

УДК 004.85

МЕТОДИКА ВЫБОРА ВХОДНЫХ ПРИЗНАКОВ ДЛЯ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

© 2022 г. В. В. Тагунов¹, К. Я. Кудрявцев¹, А. И. Петрова¹, Т. И. Возненко^{1,*}

¹ Институт интеллектуальных кибернетических систем,
Национальный исследовательский ядерный университет “МИФИ”, Москва, 115409, Россия

*e-mail: TIVoznenko@mephi.ru

Поступила в редакцию 28.04.2022 г.

После доработки 30.04.2022 г.

Принята к публикации 04.05.2022 г.

В данной работе описана методика выбора (отбора) признаков для обучения различных алгоритмов машинного обучения. Методика основана на известных методах отбора признаков, может быть использована в качестве обработки данных для решения задачи классификации с помощью алгоритмов машинного обучения. Методика состоит из нескольких этапов: вычисление оценки каждого признака с помощью существующего метода перетасовки признаков (Feature shuffling) на основе ряда метрик качества (scoring parameters) модели машинного обучения; обработка собранного массива данных для разделения на два класса (релевантные и нерелевантные признаки) с помощью алгоритма кластеризации K-средних; удаление нерелевантных признаков из общего набора данных для дальнейшего использования в обучении SVM классификатора; оценка точности классификации алгоритма. Особенность методики заключается в применении сразу нескольких метрик качества для улучшения показателя точности и гибкости модели, а также в использовании ансамбля алгоритмов машинного обучения для отбора лучших признаков. В рамках исследования был проведен ряд экспериментов для получения результатов эффективности методики. В качестве набора входных данных для классификатора использовались показания электромиографического (ЭМГ) сигнала мышечной активности, собранных специализированным датчиком, где каждый набор данных соответствует отдельному жесту (классу). В ходе обработки из сигнала был выделен и отобран с помощью разработанной методики ряд признаков для составления входного набора данных для дальнейшего обучения SVM классификатора. Обученная модель была использована для интерпретации жестов в команды управления роботизированного устройства в реальном времени. Применение методики обеспечило более высокую точность распознавания жестов по сравнению с методами, которые используют одну метрику качества модели машинного обучения для отбора признаков.

Ключевые слова: метрики качества модели, ЭМГ, SVM, алгоритм K-средних, признак, метод выбора признаков, модель машинного обучения, алгоритм машинного обучения

DOI: 10.56304/S2304487X22010114

ВВЕДЕНИЕ

В настоящее время различные алгоритмы машинного обучения находят все большее применение в различных сферах деятельности. Жизненный цикл модели машинного обучения – это сложный процесс, который требует прохождения множества этапов таких как:

- бизнес-анализ предметной области, в рамках которой будет выполняться задача;
- анализ и подготовка данных;
 - о анализ и сбор данных;
 - о нормализация;
 - о моделирование данных;
 - о выбор и конструирование признаков;
- моделирование;

- о выбор алгоритма;
- о обучение модели;
- о оценка результатов;

– внедрение и сопровождение.

В данной работе будет рассмотрен именно этап выбора признаков. Данный этап является одним из самых важных, так как выбор признаков напрямую влияет на длительность времени обучения, требования к вычислительным ресурсам, а также на точность результатов обучения модели. Существует множество различных методов выбора признаков, которые используют специальные метрики качества модели для оценки важности признаков. Как правило такие методы используют лишь одну метрику, но при таких условиях возможна ситуация, когда метод удалит признак,

который позже может оказаться очень важен при дальнейшей работе модели машинного обучения с новыми данными. Цель данной работы заключается в разработке такой методики выбора признаков, которая будет учитывать особенности множества метрик качества модели для того, чтобы получить максимально универсальный набор признаков с учетом улучшения гибкости модели и точностью результатов прогнозирования.

