

24. Bozhenyuk A., Gorbachev S., and Knyazeva M. Finding Fuzzy Sets of Bases and Antibases of Periodic Fuzzy Graph, *Lecture Notes in Networks and Systems (LNNS)*, 2024, Vol. 1089 (2), pp. 767-774.
25. Bozhenyuk A., Knyazeva M., Kosenko O., and Rozenberg I. Strong Connectivity Definition of Periodic Fuzzy Graph, *Lecture Notes in Networks and Systems (LNNS)*, 2023, Vol. 758, pp. 168-174.

Никашина Полина Олеговна – Южный федеральный университет; e-mail: nikashina@sfedu.ru; г. Таганрог, Россия; тел.: +79381453823; кафедра информационно-аналитических систем безопасности имени профессора Берштейна Леонида Самойловича; аспирант.

Nikashina Polina Olegovna – Southern Federal University; e-mail: nikashina@sfedu.ru; Taganrog, Russia; phone: +79381453823; the Department of Information and Analytical Security Systems named after Professor Bershtein Leonid Samoylovich; graduate student.

УДК 004.023

DOI 10.18522/2311-3103-2026-1-42-52

М.В. Пикалов

ВЛИЯНИЕ МЕТОДА ВЫБОРКИ НА ОЦЕНКУ ХАРАКТЕРИСТИК В АНАЛИЗЕ ЛАНДШАФТА ПОИСКА

Методы анализа ландшафта поиска (ELA) оценивают ландшафт приспособленности задачи в виде числовых дескрипторов, часто используемых для рекомендации оптимальных параметров алгоритма. Данное исследование изучает влияние методов и объёма выборки на аппроксимацию признаков ELA и последующую производительность моделей машинного обучения. Исследование демонстрирует, что эти аппроксимированные признаки не являются абсолютными характеристиками ландшафта, а существенно зависят от метода, использованного для генерации точек выборки. Хотя увеличение объёма выборки снижает дисперсию оценок признаков, сам выбор метода семплирования вносит значительное смещение, приводя к статистически различным значениям признаков для таких методов, как вихрь Мерсенна, латинский гиперкуб (LHS) и последовательности Фауре. Ключевой эксперимент включал предсказание параметров настраиваемой задачи W-model с использованием регрессионных моделей, обученных на признаках ELA. Результаты показали, что модели, обученные и протестированные на данных, полученных одним и тем же методом выборки, показали наилучшие результаты, что подчеркивает отсутствие совместимости между различными методами выборки. Примечательно, что квазислучайные последовательности Фауре давали наименьшую ошибку регрессии, превосходя распространённые методы, такие как равномерное случайное распределение и LHS. Более того, перекрёстная проверка выявила, что модели, особенно обученные на последовательностях Фауре, демонстрировали значительное падение производительности при тестировании на данных, полученных любым другим методом, что подтверждает: стратегия выборки накладывает специфический «отпечаток» на данные признаков. Полученные результаты ставят под сомнение стандартное использование распространённых методов выборки в ELA. Точность моделей машинного обучения для выбора и настройки алгоритмов сильно зависит от метода выборки, используемой для извлечения признаков. Следовательно, для обеспечения точности критически важно соблюдать согласованность между методами выборки, используемыми на этапах обучения модели и её практического применения. Высокая производительность последовательностей Фауре указывает на то, что подобные последовательности являются перспективным направлением для будущих исследований в области создания более надёжных и точных моделей на основе ELA.

Эволюционные алгоритмы; настройка параметров; анализ ландшафта поиска.

M.V. Pikalov

IMPACT OF SAMPLING TECHNIQUES ON EXPLORATORY LANDSCAPE ANALYSIS

Exploratory Landscape Analysis (ELA) features are numerical descriptors of a problem's fitness landscape, often used to recommend optimal algorithm parameters. This study investigates the critical impact of sampling methods and sample size on the approximation of ELA features and the subsequent performance of machine learning models. The research demonstrates that these feature approximations are not absolute characteristics of the landscape but are significantly influenced by the method used to

generate the sample points. While increasing the sample size reduces the variance of feature estimates, the choice of sampling strategy itself introduces substantial bias, leading to statistically different feature values across methods like Mersenne Twister, Latin Hypercube Sampling (LHS), and Faure sequences. The core experiment involved predicting the parameters of the tunable W-model problem using regression models trained on ELA features. The results showed that models trained and tested on data from the same sampling method performed best, highlighting a lack of interoperability between different sampling techniques. Notably, the Faure quasirandom sequences consistently yielded the lowest regression error, outperforming common methods like uniform random sampling and LHS. Furthermore, cross-sampling validation revealed that models, especially those trained on Faure sequences, suffered a significant performance drop when tested on data from any other method, confirming that the sampling strategy imparts a specific "fingerprint" on the feature data. In conclusion, the findings challenge the default use of common sampling methods in ELA. The accuracy of machine learning models for algorithm selection and configuration is highly sensitive to the sampling strategy employed for feature extraction. Therefore, ensuring consistency between the sampling methods used during model training and application is crucial. The superior performance of Faure sequences suggests that low-discrepancy sequences are a promising avenue for future research in making ELA-based models more robust and accurate.

Evolutionary algorithms; parameter tuning; landscape analysis.

