Human Activity Recognition from Accelerometer Data: A Probabilistic Approach

by
Walid Gomaa
Cyber-Physical Systems Lab, Egypt-Japan University of Science and
Technology, And Faculty of Engineering, Alexandria University,
Alexandria, Egypt.

Presentation: VIBHAVASU

Indian Institute of Technology Hyderabad

June 25, 2021

Index

Introduction

- Automatic recognition of human activities has become a very substantial research topic. A wide range of digital applications depend mainly on human activity recognition (HAR) in their work such as monitoring patients and elderly people, surveillance systems, robots learning and cooperating, human robot interaction, and in the military.

Data Acquisition Methods

- A typical example of such activities are Activities of Daily Living (ADL) that people have the ability for doing on a daily basis like eating, moving, hygiene, dressing, etc.
- There are many methods and data acquisition systems which are based on different sensory readings for recognizing people's actions.
- Two popular ways in which human activities are recorded:
  - Ambient/Surrounding fixed sensor systems
  - Wearable mobile sensor systems

Ambient/Surrounding fixed sensor systems

- Data is collected from distributed sensors attached to fixed locations in the surrounding environment such as surveillance cameras, microphones, motion sensors, etc.

- Alternatively, the sensors are attached to interactive objects in order to detect the type of interaction with them, such as a motion sensor attached to the cupboard doors, microwave ovens, water taps, etc.

- This method can detect complex actions efficiently.

- Drawbacks:
  - ▶ It is fixed. So, if the user leaves the place, his activities won't be detectable.
  - ▶ Privacy. Especially when considering video surveillance cameras or auditory sensors. User may not feel free and may not perform his/her activities normally when in the observable range of these systems.

Wearable mobile sensor systems

- The data is taken from mobile sensors mounted to human body parts.
- The typical sensory unit used is the Inertial Measurement Unit (IMU) which streams signals about the acceleration, Gyroscope rotational velocity, Gyroscope orientation, and the gravitational field.
- Advantages:
  - Precise
  - Portable
  - Measure data directly.
  - Can be Embedded as MEMS sensors in wearable gadgets.

Hypothesis

- The data used in this research is acquired from mobile sensor systems and is streamed while performing 14 activities using several human subjects and with varying number of samples for the different activities.

- Most of the methods targeted at activity recognition are based either on the machine learning paradigm or time series analysis.

Hypothesis: *The predictive performance can be boosted by including more sensory data such as the gyroscope data*

Data Description

- The data is collected using a wrist-worn accelerometer with sampling rate of 32Hz.
- Measurement range is [-1.5,1.5] (in g-force units)
- One data sample is a tri - axial acceleration: X - axis pointing toward the hand, Y - axis pointing toward the left, and Z - axis perpendicular to the plane of the hand.
- Acceleration data are collected for 14 activities with varying number of samples for each.

Table: List of Activities and Associated number of samples.

No.	Activity	No.of samples
1	brush teeth	12
2	climb stairs	102
3	comb hair	31
4	drink glass	100
5	getup bed	101
6	sit down in chair	100
7	use telephone	13
8	pour water	100
9	eat meat	5
10	walk	100
11	lie down on bed	28
12	stand up from chair	102
13	descend stairs	42
14	eat soup	3

Estimation using Discrete distributions

- This research deals with the statistical properties of the raw signal without taking time into consideration relying on probabilistic modelling of the signals.
- Distance among these probability distributions is measured in order to classify the given activity signal.
- The points of a time series are considered, as generated from a discrete distribution.
  Hence, the empirical distribution of the time series is estimated using the discrete distribution that corresponds directly to the observed data.

Distance measures used to measure dissimilarity

- Several distance measures are applied among the discrete distributions in order to measure the dissimilarity among different time series samples and accordingly decide on activity membership.
- *The resolution of the accelerometer data is fine enough to provide a comparable discrete distributions among the time series samples*.
- The four distance measures used here to measure dissimilarity among discrete probability distributions are:
  - ▸ Hellinger distance
  - ▸ Total variation distance
  - ▸ Cramer-von Mises distance
  - ▸ Kolmogorov distance

Hellinger distance

- It is defined in terms of the Hellinger integral.
- It can be computed for discrete distributions as:

Given two discrete distributions $P = (p1,...,pn)$ and $Q = (q1,...,qn)$, then:

$$d_H(P, Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2} \tag{1}$$

Hellinger distance is a proper metric and it is normalized by the factor $\frac{1}{\sqrt{2}}$ such that it is upper bounded by 1.

Total variation distance

- Given two probability measures P and Q on the same σ-algebra (Ω, F), the total variation distance between P and Q can be computed as:

$$d_T(P, Q) = sup\{|P(A) - Q(A)| : A \in F\} \tag{2}$$

- It is the largest difference between the two measures for any given event in the σ-algebra.

- In the assumed discrete case, as also the support is finite, this can be reduced to:

$$d_T(P, Q) = \sum_{i=1}^{n} |p_i - q_i| \tag{3}$$

- This reduces to the Manhattan distance.

Cramer-von Mises distance

Computes the distance between the empirical cumulative distribution functions(CDF) that are estimated from the given data. This distance can be computed as:

$$d_C^2(P, Q) = \int \Big(P(\{y \in [-1.5, 1.5] : y \leq x\}) $$
$$-Q(\{y \in [-1.5, 1.5] : y \leq x\}) \Big)^2 \mu(dx) \tag{4}$$

Where μ is the integration measure, which in these experiments taken to be the probability measure Q.