В настоящей работе применение и тестирование предложенной методики осуществляется в рамках задачи распознавания ЭМГ-сигнала и классификации жестов. Данная задача является актуальной в рамках реализации человеко-машинного интерфейса на основе ЭМГ, в частности жестового, и в ее решении применяются в том числе различные алгоритмы машинного обучения [1]. Для практического тестирования разработанный программный модуль был интегрирован в ПО управления роботизированным устройством с помощью ЭМГ, осуществляющее сбор и обработку ЭМГ данных и их интерпретацию в команды для роботизированного устройства. В рамках эксперимента для классификации ЭМГ паттернов в реальном времени был использован SVM – классификатор на основе радиальной базисной функции (RBF) [1], обученный при помощи предлагаемой методики. Похожая разработка была реализована в ряде работ: [2–4], где также было разработано ПО для распознавания жестов с помощью ЭМГ на основе алгоритма SVM.

1. СУЩЕСТВУЮЩИЕ МЕТОДЫ ВЫБОРА ПРИЗНАКОВ

Существует множество различных методов выбора признаков (Feature selection methods) [5]. Основные алгоритмы можно разделить на три категории, рассмотренные далее.

Метод фильтрации (Filter method) статистически оценивает взаимосвязи между каждой входной переменной и целевой переменной, и эти оценки используются в качестве основы для выбора (фильтрации) тех входных переменных, которые будут использоваться в модели. Данные методы оценивают релевантность прогнозов за пределами прогностических моделей и впоследствии моделируют только те прогнозы, которые удовлетворяют некоторому критерию.

Метод обертки (Wrapper method) создает множество моделей с различными подмножествами входных параметров и выбирает те признаки, которые приводят к наилучшей производительности модели в соответствии с метрикой качества модели.

Встроенные методы (Embedded methods). Некоторые модели машинного обучения содержат встроенный выбор признаков, где выбираются

только те прогнозы, которые помогают максимизировать точность.

Вышеперечисленные методы имеют ряд недостатков.

Методы обертки вычисляют каждую возможную комбинацию признаков на основе шага вперед, шага назад или полного перебора. Для каждой комбинации признаков эти методы будут обучать отдельную модель машинного обучения, обычно с перекрестной проверкой, и определять ее производительность. Таким образом, данные методы очень затратны в вычислительном отношении и часто невозможны для выполнения при большом наборе признаков.

Методы фильтрации используют статистические тесты, которые выбирают функции на основе их распределения. Эти методы очень быстрые в вычислительном отношении, но на практике они не дают хороших характеристик для моделей решающих задачу классификации.

Существуют также гибридные методы, которые совмещают в себе общие характеристики встроенных и оберточных методов. Одним из таких методов является перетасовка признаков (Feature shuffling) [6]. Перетасовка признаков состоит из присвоения оценки признаку на основе снижения показателя качества модели, когда значения одного признака перетасовываются случайным образом. Метод работает следующим образом:

- 1) производится обучение модели машинного обучения и определяется ее уровень качества;
- 2) метод сдвигает порядок (столбец) значений одного признака в общем наборе данных;
- 3) делаются прогнозы с моделью, обученной на шаге 1, и определяется производительность;
- 4) если производительность падает ниже порогового значения, признак сохраняется, в противном случае удаляется;
- 5) цикл повторяется с шага 2 до тех пор, пока не будут проверены все признаки.

Данный метод имеет несколько преимуществ. Во-первых, здесь нужно обучить только одну модель машинного обучения. Впоследствии важность признака назначается путем перетасовки значений признаков и прогнозирования с помощью этой модели. Во-вторых, можно выбрать признаки для любой модели машинного обучения. Именно поэтому для разработки методики был выбран метод перетасовки признаков.

2. ТЕОРИЯ

Все перечисленные ранее методы выбирают признаки на основе значения метрики качества модели (scoring parameter). Можно добиться наибольшей точности результатов прогнозирования

Таблица 1. Метрики качества модели

Название метрики	Определение	Формула
Accuracy	Показатель того, как часто модель делает правильный прогноз	$\frac{TN + TP}{TP + TN + FP + FN}$
Balanced accuracy	Позволяет избежать завышенных оценок качества признака для несбалансированных наборов данных в отличие от accuracy	$\frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$
Recall	Доля верных положительных прогнозов от суммы количества неверных отрицательных прогнозов и верных положительных	$\frac{TP}{TP + FN}$
Precision	Доля верных положительных прогнозов от суммы количества неверных положительных прогнозов и верных положительных	$\frac{TP}{TP + FP}$
F1	Среднее гармоническое значение между метриками recall и precision	$2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$

Здесь, TN = True Negative; FP = False Positive; FN = False Negative; TP = True Positive.

