

2. ГОСТ 12020- 2018 (ISO 175:2010) Пластмассы. Методы определения стойкости к действию химических сред.
3. ГОСТ 24888-81 Пластмассы, полимеры и синтетические смолы. Химические наименования, термины и определения
4. ГОСТ 32588-2013 Композиты полимерные. Номенклатура показателей.
5. ГОСТ 32794-2014 Композиты полимерные. Композиты полимерные. Термины и определения.
6. ГОСТ 33742-2016 Композиты полимерные. Классификация.
7. ГОСТ 34370-2017 (ISO 527-1: 2012) Пластмассы. Определение механических свойств при растяжении.
8. ГОСТ Р 58060-2018 Препреги наномодифицированные. Типы и основные параметры.
9. Кербер, М. Л. Полимерные композиционные материалы: структура, свойства, технология. – ЦОП Профессия, 2018.
10. Коваленко, О. А. Новые материалы: уч. пособие. – Алчевск, 2003.
11. Колосова, А. С. Современные полимерные композиционные материалы и их применение / А. С. Колосова // Международный ж-л прикладных и фундаментальных исследований. – 2018. – №5, ч. 1. – С. 245–250.

УДК 691.32:004.85

## **ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ПРОЧНОСТИ БЕТОНА**

*ЯКИМОВИЧ Г. Д.*

Белорусский национальный технический университет  
Минск, Беларусь

**Аннотация.** Данная статья посвящена вопросам применения современных технологий в науке о бетоне, и в частности, использованию технологии машинного обучения для исследования и прогнозирования прочностных свойств, а в перспективе и для проектирования составов бетона.

Одним из важнейших вопросов технологии бетона на сегодняшний день остается проектирование бетонных смесей с требуемыми

характеристиками. С одной стороны, такой состав обеспечивает долговечность эксплуатации бетонных изделий и конструкций, тем самым увеличивая период эксплуатации, снижая тем самым эксплуатационные затраты. С другой стороны, хорошо спроектированный состав предполагает снижение расхода цемента, являющегося не только самым дорогим компонентом бетонно смеси, но и наименее экологичным. Производство одной тонны цемента сопряжено с выбросом 520 килограммов углекислого газа [1] в атмосферу, что делает вопрос экономии цемента еще более острым.

На данный момент имеется довольно широкий ряд прогнозирования прочности бетона. Наибольшую точность, обеспечивающую коэффициент вариации 13,5 %, показали пять формул [2], а именно, формула НИИЖБа, формула И. Н. Ахвердова, формула В. П. Сизова, формула В. В. Бабкова и модифицированная формула Вольфа. Все эти формулы определяющее значение отдают водоцементному отношению и непосредственно прочности вяжущего. Формулы Ахвердова, Сизова и модифицированная формула Вольфа предлагают учитывать качество заполнителя согласно табличным значениям.

В связи с ограниченностью технологий, в годы проектирования формул было достаточно проблематично учитывать все факторы, которых в процессе приготовления бетонной смеси достаточно много, включая физические свойства заполнителей, их геометрические параметры и т. д. Но в последнее время, с появлением на рубеже 2010-х высокопроизводительных графических ускорителей и роста их доступности, стали развиваться и технология, оперирующая с большими объемами данных, искусственный интеллект, и, в частности, такая его область как машинное обучение.

Главным принципом машинного обучения является построение алгоритмов, способных обучаться самостоятельно на наборах данных. Это облегчает нахождение зависимостей сложных зависимостей между входными данными и получаемым результатом. Недостатком такого способа для научной деятельности может быть лишь то, что результатом обучения модели является не формула с переменными, а большой массив правил, чтение которых крайне сложно для человека.

Обучение алгоритмов может проходить с учителем, без него и с подкреплением. Обучение с учителем проходит на размеченных данных, благодаря обучению на которых программа может выда-

вать те или иные ответы на неразмеченных. Способ обычно применяется для задач классификации или предсказания.

Обучение без учителя необходимо для выявления закономерностей, объяснение которых на машинном языке является затруднительным. В этом случае алгоритм сам математически разделяет данные на категории, выявляя их общие особенности по каким-либо критериям. Разновидностью такого способа является обучение с подкреплением, когда при каждом правильном ответе алгоритм получает поощрение со стороны, тем самым понимая, в какую сторону корректировать свое поведение.

Среди основных алгоритмов машинного обучения можно выделить следующие.

Линейная регрессия. Ее суть состоит в построении на графике с исследуемыми значениями прямой линии, для которой сумма квадратов расстояний от нее до любой точки будет минимальной.

