

# Методы повышения точности алгоритмов машинного обучения при сокращении размерности набора данных

А.В. Воробьев

**Аннотация** — Ограниченная доступность сбора информации выступает фактором, препятствующим применению высокоэффективных алгоритмов машинного обучения. Разработка методов повышения точностей моделей при снижении периодов наблюдений, может стать эффективным инструментом прогнозирования в малоисследованных областях. В статье рассмотрена взаимосвязь между размерностью набора данных и прогностическими возможностями моделей машинного обучения, определено воздействие количества наблюдений на точность и устойчивость моделей, построенных на ансамблевых алгоритмах и алгоритмах регуляризованной регрессии. В ходе экспериментов рассмотрено изменение средневзвешенной абсолютной ошибки при снижении размерности набора, выявлены алгоритмы наиболее устойчивые к данному фактору. Выявлен нижний предел использования ансамблевых алгоритмов в целях обнаружения закономерностей и построения устойчивой модели, в задачах регрессии, в случаях нелинейной зависимости целевой переменной с предикторами и при условии отсутствия высокого воздействия аномалий и шумов в данных. Рассмотрено влияние автоматизированной Байесовской оптимизации гиперпараметров на точность моделей при сокращении набора данных. Определены модели, предварительная оптимизация гиперпараметров для которых, посредством древесно-структурированной оценки Парзена, наиболее предпочтительна.

**Ключевые слова**— машинное обучение, размерность набора данных, ансамблевые алгоритмы, точность модели, оптимизация гиперпараметров, оценка Парзена.

## I. ВВЕДЕНИЕ

При анализе данных с применением алгоритмов машинного обучения существенное внимание уделяется проблемам недостаточности данных, под которой понимается, что для переменной в наблюдении не сохраняются значения данных (пропуск полей в наборе) [1], а также дисбаланс классов [2] возникающий при значительном превосходстве одного класса над другим по численности. Широкое обсуждение обозначенных вопросов определило разработку методов как изолированно решающих данные задачи, так и комбинированных методов декомпозиции информации на основе нечеткости, которые рассматривают эти две

проблемы как задачу оценки недостающих данных [3]. При этом вопрос снижения точности работы моделей, определенный незначительным размером набора данных, не имеет массовой методологической проработки, ввиду отсутствия общих случаев, устанавливающих размер набора, необходимый для достижения целевой точности.

Определение влияния размерности набора данных на точность моделей более выражено в областях с ограниченной доступностью к предикторам. Например, Феликс Фабер (Институт физической химии и Национальный центр вычислительного проектирования и открытия новых материалов, химический факультет Базельского университета) обнаружил, что точность ML-модели для прогнозирования энергии образования соединений эльпасолита показала систематическое улучшение с увеличением размера обучающего набора. [4] Джонатан Шмидт (Университет Мартина Лютера в Галле-Виттенберге), в работе «Прогнозирование термодинамической устойчивости твердых тел, сочетающее теорию функционала плотности и машинное обучение», определил, что ошибка прогнозирования модели линейно уменьшалась с размером обучающего набора, удвоение обучающего набора уменьшало погрешность примерно на 20%. [5] Ин Чжан и Чэнь Лин (Научно-исследовательский институт Toyota в Северной Америке) продемонстрировали, что меньшая доступность обучающих данных не только затрудняет обнаружение закономерностей, но и ухудшает способность делать прогнозы в неисследованной области. [6]

## II. ОЦЕНКА ТОЧНОСТИ МОДЕЛЕЙ ПРИ СОКРАЩЕНИИ НАБОРА ДАННЫХ

В целях проверки работы алгоритмов под воздействием сокращения набора данных были выбраны популярные датасеты конкурсной платформы по исследованию данных Kaggle Machine Learning Competition Platform (Google) различной размерности: от 1 000 наблюдений с 25 атрибутами до 1 048 576 наблюдений с 28 атрибутами.

Учитывая уникальность датасетов, критериями оценки изменений воздействия снижения размерности набора стали: относительная средняя доля изменения средневзвешенной абсолютной ошибки (WMAPE) к минимальным показателям, достигнутом на конкретном

тестовом наборе, а также динамика коэффициента детерминации.

