Aggregation Algorithm

Coursework

Submitted in partial fulfillment of requirement
for the module of

Internet of Things (EEEM048)

Submitted by

MANISH PANDA(6614056)

Submitted to

PROF. Mohammad Shojafar

University of Surrey

School of Computer Science and Electronic Engineering

Department of Electrical and Electronic Engineering

United Kingdom

November 2023

Abstract

With the purpose of enhancing sensor data processing on IoT nodes, this project presents an intelligent aggregation technique. By using a 12-sized FIFO buffer for continuous storage, the system dynamically modifies data aggregation based on activity levels. The measurement frequency parameter (k) can be customized by users through the use of standard deviation in activity measurement. There are three different approaches to aggregation: moderate activity combines every four readings, strong activity necessitates no aggregation, and low activity initiates an average of all 12 readings. In dynamic IoT situations, the system strikes a compromise between responsiveness and resource conservation. Its adaptability is highlighted by user-defined threshold settings, and results are shown on the terminal.

Contents

Chapter 1

Aggregation & Reporting

1.1 Basic Features

1.1.1 High Activity

A environment that is dynamic and changing quickly is indicated by a large standard deviation. The algorithm gives up on data aggregation in order to preserve a high temporal resolution in these situations. For fast fluctuating events to be accurately captured and represented, this method is important.

```
reading = 821
reading = 512
reading = 205
reading = 102
reading = 359
reading = 411
reading = 281
reading = 256
reading = 256
reading = 359
reading = 411
reading = 359
B = [821, 512, 205, 102, 359, 411, 281, 256, 256, 359, 411, 359]
StdDev = 173.03
Aggregation = no aggregation
X = [821, 512, 205, 102, 359, 411, 281, 256, 256, 359, 411, 359]
```

Figure 1.1: No Aggregation.

1.1.2 Some Activity

Significant environmental variability, potentially resulting from irregular movements or changes in lighting, is shown by a moderate standard deviation. In response, the algorithm performs a 4-to-1 data consolidation, which successfully strikes a compromise between the need for compression and the preservation of data complexity.

```
reading = 50
reading = 50
reading = 77
reading = 77
reading = 50
reading = 25
reading = 25
reading = 77
reading = 77
reading = 102
reading = 25
reading = 77
B = [50, 50, 77, 77, 50, 25, 25, 77, 77, 102, 25, 77]
StdDev = 24.45
Aggregation = 4-into-1
X = [63.50, 44.25, 70.25]
```

Figure 1.2: 4-into-1 Aggregation.

1.1.3 Low Activity

An environment with little fluctuations in lighting is indicated by a low standard deviation. The algorithm thus creates a single average value from the whole data set. Based on the idea that the loss of detail caused by this averaging process is negligible and acceptable when not much is happening, this method was developed.

```
reading = 13
reading = 6
reading = 13
reading = 6
reading = 0
reading = 0
reading = 13
reading = 13
reading = 6
reading = 20
reading = 20
reading = 13
StdDev = 6.40
Aggregation = 12-into-1
X = [10.25]
B = [13, 6, 13, 6, 0, 0, 13, 13, 6, 20, 20, 13]
```

Figure 1.3: 12-into-1 Aggregation.

1.2 Advanced Features

1.2.1 Autocorrelation & Discrete cosine transform

The algorithm's improved features greatly increase its capacity to examine temporal trends in the data. It measures the degree to which the signal resembles a time-delayed version of itself using an autocorrelation function in order to identify any cyclical or repeated patterns. Additionally, these autocorrelation values are transformed into the frequency domain using the discrete cosine transform (DCT). This conversion plays a crucial role in highlighting the main frequencies, which makes it possible to portray the essential properties of the signal in a more effective and concise manner. Mathematically, Autocorrelation is computed using the formula :

$$\hat{R}(k) = \frac{1}{(n-k)\sigma^2} \sum_{t=1}^{n-k} (X_t - \mu_x)(X_{t+k} - \mu_x), \text{ for all } k \in \{0, 1, 2, \ldots, n-1\}.$$

where $\hat{R}(k)$ is the estimated autocorrelation at lag k, n is the number of observations, σ^2 is the variance, X_t is the value of the time series at time t, and μ_x is the mean of the time series. The summation is carried out from $t = 1$ to $n - k$, and k ranges from 0 to $n - 1$.

Mathematically, Discrete cosine transform of the autocorrelation vectoe is computed using the formula :

$$S(l) = \frac{1}{n} \sum_{k=0}^{n-1} \hat{R}(k) \cos\left(\frac{\pi}{n}\left(k + \frac{1}{2}\right)\left(l + \frac{1}{2}\right)\right), \text{ for all } l \in \{0, 1, 2, \ldots, n-1\}.$$

```
reading = 915
reading = 915
reading = 1181
reading = 1361
reading = 1181
reading = 1052
reading = 1258
reading = 1361
reading = 1334
reading = 1258
reading = 1309
reading = 1233
Autocorrelation Function:
R(0) = 0.99
R(1) = 0.49
R(2) = 0.23
R(3) = 0.11
R(4) = 0.39
R(5) = 0.17
R(6) = 0.45
R(7) = 0.59
R(8) = 0.56
R(9) = 0.72
R(10) = 0.91
R(11) = 0.44

Discrete Cosine Transform of Autocorrelation vector:
S(0) = 1.00
S(1) = 0.47
S(2) = 0.28
S(3) = 0.16
S(4) = 0.37
S(5) = 0.14
S(6) = 0.51
S(7) = 0.66
S(8) = 0.63
S(9) = 0.79
S(10) = -1.00
S(11) = 0.51
B = [915, 915, 1181, 1361, 1181, 1052, 1258, 1361, 1334, 1258, 1309, 1233]
StdDev = 151.22
Aggregation = no aggregation
X = [915, 915, 1181, 1361, 1181, 1052, 1258, 1361, 1334, 1258, 1309, 1233]
```

Figure 1.4: Autocorrelation & DCT.

Chapter 2

Conclusion

Each dataset's variability is measured using the statistical metric known as Standard Deviation (StdDev). large StdDev values of 151.22 and 173.03 were reported in the first two datasets, indicating a large degree of dispersion in the sensor readings. In contrast, the subsequent two datasets showed smaller StdDev values (24.45 and 6.40), which suggests a more constrained range of values. The aggregation phase of the system showed sophisticated decision-making based on the attributes of each dataset, which was guided by measured activity levels. The program avoided aggregating when there was high activity, as indicated by significant StdDev, and instead preserved individual data to record minute fluctuations. In contrast, the algorithm used aggregation (either a 4-into-1 or 12-into-1 style) to carefully select the most important information while maximizing resource use in datasets with lower activity levels.

With each dataset, the Autocorrelation Function revealed distinct patterns and significant correlations at various lag intervals. A spectrum analysis was made possible by applying the Discrete Cosine Transform to the Autocorrelation vector. This highlighted the frequency components found in the datasets. The negative coefficient at the highest frequency, which highlighted the data's fundamental structure and periodicity, was especially notable.