатия

$$CR = 1 - \frac{CompressedSamples}{UncompressedSamples},$$

где  $CompressedSamples$  размер данных после сжатия,  $UncompressedSamples$  размер данных до сжатия.

# Показатели эффективности алгоритма сжатия данных

- ▶ критерием сжатия ( $CC$ ) метрика отражает эффективность сжатия

$$CC(CE, CR) = \frac{2 \times CR \times (1 - CE)}{CR + (1 - CE)}$$

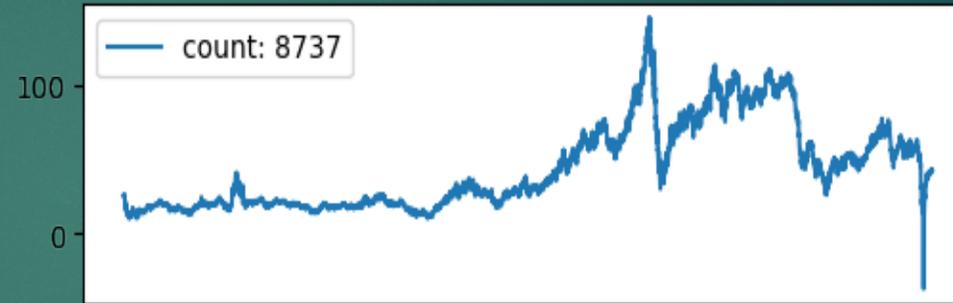
$CC[0, 1]$  хорошее сжатие имеет значение, близкое к 1, что соответствует высоким значениям  $CR$  и низким  $CE$

# Оценка эффективности алгоритма сжатия

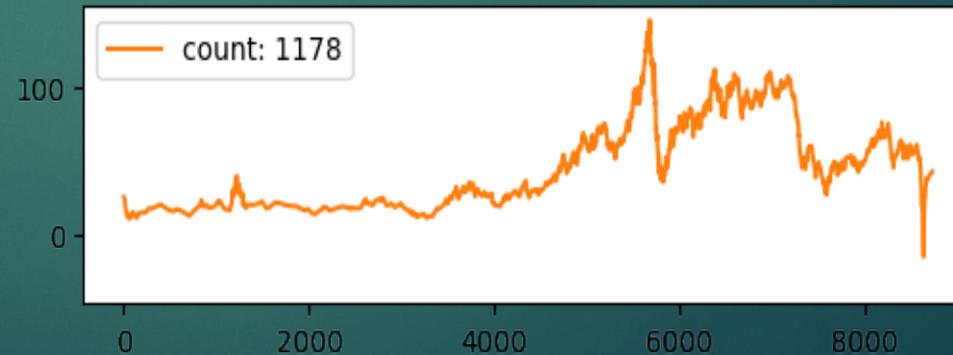
- ▶ Для оценки был выбран открытый набор данных о суточной стоимости нефти марки WTI

Алгоритм продемонстрировал высокую эффективность, сократив исходный объем данных на 87% с сохранением репрезентативности.

The operation of the Swinging Door algorithm.  
Original data

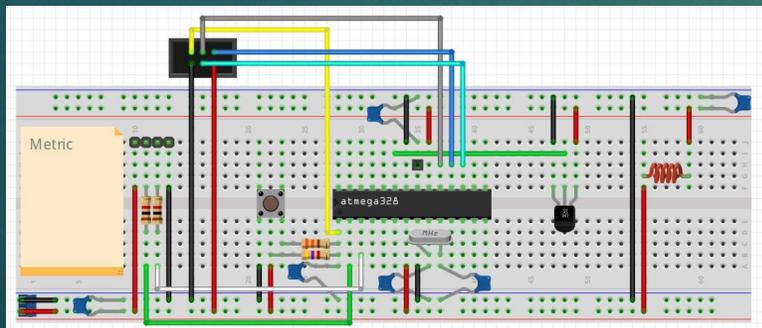
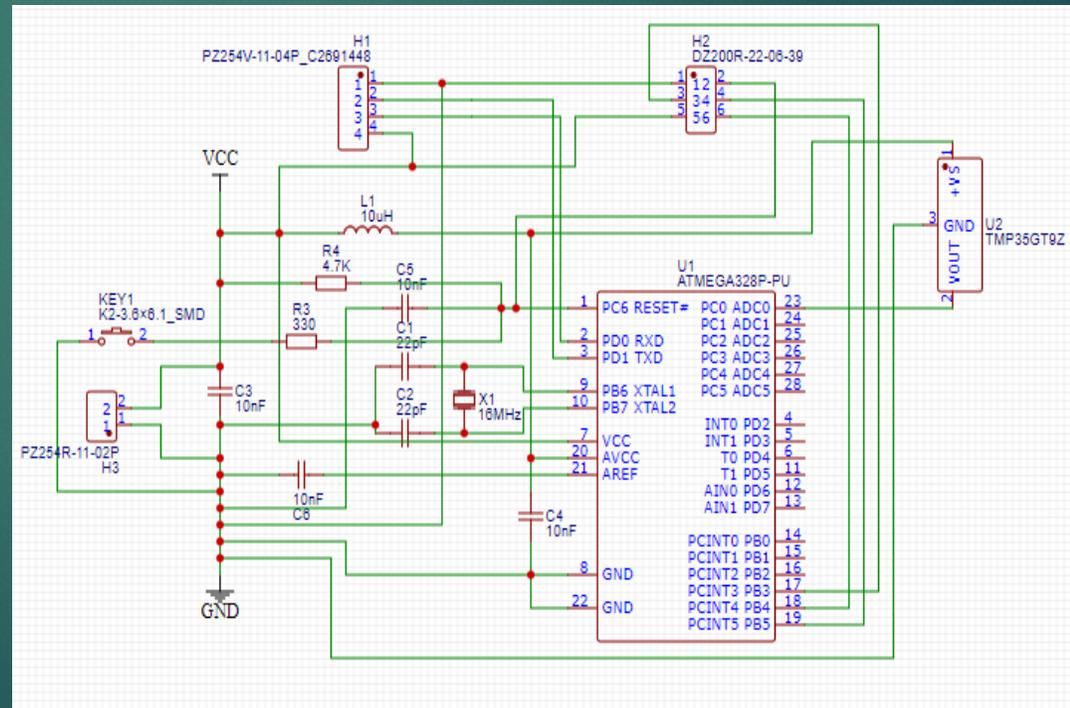
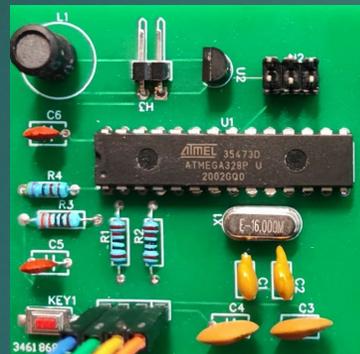
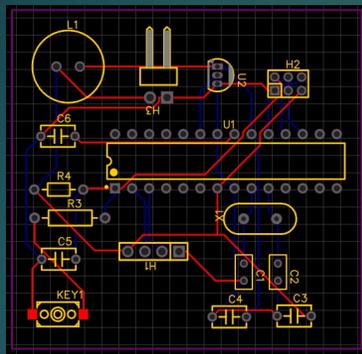