Прогностические модели построены в языковой среде Python на наиболее распространенных и популярных алгоритмах машинного обучения (таблица 1) [7].

Таблица 1 – Тестовые алгоритмы машинного обучения

Ансамблевые алгоритмы		Алгоритмы регуляризованной регрессии
деревья решений	градиентный бустинг	
Decision Tree Regressor (DTR)	XGBoost (XGB)	Ridge Regression (RR)
Bagging Regressor (BR)		Lasso (LS)
Random Forest Regressor (RFR)	LightGBM (LGBMR)	Elastic Net (EN)

#### A. Деревья решений

В основе создания обучающей прогностической модели анализ правил принятия решений, выведенных из обучающего набора. В интеллектуальном анализе данных деревья решений можно также описать как комбинацию математических и вычислительных методов, помогающих описанию, категоризации и обобщению заданного набора данных [8]. При входных данных:

$$(x, Y) = (x_1, x_2, x_3, \dots, x_k) \quad (1)$$

$Y$  - целевая переменная. Вектор  $x$  состоит из функций  $x_1, x_2, x_3, \dots, x_k$ , которые используются для поставленной задачи.

Алгоритмы Bagging Regressor и RandomForestRegressor имеют в основе мета-алгоритм бутстрэп-агрегирования, повышающий стабильность и точность моделей.

Результаты тестирования моделей, построенных на алгоритмах деревьев решений (Рис.1) демонстрируют линейный прирост WMAPE по мере сокращения объемов тренировочных наборов, содержащих от 1 000 000 наблюдений («больших наборах»). Снижение данных датасетов на 50%, определяет прирост ошибки, в среднем, на 20%. Таким образом, мы видим подтверждение выводов Джонатана Шмидта о линейном воздействии объема набора на ошибку, а также о снижении погрешности на 20% при удвоении обучающего набора. Динамика WMAPE на наборах от 700 до 1 000 наблюдений («малых наборах») имеет более выраженный прирост ошибки. В точке снижения объема набора на 50% прирост ошибки составил от 50% (Bagging Regressor) до 84% (Decision Tree Regressor). Лучшие показатели при сокращении объема наблюдений на малых наборах имеет алгоритм Bagging Regressor.

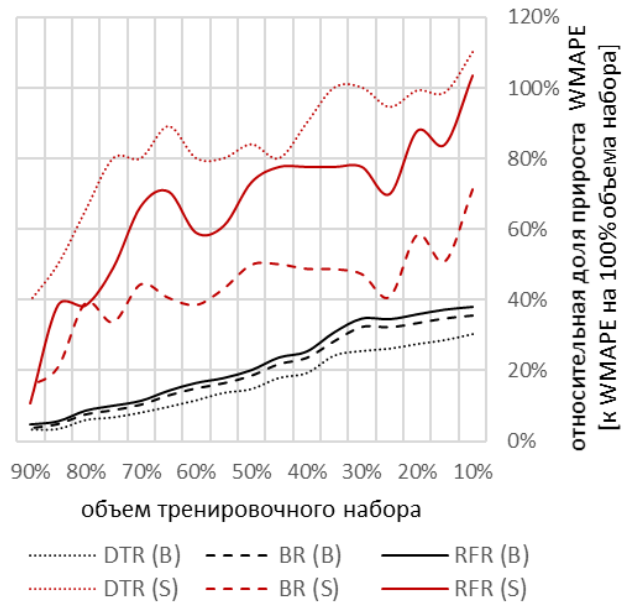


Рис.1. Усредненная динамика изменения WMAPE моделей, построенных на алгоритмах деревьев решений, на наборах, содержащих от 1 000 000 (B) и от 700 до 1 000 наблюдений (S)

#### B. Градиентный бустинг

При использовании градиентного бустинга в качестве основы алгоритма аддитивная модель строится поэтапно. В любой момент времени  $t$  результаты модели взвешиваются на основе результатов предыдущего момента времени  $t-1$ . Правильно предсказанные результаты имеют меньший вес, а те, которые не были классифицированы, имеют больший вес. Это позволяет оптимизировать произвольные дифференцируемые функции потерь. На каждом этапе деревья регрессии помещаются на отрицательный градиент биномиальной или полиномиальной функции потерь отклонений.

