Identification of regions of interest in accelerometer data

A.Y.Vasilyev

June 13, 2018

1 Acceleration

Accelerometer data represents force that acts on seismic mass,[1] which consists of gravitational force and inertial force:

$$\mathbf{F(t)} = m(\mathbf{g} + \mathbf{a(t)}) \tag{1}$$

where $\mathbf{g}$ is gravitational acceleration and $a(t)$ is acceleration of mobile device relative to the inertial frame of reference. We assume that gravitational acceleration is constant, whereas acceleration of device may change in time. Acceleration measured by accelerometer is:

$$\mathbf{a_m(t)} = \mathbf{F(t)}/m = (\mathbf{g} + \mathbf{a(t)}) \tag{2}$$

And its derivative:

$$\mathbf{da_m}(t)/dt = d(\mathbf{g} + \mathbf{a(t)})/dt = d\mathbf{a(t)}/dt \tag{3}$$

Because shaking is periodic movement of the object, acceleration will change in time with frequency of shaking. During certain stages of shaking the derivative of force will exceed threshold $\epsilon > 0$:

$$\exists t: \ |\mathbf{da_m}(t)/dt| > \epsilon \tag{4}$$

Figure 1 shows the absolute derivative of acceleration in time. We assume that unequality 4 will be satisfied on certain set

$$\forall t \in T_\epsilon : |\mathbf{da_m}(t)/dt| > \epsilon \tag{5}$$

which may differ with the value of threshold. Figure 2 demonstrate T_ϵ for $\epsilon = 10$ as a logical variable. Due to discretization T_ϵ is a finite set: $T_\epsilon = (t_1, t_2, .., t_n)$.

2 Regions of interest

On the next step we define partition of T_ϵ: $S_K = (T_{\epsilon,1}, T_{\epsilon,2}, ..., T_{\epsilon,K})$. We aim to find such a partition S_K that minimizes within-cluster variance:

$$S_K = \underset{S_K}{\arg\min} \sum_{k=1}^{K} \sum_{t \in T_{\epsilon,k}} ||t - \mu_k|| \tag{6}$$

where $\mu_k = \sum_{t \in T_{\epsilon,k}} t/(\text{Length}(T_{\epsilon,k}))$ is mean time of events in cluster $T_{\epsilon,k}$, and $\text{Length}(T_{\epsilon,k})$ is number of elements in $T_{\epsilon,k}$. We approach problem 6 with standard algorithm of k-means clustering. We set the number of clusters as a constant $K = 5$ (as it's defined within the task). The figure 3 shows the indexes of events after procedure of k-means clustering.

We define the regions of interest as a sequence of non-overlapping intervals with minimal length $R = [r_1, r_1^\star] \cup [r_2, r_2^\star] \cup ..[r_k, r_k^\star] = R_1 \cup R_2 \cup ...R_k$ each of them include corresponding cluster $T_{\epsilon,k}$: $T_{\epsilon,k} \subset R_k$. The figure 4 shows regions of interests for $\epsilon = 10$.

3 Dependence on threshold

In this section we study the stability of regions of interest. Because k-means clustering involves random initialization of starting centroid, the choice of partition may change both with trial and the threshold ϵ that define shaking events. For each value of threshold we launched 100 runs of k-means clustering and calculated probability of certain time interval to be within region of interest.

