

Анализ методов отбора признаков для задачи классификации когнитивных состояний и состояний покоя по данным ЭЭГ

М.Ш. Муртазина

Новосибирский государственный технический университет
Новосибирск, Россия
murtazina@corp.nstu.ru

Т.В. Авдеевко

Новосибирский государственный технический университет
Новосибирск, Россия
avdeenko@corp.nstu.ru

Аннотация—В работе исследуются вопросы применения методов отбора признаков, входящих в пакет программ Weka, для решения задачи классификации когнитивных состояний и состояний покоя по данным ЭЭГ. Для проведения эксперимента использован набор данных «EEG During Mental Arithmetic Tasks». Из записей ЭЭГ для 19 скальповых электродов извлечены параметры Хьорта и мощность в дельта-, тета-, альфа- и бета-частотных диапазонах. На первом этапе эксперимента на полученном признаковом пространстве выполнено обучение классифицирующих моделей, на втором – выполнено исследование изменения точности классификаторов, обученных на том же наборе алгоритмов на подмножестве признаков, полученных при помощи методов отбора признаков. Для получения подмножества признаков использованы оценщики атрибутов CfsSubsetEval и WrapperSubsetEval.

Ключевые слова— отбор признаков, классификация, состояние покоя, когнитивные состояния, ЭЭГ, Weka.

1. ВВЕДЕНИЕ

В последние годы все больше исследований посвящено вопросам применения данных электроэнцефалографии (ЭЭГ) для классификации когнитивных состояний и состояний покоя на основе признаков ЭЭГ [1,2,3]. Отбор признаков является важным шагом в классификации умственных состояний по данным ЭЭГ. На практике наиболее значимые для задачи классификации умственных состояний по данным ЭЭГ признаки априори неизвестны. По этой причине в признаковое пространство может быть включено большое количество признаков, извлеченных из сигналов ЭЭГ от каждого скальпового электрода. При этом количество электродов определяется целями исследования и используемым оборудованием.

Методы отбора признаков эффективно используются при подготовке данных для интеллектуального анализа и машинного обучения. Целью процедуры отбора признаков является поиск подмножества признаков, которые являются значимыми для прогноза. Отбор признаков считается успешным, если размерность признакового пространства уменьшена, а точность классификатора стала выше или не изменилась.

Цель данного исследования – провести анализ методов отбора признаков, предоставляемых пакетом программ Weka, которые могут быть использованы для уменьшения признакового пространства ЭЭГ высокой размерности.

Работа организована следующим образом. В разделе 1 обосновывается актуальность темы исследования. В

разделе 2 дается описание материалов и методов исследования. В разделе 3 представлены результаты экспериментальной части исследования. В разделе 4 подводятся итоги работы.

2. МАТЕРИАЛЫ И МЕТОДЫ

А. Формирование признаков

Для проведения эксперимента использован набор данных «EEG During Mental Arithmetic Tasks», который включает записи ЭЭГ в состоянии покоя (три минуты) и выполнения арифметических вычислений (одна минута) для 36 субъектов [4]. Файлы с записями электроэнцефалограмм представлены в наборе данных в формате EDF.

Для извлечения признаков использованы данные 19 скальповых электродов в соответствии с международной схемой 10-20. Из записей ЭЭГ для каждого электрода извлечены параметры Хьорта (активность, мобильность и сложность) и мощность в дельта-, тета-, альфа- и бета-частотных диапазонах. Оценка спектральной плотности мощности (PSD) произведена методом Уелча. Размер окна для извлечения признаков составил одну минуту. Из исходного набора были использованы записи ЭЭГ для 34 субъектов, поскольку длины записей ЭЭГ в состоянии покоя у двух испытуемых были менее трех минут. Для каждого субъекта было получено по четыре экземпляра данных. Три из них соответствуют состоянию покоя с закрытыми глазами, один – состоянию, связанному с когнитивной нагрузкой, обусловленной выполнением арифметических вычислений.

Для извлечения признаков из записей ЭЭГ использована Python-библиотека eeglib, которая предоставляет инструменты, позволяющие пользователю легко создать целый набор данных, совместимых с большинством Python-библиотек анализа данных. Эта возможность обеспечивается созданием объекта DataFrame с извлекаемыми признаками [5]. Для автоматизации процесса формирования набора данных в формате ARFF, используемом пакетом программ Weka, был создан Python-скрипт.

Б. Методы отбора признаков в Weka

Для отбора признаков и обучения классификаторов использовано свободное программное обеспечение для анализа данных и машинного обучения Weka. Данный программный продукт может работать на любой платформе, поскольку является java-приложением.

Процедура отбора признаков в Weka предполагает совместное использование оценщика атрибутов и метода

поиска. Оценщик атрибутов – это метод, с помощью которого каждый признак оценивается в контексте целевой переменной. Метод поиска – это метод, с помощью которого просматривается признаковое пространство для поиска подходящего подмножества признаков.

В рамках данного исследования используются оценщики подмножества атрибутов CfsSubsetEval и WrapperSubsetEval. Первый оценщик атрибутов оценивает значимость подмножества признаков, учитывая индивидуальную прогностическую способность каждого признака, а также степень избыточности между ними. Второй – оцениваем эффективность подмножества признаков, учитывая результат примененного алгоритма классификации для оценки набора признаков. В поисковой стратегии может использоваться алгоритм, отличный от того, что будет затем использован для обучения классифицирующей модели.