Compressed data use deviation 1.0



# Реализация алгоритма на базе микроконтроллера

Решение было успешно построено на базе 8-битного AVR микроконтроллера ATmega328



# Применение аппаратной реализации в сенсорной сети умного города

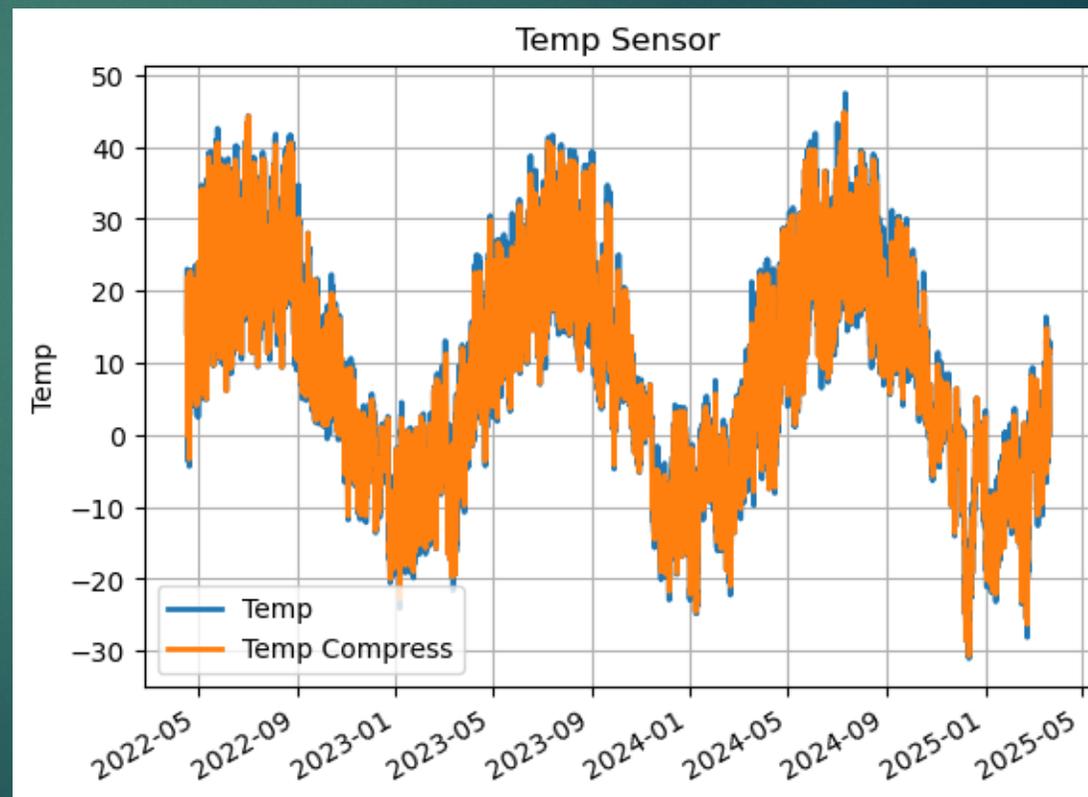
10

Данные мониторинга температуры окружающей среды

погрешность сжатия - 0.06

коэффициент сжатия - 0.84

качество сжатия - 0.89



# Результаты работы

Аппаратная реализация сенсора для умного города

Интеграция алгоритма в системы мониторинга высокопроизводительных вычислительных комплексов

Была реализована библиотека на языке программирования Python и размещена в репозитории PyPI.

- ▶ <https://pypi.org/project/swinging-door/>
- ▶ <https://github.com/chelaxe/SwingingDoor>
- ▶ <https://pypi.org/project/sausage-links/>
- ▶ <https://github.com/chelaxe/SausageLinks>



Благодарю за внимание

[chelaxe@gmail.com](mailto:chelaxe@gmail.com)



@chelaxe