Введение. Успехи методов машинного обучения (МО) в последние годы оказывают влияние практически на все научные области. Компьютерные науки – не исключение, и одна из ключевых областей применения МО – это выбор и настройка эвристик оптимизации. Как было показано в прошлом исследовании [1], анализ ландшафта приспособленности (FLA) предоставляет мощный инструмент для извлечения знаний о задаче, что позволяет строить модели машинного обучения для рекомендации параметров алгоритмов [2].

В работе [1] была успешно применена нейронная сеть, обученная на признаках ландшафта, для подбора статических параметров генетического алгоритма. Однако этот и другие подобные подходы [3], основанные на обучении с учителем, критически зависят от качества и способа получения этих самых признаков. Формальнее, признак – это отображение $f: X \subseteq R^n \rightarrow R$ в вещественное число. Такой признак может измерять, например, гладкость или мультимодальность ландшафта, наличие на нем “пиков” и “плато”.

На практике многие численные оптимизационные задачи являются задачами “чёрного ящика”, и, соответственно, не имеют явной модели и доступны только через оценку точек $x \in X$. В случаях, когда $f(x)$ задана неявно, значения признаков приходится приближать на основе набора пар $(x, f(x))$. Изучение подобных приближений лежит в основе анализа ландшафта поиска (ELA) [4]. Методы ELA успешно применялись, например, в оптимизации гиперпараметров для конкретных экземпляров задач [1] и в выборе алгоритмов [5]. При использовании ELA для задачи “чёрного ящика” в первую очередь необходимо определить, сколько точек брать и как их генерировать.

При выборе количества образцов для вычисления признаков ELA существует компромисс между точностью оценки и вычислительными затратами. Для быстрых функций рекомендуется использовать порядка $(50-100)n$ образцов, тогда как для ресурсоемких вычислений это число обычно сокращают (в некоторых случаях до $(1-10)n$) [6]. Что касается методов выборки, то здесь доминируют равномерное распределение [7] и латинский гиперкуб (LHS), хотя исследуются и альтернативы. Отдельно стоят траекторные методы [8], которые применяются, если есть возможность выбирать точки для расчёта, что выполняется не всегда.

Проведенное исследование выявило, что аппроксимации признаков, полученные разными методами выборки, не сходятся к единым значениям. Таким образом, модели машинного обучения нельзя использовать на наборе данных, собранном другим методом выборки, а при сравнении задач необходимо убедиться, что различия вызваны свойствами ландшафта, а не методом выборки. Кроме того, в работе показано, что использование последовательностей Фауре позволяет строить более качественные модели по сравнению с общепринятыми методами выборки, такими как равномерная и LHS, что подчеркивает важность исследования влияния свойств выборки на качество обучения.

Постановка задачи. Цель работы – исследовать влияние метода и объема выборки на аппроксимацию признаков анализа ландшафта поиска (ELA) и на последующую точность моделей машинного обучения. Предполагается, что метод выборки вносит в признаки ELA не только случайную ошибку, но и систематическое смещение, что ставит под сомнение объективность их значений и совместимость между разными методами семплирования.

В рамках работы проводится сравнительный анализ методов выборки (Mersenne Twister, LHS, Faure) и объемов данных ($k = n, 10n, 100n$) с целью оценки их влияния на качество регрессионных моделей, предсказывающих параметры тестовой задачи W-model, а также выполняется кросс-валидация моделей между разными методами выборки для оценки зависимости точности и надежности всей процедуры аппроксимации признаков ELA от стратегии и объема исходной выборки.

Регрессия параметров W-model. Для изучения влияния методов выборки на аппроксимацию признаков проанализировано их воздействие на качество модели в задаче регрессии параметров функции W-model. В качестве объектов исследования рассматриваются различные конфигурации параметризованной задачи W-model [9]. Каждая конфигурация задается параметрами: dummy, neutrality, epistasis, ruggedness, а также размерностью n . Для каждой конфигурации W-model используется реализация на основе битовых строк, и целью является предсказание оригинальных параметров W-model по аппроксимированным значениям признаков. Диапазоны параметров W-model и размерности соответствуют настройкам, использованным в прошлом исследовании [1] для переиспользования данных.

Для вычисления признаков каждой функции из полученного набора генерируются k случайных точек $x(1), \dots, x(k)$ и вычисляются соответствующие значения функций $f(x(1)), \dots, f(x(k))$. Полученные пары $(x(i), f(x(i)))_{i=1}^k$ обрабатываются пакетом flacco [6], который высчитывает вектор из 35 признаков. Данная процедура повторяется 2000 раз независимо с использованием одного и того же метода выборки. Таким образом для каждой функции получаются 2000 векторов признаков.

Из этого набора случайным образом отбираются 1000 векторов на функцию для обучения модели регрессора. Ее задача – по новому вектору признаков определить параметры функции W-model. Тестирование проводится на оставшихся векторах, не использованных при обучении, вычислением среднеквадратической ошибки между предсказанными и истинными значениями параметров. Процедура случайного разделения векторов на обучающую и тестовую выборки повторяется 50 раз для повышения достоверности результатов.

Для исследования влияния объема выборки эксперимент проводится при трех значениях k : $k = n, k = 10n, k = 100n$, где n – размерность соответствующей выборке функции W-model.

Конфигурации W-model охватывают широкий спектр свойств ландшафта целевой функции, что делает задачу регрессии параметров нетривиальной и показательной для оценки устойчивости признаков.

Распределения значений признаков. На рис. 1 представлено распределение аппроксимированных значений для конкретного признака, оценивающего степень соответствия задачи квадратичной модели. Результаты показаны для выборок mersenne размером $k = 10n$ и $k = 100n$.

