

УДК 544.275.3

АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРЕДСКАЗАНИЯ ПЛОТНОСТИ ГЛУБОКИХ ЭВТЕКТИЧЕСКИХ РАСТВОРИТЕЛЕЙ

Одегова В.С. (Университет ИТМО), Лавриненко А.К. (Университет ИТМО)

Научный руководитель – д.х.н. проф. Виноградов В.В.
(Университет ИТМО)

Введение. Современная промышленность невозможна без химии, но многие процессы являются экологически небезопасными. Нефтехимия, синтез лекарственных препаратов, целлюлозно-бумажная промышленность требуют использования токсичных органических растворителей или высоких энергетических затрат. Однако, проблемы данных производств можно решить с помощью глубоких эвтектических растворителей (ГЭР), которые представляют собой смесь двух или более веществ, характеризующихся пониженной температурой плавления по сравнению с чистыми исходными материалами [1]. Чаще всего ГЭР состоят из природных компонентов, таких как органические кислоты, спирты, сахара и четвертичные аммониевые соли, поэтому они биodeградируемы, их производство доступно и легко осуществимо, а эффективность экстракции сопоставима с известными органическими растворителями [1]. Из-за того, что данные вещества перспективны для применения в различных областях химии, фармацевтики, металлургии, нефтяной и пищевой промышленности, их популярность растет, однако существует проблема выбора системы из двух или более соединений, которые могут образовывать смесь с оптимальными свойствами для конкретной задачи. На данный момент эта проблема решается эмпирически, путем проведения большого количества экспериментов, что не всегда приводит к оптимальному результату [2]. Поэтому ученые все чаще прибегают к алгоритмам машинного обучения для предсказания свойств этих веществ [2]. Однако, разработанные на данный момент модели были применимы только для узкого круга двухкомпонентных ГЭР с определенной структурой [2]. В своей работе мы предлагаем создать модели для прогнозирования физико-химических свойств двух- и трехкомпонентных ГЭР с различной структурой исходных соединений.

Основная часть. Одним из наиболее важных свойств глубоких эвтектических растворителей является плотность, которая влияет на эффективность экстракции, а также является ключевым технологическим параметром при внедрении ГЭР в производство [2]. Для построения моделей машинного обучения была собрана уникальная база данных по плотностям бинарных и трехкомпонентных ГЭР, измеренных при различных температурах и мольных долях, содержащая более 4000 строк. База данных включает все выделенные на данный момент типы ГЭР, классифицируемые по структуре исходных веществ. Тем не менее, из-за большей популярности бинарных ГЭР, состоящих из органических веществ и четвертичных аммониевых солей, база данных оказалась несбалансированной, поэтому при разработке алгоритмов были использованы веса для повышения важности малочисленных типов. Построение моделей машинного обучения требует использования дескрипторов, особых параметров, описывающих геометрические свойства молекул, количества структурных групп и связей, а также мольную долю и температуру, при которых была измерена плотность. Для предсказания плотности были использованы классические модели машинного обучения на основе деревьев, бустинга, метода опорных векторов, метода k -ближайших соседей, а также многослойный персептрон. Наиболее эффективными моделями оказались Cat Boosting Regression и eXtreme Gradient Boosting, которые способны предсказывать плотность двух- и трехкомпонентных ГЭР различного состава с точностью на кросс-валидации, учитывающей уникальность смеси, $R^2 = 0.85$, а также с низким RMSE < 0.07 .

Выводы. В работе показана возможность создания моделей машинного обучения, способных предсказывать плотность ГЭР с высокой точностью для различных типов бинарных и трехкомпонентных ГЭР. Это исследование показывает возможность прогнозирования других

физико-химических свойств глубоких эвтектических растворителей, что позволит проводить направленный дизайн ГЭР.

Список использованных источников:

[1] Hansen B.B. et al. Deep Eutectic Solvents: A Review of Fundamentals and Applications // Chem. Rev. 2021. Vol. 121, № 3. P. 1232–1285.

[2] Abdollahzadeh, M., et al. Estimating the density of deep eutectic solvents applying supervised machine learning techniques. Sci Rep 2022 12, 4954