модели, выбрав одну конкретную метрику качества. Точность прогнозирования, без сомнения, является важным показателем оценки модели, но рекомендуется также принимать во внимание ряд других метрик для оценки гибкости модели относительно вариации и распределения тестовых данных. Используемые метрики представлены в таблице 1 [7].

В качестве доказательства того, почему необходимо учитывать несколько метрик качества для улучшения гибкости модели, рассмотрим некоторые из них.

Метрика точности (accuracy) говорит о том, как часто модель делает правильный прогноз. Эта метрика чаще всего используется для определения признаков и во многих методах является параметром по умолчанию.

Полнота или чувствительность (recall) – это доля релевантных экземпляров, которые были извлечены, по сравнению с общим количеством релевантных экземпляров. То есть, если recall = 1, то это означает, что модель правильно классифицировала все значения, когда фактическое значение было положительным (TP), но это также может быть в том случае, если модель предсказала все значения как положительные, что означало бы, что модель имеет низкую точность прогнозирования.

Следовательно, может произойти ситуация, когда результаты оценки качества одного и того же признака по разным метрикам кардинально отличаются по значениям. Обученная на отобранных признаках одной конкретной метрикой модель в дальнейшем будет использоваться для классификации новых данных, и точность прогнозирования может понижаться, если оценка

выбранной метрики не указала на один из релевантных признаков.

Таким образом, при построении модели необходимо найти оптимальный баланс между точностью и гибкостью модели, основываясь на результатах разных метрик качества. Цель эксперимента – доказать, что при использовании набора метрик увеличится точность прогнозирования модели на разных наборах данных.

3. ОПИСАНИЕ МЕТОДИКИ

Методика состоит из нескольких этапов:

1) на вход алгоритма подаются обработанные наборы данных, состоящие из ряда признаков. Каждый отдельный набор данных имеет принадлежность к отдельному классу и в итоге для каждого набора присваивается метка его класса (значения от 0 и далее);

2) далее производится вызов метода Feature shuffling, на вход которого подается экземпляр SVM-модели, метрика качества модели (scoring parameter) и значение кросс-валидации (cross validation parameter). На выходе метод возвращает результаты оценки качества для каждого признака;

о Вызов Feature shuffling выполняется в цикле для каждой метрики качества (таблица 1), где в каждой итерации результаты склеиваются в один двумерный массив данных значений по каждой метрике и признаку (таблица 2);

о Оценка качества признака по метрике вычисляется как разность между оценкой признака в изначальном наборе данных и оценкой после сдвига столбца признака;

Таблица 2. Структура результирующей таблицы оценки качества метрик

Название признака	Оценка качества по метрике					
	accuracy	balanced accuracy	f1	precision	recall	...
rms	-0.016230	-0.015823	-0.012484	-0.024969	-0.020807	...
wl	-0.021223	-0.007602	-0.013317	-0.018310	-0.023304	...
iemg	-0.023720	-0.027458	-0.025385	-0.020391	-0.018727	...
mav2	-0.019975	-0.024480	-0.017894	-0.017478	-0.014149	...
var	-0.012901	-0.008585	-0.016230	-0.015397	-0.020391	...
...

3) полученный набор данных подается на вход алгоритма К-средних (K-means) [7];

о Данный алгоритм вычисляет функцию для разделения данных на два класса (релевантные и нерелевантные признаки), таким образом, обобщая результаты всех метрик качества;

о В итоге функция возвращает список признаков с худшими показателями качества, и далее данные этих признаков удаляются из начального общего набора данных;

4) обновленный набор данных с отобранными признаками нормализуется и делится на наборы для обучения и тестирования. Затем данные подаются на вход функции SVM-классификатора;

о Происходит процесс обучения модели, и уже на обученной модели получается итоговый результат прогнозирования на основе тестируемых данных;

о Далее полученная модель сохраняется и может быть использована в дальнейшей работе классификации жестов в реальном времени.

4. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ И ТЕСТИРОВАНИЕ

Представленная методика была реализована в виде программного модуля на языке Python. В рамках программного модуля были использованы готовые программные реализации алгоритмов SVM и K-means из библиотеки Scikit-learn [7], а также метод Feature Shuffling из библиотеки Feature-engine [6].

Для практического тестирования данный программный модуль был интегрирован в ПО управления роботизированным устройством с помощью ЭМГ, осуществляющее сбор и обработку ЭМГ-данных и их интерпретацию в команды для роботизированного устройства. С целью тестирования методики был проведен ряд экспериментов по классификации ЭМГ-паттернов в реальном

времени с помощью SVM-классификатора, обученного при помощи предлагаемой методики.

Собранная активность подается на вход модуля обработки данных, где данные усредняются относительно нуля и фильтруются фильтром Баттерворта 4-й степени (библиотека python scipy). Затем после обработки из сигнала необходимо выделить ряд признаков. Так как сигнал снимается во временном домене, были выбраны следующие признаки [7] (таблица 3). Все признаки вычислялись в рамках скользящего окна размером в 200 значений и с шагом 100 (размер и шаг окна выбран в соответствии с исследованием [4]).

Затем обработанные данные идут на вход модуля отбора признаков. Для разработки методики был выбран метод перетасовки признаков (Feature shuffling) из библиотеки Feature-engine [6] на основе языка Python. Для вычислений использовался ряд метрик качества (таблица 1).

Далее в модуль отбора признаков подается уже обработанный набор данных. Оценка важности признаков вычисляется на основе ряда метрик качества (таблица 1) с помощью метода Feature shuffling. В конце все результаты классифицируются алгоритмом К-средних, в итоге получается список нерелевантных признаков, столбцы которых удаляются из общего набора данных.

Данные, содержащие только отобранные признаки, отправляются на вход модуля SVM-классификатора, где происходит процесс обучения модели. Далее обученная модель сохраняется и может быть использована для классификации жестов оператора и интерпретирования их в команды в реальном времени.

С использованием данного ПО был проведен ряд экспериментов, в ходе которых осуществлялся сбор данных, обучение классификатора и дальнейшее тестирование качества полученной модели путем последовательного выполнения оператором соответствующих жестов. Сбор данных осуществлялся в рамках нескольких разных экспериментов – с целью обеспечения большей универсальности модели по отношению к ЭМГ-данным. Были собраны данные о четырех жестах

Таблица 3. Набор признаков во временном домене [7]

Название признака	Перевод	Сокращение
Root Mean Square	Среднеквадратичное отклонение	RMS
Waveform length	Длина сигнала	WL
Autoregression coefficients	Коэффициенты авторегрессии	auereg
Integrated EMG	Интегрированная ЭМГ	IEMG
Mean Absolute Value 2	Среднее абсолютное значение	MAV2
Variance	Дисперсия	var
Standard Deviation	Стандартное отклонение	stdev
Number of Peaks	Количество пиковых значений	NOP
Mean Number of Peaks	Среднее количество пиковых значений	MNOP
Difference Absolute Mean Value	Разница абсолютного среднего значения	DUMV
Percentile	Процентиль	per
Maximum fractal length	Максимальная длина фрактала	MFL
Skewness	Асимметрия	skew
Kurtosis	Куртозис (меры дисперсии)	kurt
Zero Crossing	Скорость пересечения нуля	ZC
Slope Sign Changes	Изменения знака наклона	SSC
Wilson Amplitude	Амплитуда Уилсона	WA

(суммарное количество экземпляров – 40000, доля экземпляров каждого класса (жеста) составляет 25%). Полученные ЭМГ-данные были разделены на обучающий и тестовый наборы в соотношении 70:30.