Например, XGBoost (XGB) следует структуре градиентного бустинга и применяет численную оптимизацию в функциональном пространстве. Функция потерь  $l(y, \hat{y})$  определяется для измерения разницы между меткой  $y$  и оценкой  $\hat{y}$ . В качестве аддитивной модели XGB итеративно минимизирует сумму потерь всех выборок как объективную. Для построения каждого дерева алгоритм сначала вычисляет необходимую статистику, т. е. градиент  $g_i$  первого порядка и градиент  $h_i$  второго порядка выборки  $i$ :

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}), \quad h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) \quad (2)$$

где  $\hat{y}_i^{(t-1)}$  это прогноз к концу последней итерации. Затем сумма производных может быть аккумулирована для набора экземпляров  $I_j$  на каждом узле  $j$ :

$$G_j = \sum_{i \in I_j} g_i, \quad H_j = \sum_{i \in I_j} h_i \quad (3)$$

Принимается аппроксимация расширения Тейлора второго порядка цели и к модели добавляется регуляризация.

Оптимальный вес листа  $w_j^*$  и задача  $obj^*$  решаются с учетом целевой функции:

$$w_j^* = -\frac{G_j}{H_j - \lambda} \quad (4)$$

$$obj^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (5)$$

где  $\gamma$  и  $\lambda$  регуляризаторы для числа листьев и веса листьев соответственно.

Уравнение (5) применяется для измерения предложения разделения в каждом узле и определяет древовидную структуру. Уравнение (4), определяет финальные веса листьев [9].

Результаты тестирования моделей, построенных на алгоритмах градиентного бустинга (Рис.2) имеют схожий с «древесными» алгоритмами тренд по большим наборам, с менее выраженным приростом ошибки. В точке сокращения больших наборов на 50% прирост WMAPE составляет 12% для LightGBM и 15% для XGBoost.

Алгоритм LightGBM сохраняет значительно меньшие, относительно XGB, показатели прироста ошибки при сокращении малых наборов. Изменение тренда и значительный прирост WMAPE на данных наборах у алгоритмов наблюдается при сокращении малых наборов до 35%. Таким образом при сокращении количества наблюдений до 250-350, даже высокоэффективный алгоритм LightGBM существенно снижает показатели точности.

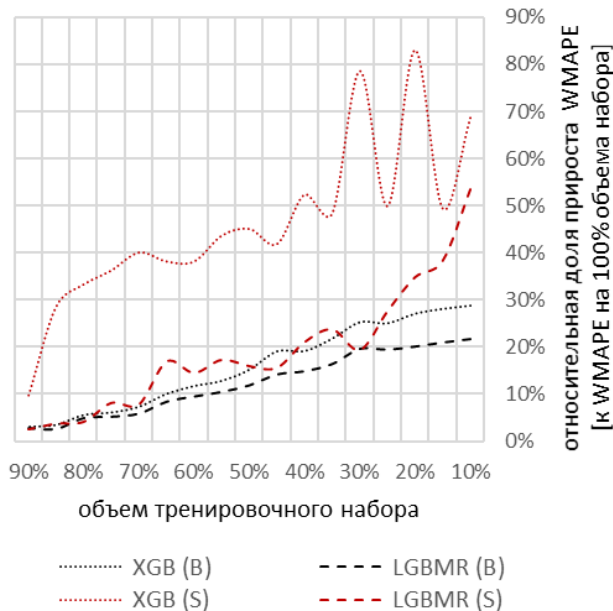


Рис.2. Усредненная динамика изменения WMAPE моделей, построенных на алгоритмах градиентного бустинга, на наборах, содержащих от 1 000 000 (B) и от 700 до 1 000 наблюдений (S)

### С. Алгоритмы регуляризованной регрессии

В основе алгоритмов регуляризованной регрессии лежит уменьшение дисперсии за счет смещения к оптимуму. Применение подобных алгоритмов обеспечивает снижение сложности модели за счет сокращения количества предикторов [10].

Исключение предикторов из модели можно рассматривать как обращение их коэффициентов в нуль. При этом, например, алгоритм гребневой или ридж-регрессии уменьшает сложность модели при сохранении всех переменных за счет «штрафа», в случае сильного смещения коэффициентов от нуля, тем самым заставляя их быть малыми непрерывным образом.

