**О датасете.**

Оригинал текста:

The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals).

These signals were used to estimate variables of the feature vector for each pattern:

'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.

tBodyAcc-XYZ

tGravityAcc-XYZ

tBodyAccJerk-XYZ

tBodyGyro-XYZ

tBodyGyroJerk-XYZ

tBodyAccMag

tGravityAccMag

tBodyAccJerkMag

tBodyGyroMag

tBodyGyroJerkMag

fBodyAcc-XYZ

fBodyAccJerk-XYZ

fBodyGyro-XYZ

fBodyAccMag

fBodyAccJerkMag

fBodyGyroMag

fBodyGyroJerkMag

The set of variables that were estimated from these signals are:

mean(): Mean value

std(): Standard deviation

mad(): Median absolute deviation

max(): Largest value in array

min(): Smallest value in array

sma(): Signal magnitude area

energy(): Energy measure. Sum of the squares divided by the number of values

iqr(): Interquartile range

entropy(): Signal entropy

arCoeff(): Autorregresion coefficients with Burg order equal to 4

correlation(): correlation coefficient between two signals

maxInds(): index of the frequency component with largest magnitude

meanFreq(): Weighted average of the frequency components to obtain a mean frequency

skewness(): skewness of the frequency domain signal

kurtosis(): kurtosis of the frequency domain signal

bandsEnergy(): Energy of a frequency interval within the 64 bins of the FFT of each window

angle(): Angle between to vectors

Additional vectors obtained by averaging the signals in a signal window sample. These are used on the angle() variable:

gravityMean

tBodyAccMean

tBodyAccJerkMean

tBodyGyroMean

tBodyGyroJerkMean

The complete list of variables of each feature vector is available in 'features.txt'

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain. See 'features\_info.txt' for more details.

Перевод:

Признаки, отобранные для этого датасета, были получены с помощью акселерометра и гироскопа по 3 осям. Необработанные сигналы названы tAcc-XYZ и tGyro-XYZ. Эти сигналы зависят от времени (префикс «t» для обозначения времени) и были захвачены с постоянной частотой 50 Гц. Затем они были отфильтрованы с помощью медианного фильтра и низкочастотного фильтра Баттерворта 3-го порядка с угловой частотой 20 Гц для удаления шума. Аналогично, сигнал акселерометра затем был разделён на сигналы ускорения тела и гравитации (tBodyAcc-XYZ и tGravityAcc-XYZ) с использованием другого фильтра Баттерворта низких частот с угловой частотой 0,3 Гц.

Впоследствии из линейного ускорения и угловой скорости тела, взятых по времени, были получены сигналы рывка (tBodyAccJerk-XYZ и tBodyGyroJerk-XYZ). Также были рассчитаны магнитуды этих трёхмерных сигналов с использованием евклидовой нормы (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

Наконец, быстрое преобразование Фурье (Fast Fourier Transform, FFT) было применено к некоторым из этих сигналов, и были получены признаки fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag (обратите внимание на букву «f» для обозначения сигналов частотной области).

Эти сигналы использовались для оценки переменных вектора признаков для каждого паттерна («-XYZ» используется для обозначения 3-осевых сигналов в направлениях X, Y и Z):

tBodyAcc-XYZ

tGravityAcc-XYZ

tBodyAccJerk-XYZ

tBodyGyro-XYZ

tBodyGyroJerk-XYZ

tBodyAccMag

tGravityAccMag

tBodyAccJerkMag

tBodyGyroMag

tBodyGyroJerkMag

fBodyAcc-XYZ

fBodyAccJerk-XYZ

fBodyGyro-XYZ

fBodyAccMag

fBodyAccJerkMag

fBodyGyroMag

fBodyGyroJerkMag

Набор переменных, которые были оценены по этим сигналам, таков:

mean(): среднее значение

std(): стандартное отклонение

mad(): медианное абсолютное отклонение

max(): наибольшее значение в массиве

min(): наименьшее значение в массиве

sma(): область величины сигнала

energy(): мера энергии. Сумма квадратов, деленная на количество значений.

iqr():интерквартильный размах

entropy(): энтропия сигналов

arCoeff(): коэффициенты авторегрессии с порядком Бурга равным 4

correlation(): коэффициент корреляции между двумя сигналами

maxInds(): индекс частотной составляющей с наибольшей величиной

meanFreq (): средневзвешенное значение частотных составляющих для получения средней частоты

skewness(): асимметрия частотной области сигнала

kurtosis(): коэффициент эксцесса сигнала частотной области

bandsEnergy(): энергия частотного интервала в пределах 64 ячеек FFT каждого окна.

angle(): угол между векторами.

