

А.Л. ЕРОХИН, д-р техн. наук, проф, ХНУРЭ, Харьков,
С.Н. ЛЕДНЁВ, магистр, ХНУРЭ, Харьков

МЕТОДЫ РАСПОЗНАВАНИЯ ЖЕСТОВ НА ОСНОВЕ ДАНЫХ ТРЕХОСЕВЫХ АКСЕЛЕРОМЕТРОВ ANDROID УСТРОЙСТВ

Статья посвящена анализу методов распознавания жестов на основе показаний трехосевых акселерометров мобильных устройств. Рассмотрены основные методы машинного обучения применительно к задаче классификации жестов. Произведена начальная обработка данных, отбор признаков и выполнено сравнение алгоритмов с приведением качества распознавания. Ил.: 4. Табл.: 1. Библиогр.: 11 назв.

Ключевые слова: распознавание жестов; трехосевые акселерометры; мобильные устройства; классификация.

Постановка проблемы. Распознавание жестов является перспективным направлением развития интерфейсов взаимодействия человека с техникой в режиме реального времени, которые в будущем вполне смогут заменить текстовые и графические интерфейсы. Большая часть способов, направленных на распознавание жестов, применяет техники компьютерного зрения и обработки изображений [1]. Для этого необходимо наличие камер и присутствие человека в зоне видимости. Еще одним способом является анализ показаний акселерометров и гироскопов с устройств. Самым известным контроллером, использующих такой тип распознавания жестов, является Wii Remote [2] для игровой консоли Wii от компании Nintendo.

Развитие мобильных технологий и усовершенствование технологических процессов ведут к уменьшению размеров и увеличению мощностей мобильных устройств, появлению носимой электроники, оснащенной разными датчиками. Большинство современных смартфонов обладают встроенным трехосевым акселерометром, с помощью которого, например, реализуется всем привычная функция авто-поворота дисплея. Основное предназначение такого датчика – предоставление информации о текущем ускорении устройства (разности ускорения устройства и ускорения свободного падения). В состоянии покоя показания датчика совпадают с вектором ускорения свободного падения.

Все более распространенными и доступными на рынке становятся "умные" часы, которые тоже представляют собой Android устройства,

также часто оснащены акселерометром и другими датчиками. Также беспроводные устройства малых размеров, оснащенные акселерометром, могут быть встроены в одежду либо другие персональные электронные устройства. Распознавание жестов позволяет добавлять разные функции не только для работы с самими устройствами, например, управление медиаплеером или включение камеры, но и для работы с другими устройствами и приложениями на компьютере, например рисование или использование мобильного устройства как контроллера в играх.

Поскольку распознавание жестов является перспективным направлением взаимодействия интерфейсов, то необходимо провести дополнительные исследования моделей и методов распознавания двигательной активности пользователя. При изучении предметной области не было выявлено исследований, связанных с применением групповых (ensemble) методов классификации относительно задачи распознавания жестов, кроме Random Forest [3]. Поэтому было принято решение исследовать групповые алгоритмы, такие как Extra Trees (Extremely Randomized Trees), Random Forest, Gradient Boosting, а также использовать метод Bagging (сокращение от "bootstrap aggregation") для общепринятых для данного класса задач алгоритмов KNN и SVM. Bagging представляет собой классификатор, состоящий из нескольких классификаторов, обученных на случайным образом сгенерированных выборках, данные в которых могут пересекаться. Обычно применяют данный метод для деревьев решений, поскольку их комбинации хорошо работают при построении моделей, обученных на различных частях выборки, но также Bagging можно применить и к любому другому алгоритму классификации.

Анализ литературы. Платформа Android предоставляет несколько датчиков, которые позволяют отследить движение устройства [4]. Акселерометр и гироскоп всегда основаны на аппаратной составляющей, а датчик гравитации, линейного ускорения и вектор вращения могут быть как аппаратными, так и программными. Датчики движения полезны для мониторинга движений устройства, таких как наклон, встряска, поворот или взмах.

