

РОЛЬ МАСШТАБИРОВАНИЯ ДАННЫХ ПРИ ПОСТРОЕНИИ РАСЧЁТНЫХ МОДЕЛЕЙ В ЗАДАЧАХ МЕХАНИЧЕСКОЙ ОБРАБОТКИ

Зотов В.А.^{1,2}, Евдокимов Д.В.^{1,2}, Скуратов Д.Л.¹

¹Самарский университет, г. Самара, vlad198189@gmail.com

²СамНЦ РАН, г. Самара

Ключевые слова: шероховатость поверхности, машинное обучение, масштабирование признаков, стандартизация, точение

Введение. При точении и фрезеровании качество поверхности определяется правильным выбором режимов резания, геометрии инструмента и динамической жёсткости технологической системы. Традиционно связь между этими факторами и шероховатостью описывают эмпирическими степенными формулами [1] или имитационными моделями, требующими идентификации большого числа параметров [2-7]. Однако в производственных и лабораторных условиях объём экспериментальных данных часто ограничен, что ограничивает применимость сложных расчётных схем. Современные методы машинного обучения позволяют выявлять устойчивые закономерности даже на основе таких ограниченных данных [2], но требуют аккуратного обращения с исходными признаками [3].

Предмет. В работе анализируется, как предварительное масштабирование технологических параметров: скорость резания, подача, глубина, радиус при вершине, тип пластины влияет на точность расчётных моделей шероховатости, построенных с использованием машинного обучения.

Цель работы. Показать, что пренебрежение стандартизацией признаков способно не просто ухудшить, а полностью обесценить предсказания ряда широко распространённых алгоритмов, и предложить простой методический приём, позволяющий извлечь максимальную точность даже из ограниченного набора экспериментальных данных.

Метод и методология. Исследование проведено на эталонном датасете из 72 опытов по точению стали С45. Датасет включает пять входных факторов, а в качестве выходной переменной выступает шероховатость R_a , описанный в исследовании [1]. Для тринадцати алгоритмов, а именно: линейная и полиномиальная регрессия, Ridge, Lasso, BayesianRidge, RandomForest, Gaussian Process, SVR, XGBoost, LightGBM, CatBoost, GradientBoosting – выполнено сравнение двух подходов к предобработке данных. В первом случае использовались исходные значения факторов в их естественных единицах измерения. Во втором – стандартизованные значения, для чего каждый признак был преобразован к нулевому среднему и единичной дисперсии с помощью метода StandardScaler. Стандартизация осуществлялась строго внутри каждой итерации кросс-валидации, что исключает утечку информации. Качество прогноза оценивалось методом Leave-One-Out и многократной Shuffle-валидацией; основная метрика – коэффициент детерминации R^2 .

Результаты и обсуждение. Установлено, что древовидные ансамбли XGBoost, CatBoost, RandomForest практически нечувствительны к масштабированию и показывают высокую точность $R^2 \approx 0,99$ в обоих вариантах. Линейная и полиномиальная регрессия без регуляризации также не изменили свои показатели. Однако для регуляризованных линейных моделей Ridge, Lasso, BayesianRidge стандартизация дала заметный прирост точности: коэффициент детерминации R^2 , полученный методом скользящего экзамена Leave-One-Out, вырос на 0,9–25 процентных пункта. Наиболее выраженный эффект проявился у методов, которые способны выявлять сложные нелинейные зависимости без явного задания формулы. Метод опорных векторов (SVR) без стандартизации дал отрицательное значение $R^2 = -0,16$. Отрицательный коэффициент

детерминации означает, что модель предсказывает хуже, чем если бы в качестве прогноза всегда использовалось среднее арифметическое измеренных значений шероховатости. Иными словами, такая модель совершенно непригодна для практического применения. После приведения признаков к единому масштабу значение R^2 достигло 0,71. Гауссовский процесс (Gaussian Process) в тех же условиях улучшился с $R^2 \approx -0,10$ до $R^2 = 0,9947$, что означает практически идеальное совпадение расчётных значений с экспериментальными.

Наглядная интерпретация полученного выигрыша в точности представлена на рис.1, где по оси ординат отложен прирост LOO R^2 при переходе от исходных данных к стандартизованным.

