

И. Ю. НЕБОГАТИКОВ, И. П. СОЛОВЬЕВ
Санкт-Петербургский государственный университет, Санкт-Петербург

РАСПОЗНАВАНИЕ АКТИВНОСТИ ЧЕЛОВЕКА ПО НОСИМЫМ ДАТЧИКАМ В ЗАДАЧЕ УПРАВЛЕНИЯ УМНЫМ ДОМОМ

В данной работе проводятся анализ и сравнение методов машинного обучения, используемых для распознавания активности человека в контексте умного дома. Данные получены из оптического датчика сердцебиения и акселерометра, встроенных в умные часы. Делается вывод о наибольшей универсальности метода случайного леса для задач данной категории.

Введение. Проблематика «умного дома» включает в себя много различных задач, диапазон которых не ограничивается мониторингом и управлением бытовыми приборами, но подразумевает также создание достаточно сложных стратегий поддержки различных типов обитателей этого дома, опирающихся все же как на мониторинг приборов, так и на мониторинг активности обитателя, в частности, использование технологии мониторинга активности человека по носимым датчикам. Это могут быть довольно сложные системы с электрокардиографами, пульсоксиметрами и другими приборами, или одно «умное» устройство, например, часы или фитнес-браслет, оснащенное оптическим датчиком пульса и (или) акселерометром. Последние устройства являются наиболее популярными из-за относительно меньшей стоимости. Основное их применение – отслеживание физической активности. Постоянный мониторинг различных параметров позволяет делать предположения о текущем виде деятельности человека в предположении существования известных шаблонов поведения. В системах умного дома такой подход можно применять в сценариях управления устройствами на основе текущего вида деятельности. Выделяют два основных подхода к распознаванию шаблонов поведения: классификация, распознавание некоторого заранее заданного набора действий, и кластеризация, автоматическое извлечение шаблонов [1].

Предлагаемый доклад посвящен анализу и сравнению методов машинного обучения, используемых для распознавания шаблона поведения человека в задаче управления умным домом, на основе данных, находящихся в открытом доступе. В данной работе рассматривается задача классификации, поскольку для реализации процесса автоматизации умного дома предполагается использование заранее заданных видов деятельности обитателя. Для упрощения рассуждений мы здесь рассматриваем модель, связанную с активностью одного человека внутри квартиры.

Описание данных. Для анализа используются два набора данных.

- Данные сердцебиения и дыхания. Выборка состоит из 5 классов активности: бег, шаг, нет активности, подъем по лестнице, спуск по лестнице. Всего выборка содержит 128694 элемента (87235 без учета спуска/подъема по лестнице). Данные не размечены временными метками, поэтому нельзя положиться на их упорядоченность и применить методы [2].
- Открытые данные об активности человека в течении трех дней. Набор содержит 4742 записи, разделённых на 10 видов деятельности: приготовление еды, еда, в компьютере, бег, сидит, сон, тренировка, видео игры, шаг, просмотр ТВ [3].

Описание методов. Для классификации были выбраны следующие алгоритмы и их параметры:

- нейронная сеть (Решатель: Adam) [4];
- метод K ближайших соседей (K = 10) [5];
- случайный лес (100 деревьев) [6];
- байесовский классификатор [7];
- AdaBoost (стандартный метод: решающие деревья) [8];
- метод опорных векторов [9].

Основные метрики измерения качества:

- Точность классификации (ассигасу): отношение правильно предсказанных значений к общему размеру проверочной выборки;

- Точность для класса (precision) – отношение правильно определенных значений класса к сумме правильно и ложноположительно определенных значений;
- ROC AUC (area under receiver operating characteristic curve) – площадь подграфика кривой ошибок.

Анализ результатов. Каждый набор данных был разбит на две части: обучающая и тестовая выборки в отношении 85 % и 15 %.

Для набора данных сердцебиения использовались алгоритмы случайного леса и метода опорных векторов как наиболее эффективные в решении похожих задач [10, 11].

Значения точности классификации (ассигасы) составили 0.438487 для случайного леса и 0.386946 для метода опорных векторов. Значения точности для класса (precision) и площадь подграфика (AUC) представлены в табл. 1.

Таблица 1

Точность распознавания с использованием датчика пульса (5 классов)

Вид активности	Случайный лес		Метод опорных векторов	
	Precision	AUC	Precision	AUC
Вниз по лестнице	0,61919	0,564738	0,160682	0,609572
Нет активности	0,487542	0,926043	0,268873	0,930411
Бег	0,599251	0,771685	0,639908	0,789139
Вверх по лестнице	0,243863	0,595384	0,265470	0,614261
Шаг	0,370785	0,624040	0,392319	0,636381

Значения AUC для метода опорных векторов превосходят соответствующие значения для случайного леса, поскольку считается во время построения классификатора на тренировочных данных, которых значительно больше, чем тестируемых. Это показывает достаточную разделяемость данных для небольшого количества классов. Точность (precision) значительно ниже в случае отсутствия активности для метода опорных векторов.

Для увеличения различий между классами и уменьшения количества классов были проанализированы данные без учета подъема/спуска по лестнице. В данном случае значения точности классификации составили 0,666590 для случайного леса и 0,614473 для метода опорных векторов, что в 1,5–1,6 раз выше, чем с учетом пяти классов. Метрики показаны в табл. 2.

