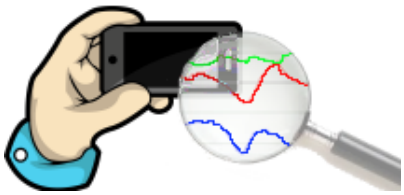

Accelerometer Biometric Competition

Recognize users of mobile devices from accelerometer data

$5,000 · 633 teams · 4 years ago

Competition Data

Edit

- questions.csv
- sampleSubmission.csv
- test.csv.zip
- train.csv.zip

questions.csv 1.43 MB

Download

Data Description

Your task is to determine whether the accelerometer recordings in the test set (test.csv) belong to the proposed devices in the question set (questions.csv).

The following files are provided:

File name	Description
train.zip	30m samples for training, labeled with DeviceId
test.zip	30m samples for test, split into 90k sequences, labeled with SequenceId
questions.csv	90k questions that match test sequences to DeviceId

train.csv

Field name	Description
T	Unix time (miliseconds since 1/1/1970)
X	Acceleration measured in g on x co-ordinate
Y	Acceleration measured in g on y co-ordinate
Z	Acceleration measured in g on z co-ordinate
DeviceId	Unique Id of the device that generated the samples

test.csv

Field name	Description
T	Unix time (miliseconds since 1/1/1970)
X	Acceleration measured in g on x co-ordinate
Y	Acceleration measured in g on y co-ordinate
Z	Acceleration measured in g on z co-ordinate
SequenceId	Unique sequence number assigned to each question. Each group of samples is labeled with a unique SequenceId. Each SequenceId is matched to a professed DeviceId in the questions.csv file.

questions.csv

Field name	Description
QuestionId	Id of question
SequenceId	Unique number assigned to each sequence of samples
QuizDevice	Professed device that generated the sequence of accelerometer data in the test file

Data Preparation

1. After removing from the data set samples from devices with fewer than 6000 samples collected and sequences of samples during periods of rest exceeding 10 seconds, 60 million samples from 387 users were made available for the competition.
2. Samples from each device were split chronologically into train and test sets, with the earliest samples assigned to the train set.
3. The test set was demarcated into 90,000 consecutive sequences of 300 samples each from a same device and each sequence assigned a unique Sequence Id.
4. Each sequence Id was associated with a professed device Id such that in some cases a true device id was assigned and in others a false device id was assigned.