5. ОЦЕНКА РЕЗУЛЬТАТОВ

Был проведен ряд экспериментов для сравнения результатов точности и гибкости полученной модели после обработки и отбора признаков с помощью разработанной методики. На вход подавались данные о четырех жестах, разделенные на четыре класса соответственно (Class0, Class1, Class2, Class3). Результаты показаны в виде мат-

Таблица 4. Результаты прогнозирования на основе признаков, отобранных метрикой accuracy

	Class0	Class1	Class2	Class3
Class0	49	0	0	0
Class1	0	40	5	5
Class2	0	5	40	0
Class3	0	12	2	43

Результат прогнозирования равен 85.57%. Общее количество выбранных признаков – 9 (“wl”, “areg3”, “iemg”, “var”, “stdev”, “nop”, “mnop”, “dumv”, “wa”).

риц ошибок (Confusion matrix) и процента точности классификации модели (процент правильных прогнозов от общего количества сделанных прогнозов). В рамках экспериментов на всех этапах обучения моделей и процесса выбора параметров был использован обучающий датасет. Результирующие метрики были получены на уже обученных моделях с использованием тестового набора.

Для сравнения были получены результаты при использовании алгоритма отбора признаков на основе метрики качества accuracy (таблица 4) и при использовании признаков, отобранных с помощью разработанной методики (таблица 5).

Согласно полученным результатам, точность с метрикой accuracy ниже всего на несколько процентов по сравнению с разработанным алгоритмом. Далее модель была протестирована на новых данных, снятых в рамках отдельного эксперимента (тестовый набор), при этом использовались те же отобранные признаки. Новые данные тоже разделены на 4 класса (жеста). Результаты также приведены в виде матриц ошибок для использования алгоритма отбора признаков на основе accuracy (таблица 6), и для использования предлагаемой методики (таблица 7).

Таблица 5. Результаты прогнозирования на основе признаков, отобранных алгоритмом

	Class0	Class1	Class2	Class3
Class0	49	0	0	0
Class1	0	42	3	5
Class2	0	5	40	0
Class3	0	9	2	46

Результат прогнозирования равен 88,06%. Общее количество выбранных признаков – 13 (“rms”, “wl”, “areg2”, “areg3”, “iemg”, “ssi”, “var”, “stdev”, “nop”, “dumv”, “per”, “skew”, “wa”).

Таблица 6. Результаты прогнозирования на основе признаков новых данных, отобранных метрикой assu-gasy

	Class0	Class1	Class2	Class3
Class0	49	0	0	0
Class1	0	10	12	33
Class2	0	7	38	0
Class3	0	12	1	39

Результат прогнозирования равен 67,66%.

Таблица 7. Результаты прогнозирования на основе признаков новых данных, отобранных алгоритмом

	Class0	Class1	Class2	Class3
Class0	49	0	0	0
Class1	0	28	7	19
Class2	0	6	39	0
Class3	0	9	0	43

Результат прогнозирования равен 79,5%.

ВЫВОДЫ

Результаты эксперимента показали, что разработанная методика выбора признаков не только увеличивает точность классификации, но и лучше адаптируется к новым поступающим данным, таким образом, улучшая гибкость модели.

Преимущества методики:

- повышение точности прогнозирования для задачи классификации;
- повышение гибкости модели машинного обучения;
- снижение требований к вычислительным ресурсам при дальнейшем использовании модели за счет уменьшения количества признаков;
- универсальность, возможность использования в любой предметной области с разными видами признаков.

Недостатки методики:

- требует больших вычислительных ресурсов и времени;

– не подходит для задач с очень большим массивом данных;

– для некоторых данных повышение точности классификатора может быть незначительным.

Относительно небольшое увеличение точности объясняется тем, что все признаки образованы путем интегрирования одного и того же массива данных (собранный на основе ЭМГ-сигнала).

