

Сравнение эффективности различных методов биометрической идентификации личности с помощью электрокардиограмм

Научный руководитель – Богданов Марат Робертович

Мухаметьянов Руслан Альбертович

Студент (магистр)

Уфимский государственный авиационный технический университет, Уфа, Россия

E-mail: ffewcher@gmail.com

Актуальность исследования обусловлена возросшей популярностью различных методов биометрической идентификации личности. В работе сравниваются три метода классификации, широко применяемые в машинном обучении: смешанные гауссовские модели, случайные леса и метод опорных векторов.

Метод опорных векторов (Support vector machine - SVM) — это метод машинного обучения, целью которого является попытка классифицировать входные наборы данных в один из двух классов. Для эффективной работы метода сначала необходимо использовать обучающую выборку, состоящую из входных и выходных данных, которая необходима для построения модели метода опорных векторов, и которую в дальнейшем можно будет использовать для классификации новых данных.

Случайный лес (Random forest) — алгоритм машинного обучения, предложенный Лео Брейманом и Адель Катлер, заключающийся в использовании комитета (ансамбля) решающих деревьев.

Метод основан на построении большого числа (ансамбля) деревьев решений (это число является параметром метода), каждое из которых строится по выборке, получаемой из исходной обучающей выборки с помощью бутстрепа (т. е. выборки с возвращением).

Смешанная модель Гаусса (Gaussian mixture models - GMM) является методом статического моделирования показывающая нормально распределенные субпопуляции (совокупность) в общей субпопуляции. Смешанные модели в целом не требуют знания в какой совокупности относится точка данных, позволяя модели изучать субпопуляции автоматически. Поскольку назначение субпопуляции неизвестно представляет, она собой форму неконтролируемого обучения.

Мы использовали базу данных оцифрованных ЭКГ РТВ с сайта <http://www.physionet.org/>. Сравнивались электрокардиограммы 51 здорового испытуемого и 51 представителя смешанной группы, в которую входили люди с различными нарушениями сердечно-сосудистой системы. На первом этапе была проведена предварительная обработка электрокардиосигналов и отбор признаков (амплитудные и временные характеристики P, Q, S, и T-областей кардиоцикла). Результаты экспериментов приведены в таблице ниже.

Таблица 1.

Сравнение эффективности различных методов распознавания (коэффициенты распознавания)

Испытуемые

Смешанные гауссовские модели

Случайные леса

Метод опорных векторов

Здоровые

0.8517

1.0

0.99

Смешанная группа

0.9413

1.0

0.96

Из таблицы видно, что наилучшие результаты дал метод случайных лесов, остальные методы дали несколько худшие результаты. Наличие сердечно-сосудистых заболеваний существенного влияния на распознавание не оказало. Кластеризация методом k-ближайших соседей оказалась очень нестабильной.

Источники и литература

- 1) Fratini A., Sansone M., Bifulco P and Cesarelli M. Individual identification via electrocardiogram analysis. BioMed Eng OnLine (2015) 14:78 DOI 10.1186/s12938-015-0072-y
- 2) Carreiras C, Lourenço A, Fred A, Ferreira R. ECG signals for biometric applications: are we there yet? In: ICINCO. 2014—Proceedings of the 11th international conference on informatics in control, automation and robotics. 2014. p. 765–72.
- 3) Shen T-W, Tompkins WJ, Hu YH. Implementation of a one-lead ECG human identification system on a normal population. J Eng Comput Innov. 2011;2(1):12–21.
- 4) Israel SA, Irvine JM, Cheng A, Wiederhold MD, Wiederhold BK. ECG to identify individuals. Pattern Recognit. 2005;38(1):133–42.
- 5) Tantawi M, Salem A, Tolba MF. Fiducial based approach to ECG biometrics using limited fiducial points. Commun Comput Inf Sci. 2014. p. 199–210.