Медианные значения признака (белые линии) демонстрируют существенные межконфигурационные различия между различными параметризациями W-model, что свидетельствует о чувствительности признака к свойствам ландшафта. При этом для каждой отдельной конфигурации наблюдается близкое соответствие медианных оценок, полученных при различных объемах выборки. Однако дисперсия аппроксимации признака значительно снижается с увеличением объема выборки: распределения для $k = 100n$ характеризуются существенно меньшим разбросом по сравнению с $k = 10n$.

Полученные результаты согласуются с данными работы [6] и подтверждают, что увеличение объема выборки повышает устойчивость оценок признаков, хотя межконфигурационные различия сохраняются независимо от объема данных.

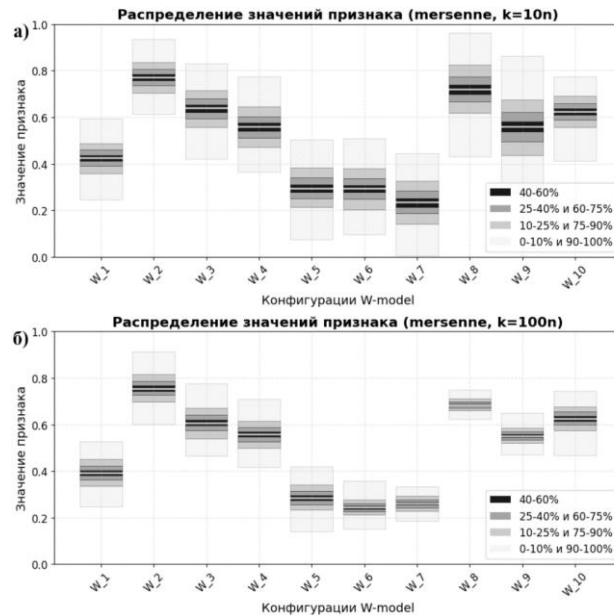


Рис. 1. Распределение значений признака *flasso* на разных конфигурациях *W-model*

Линейная регрессия и градиентный бустинг для задачи регрессии. Все эксперименты выполнены с использованием пакетов *statsmodels* и *XGBoost*. Для задачи регрессии параметров *W-model* применяются: линейная регрессия с L2-регуляризацией (Ridge) [10] и градиентный бустинг (Gradient Boosting) [11]. Поскольку не стоит задачи изучения точности самих моделей, используются стандартные параметры их реализации в соответствующих пакетах.

Использование двух различных регрессоров, линейного с регуляризацией и нелинейного ансамблевого метода, позволяет проверить, наблюдаются ли выявленные эффекты независимо от метода машинного обучения, что помогает исключить специфические артефакты отдельных алгоритмов. Дополнительно была проведена кросс-валидация с 10 запусками модели случайного леса (Random Forest) [12]. Структура основных результатов (в частности, представленных в табл. 1) осталась схожей с полученной при использовании двух основных регрессоров.

Методы выборки. В задачах извлечения признаков наиболее распространенными методами выборки являются равномерное случайное распределение и латинский гиперкуб (LHS). Для исследования зависимости устойчивости аппроксимации признаков от метода выборки, были рассмотрены пять различных методов выборки:

- ◆ *mersenne*: вихрь Мерсенна [13], являющийся стандартным генератором во многих языках программирования и информационных системах. Считается достаточно надежным решением и в данной работе рассматривается в качестве основного генератора для равномерного распределения.

- ◆ *gandu*: линейный конгруэнтный генератор псевдослучайных чисел [14]. Данный генератор считается неудачным решением и не проходит спектральный тест. В работе рассматривается для оценки влияния качества генератора на результаты задачи регрессии параметров *W-model* и, соответственно, аппроксимированные признаки ландшафта.

- ◆ *lhs*: это статистический метод для создания псевдослучайной выборки значений параметров из многомерного распределения [15]. В LHS новые точки выбираются таким образом, чтобы избежать совпадения координат с ранее отобранными точками. Диапазон каждой координаты разбивается на *k* равных интервалов. Из образованной сетки точки отбираются так, чтобы каждая одномерная проекция содержала ровно одну точку в каждом интервале.

♦ *ilhs*: оптимизация генератора LHS [16], использующая жадную эвристику для выбора точек. На каждом шаге из нескольких случайных кандидатов выбирается та точка, которая максимизирует минимальное расстояние до уже выбранных точек.

♦ *faure*: последовательности, предложенные Фауре в [17] являются развитием последовательностей Холтона [18] и известны своим равномерным покрытием пространства [19]. Данный метод основывается на перестановках цифр в разложении чисел по основанию, которым служит наименьшее простое число, не меньшее размерности пространства, что гарантирует отсутствие корреляций между проекциями.

Качество регрессии параметров W-model. На рис. 2 представлено распределение ошибки регрессии, достигнутой для каждого из пяти методов выборки, при условии использования одинаковых методов для построения и оценки регрессионных моделей. Результаты слева соответствуют модели линейной регрессии с L2-регуляризацией, справа – градиентному бустингу.