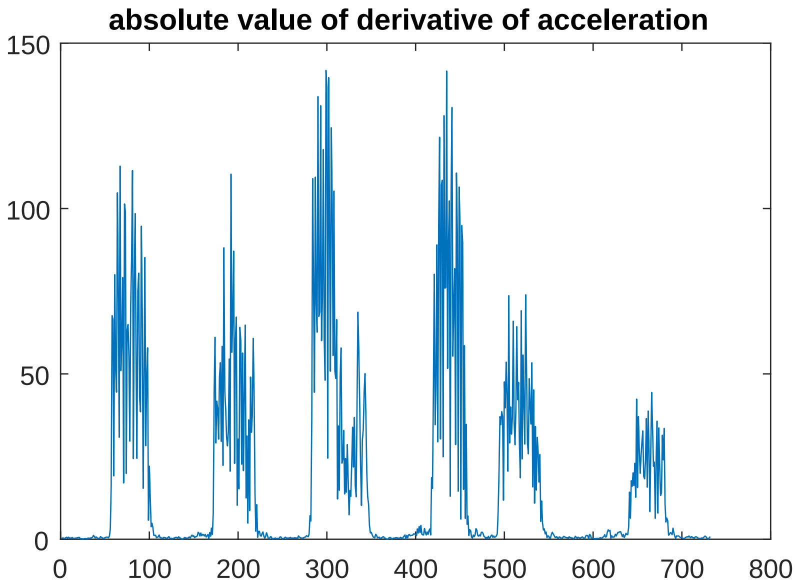

Figure 1: Absolute value of derivative of acceleration

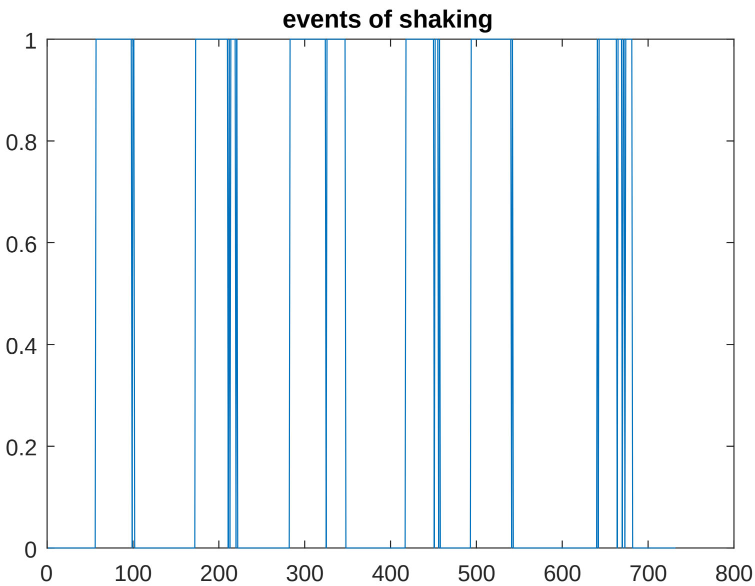

Figure 2: Shaking events

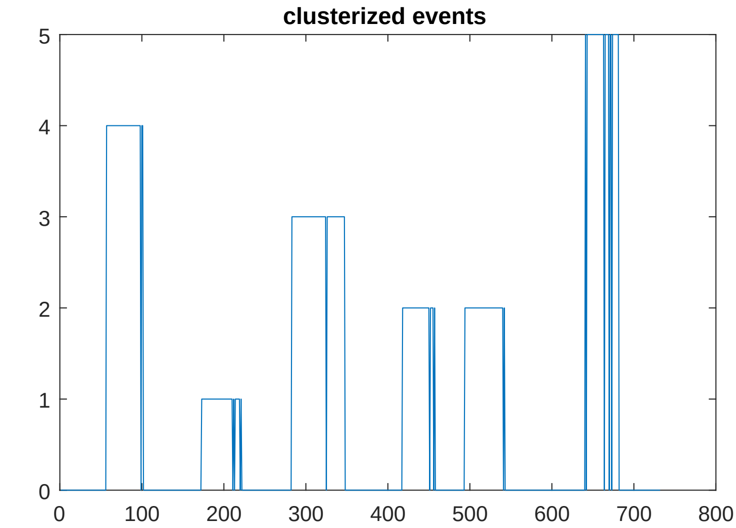

Figure 3: Clustered events

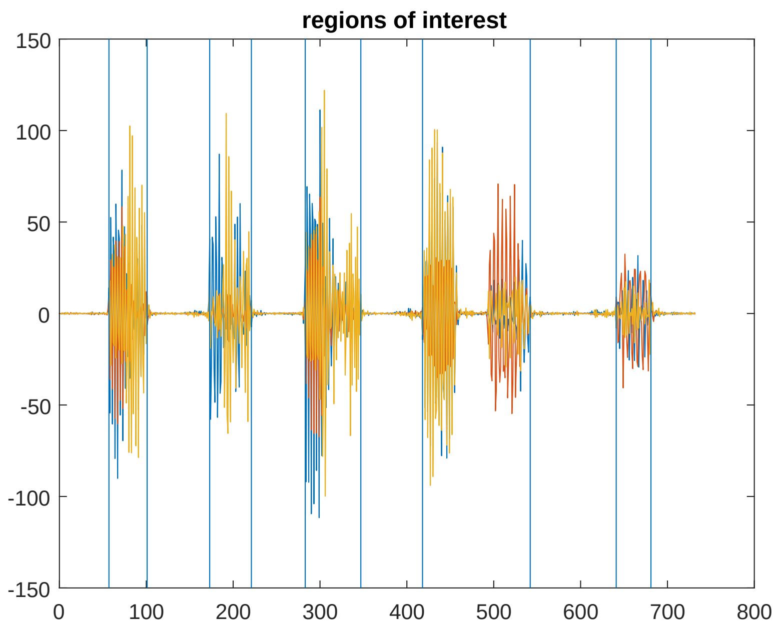

Figure 4: Regions of interest

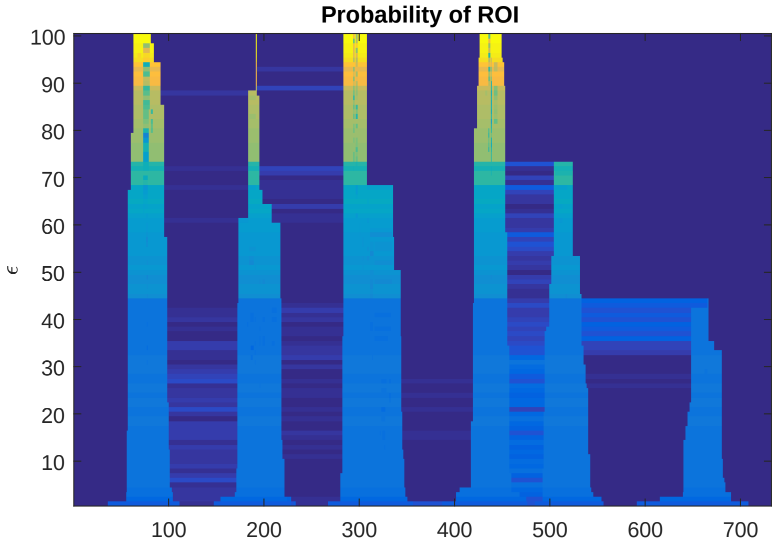

Figure 5: Probability of region of interest

Figure 5 demonstrate the probability of certain time interval to be within region of interest. We can see that for threshold values above 40 the shaking that happens around 650 ms is not included into region of interest due to low amplitude of shake. Between 30 and 40 we can see that there is high probability of merge between ROI on 500ms and 650ms due low number of events in the last one.

4 Conclusion

Developed method of identification of ROI in accelerometer data demonstrate stability for low values of threshold, however usage of such method may result in loss of information in the case of high values of threshold.

References

[1] Paresh Gujarati, *What is Accelerometer and how does it work on smartphones*, http://www.techulator.com/resources/8930-How-does-smart-phone-accelerometer-work.aspx