Machine Learning for Internet of Things 2024-2025 Homework 1, Group 4

Tanguy Marie Yvan Dugas du Villard, Muhammad Nouman Siddiqui, Shadi Mahboubpardahi
Student id: s321277, s329112, s329057
s321277@studenti.polito.it, s329112@studenti.polito.it, s329057@studenti.polito.it
Politecnico di Torino

I. Exercise 1

To provide the monitoring service for 1000 clients, we need to compute the maximum number of records that will be stored for each client.

To store the temperature and humidity every two seconds for 30 days, at most 2,592,000 records are necessary. Then, six aggregations are to be stored every hour, for 365 days: minimum, maximum, and average temperature, minimum, maximum, and average humidity, which represent 52,560 records per client. Overall, each client requires at most 2,644,560 records to be stored, which represents 42,312,960 bytes, as 16 bytes are required for each record.

For 1000 clients, 42,312,960,000 bytes (or 40.353 GB) are required.

As Redis can compress data with an average compression rate of 90%, the storage requirement for this application is about **4.035 GB**.

II. Exercise 2

A. *Range of parameters*

Optimizing Voice Activity Detection (VAD) hyperparameters requires choosing values within specific ranges to meet maximum accuracy and minimum latency in the search space. Important factors comprised of frame duration(frame_length_in_s) ranging from 10 to 60 ms to record short and steady audio characteristics, and frame steps(frame_steps_in_s) of 10 to 30 ms to preserve temporal precision. The sensitivity of the dB threshold (dBthres) from 5 to 15 dB was adjusted to filter out noise, while the duration threshold(duration_thres) from 0.1 to 0.4 seconds ensured that only significant non-silence segments were identified.

B. *Parameters selected*

Parameter	Value
Frame length	10 ms
Frame step	20 ms
dB threshold	15
Duration threshold	10 ms

TABLE I
Selected Values

Table I displays the chosen values that achieved required accuracy of **97.78%** and the minimal latency of **22.1 ± 0.03 ms**

C. *Parameters influence on accuracy and latency*

The value of the parameters dBthres and duration_thres do not affect the latency, as the amount of calculation needed to process the time signal does not depend on the value of these parameters. However, increasing the duration threshold will decrease the accuracy, while dBThres has a positive correlation with the accuracy.

The frame step and frame length show an inverse relationship with accuracy. As observed in Fig. 1, where the accuracy and latency of the VAD are displayed with a dB threshold of 15 and a duration threshold of 10 milliseconds, the highest accuracy is achieved with the smallest frame step and frame length. However, the latency is influenced by both frame steps and frame length, as shown on Fig. 1. The lowest latency is achieved by selecting the highest frame steps and reducing the frame length.

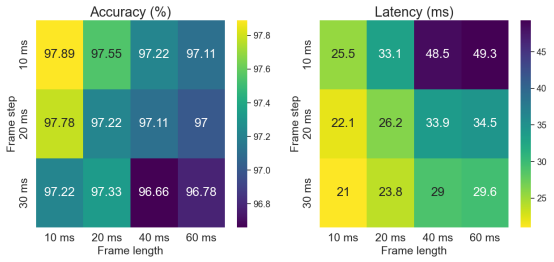

Fig. 1. The impact of FS and FL on accuracy and latency

D. *Discussion of the chosen parameters*

According to Tab. I, higher accuracy and minimal latency are obtained with the following parameters:

- **Frame Length (frame_length_in_s): 10 miliseconds** has been chosen, balancing between high accuracy and minimal latency. Shorter frames capture quick changes in speech, improving accuracy but also requiring fast processing to maintain low latency.
- **Frame Step (frame_steps_in_s)** : **20 miliseconds** has been selected, offering a good compromise. It provides sufficient temporal resolution to detect speech activity while ensuring the system remains efficient and fast enough to meet the latency target.
- **dB Threshold (dBthres) :** The choice of **15 dB** as the threshold helps ensure that the system remains accurate

while not overfitting to low-level noise, which also helps keep the system's processing requirements manageable and latency low.

- **Duration Threshold (duration_thres) : 10 miliseconds** was chosen, striking a balance by preventing the VAD system from falsely classifying short, irrelevant sounds as speech, while ensuring that speech segments of reasonable length are detected. This also helps with reducing unnecessary computations and keeps latency low.