Kolmogorov distance

It computes the supremum distance between the two given probability distributions evaluated over the sets in the $\sigma-$ algebra defined over $[-1.5, 1.5]$:

$$d_K(P, Q) = sup\{|P(\{y \in [-1.5, 1.5] : y \leq x\}) \\ -Q(\{y \in [-1.5, 1.5] : y \leq x\})| : x \in [-1.5, 1.5]\} \tag{5}$$

Phases

- For each activity A, some of its collected samples s, called 'Support Samples', are randomly selected to act as templates for the activity they are drawn from. The remaining samples are used for testing.
- In experimentation *65%* of the samples are used as 'support samples'.
- Offline modeling phase:
  Easy and computationally efficient estimation of the discrete probability distributions of the given time series samples.
- Testing phase:
  Computing the distance, using one of the four distance metrics, between the test sample and all of the support samples of the 14 activities.

Classification of samples

- When a new test sample s is classified, the following is done:
- For each activity A, the distance between s and every member of the supporting samples of A is found. Each test sample s has three empirical distributions associated with it for the tri-axial directions.
- Each such distribution is compared against the corresponding distribution in the supporting sample s. So we have three scores. These scores are combined together by taking their summation.

$$d(s, s') = d_X(s_x, s'_x) + d_X(s_y, s'_y) + d_X(s_z, s'_z)$$

d_X is either of the four metrics described previously.

- To compute the distance scores of the given test sample s against all the supporting samples of activity A, a single score must be chosen to represent the distance between the test sample s and activity A. In this case, the minimum score is chosen:

$$d(s,A) = \min_{s' \in Supp(A)} d_X(s, s')$$

- The decision is then taken by assigning s to the activity A with minimum distance:

$$A' = \operatorname{argmin}_A d(s, A)$$

This procedure is equivalent to a k-NN classifier with $k = 1$
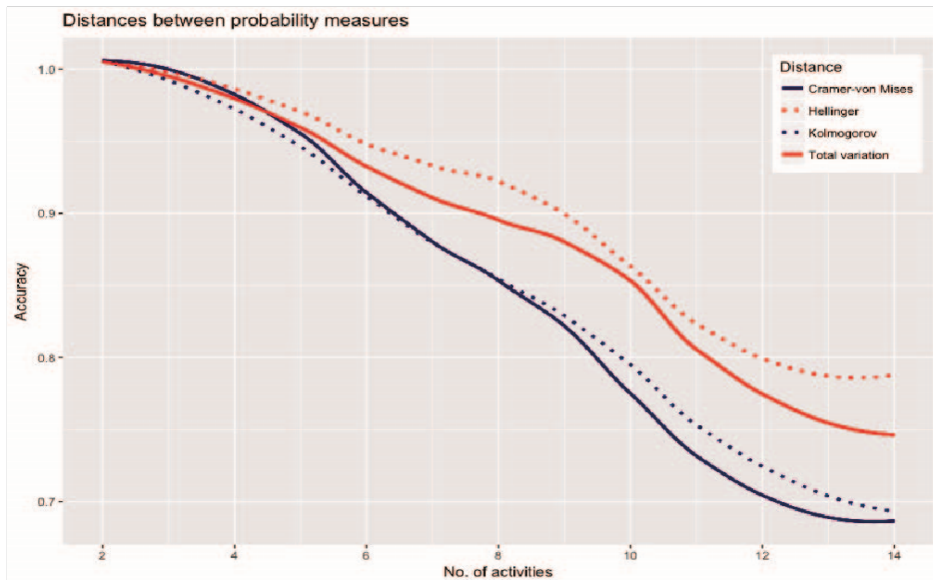
Figure: Comparing the accuracy using different distance metrics between discrete empirical distributions.

Inferences

- Working with probability mass functions, (Hellinger and Total variation), outperforms working with cumulative distribution functions, (Cramervon Mises and Kolmogorov).

- Working with an L2 norm style as in the Hellinger case outperforms working with an L1 norm style as in the total variation case. The former is smoother and more uniform in accounting for distances among the various parts of the two input signals.

- The best performance is achieved using the Hellinger distance, whose accuracy is about 80%.

Prospective future work

- The future research can investigate more diverse methods within the same paradigm of using probability distributions as representations of the signals along with multitude of distance metrics. It may also experiment with other modes of sensory data, in particular, the gyroscope.

Summary

- This research aimed at recognition of human activity based on probabilistic modelling of the underlying tri-axial accelerometer signal that describes the activity.
- The signal is assumed to be streamed from a wearable device.
- Time series is represented as a probability distribution, which loses the temporal succession of the streaming.
- Despite of this loss, the approach has shown effectiveness in the predictive performance as well as computationally.
- This can be attributed in part to the stationarity of the underlying time series as well as to a hypothesized short-term correlations in the given studied activities.
- These claims need to be verified using more datasets as well as more thorough investigation of the time series.

Summary

- For each activity, a set of samples from the training set, called 'support samples' are chosen *at random* to act as templates of the activity. Then, at test time a new test sample is then compared using their representations as probability distributions, against each template set of each activity to decide its activity membership.
- The comparison is based on some typical dissimilarity metrics.
- It turned out that the Hellinger distance gives the best stable performance.

Thank You

Vibhavasu Pasumarti
EP20BTECH11015