Дополнительные векторы получаются путём усреднения сигналов в выборке окна сигнала. Они используются в переменной angle():

gravityMean

tBodyAccMean

tBodyAccJerkMean

tBodyGyroMean

tBodyGyroJerkMean

Полный список переменных каждого вектора объектов доступен в разделе «features.txt».

Эксперименты проводились с группой из 30 добровольцев в возрасте от 19 до 48 лет. Каждый человек выполнял шесть видов деятельности (ходьба, ходьба вверх по лестнице, ходьба вниз по лестнице, положение сидя, положение стоя, положение лёжа), нося смартфон (Samsung Galaxy S II) на талии. Используя его встроенный акселерометр и гироскоп, мы захватили 3-осевое линейное ускорение и 3-осевую угловую скорость с постоянной скоростью 50 Гц. Эксперименты были записаны на видео, чтобы маркировать данные вручную. Полученный набор данных был случайным образом разделен на два набора, где 70% добровольцев были отобраны для генерации обучающих данных и 30% - тестовых данных.

Сигналы датчиков (акселерометр и гироскоп) предварительно обрабатывались с применением шумовых фильтров, а затем отбирались в скользящих окнах фиксированной ширины с перекрытием 2,56 сек и 50% (128 показаний/окно). Сигнал акселерометра, который имеет гравитационную составляющую и компонент движения тела, был разделен с помощью фильтра нижних частот Баттерворта на ускорение и гравитацию тела. Предполагается, что гравитационная сила имеет только низкочастотные составляющие, поэтому был использован фильтр с частотой отсечения 0,3 Гц. Из каждого окна был получен вектор признаков путем вычисления переменных из временной и частотной областей.

**Алгоритм:**

1. Получение из csv-файлов информации от акселерометра и гироскопа в массивы.
2. Отфильтровать данные через медианный фильтр.
   1. [Медианный фильтр](https://ru.wikipedia.org/wiki/Медианный_фильтр)
   2. [Обобщение медианного фильтра](https://habr.com/ru/post/114551/)
   3. [scipy.ndimage.median\_filter](https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.median_filter.html)
3. Отфильтровать данные через низкочастотный фильтр Баттерворта 3-го порядка с угловой частотой 20 Гц для удаления шума.
   1. [Фильтр Баттерворта](https://ru.wikipedia.org/wiki/Фильтр_Баттерворта)
   2. [О ФЧХ и фильтрах](https://web.archive.org/web/20070312233719/http:/www.dsp.sut.ru/rus/training/book/lections/l3/l3_2.htm)
   3. [scipy.signal.butter](https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.butter.html)
   4. [Как реализовать полосовой фильтр Баттерворта с помощью scipy.signal.butter](https://stackoverflow.com/questions/12093594/how-to-implement-band-pass-butterworth-filter-with-scipy-signal-butter)
   5. [Butterworth filter in python](https://stackoverflow.com/questions/26757271/butterworth-filter-in-python)
   6. [Применение фильтра нижних частот Баттерворта в Python](https://switch-case.ru/60840039)
   7. [scipy.signal.filtfilt](https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.filtfilt.html)
4. Выделение из сигнала акселерометра сигналов ускорения тела и гравитации (tBodyAcc-XYZ и tGravityAcc-XYZ) с использованием другого фильтра Баттерворта низких частот с угловой частотой 0,3 Гц.
   1. [How to decompose raw acceleration signals into gravity and dynamic acceleration?](https://dsp.stackexchange.com/questions/50046/how-to-decompose-raw-acceleration-signals-into-gravity-and-dynamic-acceleration)
   2. [Программирование&Музыка: Частотный фильтр Баттервота. Часть 3](https://habr.com/ru/post/313062/)
5. Получение из линейного ускорения и угловой скорости тела сигналов рывка (tBodyAccJerk-XYZ и tBodyGyroJerk-XYZ).
   1. [Рывок (кинематика)](https://ru.wikipedia.org/wiki/Рывок_(кинематика))
   2. [scipy.signal.ellip](https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.ellip.html)
   3. [Эллиптический фильтр](https://ru.wikipedia.org/wiki/Эллиптический_фильтр)
6. Получение магнитуд трёхмерных сигналов с использованием евклидовой нормы (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).
   1. [Евклидова норма](https://ru.wikipedia.org/wiki/Норма_(математика))
7. Применение быстрого преобразования Фурье (Fast Fourier Transform, FFT) к некоторым из сигналов для получения признаков fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag.
   1. [Быстрое преобразование Фурье](https://ru.wikipedia.org/wiki/Быстрое_преобразование_Фурье)
   2. [Дискретное преобразование Фурье](https://ru.wikipedia.org/wiki/Дискретное_преобразование_Фурье)
   3. [Реализации алгоритма быстрого преобразования Фурье](https://ru.wikibooks.org/wiki/Реализации_алгоритмов/Быстрое_преобразование_Фурье)
   4. [scipy.fft.fft](https://docs.scipy.org/doc/scipy/reference/generated/scipy.fft.fft.html)
   5. [scipy.fft.fftshift](https://docs.scipy.org/doc/scipy/reference/generated/scipy.fft.fftshift.html)
   6. [scipy.fftpack.fft](https://docs.scipy.org/doc/scipy/reference/generated/scipy.fftpack.fft.html)
8. Вычисления средних значений, отклонений, коэффициентов и т.д.
   1. mean – среднее значение.
      1. [numpy.mean](https://numpy.org/doc/stable/reference/generated/numpy.mean.html)
   2. std – стандартное отклонение.
      1. [numpy.std](https://numpy.org/doc/stable/reference/generated/numpy.std.html)
   3. mad – медианное абсолютное отклонение.
      1. [scipy.stats.median\_absolute\_deviation](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.median_absolute_deviation.html)
   4. max – наибольшее значение в массиве.
      1. [numpy.amax](https://numpy.org/doc/stable/reference/generated/numpy.amax.html)
   5. min – наименьшее значение в массиве.