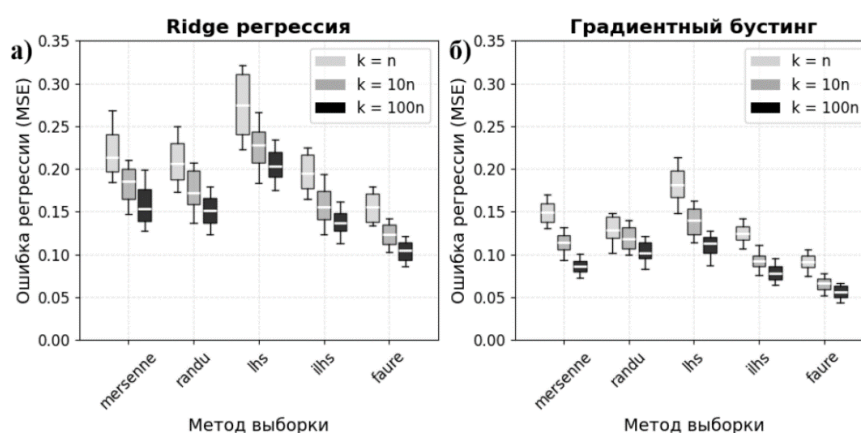


Рис. 2. Ошибка моделей регрессии, обученных на данных различных методов выборки

Как и ожидалось, увеличение объема выборки приводит к улучшению качества регрессии. Можно также заметить, что результаты Ridge регрессии хуже результатов градиентного бустинга. Этот результат закономерен, поскольку линейная модель Ridge, как правило, уступает в точности нелинейному градиентному бустингу на задачах сложной структуры. Однако, сравнение самих моделей между собой выходит за рамки этой работы.

Для регрессора на основе градиентного бустинга четко наблюдается уменьшение разброса ошибки предсказания с ростом объема выборки от $k = n$ к $k = 100n$. Это согласуется с теоретическим снижением дисперсии аппроксимации признаков, обсуждавшимся ранее. В то же время, для линейной регрессии дисперсия ошибки существенно уменьшается при переходе от $k = n$ к $k = 10n$, но затем стабилизируется при дальнейшем увеличении объема данных до $k = 100n$, что может указывать на принципиальное ограничение этой модели в рамках данной задачи.

Не наблюдается существенных различий между двумя генераторами псевдослучайных чисел (*mersenne* и *randu*) по их влиянию на ошибку регрессии. В случае с LHS, напротив, базовый метод (*lhs*) показывает значительно большую ошибку предсказания по сравнению с его улучшенной версией (*ilhs*) на всех объемах выборки.

Для метода выборки на основе последовательностей Фауре (*faure*) в большинстве рассмотренных комбинаций (модель регрессии, объем выборки) наблюдается наименьшая медианная ошибка регрессии. Графики для этого метода, особенно при объемах выборки $k = n$ и $k = 10n$, хорошо отделены от других методов, что визуально свидетельствует о его эффективности. Статистические тесты подтверждают значимость этих различий практически во всех случаях.

Результаты проведенного экспериментального исследования свидетельствуют о существенном влиянии метода выборки на качество аппроксимации признаков и последующую точность моделей машинного обучения. Наблюдаемое устойчивое превосходство детерминированных последовательностей Фауре (faure) над общепринятыми методами (равномерное распределение и LHS) ставит под сомнение целесообразность их использования по умолчанию в задачах анализа ландшафта поиска (ELA). Однако, применение последовательностей Фауре в машинном обучении требует больше вычислительных ресурсов для их генерации, по сравнению с общепринятыми методами.

Статистический анализ не выявил значимых преимуществ методов, основанных на псевдослучайных числах, над базовым LHS – наблюдаемые различия были статистически незначимыми. В то же время метод на основе последовательностей Фауре продемонстрировал статистически значимое снижение ошибки регрессии по всему диапазону исследуемых объемов выборки (от $k = n$ до $k = 100n$) и для всех протестированных алгоритмов машинного обучения. Устойчивость данного результата, подтвержденная различными регрессорами, указывает на то, что низкое отклонение квазислучайных последовательностей обеспечивает более точную и стабильную аппроксимацию признаков ландшафта.

Влияние метода выборки. Из предыдущего анализа следует, что вариативность ошибки регрессии при изменении объема выборки в значительной степени объясняется дисперсией аппроксимированных значений признаков. Вместе с тем сравнительный анализ различных методов демонстрирует, что одна лишь дисперсия не может служить основным фактором, объясняющим наблюдаемые между ними различия. Следовательно, на эффективность методов влияют иные механизмы, требующие дополнительного изучения.

На рис. 3 представлено распределение аппроксимированных значений для двух из 35 рассматриваемых признаков. Анализ визуализации выявил существенную зависимость оценок признаков от метода выборки, что противопоставляет традиционному представлению об их абсолютном характере. Для признака, представленного на левом графике, распределения не только обладают различными медианами и средними значениями, но и являются полностью разделимыми. Данная закономерность устойчиво проявляется при различных объемах выборки.