Необработанные значения, полученные от акселерометра смартфона, представляют собой массив из трех элементов, соответствующих трем координатным осям, приведенным на рис. 1.

Значения элементов массива являются ускорением вдоль оси (с учетом гравитации) и могут быть как положительными, так и отрицательными. Измеряются в m/s^2 . Стоит отметить, что важной особенностью координатной системы Android устройств является тот факт, что при смене положения устройства оси координат не меняются,

т.е. если сначала устройство держали в книжной ориентации, а затем перевернули в альбомную, то оси x и y не меняют своего положения.

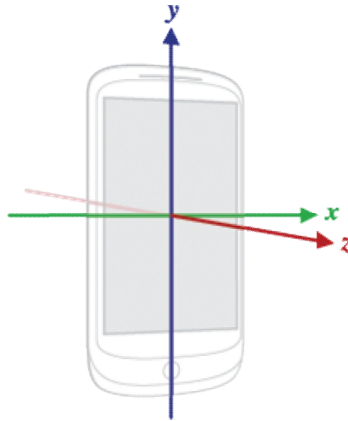


Рис. 1. Система координат датчиков устройства Android

Особенность датчиков мобильного устройства также в том, что в состоянии покоя акселерометр считывает величину $g = 9.81 \text{ м/с}^2$ [4], а когда устройство находится в состоянии свободного падения и быстро ускоряется к земле с ускорением 9.81 м/с^2 , то акселерометр считывает величину $g = 0 \text{ м/с}^2$. Таким образом, для измерения реального ускорения необходимо устранить вклад силы тяжести из данных акселерометра.

Акселерометры в мобильных устройствах обычно не являются очень качественными, поэтому сигнал содержит большое количество шумов. Убрать вклад силы тяжести и шумы из исходного сигнала можно с помощью фильтра низких (low-pass) и высоких (high-pass) частот, простейшие реализации которых (формулы 1 и 2) предлагаются в документации Android [4]:

$$g_i = \alpha \cdot g_{i-1} + (1 - \alpha) \cdot v_i, \quad (1)$$

$$acc_i = v_i - g_i, \quad (2)$$

где g_i – изолированная сила гравитации; v_i – необработанные, "сырые" данные, полученные от акселерометра; α – константа; $0 \leq \alpha \leq 1$, acc_i – чистое ускорение (без учета гравитации).

Фильтры высоких и низких частот довольно часто используются при обработке данных с датчиков, благодаря простоте и эффективности. Также популярны скользящее среднее и фильтр Калмана.

Все предыдущие работы в области распознавания жестов можно разделить на две группы: основанные на шаблонах и основанные на моделях. Принцип работы первых заключается в хранении "эталонных"

жестов и дальнейшем их сравнении с выполненным жестом при помощи различных метрик расстояния, например евклидовой метрики. К таким алгоритмам можно отнести uWave [5], который по заявлениям авторов позволяет с точностью до 98% распознать совершенный жест. Данный алгоритм использует в своей основе другой алгоритм – DTW [6, 7] (Dynamic Time Wrapping, Динамическая Шкала Трансформации Времени), который часто применяется в обработке видео, аудио, распознавании речи и использует расстояние Левенштейна.

Статистические методы, например, такие как широко известные Скрытые Марковские Модели (Hidden Markov Model, HMM), базируются на вероятностной интерпретации образцов жестов для моделирования траектории жеста. В довольно известной работе Schlomer [8] использовалась HMM, оптимизированная алгоритмом Баума-Вельша с 8 состояниями. Данные жеста перед тем, как попасть в HMM, предварительно обрабатывались двумя фильтрами. Первый, так называемый "фильтр простоя", представляет собой пороговый фильтр, отбрасывающий вектора данных, не вносящих значимый вклад в характеристику жеста. Второй фильтр устранял векторы, которые приблизительно похожи на предыдущий.