Таким образом, эксперимент показал, что методику лучше использовать в ситуации, когда значения признаков сильно варьируются и имеют разный характер распределения, также методика хорошо себя показала на относительно небольших объемах данных. Исследование работы методики для больших данных и данных с различным характером распределения, а также выявление детальных ограничений использования методики будет проведено в будущих работах. Также будущие исследования методики будут включать в себя рассмотрение ее работы с более широким спектром метрик качества, включая метрику MTnP [8–10]. Методика показала высокую эффективность на примере с алгоритмом SVM. В будущем планируется применить предложенную методику с моделями, использующими другие алгоритмы машинного обучения, например, искусственные нейронные сети, такие как нейронная сеть прямого распространения (feed forward) на основе радиальных базисных функций и сверточная нейронная, аналогично с исследованиями [2, 11].

СПИСОК ЛИТЕРАТУРЫ

1. *Quitadamo L.R., Cavrini F., Sberini L., Riillo F., Bianchi L., Seri S., Saggio G.* Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review. //Journal of neural engineering, 2017. V. 14. № 1. P. 011001.
2. *Будко Р.Ю., Чернов Н.Н., Будко А.Ю.* Распознавание мышечных усилий по сигналу лицевой электромиограммы в режиме реального времени // Научный вестник Новосибирского государственного технического университета, 2018. № 2. С. 59–74.
3. *Семендаров А.В.* Выбор структуры и гиперпараметров svm для эффективного решения задач классификации сигналов электромиографии //Научный взгляд в будущее, 2019. Т. 1. № 14. С. 23–33.
4. *Khokhar Z.O., Xiao Z.G., Menon C.* Surface EMG pattern recognition for real-time control of a wrist exoskeleton. // Biomedical engineering online, 2010. V. 9. № 1. P. 1–17. <https://doi.org/10.1186/1475-925X-9-41>
5. *Brownlee J.* How to choose a feature selection method for machine learning. //Machine Learning Mastery, 2019. V. 10.
6. *Soledad G.* Alternative Feature Selection Methods in Machine Learning. Available at: <https://www.kdnuggets.com/2021/12/alternative-feature-selection-methods-machine-learning.html> (accessed 04.04.2022)

7. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. Scikit-learn: Machine learning in Python // *Journal of machine Learning research*, 2011. V. 12. P. 2825–2830.
8. Voznenko T.I., Gridnev A.A., Kudryavtsev K.Y., Chepin E.V. The Decomposition Method of Multi-channel Control System Based on Extended BCI for a Robotic Wheelchair // *Advances in Intelligent Systems and Computing*, 2020. V. 948. P. 562–567. https://doi.org/10.1007/978-3-030-25719-4_73
9. Voznenko T.I., Gridnev A.A., Chepin E.V., Kudryavtsev K.Y. The command interpretation in decomposition method of multi-channel control for a robotic device // *Procedia Computer Science*, 2020. V. 169. P. 152–157. <https://doi.org/10.1016/j.procs.2020.02.127>
10. Voznenko T.I., Gridnev A.A., Chepin E.V., Kudryavtsev K.Y. Comparison Between Coordinated Control and Interpretation Methods for Multi-channel Control of a Mobile Robotic Device // *Advances in Intelligent Systems and Computing*, 2021. V. 1310. P. 558–564. https://doi.org/10.1007/978-3-030-65596-9_68
11. Kapur A., Kapur S., Maes P. AlterEgo: A personalized wearable silent speech interface // In 23rd International conference on intelligent user interfaces, 2018. P. 43–53. <https://doi.org/10.1145/3172944.3172977>

Vestnik Natsional'nogo Issledovatel'skogo Yadernogo Universiteta "MIFI", 2022, vol. 11, no. 1, pp. 51–58

Feature Selection Method for Machine Learning Algorithms

V. V. Tagunov^a, K. Y. Kudryavtsev^a, A. I. Petrova^a, and T. I. Voznenko^{a, #}

^a *Institute of Cyber Intelligence Systems, National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, 115409 Russia*