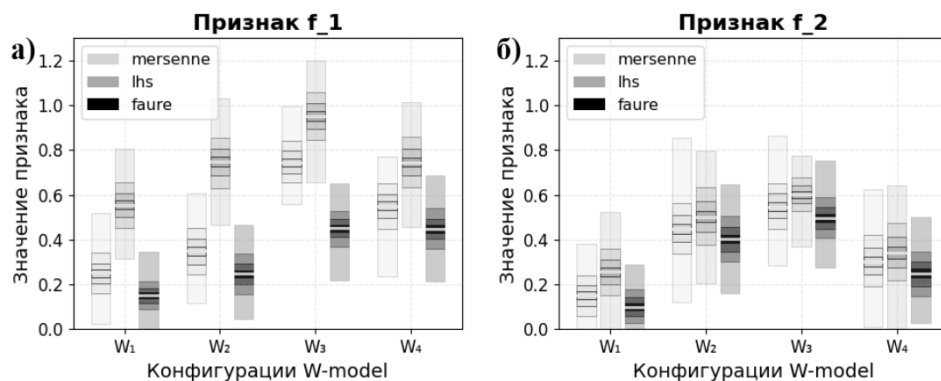


Рис. 3. Распределения двух признаков фласко для разных конфигураций W-model и методов выборки

Полученные результаты свидетельствуют против предположения о возможности прямого переноса оценок признаков между различными методами выборки. В качестве иллюстрации рассмотрим две конфигурации W-model: W_1 ($n = 256, dummy = 0.9, neutrality = 2, epistasis = 2, ruggedness = 5$) и W_2 ($n = 256, dummy = 0.8, neutrality = 6, epistasis = 3, ruggedness = 10$). Анализ показывает, что методы случайной выборки (mersenne) и квазислучайных последовательностей Фауре (faure) дают согласованные оценки признаков для каждого ландшафта в отдельности. В то же время аппроксимации на основе латинского гиперкуба (LHS) демонстрируют статистически значимое за-

вышение значений определенных признаков для W_1 относительно W_2 . Следовательно, корректная интерпретация значений признаков требует обязательного учета использованного метода выборки, поскольку наблюдаемые межзадачные различия могут быть обусловлены не фундаментальными свойствами ландшафта, а артефактами процедуры семплирования.

Зависимость аппроксимации признаков от выборки. С целью исследования зависимости устойчивости аппроксимаций признаков от методов выборки был реализован специализированный эксперимент в контексте задачи регрессии. В рамках данной схемы регрессоры обучались на векторах признаков, вычисленных по выборкам, сгенерированным одним методом, а их обобщающая способность оценивалась на тестовых наборах, полученных с применением альтернативных методов выборки. Используемый набор признаков и общая методология подготовки данных соответствуют подходу, изложенному в предыдущих разделах, с единственным изменением, заключающимся в кросс-методологическом разделении исходных данных на обучающую и тестовую выборки.

В табл. 1 представлены данные, отражающие медианную ошибку регрессионных моделей для всех 25 комбинаций методов выборки, использованных при формировании обучающих и тестовых наборов данных. Таблица демонстрирует результаты для трех объемов выборки: $k = n$, $k = 10n$ и $k = 100n$. Элементы главной диагонали, соответствующие случаю совпадения методов на этапах обучения и тестирования, были проанализированы ранее. Левая часть таблицы содержит результаты для Ridge-регрессии, правая – для градиентного бустинга.

Таблица 1

Медианные ошибки регрессии параметров W-model

a) Ridge регрессия (k = n)					
	mers	randu	lhs	ilhs	faure
mers	0.245	0.312	0.298	0.285	0.401
randu	0.315	0.248	0.305	0.292	0.408
lhs	0.301	0.308	0.285	0.275	0.395
ilhs	0.288	0.295	0.278	0.195	0.382
faure	0.405	0.412	0.398	0.385	0.152

б) Gradient регрессия (k = n)					
	mers	randu	lhs	ilhs	faure
mers	0.145	0.212	0.198	0.185	0.301
randu	0.215	0.148	0.205	0.192	0.308
lhs	0.201	0.208	0.185	0.175	0.295
ilhs	0.188	0.195	0.178	0.125	0.282
faure	0.305	0.312	0.298	0.285	0.092

в) Ridge регрессия (k = 10n)					
	mers	randu	lhs	ilhs	faure
mers	0.198	0.245	0.232	0.218	0.335
randu	0.248	0.201	0.238	0.225	0.342
lhs	0.235	0.242	0.235	0.208	0.328
ilhs	0.221	0.228	0.215	0.165	0.315
faure	0.338	0.345	0.331	0.318	0.128

г) Gradient регрессия (k = 10n)					
	mers	randu	lhs	ilhs	faure
mers	0.098	0.145	0.132	0.118	0.235
randu	0.148	0.101	0.138	0.125	0.242
lhs	0.135	0.142	0.135	0.108	0.228
ilhs	0.121	0.128	0.115	0.095	0.215
faure	0.238	0.245	0.231	0.218	0.068

д) Ridge регрессия (k = 100n)					
	mers	randu	lhs	ilhs	faure
mers	0.175	0.212	0.198	0.185	0.285
randu	0.215	0.178	0.205	0.192	0.292
lhs	0.202	0.208	0.210	0.178	0.278
ilhs	0.188	0.195	0.185	0.148	0.265
faure	0.288	0.295	0.281	0.268	0.115

е) Gradient регрессия (k = 100n)					
	mers	randu	lhs	ilhs	faure
mers	0.075	0.112	0.098	0.085	0.185
randu	0.115	0.078	0.105	0.092	0.192
lhs	0.102	0.108	0.110	0.078	0.178
ilhs	0.088	0.095	0.085	0.078	0.165
faure	0.188	0.195	0.181	0.168	0.055

Для объемов выборки $k = 10n$ и $k = 100n$ наименьшая, или близкая к наименьшей, ошибка регрессионной модели достигается при условии совпадения методов формирования выборок на этапах обучения и тестирования. Указанная закономерность является статистически устойчивой и проявляется независимо от типа используемого регрессора (Ridge-регрессия или градиентный бустинг), что визуально подтверждается наименьшими значениями ошибки на элементах главной диагонали матриц, представленных в табл. 1.

