

Методы интеллектуального анализа данных в управлении химико-технологическими процессами

Н. В. Звягинцев, email: n.zvyagintsev@gmail.com

В. А. Биллиг, email: Vladimir-Billig@yandex.ru

Тверской государственный технический университет

***Аннотация.** Работа посвящена анализу управляющих воздействий, повышающих эффективность протекания химико-технологических процессов. Для прогнозирования значений целевых параметров технологического процесса строятся деревья решений на основе данных, полученных как экспериментально, так и по результатам компьютерного моделирования. Предложенный подход апробирован на процессе карбонилирования олефинов, имеющем важное практическое значение (в том числе в фармацевтической промышленности как способ производства нестероидных жаропонижающих препаратов группы ибупрофена).*

***Ключевые слова:** Объект управления, Data Mining, интеллектуальный анализ данных, деревья решений, химико-технологические процессы, карбонилирование олефинов.*

Введение

Управление химико-технологическими процессами (ХТП) является важной задачей, позволяющей существенно повысить эффективность протекания процессов. ХТП включают множество этапов, в том числе подвод реагирующих процессов, непосредственно химическое превращение, отвод и очистка продуктов [1, 2]. Для многих ХТП решающим этапом являются именно химические реакции превращения, на протекании которых влияет ряд факторов: давление, температура, концентрации компонентов, природа растворителя и состав каталитической системы.

Объектами управления ХТП являются выход целевого продукта (отношение массы целевого продукта к теоретически возможному), конверсия исходного вещества (отношение массы израсходованного субстрата к начальной массе), а также средняя скорость самого процесса, которая выражается через изменение концентраций компонентов химического превращения. В литературе накоплено значительное количество экспериментальных данных о протекании многих процессов, поэтому использование методов и алгоритмов Data

Mining (Интеллектуального анализа данных) позволяет провести анализ влияния различных факторов на протекание процесса. Кроме того, формирование выборок для дальнейшего анализа с применением методов Data Mining мы реализовали на основе кинетических моделей.

1. Подготовка и обработка данных

Подготовка данных для обработки является важным этапом, на котором производится кодирование информации о химическом строении компонентов, масштабирование данных, а также загрузка необходимых справочных данных. Кодирование информации о химическом строении подразумевает формирование вектора признаков о наличии в соединении некоторых атомов, фрагментов или химических связей, с последующим добавлением в результирующую выборку. Для автоматизации кодирования полезно использовать различные библиотеки, в частности ChemFormula [3], реализованную на Python.

Результирующие выборки для дальнейшего исследования мы формируем как на полностью экспериментальных данных, взятых в работах [4-9], так и на основе кинетических моделей. При работе с данными на основе кинетических моделей формируется сетка концентраций, выбирается время интегрирования и производится расчет, например методом Эйлера [10]. В данной работе рассматривается модельная выборка, рассчитанные на основе кинетической модели из [11], алгоритм формирования детально рассмотрен в [12]. В зависимости от исходных данных, в результирующую выборку может быть добавлено время проведения процесса (если оно различается для разных записей в выборке).

2. Обработка данных методами Data Mining

После подготовки результирующей выборки производится анализ данных, в результате которого определяются наиболее значимые признаки, формируются ассоциативные правила, позволяющие влиять на протекание химического превращения.

В данной работе исследуется протекание процесса карбонилирования ряда олефинов. Данное химическое превращение хорошо изучено, поэтому удалось собрать достаточно представительную выборку [4-9], содержащую экспериментальные данные о влиянии различных условий на получение целевого продукта.

В состав результирующей выборки включены:
конверсия исходного вещества (целевой параметр),
селективность (целевой параметр),
давление,

температура,
состав исходного олефина,
состав катализатора (для входных параметров).

Давление пересчитано в атмосферах, температура в градусах Цельсия. На основе информации о химическом составе были сформированы атрибуты булева типа:

`l_has_p` — в состав лиганда входит атом фосфора,
`substr_n_c` — количество атомов углерода в олефине,
`substr_o1` — субстрат является спиртом,
`prec_cl` — катализатор содержит атом хлора,
`acid_type` — тип кислоты (органическая или не органическая).

Построение дерева классификации является полезным при выявлении управляющих воздействий на ХТП. При построении дерева классификации используются следующие критерии – примесь Джини и энтропия.

Примесь Джини считается по формуле:

$$gini = \sum_{i=1}^N (1 - p_i^2) \quad (1)$$

Энтропия считается по формуле Шеннона:

$$entropy = - \sum_{i=1}^N p_i \log_2(p_i) \quad (2)$$

В приведенных выше формулах p_i — вероятность появления классов.