[#]*e-mail: TIVoznenko@mephi.ru*

Received April 28, 2022; revised April 30, 2022; accepted May 4, 2022

Abstract—A feature selection method for training various machine learning algorithms is described. It is based on well-known feature selection methods and can be used for data processing to solve the classification problem using machine learning algorithms. The method consists of several stages: calculation of the score of each feature using the existing feature shuffling method based on several scoring parameters of the machine learning model, processing the collected data array for division into two classes (relevant and irrelevant features) using the K-means clustering algorithm, removal of irrelevant features from the general dataset to train the SVM classifier, and assessment of the classification accuracy of the algorithm. A uniqueness of the method lies in the use of several scoring parameters at once to improve the accuracy and flexibility of the model, as well as in the use of an ensemble of machine learning algorithms to select the best features. A number of experiments have been carried out to determine the effectiveness of the method. As a set of input data for the classifier, electromyographic muscle activity signal readings have been collected by a specialized sensor, where each data set corresponds to a special gesture (class). During the processing, a number of features have been extracted from the signal and selected using the developed method to compile the input dataset for further training of the SVM classifier. The trained model has been used to interpret gestures into control commands for the robotic device in real time. The application of the technique provides a higher accuracy of gesture recognition compared to methods that involve only one scoring parameter of the machine learning model for feature selection.

Keywords: scoring parameters, EMG, SVM, K-means algorithm, feature, feature selection method, machine learning model, machine learning algorithm

DOI: 10.56304/S2304487X22010114

REFERENCES

1. Quitadamo L.R., Cavrini F., Sberini L., Riillo F., Bianchi L., Seri S., Saggio G. Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review. *Journal of neural engineering*, 2017, vol. 14, no. 1, pp. 011001.
2. Budko R.Yu., Chernov N.N., Budko, A.Yu. Raspoznavanie myshechnykh usilij po signalu licevoj elektromiogrammy v rezhime real'nogo vremeni [Recognition of muscle efforts by the signal of the facial electromyogram in real time]. *Nauchnyj vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta*, 2018, no. 2, pp. 59–74. (in Russian).
3. Semendarov A.V. Vybory struktury i giperparametrov SVM dlya effektivnogo resheniya zadach klassifikacii signalov elektromiografii [Selection of structure and

- hyper-parameters of SVM for efficient solution of the tasks of classification of electromyography signals]. *Nauchnyj vzglyad v budushchee*, 2019, vol. 1, no.14, pp. 23–33. (in Russian).
4. Khokhar Z.O., Xiao Z.G., Menon C. Surface EMG pattern recognition for real-time control of a wrist exoskeleton. *Biomedical engineering online*, 2010, vol. 9, no. 1, pp. 1–17. doi: 10.1186/1475-925X-9-41
 5. Brownlee J. How to choose a feature selection method for machine learning. *Machine Learning Mastery*, 2019, vol. 10.
 6. Soledad G. Alternative Feature Selection Methods in Machine Learning. Available at: <https://www.kdnuggets.com/2021/12/alternative-feature-selection-methods-machine-learning.html> (accessed 04.04.2022)
 7. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. Scikit-learn: Machine learning in Python. *Journal of machine Learning research*, 2011, vol. 12, pp. 2825–2830.
 8. Voznenko T.I., Gridnev A.A., Kudryavtsev K.Y., Chepin E.V. The Decomposition Method of Multi-channel Control System Based on Extended BCI for a Robotic Wheelchair. *Advances in Intelligent Systems and Computing*, 2020, vol. 948, pp. 562–567. doi: 10.1007/978-3-030-25719-4_73
 9. Voznenko T.I., Gridnev A.A., Chepin E.V., Kudryavtsev K.Y. The command interpretation in decomposition method of multi-channel control for a robotic device. *Procedia Computer Science*, 2020, vol. 169, pp. 152–157. doi: 10.1016/j.procs.2020.02.127
 10. Voznenko T.I., Gridnev A.A., Chepin E.V., Kudryavtsev K.Y. Comparison Between Coordinated Control and Interpretation Methods for Multi-channel Control of a Mobile Robotic Device. *Advances in Intelligent Systems and Computing*, 2021, vol. 1310, pp. 558–564. doi: 10.1007/978-3-030-65596-9_68
 11. Kapur A., Kapur S., Maes P. Alterego: A personalized wearable silent speech interface. *In 23rd International conference on intelligent user interfaces*, 2018, pp. 43–53. doi: 10.1145/3172944.3172977