Для минимального объема выборки $k = n$ регистрируется значительный разброс значений ошибки регрессии, что свидетельствует о низкой устойчивости аппроксимации признаков. При этом, несмотря на высокую сложность задачи прогнозирования парамет-

ров W-model, обусловленную многомерностью пространства признаков и разнообразием конфигураций ландшафта, абсолютное значение ошибки предсказания оказывается приемлемым, особенно в случае применения нелинейного метода градиентного бустинга.

Анализ случаев с объемом выборки $k \geq 10n$ показывает значительное уменьшение ошибки регрессии, что согласуется с теоретическим снижением дисперсии аппроксимации признаков. Ключевым фактором, определяющим эффективность моделей, является степень согласованности методов выборки на этапах обучения и тестирования. Наибольшая ошибка наблюдается при принципиальном различии этих методов (например, *mersenne-lhs*, *lhs-faure*, *randu-ilhs*). Промежуточные значения ошибки достигаются в случае использования схожих методов *lhs-ilhs*. При этом минимальная ошибка регрессии обеспечивается только при полном совпадении методов генерации выборок для обучающей и тестовой выборок, что подтверждается результатами для всех пар типа *mersenne-mersenne*, *faure-faure* и т.д. Данная закономерность имеет систематический характер и наблюдается для обоих регрессоров.

Анализ перекрестного тестирования выявил специфическую особенность метода Фауре. В то время как регрессоры, обученные на признаках, аппроксимированных по другим методам, демонстрируют сопоставимый уровень ошибки при тестировании на данных иного типа, модель, обученная на выборках Фауре, показывает статистически значимо более высокую ошибку при валидации на любом из альтернативных методов. Данный эффект наблюдается для обоих регрессоров, однако является более выраженным для градиентного бустинга по сравнению с моделью линейной регрессии.

Результаты детального анализа распределений аппроксимированных признаков позволяют выявить кластеризацию методов выборки по степени схожести генерируемых ими оценок. Наблюдается высокая степень согласованности между методами *lhs* и *ilhs*, значения признаков для которых практически не отличаются. В то же время, методы *faure* и *randu* формируют обособленные группы, существенно отклоняясь не только от пары *lhs/ilhs*, но и от стандартной случайной выборки *mersenne*. Количественная оценка данного эффекта показывает, что почти для половины всех признаков на всех конфигурациях W-model при $k = 100n$, медианное значение признаков, полученных с использованием *faure* или *randu* статистически значимо отличается от медиан, полученных другими методами.

Таким образом, результаты кросс-методологической валидации однозначно демонстрируют, что аппроксимированные признаки не являются инвариантными относительно метода выборки. Их значения и, как следствие, информативность для моделей машинного обучения критически зависят от метода выборки. Это означает, что признаки, полученные разными методами, несут в себе не только информацию о свойствах ландшафта, но и специфический «отпечаток» способа их генерации. Следовательно, для построения корректных прогнозных моделей в рамках ELA необходимо строгое соответствие между методами выборки, использованными на этапах обучения и применения.

Ошибки моделей регрессии. Для углубленного анализа характера ошибок, допускаемых регрессионными моделями, был проведен детальный разбор предсказаний. Результаты анализа показывают, что для большинства конфигураций W-model медианная ошибка предсказания отдельных параметров является достаточно низкой. Однако наблюдаются систематические смещения для определенных диапазонов значений параметров. Так, для конфигураций с высокими значениями параметра *ruggedness* (например, > 8) модель имеет тенденцию к устойчивому занижению предсказаний. Аналогичным образом, для параметра *dimmy* в области значений около 0.5 часто наблюдается бимодальное распределение ошибок, где значительная часть предсказаний образует кластер вокруг истинного значения, а другая – смещенный кластер в области более высоких значений.

Эти систематические ошибки проявляются согласованно для множественных повторений эксперимента на одних и тех же конфигурациях W-model. Это указывает на то, что они обусловлены не случайным шумом, а фундаментальными особенностями взаимодействия между конкретными свойствами ландшафта и способом их отражения в аппроксимированных признаках при данном методе выборки.

Подобный анализ ошибок полезен для выявления слабых мест как в наборе признаков, так и в методах их оценки. Устойчивое смещение предсказаний для конкретных параметров может указывать на необходимость введения новых, более специализированных признаков, чувствительных к соответствующим свойствам ландшафта [20]. Кроме того, различия в структуре ошибок между разными методами выборки подчеркивают, что выбор метода семплирования влияет не только на общую точность модели, но и на то, какие именно конфигурации будут распознаваться хуже или лучше.

Выводы. В работе проведено исследование влияния метода и объема выборки на точность аппроксимации признаков в рамках анализа ландшафта поиска (ELA) и на последующее качество моделей машинного обучения.

Установлено, что аппроксимированные значения признаков ELA не являются абсолютными характеристиками ландшафта и существенно зависят от примененного метода выборки. Различные методы выборки приводят к статистически значимо различным оценкам одних и тех же признаков даже при больших объемах данных. Следствием этого является критическая важность использования единого метода генерации выборок на этапах обучения и применения моделей машинного обучения для обеспечения их корректной работы.