Довольно часто можно встретить использование метода опорных векторов (SVM) [3, 9] в качестве классификатора. Методы, базирующиеся на SVM, обычно обладают меньшими вычислительными требованиями во время классификации, что делает их привлекательными для использования на мобильных устройствах.

Поскольку один жест представляет собой набор данных из трех векторов по осям x , y и z , а также времени получения каждой координаты в наносекундах, для дальнейшей работы его необходимо представить в виде одного вектора, при этом сохранив наиболее важные характеристики жеста. Для решения проблемы уменьшения размерности можно воспользоваться алгоритмами кластеризации, например K-Means [10] (K-Средних), либо PCA [11] (Principal Component Analysis, Метод главных компонент) – статистической процедуры, направленной на уменьшение размерности с минимальной потерей информации.

В работе [8] для этих целей был использован алгоритм K-Means, с $K = 14$. Число K было выбрано эмпирически, путем расположения центров кластеров в трехмерной системе координат по осям x , y и z , формируя таким образом сферу.

Целью исследования является сравнительное исследование групповых алгоритмов классификации Random Forest, Gradient Boosting и Extra Trees и применение техники Bagging для алгоритмов KNN (K-ближайших соседей) и SVM (метод опорных векторов),

применительно к задаче распознавания жестов на основе данных трехосевых акселерометров Android устройств.

Система распознавания жестов. Работа системы по распознаванию жестов (рис. 2) состоит из следующих этапов: получение данных, фильтрация, уменьшение размерности и/или выделение признаков, классификация.

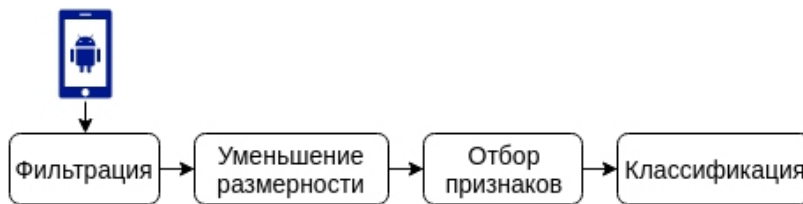


Рис. 2. Схема работы системы распознавания жестов

При работе с датчиками в Android есть возможность выставить, с какой частотой будут опрашиваться датчики. В данной работе была выбрана частота опроса в 20,000 микросекунд (что соответствует значению `SENSOR_DELAY_GAME` в документации).

Фильтрация является необходимым этапом, поскольку в исходном сигнале присутствуют шумы и разные датчики обладают разной калибровкой. Для фильтрации в данной работе были использованы фильтры, предложенные в документации Android со значением параметра $\alpha = 0.95$.

Для уменьшения размерности был использован алгоритм *K*-Means с одним кластером для каждой оси. Малое количество выбранных кластеров обосновывается использованием дополнительных признаков, сформированных на основе исходных данных.

Отбор признаков представляет собой процесс выбора характеристик из данных, которые имеют наиболее тесную связь с целевой переменной. Признаки формируются на основе уже существующих переменных в данных, либо являются комбинацией нескольких переменных. Данный шаг является важным, поскольку при грамотном подходе позволяет уменьшить переобучение модели, извлекая таким образом ненужные переменные, сбивающие с толку модель и заставляющие ее делать ложные выводы, повысить точность, а также сократить время обучения, при условии уменьшения признаков.

Признаков, полученных после этапа уменьшения размерности, может быть не достаточно для обеспечения приемлемого уровня распознавания жестов. Для улучшения качества были добавлены следующие признаки (в общем итоге 28 признаков):

- центры кластеров по осям x , y , z ;

- сумма по осям x, y, z ;
- среднее по осям x, y, z ;
- максимум по осям x, y, z ;
- минимуму по осям x, y, z ;
- стандартное отклонения по осям x, y, z ;
- сумма общего усиления (формула 3) [3];
- минимум, максимум, стандартное отклонение и среднее для продольной оси (движение "крен", roll, формула 4) и поперечной оси (движение "тангаж", pitch формула 5);
- время выполнения жеста в наносекундах;

$$a_c(i) = \sqrt{a_x(i)^2 + a_y(i)^2 + a_z(i)^2}, \quad (3)$$

где $a_c(i)$ – общее ускорение, а $a_x(i)^2, a_y(i)^2, a_z(i)^2$ – квадраты ускорений по осям x, y, z соответственно.