8.5.1. [numpy.amin](https://numpy.org/doc/stable/reference/generated/numpy.amin.html)

* 1. sma – область величины сигнала.
     1. [Signal magnitude area](https://dsp.stackexchange.com/questions/18649/signal-magnitude-area)
  2. energy – мера энергии. Сумма квадратов, делённая на количество значений.
  3. iqr – интерквартильный размах.
     1. [scipy.stats.iqr](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.iqr.html)
  4. entropy – энтропия сигналов.
     1. [scipy.stats.entropy](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.entropy.html)
     2. [scipy.stats.norm](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html)
     3. [Interpreting scipy.stats.entropy values](https://stackoverflow.com/questions/26743201/interpreting-scipy-stats-entropy-values)
  5. arCoeff – коэффициенты авторегрессии с порядком Бурга равным 4.
     1. [Autoregressive model](https://en.wikipedia.org/wiki/Autoregressive_model)
     2. [What are Autoregressive Coefficients?](https://dsp.stackexchange.com/questions/9518/what-are-autoregressive-coefficients)
     3. [Range of Autoregressive Coefficients](https://stackoverflow.com/questions/33135802/range-of-autoregressive-coefficients)
     4. [scipy.signal.iirfilter](https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.iirfilter.html)
  6. correlation – коэффициент корреляции между двумя сигналами.
     1. [scipy.signal.correlate](https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlate.html)
  7. maxInds – индекс частотной составляющей с наибольшей величиной.
  8. meanFreq – средневзвешенное значение частотных составляющих для получения средней частоты.
  9. skewness – асимметрия частотной области сигнала.
     1. [scipy.stats.skew](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.skew.html)
  10. kurtosis – коэффициент эксцесса сигнала частотной области.
      1. [scipy.stats.kurtosis](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kurtosis.html)
      2. [How to correctly use scipy's skew and kurtosis functions?](https://stackoverflow.com/questions/45483890/how-to-correctly-use-scipys-skew-and-kurtosis-functions)
  11. bandsEnergy – энергия частотного интервала в пределах 64 ячеек FFT каждого окна.
  12. angle – угол между векторами.

1. Очень полезные алгоритмы.
   1. [Add single element to array in numpy](https://stackoverflow.com/questions/7332841/add-single-element-to-array-in-numpy)
   2. [How can I represent an infinite number in Python?](https://stackoverflow.com/questions/7781260/how-can-i-represent-an-infinite-number-in-python)
   3. [numpy.concatenate](https://numpy.org/devdocs/reference/generated/numpy.concatenate.html)