В гребневой регрессии функция потерь обычного метода наименьших квадратов дополняется таким образом, что минимизируется сумма квадратов отклонений и штрафуются размер оценок параметров:

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2 + \lambda \sum_{j=1}^m \hat{\beta}_j^2 = \|y - X\hat{\beta}\|^2 + \lambda \|\hat{\beta}\|^2 \quad (6)$$

Решение для  $\hat{\beta}$  дает оценку гребневой регрессии  $\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T Y$ , где  $I$  обозначает матрицу идентичности, а параметр  $\lambda$  — штраф за регуляризацию.

По мере увеличения  $\lambda$  дисперсия уменьшается, а смещение увеличивается. Одним из оптимальных способов установления параметра регуляризации выступает метод перекрестной проверки и выбора значения  $\lambda$ , которое минимизирует перекрестную проверку суммы квадратов отклонений.

Схожие в динамике тренды среди алгоритмов-представителей регуляризованной регрессии наблюдаются по результатам тестирования как на малых, так и на больших наборах (Рис.3). При сокращении объемов больших наборов на 50% прирост ошибки определяется на уровне 20%, аналогично показателям алгоритмов деревьев решений. Значительный прирост WMAPE определяется, как и в случае с алгоритмами на основе градиентного бустинга, при сокращении малых наборов до 35%.

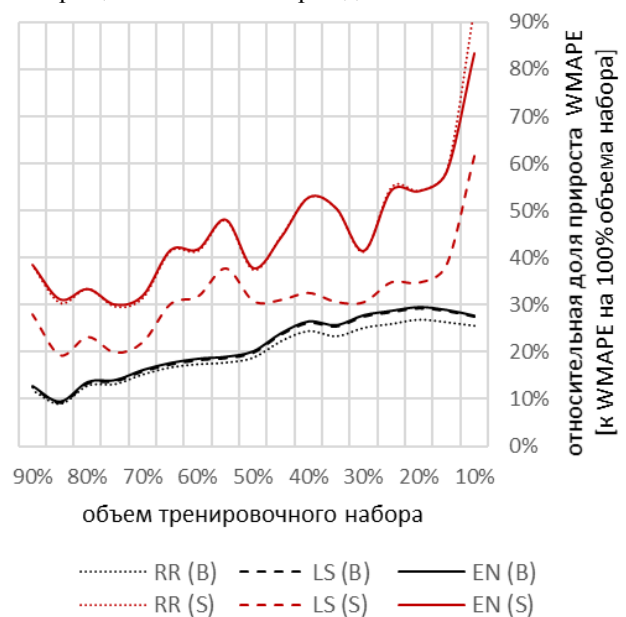


Рис.3. Усредненная динамика изменения WMAPE моделей, построенных на алгоритмах регуляризованной

регрессии, на наборах, содержащих от 1 000 000 (B) и от 700 до 1 000 наблюдений (S)

### III. ОПТИМИЗАЦИИ ГИПЕРПАРАМЕТРОВ

Последние достижения в области повышения точности машинного обучения были достигнуты не только за счет новых подходов к изучению атрибутов, но и посредством более совершенной конфигурации существующих методов, в том числе за счет оптимизации гиперпараметров. Данные методы позволяют существенно улучшить качество моделей, в том числе их точность.

Одним из эффективных методов автоматизированной оптимизации гиперпараметров выступает Байесовская оптимизация [11]. Суть метода заключается в построении вероятностной модели целевой функции и использование ее для выбора наиболее значимых гиперпараметров с последующей оценкой в истинной целевой функции. В основе алгоритма древесно-структурированная оценка Парзена.