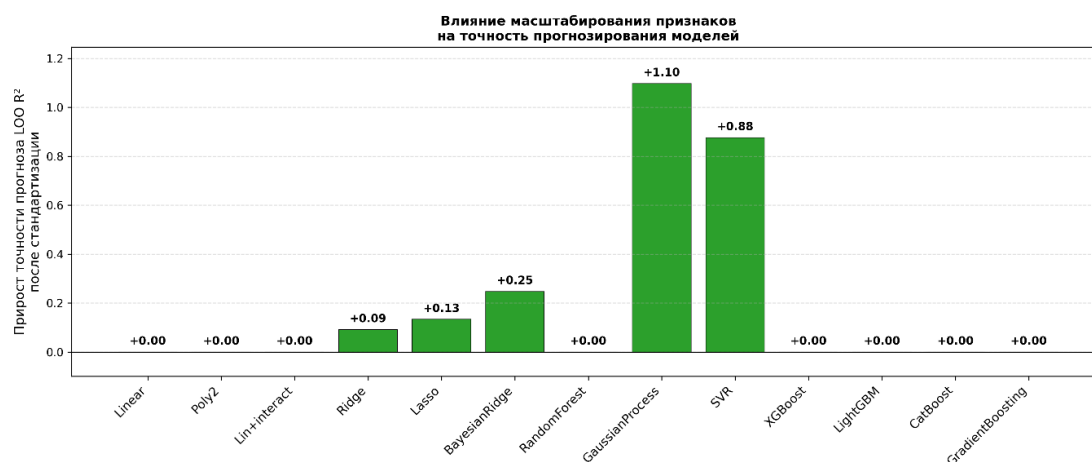


Рис. 1 – Влияние масштабирования признаков на точность прогнозирования моделей

Выводы. Стандартизация входных факторов не требует проведения дополнительных натурных испытаний, но является критически необходимой операцией для ядерных и регуляризованных алгоритмов. На малых выборках, типичных для лабораторной и заводской практики, игнорирование этого шага превращает высокоточные методы Gaussian Process, SVR в неработоспособные. Предложенная методика позволяет без увеличения числа опытов получить практически идеальную расчётную модель шероховатости. Рекомендуется при публикации работ по машинному обучению в области технологии машиностроения в обязательном порядке указывать, применялась ли предварительная стандартизация данных.

Благодарности:

Работа выполнена сотрудниками лаборатории «Инженерии поверхности» в рамках государственного задания Минобрнауки России для СамНЦ РАН по теме № FMRW-2025-0052.

Список литературы

1. Milosevic A., Simunovic G., Kanovic Z., Simunovic K., Santos Z., Sokac M., Vukelic D. Comprehensive evaluation of dimensional deviation, flank wear, surface roughness and material removal rate in dry turning of C45 steel // Facta Universitatis, Series: Mechanical Engineering. – 2024. – Vol. 22, No. 3. – P. 547–566. – DOI: 10.22190/FUME240403024M.
2. Hu L., Gao R., Guo Y.B. Transfer Learning Enhanced Gaussian Process Model for Surface Roughness Prediction with Small Data // Manufacturing Letters. – 2023. – Vol. 35, suppl. – P. 1103-1108. – DOI: 10.1016/j.mfglet.2023.08.094.
3. Pinheiro J.M.H., Oliveira S.V.B. de, Silva T.H.S. et al. The Impact of Feature Scaling in Machine Learning: Effects on Regression and Classification Tasks // IEEE Access. – 2025. – Vol. 13. – P. 199903-199931. – DOI: 10.1109/ACCESS.2025.3635541.

4. Liu R., Tian W. A Novel Approach to Surface Roughness Virtual Sample Generation to Address the Small Sample Size Problem in Ultra-Precision Machining // *Sensors*. – 2024. – Vol. 24, № 11. – P. 3621. – DOI: 10.3390/s24113621.

5. Pasandidehpour M., Nogueira A.R., Mendes-Moreira J. et al. Survey on machine learning applied to CNC milling processes // *Adv. Manuf.* – 2025. – Vol. 13. – P. 847–885. – DOI: 10.1007/s40436-025-00564-x.

6. Ko J.H., Yin C. A review of artificial intelligence application for machining surface quality prediction: from key factors to model development // *J Intell Manuf.* – 2026. – Vol. 37. – P. 775–798. – DOI: 10.1007/s10845-025-02571-y.

7. Гвинджилия В.Е., Фоминов Е.В., Моисеев Д.В., Гамалеева Е.И. Влияние динамических характеристик процесса резания на шероховатость поверхности детали при токарной обработке // *Обработка металлов (технология, оборудование, инструменты)*. – 2024. – Т. 26, № 2. – С. 143–157. – DOI: 10.17212/1994-6309-2024-26.2-143-157.

Сведения об авторах

Зотов Владислав Александрович – студент ПИАШ; лаборант лаборатории «Инженерия поверхности».

Евдокимов Дмитрий Викторович – к.т.н., доцент кафедры технологий производства двигателей; с.н.с. лаборатории «Инженерия поверхности».

Скуратов Дмитрий Леонидович – д.т.н., профессор кафедры технологий производства двигателей.

THE ROLE OF DATA SCALING IN BUILDING COMPUTATIONAL MODELS FOR MACHINING TASKS

Zotov V.A.^{1,2}, Evdokimov D.V.^{1,2}, Skuratov D.L.¹

¹Samara University, Samara, Russia, vlad198189@gmail.com

²Samara Federal Research Center of the Russian Academy of Sciences, Samara, Russia

Keywords: surface roughness, machine learning, feature scaling, standardization, turning.

The paper examines the effect of feature scaling on the accuracy of machine learning models for surface roughness prediction in turning. Using a dataset of 72 turning experiments on C45 steel with five input parameters, thirteen algorithms were tested with raw versus standardized data. Standardization did not affect tree-based ensembles or non-regularized linear models but improved regularized linear models Ridge, Lasso, BayesianRidge by 0.09–0.25 in LOO R^2 and dramatically rescued kernel methods: Gaussian process from –0.10 to 0.99, SVR from –0.16 to 0.71. The study concludes that scaling is critical for kernel and regularized algorithms, especially on small machining datasets.