Логистическая же регрессия прогнозирует вероятность того или иного бинарного состояния путем сравнения события с логистической кривой, и служит для решения задач классификации.

Наивная байесовская классификация, основана на строгом (иначе говоря, наивном) предположении о независимости признаков объектов друг от друга и применении к ним формулы Байеса. Позволяет решать задачи классификации данных.

Дерево решений – иерархическая древовидная структура, состоящая из узлов, являющихся последовательными решающими правилами, и листьев, являющихся финальным подмножеством (ответом) каждой ветви. Данный алгоритм позволяет решать задачи как классификации, так и регрессии.

Случайный лес. Представляет собой ансамбль, собранный из решающих деревьев. Каждое из деревьев учится на отдельной части обучающей выборки. Усредненный ответ в задаче регрессии, либо результат голосования большинства деревьев в задаче классификации, является более точным, чем результаты отдельных алгоритмов, обученных на выборке целиком.

Градиентный бустинг развивает идею случайного леса, и соединяет модели последовательно [3]. Каждая новая модель учится на ошибках предыдущей. При этом целью является минимизация среднеквадратичной ошибки, осуществляемая методом поиска ло-

кального минимума, известного как градиентный спуск. На данный момент является одним из самых эффективных.

Все задачи, решаемые методами машинного обучения, базируются на больших наборах данных, и чем больше их – тем более качественной получается модель. Для оценки перспективности машинного прогнозирования как такового мы использовали архивный набор данных составов бетона с их прочностями, предоставленный в свободный доступ [4] профессором I-Cheng Yeh из тайваньского университета Чунг Хуа.

Набор данных состоит из 1030 записей и собран из 17 различных источников. Набор имеет 8 атрибутов с расходами цемента, доменного шлака, золы-уноса, воды, суперпластификатора, крупного и мелкого заполнителя. Также имеются пометки о возрасте (1–365 дней) и прочность каждой партии. Несмотря на то, что известен класс цемента, 42,5, проблемой набора данных является отсутствие данных об используемых суперпластификаторах, и алгоритм может некорректно работать на бетонах с добавками.

Анализ набора данных осуществлялся при помощи языка программирования Python, с использованием модулей предобработки и визуализации данных, таких как Pandas, NumPy, Seaborn, Matplotlib. Анализ проводился при помощи модуля Scikit Learn.

Как правило, выбор модели для анализа и предсказания происходит уже в ходе исследования, так как нельзя заранее понять на каком из наборов данных как будет вести себя та или иная модель. В рамках данного исследования наилучшим способом себя показали две модели: случайный леса и градиентный бустинг.

Подбор гиперпараметров моделей осуществлялся при помощи встроенного в Scikit Learn модуля GridSearchCV, позволяющего путем кросс-валидации сравнивать эффективность моделей, имеющих разные параметры настройки. Данный модуль сохраняет в память модель с наилучшим качеством предсказаний.

Во всех случаях набор данных разделялся на обучающую (67 %) и контрольную (33 %) выборки. Для оценки точности предсказаний кросс-валидатор использует коэффициент детерминации. Для случайного леса точность модели на обучающей выборке составила 98,7 %, а на тестовой 89,4 %. Для более полной оценки была построена таблица, содержащая предсказанные значения (Predict), реальные (Real), модуль абсолютного отклонения (ASE), а также отно-

сительное стандартное отклонение (MSE). Для оценки разброса значений использована функция describe(), данные приведены в табл. 1.

Таблица 1

Оценка разброса предсказаний «случайного леса» по колонкам

	Predict	Real	ASE	MSE
Среднее	34,83	35,33	3,66	13,04
Стандартное отклонение	14,97	16,07	3,74	17,29
Минимум	4,5	4,0	0,01	0,035
Квантиль 25 %	24,82	24,0	0,92	3,2
Квантиль 50 %	34,56	35,0	2,62	7,6
Квантиль 75 %	43,74	45,0	4,87	17,29
Максимум	78,62	82,0	23,53	150,4

Из таблицы видно, что коэффициент вариации для 75 % ответов не превышает 17,3 %, то есть модель обладает высокой, но не достаточной точностью. Так же, имеются выбросы до 150 %, которым в наборе данных соответствуют составы с пластификаторами. Как отмечалось ранее, для возникновения таких неточностей имеются все предпосылки, и связаны они как с методологией создания набора данных, так и с недостаточной степенью обученности модели. В области коэффициента вариации 13,5 % лежат 67 % предсказаний.