Ожидаемое улучшение  $EI$  — это ожидание при некоторой модели  $M$  для  $f : x \rightarrow \mathbb{R}^N$ , где  $f(x)$  превышает порог  $y^*$ :

$$EI_{y^*}(x) := \int_{-\infty}^{\infty} \max(y^* - y, 0) p_M(y|x) dy. \quad (7)$$

Древесно-структурированная оценка Парзена (TRE) моделирует  $p(x|y)$  путем преобразования этого генеративного процесса, заменяя распределения предшествующей конфигурации непараметрическими плотностями. Пространство конфигурации описывается с использованием однородных, логарифмически однородных, квантованных логарифмически однородных и категориальных переменных. В этих случаях алгоритм TRE выполняет следующие замены: равномерная  $\rightarrow$  усеченная гауссова смесь, логарифмическая однородная  $\rightarrow$  экспоненциальная усеченная гауссова смесь, категориальная  $\rightarrow$  переоцененная категориальная. Используем различные наблюдения  $\{x^{(1)}, \dots, x^{(k)}\}$  в непараметрических плотностях. Эти замены позволяют создавать различные плотности в конфигурационном пространстве  $X$ . TRE определяет  $p(x|y)$  используя две такие плотности:

$$p(x|y) = \begin{cases} l(x) & \text{if } y < y^* \\ g(x) & \text{if } y \geq y^*, \end{cases} \quad (8)$$

где  $l(x)$ - плотность, сформированная с помощью наблюдений  $\{x^{(i)}\}$  такой, что соответствующая функция потерь  $f(x^{(i)})$  была меньше, чем  $y^*$ , а  $g(x)$  - плотность, сформированная с использованием оставшихся наблюдений.

Алгоритм TRE зависит от  $y^*$  больше, чем наблюдаемое  $f(x)$ , ввиду этого отдельные точки могут быть использованы для формирования  $l(x)$ . Алгоритм определяет  $y^*$  быть некоторым квантилем  $\gamma$  наблюдаемых значений  $y$ , так что  $p(y < y^*) = \gamma$ , но никакой конкретной модели для  $p(y)$  не требуется. Поддерживая упорядоченные списки наблюдаемых переменных в  $H$ , время выполнения каждой итерации

алгоритма TRE может линейно масштабироваться в  $|H|$  и в количестве оптимизируемых переменных (измерений).

Параметризация  $p(x, y)$  как  $p(y)p(y|x)$  в алгоритме TRE была выбрана для облегчения оптимизации ожидаемого улучшения  $EI$ .

$$\begin{aligned} EI_{y^*}(x) &:= \int_{-\infty}^{y^*} (y^* - y) p(y|x) dy = \\ &= \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} dy \end{aligned} \quad (9)$$

принимая  $\gamma = p(y < y^*)$  и  $p(x) = \int \mathbb{R} p(x|y)p(y) dy = \gamma l(x) + (1 - \gamma)g(x)$ .

Следовательно,

$$\begin{aligned} \int_{-\infty}^{y^*} (y^* - y) p(x|y)p(y) dy &= l(x) \int_{-\infty}^{y^*} (y^* - y) p(y) dy = \\ &= \gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y) dy, \end{aligned}$$

таким образом получаем, что

$$EI_{y^*}(x) := \frac{\gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y) dy}{\gamma l(x) + (1 - \gamma)g(x)} \propto \left( \gamma + \frac{g(x)}{l(x)} (1 - \gamma) \right)^{-1}.$$

Последнее выражение показывает, что для максимального улучшения необходимы точки  $x$  с высокой вероятностью при  $l(x)$  и низкой вероятностью при  $g(x)$ . Древовидная форма  $l$  и  $g$  позволяет построить множество кандидатов в соответствии с  $l$  и оценить их в соответствии с  $g(x)/l(x)$ . На каждой итерации алгоритм возвращает кандидата  $x^*$  с наибольшим  $EI$  [11].

В ходе применения Байесовской оптимизации были определены улучшения по всем рассматриваемым алгоритмам (Рис.4). Улучшение точности, определенное в снижении прироста WMAPE более выражено в моделях, построенных на алгоритмах деревьев решений, в том числе содержащих мета-алгоритм бутстрэп-агрегирования, а также алгоритмах регуляризованной регрессии.