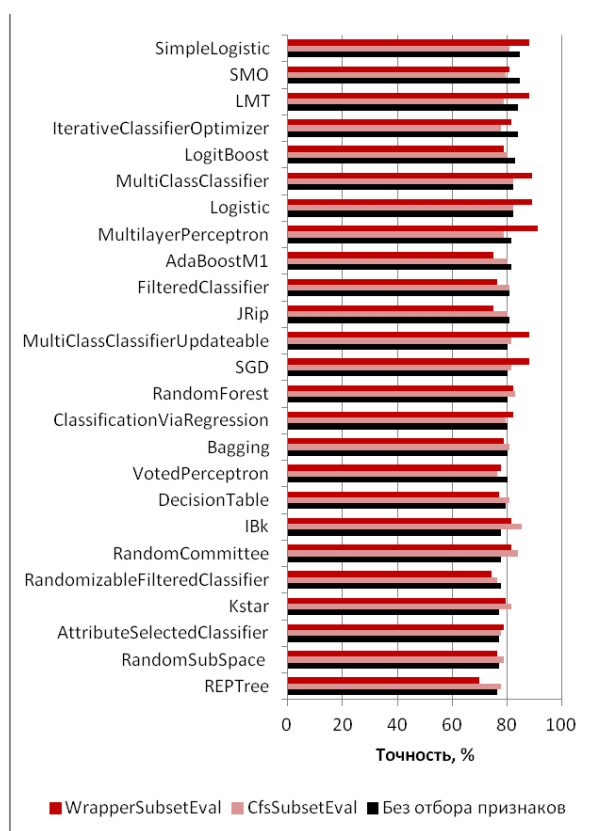


Рис. 1. Точность классификаторов

3. РЕЗУЛЬТАТЫ

В рамках эксперимента было отобрано 25 алгоритмов машинного обучения, входящих в пакет программ Weka, для которых точность классификации превысила 75 % при обучении на исходном признаковом пространстве (133 признака). Далее был произведен отбор признаков при помощи двух методов. Первый использует комбинацию оценщика атрибутов CfsSubsetEval и метода поиска GreedyStepwise, второй – комбинацию оценщика атрибутов WrapperSubsetEval с классификатором Logistic и метода поиска GreedyStepwise. На полученных подмножествах признаков вновь выполнено обучение классификаторов. Результаты показаны на Рис. 1.

При использовании оценщика атрибутов CfsSubsetEval было получено следующее подмножество признаков: активность Хьорта для канала O1, сложность Хьорта для канала F8, мощность дельта диапазона для канала Fp1, мощность тета диапазона для канала Fp1, мощность дельта диапазона для канала F8, мощность бета диапазона для канала C3 и мощность альфа диапазона для канала P4.

При использовании оценщика атрибутов WrapperSubsetEval были отобраны: активность Хьорта для канала T4, активность Хьорта для канала C4, сложность Хьорта для канала F8, мобильность Хьорта для канала F8, мобильность Хьорта для канала P4, мощность дельта диапазона для канала Fp1, мощность дельта диапазона для канала Fp2, мощность дельта диапазона для канала F8, мощность альфа диапазона для канала T6.

Отбор признаков первым методом оказался успешным для 12 алгоритмов классификации, вторым – для 13 алгоритмов. При этом наилучший результат первым методом был получен для алгоритма IBk (точность классификатора изменилась с 77,94 % до 85,29%). Наилучший результат вторым методом был получен для алгоритма MultilayerPerceptron (точность классификатора изменилась с 81,61% до 91,17 %).

4. ЗАКЛЮЧЕНИЕ

В данной работе исследованы входящие в пакет программ Weka методы отбора признаков, использующие оценщики подмножества атрибутов для сокращения размерности признакового пространства. В качестве примера для анализа выбрана задача классификации когнитивных состояний и состояний покоя по данным ЭЭГ. Дальнейшие исследования целесообразно связать с поиском оптимальных настроек для оценщиков подмножества атрибутов, учитывающих алгоритма классификации при оценке значимости признаков.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Министерства Науки и Высшего Образования Российской Федерации в рамках Госзадания (проект № FSUN-2020-0009).

ЛИТЕРАТУРА

- [1] Ahmad, R.F. Classification of cognitive and resting states of the brain using EEG features / R.F. Ahmad, A.S. Malik, H.U. Amin, N. Kamel, F. Reza // IEEE International Symposium on Medical Measurements and Applications (MeMeA). – 2016. – P. 1-5.
- [2] Mazher, M. Classification of resting and cognitive states using EEG-based feature extraction and connectivity approach / M. Mazher, I. Faye, A. Qayyum, A. S. Malik // IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES). – 2018. – P.184-188.
- [3] Choi, S. Classification of mental arithmetic and resting-state based on Ear-EEG / S. Choi, G. Choi, H. Lee, H. Hwang, J. Shin // 6th International Conference on Brain-Computer Interface (BCI). – 2018. – P. 1-4.
- [4] Zyma, I. Electroencephalograms during mental arithmetic task performance / I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, O. Shpenkov // Data. – 2019. – Vol. 4(1). – P.14.
- [5] Cabañero-Gomez, L. Eeglib: A Python module for EEG feature extraction / L. Cabañero-Gomez, R. Hervás, I. González, L. Rodríguez-Benitez // SoftwareX. – 2021. – Vol. 15. – P. 100745.