Градиентный бустинг показал более высокую точность как на обучающей выборке составила 99,2 %, а на тестовой 92,8 %.

В таблице 2 приведены результаты оценки предсказаний градиентного бустинга. Несмотря на то, что точность предсказаний оказалась выше, значению коэффициента вариации удовлетворяет всего лишь 75 % предсказаний.

Таблица 2

Оценка разброса предсказаний градиентного бустинга»

	Predict	Real	ASE	MSE
Среднее	34,83	35,33	2,88	10,22
Стандартное отклонение	14,97	16,07	3,21	13,35
Минимум	4,5	4,0	0,02	0,06
Квантиль 25 %	24,82	24,0	0,81	2,45
Квантиль 50 %	34,56	35,0	1,87	5,7
Квантиль 75 %	43,74	45,0	3,61	13,0
Максимум	78,62	82,0	20,35	119,7

Для того чтобы соотнести эффективность предсказаний модели с имеющимися на данный момент моделями прогнозирования, были проведены вычисления прочности по модифицированной формуле Вольфа [2]. Так как существующие формулы могут корректно предсказывать прочность лишь бездобавочного бетона в возрасте 28 суток, сравнение эффективности «случайного леса», градиентного бустинга и формулы Вольфа осуществлялось на выборке составов возрастом 28 суток, не имеющих добавок шлака, золы-уноса и суперпластификаторов. Оценка точности осуществлялась по величине относительного стандартного отклонения. Данные приведены в таблице 3.

Таблица 3

Сравнение точности предсказаний моделей машинного обучения и модифицированной формулы Вольфа

	MSE Случайны лес	MSE Градиентный бустинг	MSE Формула Вольфа
Среднее	10,67	7,79	18,83
Стандартное отклонение	20,29	16,47	13,03
Минимум	0,16	0,38	2,25
Квантиль 25 %	2,32	1,76	12,30
Квантиль 50 %	3,31	2,99	16,66
Квантиль 75 %	10,43	6,89	21,11
Максимум	101,00	84,03	73,48

Полученные результаты говорят о том, что алгоритм градиентного бустинга наилучшим образом предсказывает прочность бездобавочного бетона в возрасте 28 суток. Из 25 объектов тестовой выборки найдено лишь два локальных выброса, на которых алгоритм ошибся более чем на 13,5 %. Данные выбросы могут являться ошибкой составителя набора данных, либо следствием недостаточной обученности алгоритма. Формула Вольфа на этих составах показала более точный результат, 28,74 % против 84,03 %, что подкрепляет идею недообученности алгоритма. Тем не менее, на остальном наборе ее точность оказалась существенно ниже, лишь 25 % предсказаний оказались точнее 12,3 %.

Помимо высокой точности прогнозирования прочности бездобавочного бетона, в качестве преимущества градиентного бустинга



отсутствуют идентификаторы использованных материалов и их характеристики (кроме прочности цемента), прогноз прочности реального состава оказался близок к реальной прочности во всех возрастах. Это в очередной раз подтверждает эффективность алгоритмов машинного обучения, и создает предпосылки для дальнейшего изучения данной темы. Помимо этого, возникают предпосылки для пересмотра самого подхода к прогнозированию прочностных и иных характеристик бетона.

**Выводы. 1.** В ходе проведенной работы у нас получилось обучить модель машинного обучения на наборе данных о составах бетона и их прочностных характеристиках, при помощи алгоритмов «случайного леса» и градиентного бустинга. **2.** Наилучшим образом при прогнозировании прочности бездобавочного бетона в возрасте 28 суток показал себя градиентный бустинг. **3.** Машинное обучение продемонстрировало свою эффективность в вопросах прогнозирования и перспективность для дальнейшей работы в данном направлении.

#### СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Руководящие принципы национальных инвентаризаций парниковых газов МГЭИК. Промышленные процессы и использование продуктов; под ред. С. Игглестона [и др.]. – Т. 1–5. – ИГЕС: Япония, 2006.

2. Суходоева, Н. В. Анализ формул для расчета прочности бетона на сжатие / Н. В. Суходоева, В. В. Бабицкий // Вестник белорусско-российского университета. – № 3. – 2007. – С. 139–147.

3. Concrete Compressive Strength Data Set [Электронный ресурс] режим доступа: <https://archive.ics.uci.edu/ml/data-sets/Concrete+Compressive+Strength>.

4. Friedman, Jerome H. Greedy function approximation: a gradient boosting machine / Jerome H. Friedman // Annals of statistics. – 2001. – P. 1189–1232.