Из показаний акселерометра можно извлечь такие данные, как угол наклона для продольной (roll) и поперечной (pitch) осей:

$$\text{pitch} = \arctan\left(\frac{a_y}{\sqrt{a_x^2 + a_z^2}}\right); \quad (4)$$

$$\text{roll} = \arctan\left(\frac{-a_x}{a_z}\right). \quad (5)$$

Полученный по формулам выше результат будет углом в радианах. Для перевода в градусы результат необходимо умножить на 180 и разделить на число ПИ.

Такие простые статистические признаки, как сумма, среднее, медиана, максимум, минимум и т.д. обычно хорошо работают и не требуют особых вычислительных ресурсов.

В данной работе исследовались следующие типы жестов (рис.3):

- влево;
- вправо;
- вверх;
- вниз;
- круг по часовой стрелке;
- круг против часовой стрелки;
- квадрат по часовой стрелке;
- квадрат против часовой стрелки;
- латинская буква Z.

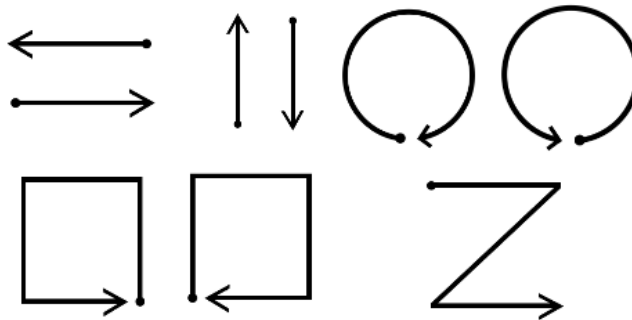


Рис. 3. Типы исследуемых жестов

В испытании приняли участие 4 человека, суммарно было собрано 1800 повторений жестов – по 200 для каждого типа. Для сбора данных было разработано Android-приложение. Начало жеста обозначалось нажатием кнопки в приложении на дисплее, затем следовал процесс выполнения жеста и отжатие кнопки означало завершение жеста.

Для каждого алгоритма классификации были подобраны наилучшие параметры путем кросс-валидации по сетке параметров. Все расчеты выполнялись с использованием языка Python и библиотеки машинного обучения Scikit-Learn с использованием пакетов Pandas и Numpy.

Алгоритмы классификации KNN и SVM очень чувствительны к входным параметрам. На наборе данных с отобранными признаками они давали очень низкое качество, поэтому для них было применена стандартизация признаков путем удаления среднего и масштабирования с единичной дисперсией (использовался класс StandardScaler из библиотеки).

Вся выборка делилась стратифицировано (для гарантии наличия одинакового количества типов жестов в каждой части) по принципу 70/30, 70% из которой использовались на этапе кросс-валидации (по 5 блокам), а затем 30% тестовых данных делились по принципу 60/40 для обучения и тестирования. Таким образом итоговая проверка качества работы алгоритмов проводилась на той части выборки, которая не принимала участие в кросс-валидации и обучении. Для оценки качества кросс-валидации использовалась взвешенная F-мера.

При вычислении итоговой оценки учитывались несколько метрик – точность, полнота и F-мера, которые вычислялись для каждого класса жеста. В таблице ниже приведены результаты работы алгоритмов (отсортированы по убыванию качества).