Таблица 2

Точность распознавания с использованием датчика пульса (3 класса)

Вид активности	Случайный лес		Метод опорных векторов	
	Precision	AUC	Precision	AUC
Нет активности	0,616322	0,923518	0,386822	0,931720
Бег	0,754862	0,802214	0,805437	0,816909
Шаг	0,612167	0,724526	0,633569	0,730616

Как и в предыдущем случае, точность распознавания отсутствия активности значительно ниже для метода опорных векторов, поскольку отличие значений пульса для ходьбы и отсутствия активности меньше, чем отличие бега от ходьбы и отсутствия активности. Алгоритм случайного леса показывает большую чувствительность к различиям признаков, что уменьшает разброс в точности для различных типов активности.

Второй набор данных обрабатывался с использованием всех описанных методов классификации. Результаты точности классификации показаны в табл. 3.

Из-за значительно меньшего количества данных, и как следствия большего влияния выбросов данных в выборках на обучение, при использовании только одного датчика методы К ближайших соседей, байесовский классификатор, AdaBoost и нейронная сеть показывают большую точность классификации. Однако, только алгоритм случайного леса показывает значительный рост точности при рассмотрении нескольких датчиков и усложнения модели данных. Точность классификации при использовании метода опорных векторов с использованием обоих датчиков увеличивается незначительно из-за усложнения структуры данных.

Точность распознавания с различными вариантами использования

Метод	Точность (датчик сердцебиения)	Точность (акселерометр)	Точность (оба датчика)
Нейронная сеть	0,286517	0,351124	0,369382
К ближайших соседей	0,237360	0,419944	0,433989
Случайный лес	0,147472	0,428371	0,501404
Байесовский классификатор	0,290730	0,271067	0,300562
AdaBoost	0,308989	0,275281	0,289326
Метод опорных векторов	0,116573	0,137640	0,143258

Для случая использования обоих датчиков при использовании алгоритма случайного леса была вычислена важность признаков (сколько раз каждый признак использовался в деревьях для определения перехода [12]):

- 0,479728 для датчика пульса;
- 0,520272 для акселерометра.

Различие в значениях незначительно и зависит от выбора тренировочной выборки, поэтому можно считать, что важность признаков одинакова.

Заключение. В представленной работе рассмотрены методы машинного обучения, используемые для распознавания шаблона поведения человека в задаче управления умным домом. Решалась задача классификации данных на примере модели активности одного человека внутри квартиры. Анализ результатов позволяет сделать вывод о том, что использование только одного датчика для распознавания активности человека возможно только при небольшом количестве классов и сильно различающихся видов деятельности, например, сон, умеренная активность или бег. Для распознавания большего числа классов, порядка 10, необходимо использовать данные сердцебиения и акселерометра.

В продолжении исследования предлагается выполнить сбор достаточного объема собственных данных для получения возможности их фильтрации и применении к ним метода случайного леса для классификации, как показавшего себя в качестве наиболее универсального алгоритма для задач рассмотренного типа.

ЛИТЕРАТУРА

1. Кириенко А.С., Соловьев И.П. Анализ активности человека в задаче автоматизированного управления умным домом // Компьютерные инструменты в образовании, 2017. № 1. С. 15-29.
2. URL: https://www.kaggle.com/onnsbrd/heart-rate-and-breathing-data?select=final_ecg_data.csv (дата обращения: 28.07.2020).
3. URL: <https://www.kaggle.com/sasanj/human-activity-smart-devices> (дата обращения: 28.07.2020).
4. Diederik P. Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization. ICLR. 2015.
5. Pdraig Cunningham, Sarah Jane Delany. k-Nearest neighbour classifiers. 2017.
6. Thais Mayumi Oshiro, Pedro Santoro Perez, Jos'e Augusto Baranauskas. How Many Trees in a Random Forest? 8th International Conference on Machine Learning and Data Mining in Pattern Recognition. 2012.
7. Irina Rish. An empirical study of the naive Bayes classifier. 2001.
8. Yoav Freund Robert E. Schapire. A Short Introduction to Boosting. Journal of Japanese Society for Artificial Intelligence. 1999. P. 771-780.
9. Theodoros Evgeniou, Massimiliano Pontil. Support Vector Machines: Theory and Applications. Machine Learning and Its Applications, Advanced Lectures. 2001.
10. Saeed Mehrang, Julia Pietila, Johanna Tolonen, Elina Helander, Holly Jimison, Misha Pavel, Ilkka Korhonen. Human Activity Recognition Using A Single Optical Heart Rate Monitoring Wristband Equipped with Triaxial Accelerometer. EMBEC & NBC 2017. 2017. P. 587-590.
11. Tao Tang, Lingxiang Zheng, Shaolin Weng, Ao Peng and Huiru Zheng. Human Activity Recognition with Smart Watch based on H-SVM. International Conference on Frontier Computing. 2018. P. 179-186.
12. Bjoern H Menze, B Michael Kelm, Ralf Masuch, Uwe Himmelreich, Peter Bachert, Wolfgang Petrich, and Fred A Hamprecht. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics. 2009.

I. Y. Nebogatikov, I. P. Soloviev (Saint Petersburg State University, Saint Petersburg)

Human Activity Recognition by Wearable Sensors in the Smart Home Control Problem

In this work we analyze and compare machine learning methods of recognizing human activity in the context of smart home by using data obtained from an optical heartbeat sensor and an accelerometer embedded in a smart watch. The conclusion is made that for problems of such type the random forest method is the most universal.