

МОДЕЛЬ ПРЕДСКАЗАНИЯ СКОРОСТИ АТМОСФЕРНОЙ КОРРОЗИИ СТАЛИ НА ОСНОВЕ АЛГОРИТМА СЛУЧАЙНОГО ЛЕСА

М.А. Малеева, А.И. Маршаков, Ю.М. Панченко

*Институт физической химии и электрохимии им. А.Н.Фrumкина
Российской академии наук, Москва, Россия*

e-mail: marina.maleeva@gmail.com

Коррозия металлов в атмосферных условиях зависит от многочисленных климатических и аэрохимических факторов природной среды. На основании проведенных натурных испытаний по всему миру была получена база данных со среднегодовыми параметрами атмосферы и откликом в виде потерь металла за первый год испытаний. На основании этой базы данных были разработаны функции «доза-ответ» (ФДО), которые стали стандартом в области прогнозирования атмосферной коррозии [1]. Однако продолжают попытки решения задачи прогноза с помощью различных алгоритмов машинного обучения. В данной работе использовался алгоритм «случайного леса».

Алгоритм «случайного леса» был разработан в 2001 году Брейманом [2]. Случайные леса - это ансамблевый метод машинного обучения для проведения классификации (и регрессии), суть которого заключается в построении на стадии обучения множества решающих деревьев, каждое из которых способно дать ответ на стадии оценки нового набора данных. Финальное решение принимается голосованием.

В работе использована база данных RU + UN/ECE + Micat, в которой содержатся данные о среднегодовой температуре, влажности, содержании SO₂ и количестве выпавших осадков, а также потери металла вследствие коррозии в течение одного года в разных точках земного шара. Предварительная оценка данных показала, что наибольшее влияние на целевой показатель оказывает концентрация SO₂ в атмосфере.

Данные для обучения и тестирования делились на две части в пропорции 3:1 случайным образом методом бутстрепа. Критерием разбиения служило значение среднеквадратичной ошибки. Случайный лес содержал в себе 100 деревьев. Разделение вели по двум случайным признакам. Корреляцию между реальными и предсказанными данными оценивали с помощью параметра R², который для тестового набора составил 0,75. Наибольшую важность при прогнозе алгоритм присвоил концентрации SO₂ в атмосфере и температуре. Остальные два признака имеют меньшее влияние на результат предсказания.

Результат предсказания случайного леса сравнивали с результатами модели ФДО на одной и той же выборке данных с помощью средней абсолютной ошибки (MAE). Данный параметр показывает, на сколько г/м² модель ошибается в своем предсказании. Получили, что модель случайного леса предсказывает потери металла в первый год испытаний с меньшей ошибкой.

Литература

1. ISO 9223:2012(E). Corrosion of metals and alloys. Corrosivity of atmospheres. Classification, determination and estimation; 2012.
2. Breiman L. // Machine Learning. – 2001. V. 45. – P. 5.