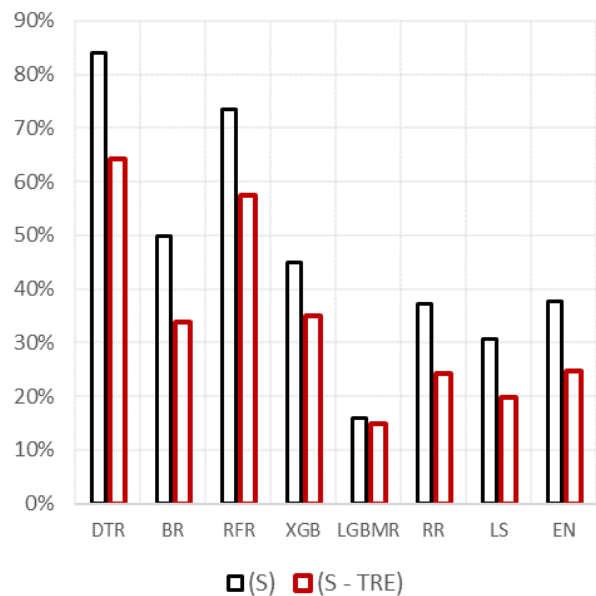


Рис.4. Изменение прироста WMAPE моделей, построенных на наборах, содержащих и от 700 до 1 000

наблюдений, при сокращении набора на 50%, с применением Байесовской оптимизации гиперпараметров (S-TRE) и без применения (S)

#### IV. ЗАКЛЮЧЕНИЕ

В ходе исследования было определено, что, несмотря на разность задач, источников данных и применяемых алгоритмов, наблюдается устойчивое снижение точностей моделей машинного обучения при сокращении количества наблюдений в тренировочных наборах.

В задачах регрессии, в случаях нелинейной зависимости целевой переменной с предикторами и невысокими показателями корреляции между ними, при сокращении количества наблюдений до 250-350, даже высокоэффективные алгоритмы существенно снижают показатели точности. Количество наблюдений менее 100, исключительно в подобных случаях, может быть определено как нижний предел использования ансамблевых алгоритмов машинного обучения в целях обнаружения закономерностей и построении устойчивой модели.

Использование Байесовской оптимизации гиперпараметров определяет нивелирование потери точности моделей при сокращении набора данных до 50% на ансамблевых алгоритмах деревьев решений. Минимальное улучшение за счет применения оптимизации наблюдается на алгоритме градиентного бустинга LightGBM, при этом данный алгоритм наименее подвержен воздействию сокращения данных в наборе и демонстрирует лучшие показатели в ряду исследуемых алгоритмов, как с применением оптимизации, так и без нее. Следующим, по наблюдаемой в ходе экспериментов устойчивости к снижению размерности набора, выступает алгоритм регуляризованной регрессии Lasso.

Применение алгоритмов LightGBM и Lasso в совокупности с Байесовской оптимизацией может рассматриваться как один из эффективных методов повышения точностей моделей при снижении периодов наблюдения в наборе данных, при решении задач регрессии без выраженной корреляционной зависимости целевой переменной и предикторов.

#### БИБЛИОГРАФИЯ

- [1] R. J. Little and D. B. Rubin. *Statistical Analysis With Missing Data*. Hoboken, NJ, USA: Wiley, 2014. DOI:10.1002/9781119013563
- [2] H. He and E. A. Garcia. Learning from imbalanced data, *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284. 2009. DOI:10.1109/TKDE.2008.239
- [3] C.Subhashri, J.Maruthupandi, K.Vimala Devi. Recovering Insufficient and Inconsistent Data using Fuzzy-Based Information Decomposition Technique. *International Journal of Pure and Applied Mathematics*. Volume 119 No. 12 2018.
- [4] Faber, F. A., Lindmaa, A., Lilienfeld, O. A. V. & Armiento, R. Machine learning energies of 2 million Elpasolite (ABC2D6) crystals. *Phys. Rev. Lett.* 117, 135502. 2016. DOI:10.1103/PhysRevLett.117.135502
- [5] Schmidt, J. et al. Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chem. Mater.* 5090–5103. 2017. DOI:10.1021/acs.chemmater.7b00156