Обнаружено, что модели, обученные на признаках, полученных с помощью квазислучайных последовательностей Фауре, демонстрируют меньшую ошибку в задаче регрессии параметров *W-model* по сравнению с моделями, использующими признаки на основе стандартных методов (равномерное распределение, латинский гиперкуб). Это указывает на перспективность дальнейшего исследования детерминированных последовательностей с низким отклонением для задач ELA.

Проведенная кросс-методологическая валидация подтвердила, что наилучшее качество прогноза достигается только при строгом совпадении метода выборки для обучающего и тестового наборов. Наблюдаемые систематические ошибки предсказания для отдельных диапазонов параметров *W-model* подчеркивают сложность взаимодействия между свойствами ландшафта и методом их оценки.

Полученные результаты свидетельствуют о необходимости учета специфики методов выборки при интерпретации признаков ELA и разработке прогнозных моделей. Для дальнейшего повышения надежности методов ELA перспективными направлениями являются исследование функциональной связи между методами выборки и значениями признаков, а также разработка методов коррекции полученных оценок.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Пикалов М.В., Письмеров А.М. Настройка параметров генетического алгоритма при помощи анализа ландшафта функции приспособленности и машинного обучения // Известия ЮФУ. Технические науки. – 2024. – № 2 (238). – С. 221-228.
2. Lobo F.J., Lima C.F., Michalewicz Z. (ed.). Parameter setting in evolutionary algorithms. – Springer Science & Business Media, 2007. – Т. 54.
3. Mersmann O. et al. Exploratory landscape analysis // Proceedings of the 13th annual conference on Genetic and evolutionary computation. – 2011. – P. 829-836.
4. Ochoa G., Malan K. Recent advances in fitness landscape analysis // Proceedings of the genetic and evolutionary computation conference companion. – 2019. – P. 1077-1094.
5. Kerschke P., Trautmann H. Automated algorithm selection on continuous black-box problems by combining exploratory landscape analysis and machine learning.
6. Kerschke P., Trautmann H. The R-Package FLACCO for exploratory landscape analysis with applications to multi-objective optimization problems // 2016 IEEE Congress on Evolutionary Computation (CEC). – IEEE, 2016. – P. 5262-5269.
7. Mersmann O., Preuss M., Trautmann H. Benchmarking evolutionary algorithms: Towards exploratory landscape analysis. – 2010.
8. Huang C., Li Y., Yao X. A survey of automatic parameter tuning methods for metaheuristics // IEEE transactions on evolutionary computation. – 2019. – Vol. 24, No. 2. – P. 201-216.
9. Weise T., Wu Z. Difficult features of combinatorial optimization problems and the tunable w-model benchmark problem for simulating them // Proceedings of the Genetic and Evolutionary Computation Conference Companion. – 2018. – P. 1769-1776.

10. McDonald G.C. Ridge regression // *Wiley Interdisciplinary Reviews: Computational Statistics*. – 2009. – Vol. 1, No. 1. – P. 93-100.
11. Friedman J.H. Greedy function approximation: a gradient boosting machine // *Annals of statistics*. – 2001. – P. 1189-1232.
12. Breiman L. Random forests // *Machine learning*. – 2001. – Vol. 45, No. 1. – P. 5-32.
13. Matsumoto M., Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator // *ACM Transactions on Modeling and Computer Simulation (TOMACS)*. – 1998. – Vol. 8, No. 1. – P. 3-30.
14. Marsaglia G. Random numbers fall mainly in the planes // *Proceedings of the National Academy of sciences*. – 1968. – Vol. 61, No. 1. – P. 25-28.
15. Helton J.C., Davis F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems // *Reliability Engineering & System Safety*. – 2003. – Vol. 81, No. 1. – P. 23-69.
16. Huntington D.E., Lyrantzis C.S. Improvements to and limitations of Latin hypercube sampling // *Probabilistic engineering mechanics*. – 1998. – Vol. 13, No. 4. – P. 245-253.
17. Faure H., Lemieux C. Generalized Halton sequences in 2008: A comparative study // *ACM Transactions on Modeling and Computer Simulation (TOMACS)*. – 2009. – Vol. 19, No. 4. – P. 1-31.
18. Wang X., Hickernell F.J. Randomized halton sequences // *Mathematical and Computer Modelling*. – 2000. – Vol. 32, No. 7-8. – P. 887-899.
19. Faure H., Lemieux C. Generalized Halton sequences in 2008: A comparative study // *ACM Transactions on Modeling and Computer Simulation (TOMACS)*. – 2009. – Vol. 19, No. 4. – P. 1-31.
20. Pitzer E., Affenzeller M. A comprehensive survey on fitness landscape analysis // *Recent advances in intelligent engineering systems*. – 2012. – P. 161-191.