Таблиця

Качество классификации алгоритмов

Алгоритм	Точность	Полнота	F-мера
Extra Trees	95,1 %	94,8 %	94,8 %
SVM + Bagging	94,4 %	93,8 %	93,9 %
SVM	94,3 %	93,9 %	93,8 %
KNN	92,6 %	92,0 %	92,0 %
KNN + Bagging	92,3 %	91,5 %	91,6 %
Random Forest	91,9 %	91,5 %	91,6 %
Gradient Boosting	87,6 %	87,3 %	87,2 %

На рис. 4 приведена матрица неточностей для алгоритма, получившего самую высокую оценку – Extra Trees.

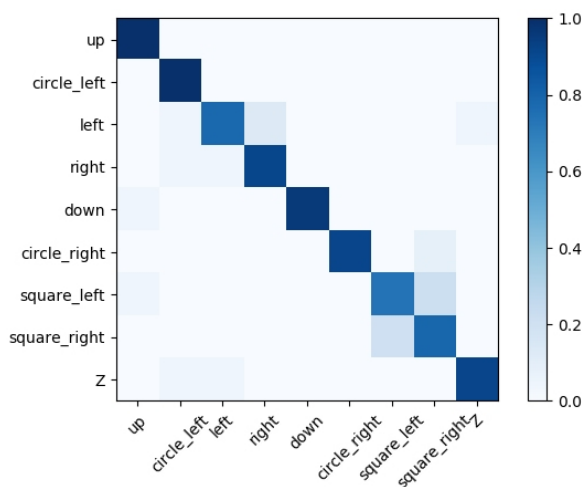


Рис. 4. Матрица неточностей для алгоритма Extra Trees

По матрице видно, что значимых отклонений не наблюдается и алгоритм редко ошибается, в основном все жесты распознаются верно. Труднее всего, по сравнению с другими жестами, поддается распознавание жестов Z и вниз – для них точность составляет 88%. Самыми успешно распознаваемыми жестами (точность 100%) являются жесты вверх, круги и квадрат против часовой стрелки.

Выводы. Проведенное в работе исследование показало, что вполне возможно добиться качества распознавания жестов близкого к 95%

используя лишь показания акселерометра путем извлечения несложных признаков и применив один из популярных алгоритмов классификации.

В работе проведено сравнение групповых методов классификации и использована техника беггинг для алгоритмов KNN и SVM, что, в свою очередь, приблизило их к разряду групповых. Ранее не используемый, применительно к задаче классификации жестов на основе датчиков мобильных устройств, алгоритм Extra Trees показал наилучшие результаты в сравнении с другими, и в особенности, с близким по своей сути алгоритмом Random Forest.

При использовании в дополнении к акселерометру показаний таких датчиков, как гироскоп и магнитометр, и извлечения из них дополнительных признаков, вполне осуществимо добиться более высокого качества распознавания. Исследования по данной тематике позволяют использовать мобильные устройства как средства для работы с дополненной реальностью, а также как средства управления интерфейсами, в особенности на носимых устройствах, таких как умные часы и фитнес-трекеры.

References:

1. "Gesture Recognition" (2016), available at: http://en.wikipedia.org/wiki/Gesture_recognition (accessed 10 March 2017).
2. "Wii | Nintendo", (2016), available at: www.nintendo.ru/Wii/Wii-94559.html (accessed 10 March 2017).
3. Feng, Hong, Shujuan, You, Meiyu, Wei, Yongtuo, Zhang, and Zhongwen, Guo (2016), *MGRA: Motion Gesture Recognition via Accelerometer*, Vol. 16 (4), 530 p.
4. "Motion Sensors and Android Developers", (2017), available at: http://developer.android.com/intl/ru/guide/topics/sensors/sensors_motion.html (accessed 12 March 2017).
5. Liu, J., Zhong, L., Wickramasuriya, J., and Vasudevan, V. (2009), *uWave: Accelerometer-based personalized gesture recognition and its applications*. *Pervasive Mob. Comput*, No. 5, pp. 657-675.
6. Berndt, D.J., Clifford, J. (1994), *Using Dynamic Time Warping to Find Patterns in Time Series*. *KDD workshop*, No. 16, pp. 359-370.
7. Senin P. (2008), *Dynamic time warping algorithm review*. Information and Computer Science Department, University of Hawaii at Manoa, Honolulu, USA, 23 p.
8. Schlömer, T. Poppinga, B. Henze, N., and Boll, S. (2008), *Gesture recognition with a Wii controller*. *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction*, Bonn, 18-21 February 2008, Germany, pp. 11-14.
9. Pati, I. M.A.R., Subbaraman, S.S. (2013), *A Review On Vision Based Hand Gesture Recognition Approach Using Support Vector Machines*, *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, pp. 7-12.
10. "K-Means clustering" (2016), available at: http://en.wikipedia.org/wiki/K-means_clustering (accessed 15 March 2017).
11. "Principal component analysis" (2016), available at: http://en.wikipedia.org/wiki/Principal_component_analysis (accessed 15 March 2017).

Статью представил д-р техн. наук, зав. каф. искусственного интеллекта ХНУРЭ Филатов В.А.

Поступила (received) 28.03.2017

Andriy Yerokhin, Dr.Sci.Tech, Professor
Kharkiv National University of Radio Electronics
Nauky Ave 14, Kharkiv, Ukraine, 61166
Tel: +38 067 575-84-81, e-mail: andriy.yerokhin@nure.ua
ORCID ID: 0000-0002-8867-993X

Serhii Lednov, master
Kharkiv National University of Radio Electronics
Nauky Ave 14, Kharkiv, Ukraine, 61166
Tel: +38 050 802 96 36, e-mail: serge.ledniiov@gmail.com
ORCID ID: 0000-0003-2542-8874

УДК 004.94

Методи розпізнавання жестів на основі даних трьохосьових акселерометрів Android пристроїв / Єрохін А.Л., Ледньов С.М. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2017. – № 21 (1243). – С. 54 – 64.

Статтю присвячено аналізу методів розпізнавання жестів на основі показань трьохосьових акселерометрів мобільних пристроїв. Розглянуто основні методи машинного навчання відносно задачі класифікації жестів. Виконано початкову обробку даних, відбір ознак і виконано порівняння алгоритмів з наведенням якості розпізнавання. Іл.: 4. Табл.: 1. Бібліогр.: 11 назв.

Ключові слова: розпізнавання жестів; трьохосьові акселерометри; мобільні пристрої; класифікація.

УДК 004.94

Методы распознавания жестов на основе данных трехосевых акселерометров Android устройств / Ерохин А.Л., Леднёв С.Н. // Вестник НТУ "ХПИ". Серія: Інформатика и моделирование. – Харьков: НТУ "ХПИ". – 2017. – № 21 (1243). – С. 54 – 64.

Статья посвящена анализу методов распознавания жестов на основе показаний трехосевых акселерометров мобильных устройств. Рассмотрены основные методы машинного обучения применительно к задаче классификации жестов. Произведена начальная обработка данных, отбор признаков и выполнено сравнение алгоритмов с приведением качества распознавания. Ил.: 4. Табл.: 1. Библиогр.: 11 назв.

Ключевые слова: распознавание жестов; трехосевые акселерометры; мобильные устройства; классификация.

UDC 004.94

Gesture recognition methods based on three-axis Android accelerometers readings / Yerokhin A.L., Lednov S.M. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2017. – №. 21 (1243). – P. 54 – 64.

The article is devoted to gesture recognition methods analysis based on the readings of three-axis accelerometers of mobile devices. The basic methods of machine learning applied to the problem of classification of gestures. Conducted initial data processing, feature selection and comparison of algorithms providing recognition quality results. Figs.: 4. Tab.: 1. Refs.: 11 titles.

Keywords: gesture recognition; three-axis accelerometers; mobile devices; classification.