- [6] Ying Z., Chen L. A strategy to apply machine learning to small datasets in materials science. *npj Computational Materials* volume 4, Article number: 25. 2018. DOI:10.1038/s41524-018-0081-z
- [7] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems* 30. NIPS. 2017.
- [8] Shalev-Shwartz, Shai, Ben-David, Shai. *Decision Trees. Understanding Machine Learning*. Cambridge University Press. 2014.
- [9] Wenjing Fang, Chaochao Chen, Jin Tan, Chaofan Yu, Yufei Lu, Li Wang, Lei Wang, Jun Zhou and Alex X. Liu. A Hybrid-Domain Framework for Secure Gradient Tree Boosting. In *The 29th ACM International Conference on Information and Knowledge Management (CIKM'20)*, Galway, Ireland. ACM, New York, NY, USA, 2020.
- [10] Donald W. Marquardt & Ronald D. Snee. Ridge Regression in Practice, *The American Statistician*, 29:1, 3-20, 1975. DOI: 10.1080/00031305.1975.10479105
- [11] James Bergstra, R. Bardenet, Balázs Kégl, Y. Bengio. Algorithms for Hyper-Parameter Optimization. *Conference: Advances in Neural Information Processing Systems*. 2011.

# Methods to improve the accuracy of machine learning algorithms while reducing the dimensionality of the data set

A.V. Vorobyev

**Abstract** — The limited availability of information collection is a factor hindering the application of high-performance machine learning algorithms. The development of methods to improve the accuracy of models while reducing the observation periods, can be an effective tool for prediction in understudied areas. The paper considers the relationship between the dimensionality of the data set and the predictive capabilities of machine learning models, and determines the impact of the number of observations on the accuracy and robustness of models built on ensemble algorithms and regularized regression algorithms. In the course of the experiments, the change in the weighted average absolute error with decreasing the dimensionality of the set was considered, and the algorithms most resistant to this factor were identified. The lower limit of use of ensemble algorithms for detection of regularities and construction of stable model, in regression tasks, in cases of non-linear dependence of target variable with predictors and under condition of absence of high impact of anomalies and noises in data was revealed. The effect of automated Bayesian hyperparameter optimization on model accuracy when the data set is reduced is considered. The models for which pre-optimization of hyperparameters, by means of wood-structured Parzen estimation, is the most preferable are determined.

**Key words** — machine learning, dataset dimensionality, ensemble algorithms, model accuracy, hyperparameter optimization, Parzen's estimation.

## REFERENCES

- [1] R. J. Little and D. B. Rubin. Statistical Analysis With Missing Data. Hoboken, NJ, USA: Wiley, 2014. DOI:10.1002/9781119013563
- [2] H. He and E. A. Garcia. Learning from imbalanced data, IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284. 2009. DOI:10.1109/TKDE.2008.239
- [3] C.Subhashri, J.Maruthupandi, K.Vimala Devi. Recovering Insufficient and Inconsistent Data using Fuzzy-Based Information Decomposition Technique. International Journal of Pure and Applied Mathematics. Volume 119 No. 12 2018.
- [4] Faber, F. A., Lindmaa, A., Lilienfeld, O. A. V. & Armiento, R. Machine learning energies of 2 million Elpasolite (ABC2D6) crystals. Phys. Rev. Lett. 117, 135502. 2016. DOI:10.1103/PhysRevLett.117.135502
- [5] Schmidt, J. et al. Predicting the thermodynamic stability of solids combining density functional theory and machine learning. Chem. Mater. 5090–5103. 2017. DOI:10.1021/acs.chemmater.7b00156
- [6] Ying Z., Chen L. A strategy to apply machine learning to small datasets in materials science. npj Computational Materials volume 4, Article number: 25. 2018. DOI:10.1038/s41524-018-0081-z
- [7] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems 30. NIPS. 2017.
- [8] Shalev-Shwartz, Shai, Ben-David, Shai. Decision Trees. Understanding Machine Learning. Cambridge University Press. 2014.
- [9] Wenjing Fang, Chaochao Chen, Jin Tan, Chaofan Yu, Yufei Lu, Li Wang, Lei Wang, Jun Zhou and Alex X. Liu. A Hybrid-Domain Framework for Secure Gradient Tree Boosting. In The 29th ACM International Conference on Information and Knowledge Management (CIKM'20), Galway, Ireland. ACM, New York, NY, USA, 2020.
- [10] Donald W. Marquardt & Ronald D. Snee. Ridge Regression in Practice, The American Statistician, 29:1, 3-20, 1975. DOI: 10.1080/00031305.1975.10479105
- [11] James Bergstra, R. Bardenet, Balázs Kégl, Y. Bengio. Algorithms for Hyper-Parameter Optimization. Conference: Advances in Neural Information Processing Systems. 2011.