REFERENCES

1. Pikalov M.V., Pis'merov A.M. Nastroyka parametrov geneticheskogo algoritma pri pomoshchi analiza landshafta funktsii prispособlennosti i mashinnogo obucheniya [Tuning genetic algorithm parameters using fitness landscape analysis and machine learning], *Izvestiya YuFU. Tekhnicheskie nauki [Izvestiya SFedU. Engineering Sciences]*, 2024, No. 2 (238), pp. 221-228.
2. Lobo F.J., Lima C.F., Michalewicz Z. (ed.). Parameter setting in evolutionary algorithms. Springer Science & Business Media, 2007, Vol. 54.
3. Mersmann O. et al. Exploratory landscape analysis, *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, 2011, pp. 829-836.
4. Ochoa G., Malan K. Recent advances in fitness landscape analysis, *Proceedings of the genetic and evolutionary computation conference companion*, 2019, pp. 1077-1094.
5. Kerschke P., Trautmann H. Automated algorithm selection on continuous black-box problems by combining exploratory landscape analysis and machine learning.
6. Kerschke P., Trautmann H. The R-Package FLACCO for exploratory landscape analysis with applications to multi-objective optimization problems, *2016 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2016, pp. 5262-5269.
7. Mersmann O., Preuss M., Trautmann H. Benchmarking evolutionary algorithms: Towards exploratory landscape analysis, 2010.
8. Huang C., Li Y., Yao X. A survey of automatic parameter tuning methods for metaheuristics, *IEEE transactions on evolutionary computation*, 2019, Vol. 24, No. 2, pp. 201-216.
9. Weise T., Wu Z. Difficult features of combinatorial optimization problems and the tunable w-model benchmark problem for simulating them, *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2018, pp. 1769-1776.
10. McDonald G.C. Ridge regression, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2009, Vol. 1, No. 1, pp. 93-100.
11. Friedman J.H. Greedy function approximation: a gradient boosting machine, *Annals of statistics*, 2001, pp. 1189-1232.
12. Breiman L. Random forests, *Machine learning*, 2001, Vol. 45, No. 1, pp. 5-32.
13. Matsumoto M., Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 1998, Vol. 8, No. 1, pp. 3-30.
14. Marsaglia G. Random numbers fall mainly in the planes, *Proceedings of the National Academy of sciences*, 1968, Vol. 61, No. 1, pp. 25-28.
15. Helton J.C., Davis F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliability Engineering & System Safety*, 2003, Vol. 81, No. 1, pp. 23-69.

16. Huntington D.E., Lyrintzis C.S. Improvements to and limitations of Latin hypercube sampling, *Probabilistic engineering mechanics*, 1998, Vol. 13, No. 4, pp. 245-253.
17. Faure H., Lemieux C. Generalized Halton sequences in 2008: A comparative study, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2009, Vol. 19, No. 4, pp. 1-31.
18. Wang X., Hickernell F.J. Randomized halton sequences, *Mathematical and Computer Modelling*, 2000, Vol. 32, No. 7-8, pp. 887-899.
19. Faure H., Lemieux C. Generalized Halton sequences in 2008: A comparative study, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2009, Vol. 19, No. 4, pp. 1-31.
20. Pitzer E., Affenzeller M. A comprehensive survey on fitness landscape analysis, *Recent advances in intelligent engineering systems*, 2012, pp. 161-191.

Пикалов Максим Вадимович – Национальный исследовательский университет ИТМО; e-mail: pikmaksim@gmail.com; г. Санкт-Петербург, Россия; тел.: 88126070283; аспирант.

Pikalov Maxim Vadimovich– ITMO University; e-mail: pikmaksim@gmail.com; St. Petersburg, Russia; phone: +78126070283; graduate student.

УДК 004.032.26

DOI 10.18522/2311-3103-2026-1-52-64

М.О. Доброхвалов, А.Ю. Филатов, Е.А. Чегодаева

НОВАЯ МЕТРИКА ВОСПРОИЗВОДИМОСТИ ДЛЯ СРАВНЕНИЯ КЛАССИФИКАТОРОВ ВРЕМЕННЫХ РЯДОВ

Воспроизводимость результатов экспериментов является критическим аспектом современного машинного обучения, однако выбор случайного инициализирующего сета существенно влияет на итоговое качество моделей, что создает проблему корректного сравнения различных архитектур и методов. Цель исследования заключалась в оценке влияния выбора случайного сета на результаты классификации временных рядов сверточными нейросетями и в разработке корректного способа сравнения моделей. Задачи включали измерение разброса метрик при множественных повторных запусках в рамках варьирующихся инициализаций, проверку нормальности распределений, введение метрики воспроизводимости RM и подбор ее параметра λ , а также проверку переносимости подхода на альтернативных архитектурах. Проведены эксперименты с двумя одномерными архитектурами (FCN, ResNet) сверточных нейронных сетей на семи открытых наборах данных временных рядов разной природы. Для каждой пары модель–датасет выполнено по 55 независимых запусков с фиксацией источников случайности и идентичными настройками обучения в PyTorch. Статистический анализ включал критерии Шапиро Уилка и Андерсона Дарлинга. Показано, что распределения аккуратности чаще всего не соответствуют нормальному закону, поэтому интервальные оценки, основанные на нормальности, некорректны. Варьирование сета приводит к различиям аккуратности до 12 процентных пунктов, причем величина разброса зависит от датасета и архитектуры. Предложенная метрика воспроизводимости (RM), штрафующая за дисперсию, при малом числе запусков на различных инициализирующих значениях приближает нижнюю наблюдаемую границу, при большом числе стремится к среднему. Предложенная RM позволяет сравнивать модели с учетом случайных “удачных” и “неудачных” инициализаций и ранжировать модели по устойчивости, задает стандарт отчетности, повышающий воспроизводимость экспериментов и надежность выводов. Эмпирическая проверка на архитектуре DenseNet подтвердила, что RM адекватно реагирует как на стабильные, так и на нестабильные наборы. Методика легко переносится на новые датасеты и архитектуры. Предложенная метрика может быть использована для стандартизации отчетности и повышения воспроизводимости исследований.

Классификация временных рядов; сверточные нейронные сети; случайный сет; чувствительность к инициализации; воспроизводимость экспериментов.