При построении дерева классификации предварительная классификация выполняется только для целевого параметра, представленного тремя классами (для разбиения на классы используются одинаковые интервалы). Для входных параметров сохраняются исходные данные непрерывного типа.

Исходная выборка разбивается на две – обучающую выборку и контрольную выборку, позволяющую оценить качество классификации.

На рис. 1 представлено дерево классификации (ограниченное четырьмя уровнями), сформированное с применением алгоритма `DecisionTreeClassifier`, реализованного в библиотеке `sclearn` для Python.

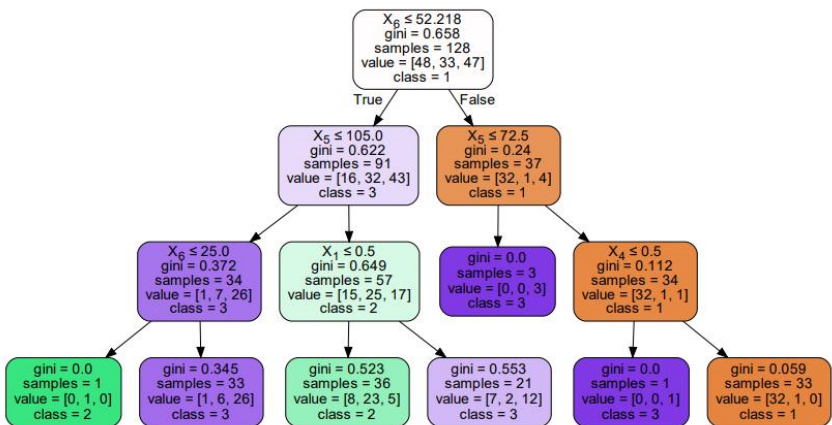


Рис. 1. Дерево классификации, сформированное на основе экспериментальной выборки с применением алгоритма DecisionTreeClassifier

При построении дерева найдены два наиболее информативных параметра – давление и температура. Интерпретация результатов показывает, что максимальная селективность процесса достигается в интервале температур от 72 до 105 °С и при давлении до 25 атм.

На основе кинетической модели процесса карбонилирования стирола [11] сформирована выборка, содержащая 30420 записей. Данная выборка содержит следующие параметры:

- Концентрацию стирола,
- Концентрацию воды,
- Концентрацию катализатора,
- Давление,
- Температуру.

На основе данной выборки построено дерево регрессии с использованием алгоритма DecisionTreeRegressor из пакета tree, входящего в библиотеку sklearn для Python. Данный алгоритм строит бинарное дерево, отличаясь от алгоритма построения дерева классификации DecisionTreeClassifier критерием разбиения множества примеров на два подмножества. В данном алгоритме используется критерий MSE (Mean Square Error), минимизирующий среднеквадратичную ошибку, вычисляемую для выходного параметра на множестве примеров, получаемых в результате разбиения. Модельная выборка также была разбита на две - обучающую (21293 записей) и

тестовую. На обучающей выборке было построено дерево регрессии (рис. 2).

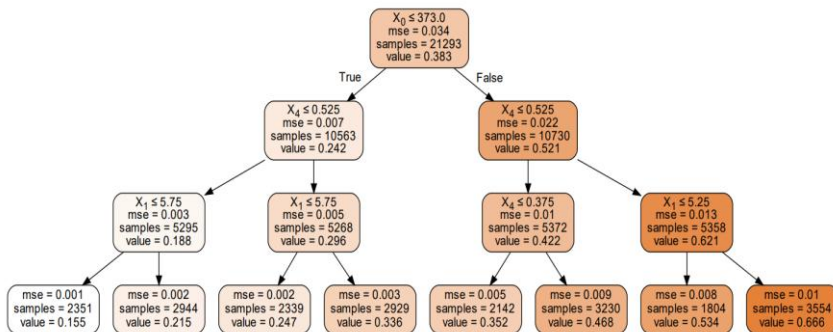


Рис. 2. Дерево регрессии, сформированное с применением алгоритма DecisionTreeRegressor по модельной выборке

3. Управление химико-технологическими процессами

Деревья решений позволяют прогнозировать значения целевых параметров при тех или иных управляющих воздействиях. Важной задачей является получение оценки стоимости управляющих воздействий на технологический процесс. Управление ХТП может сталкиваться с рядом ограничений, которые могут влиять на выбор целевых значений объекта управления. Так, проведение ХТП с выходом более 90% при высокой температуре может быть менее оправданным, чем проведение этого же ХТП при более низкой температуре и давлении. Например, для модельной выборки, описанной в данной работе, давление меняется в интервале от 4 до 8 атмосфер, температура в интервале от 95 до 105 °С. Исходя из правил Зеленой химии, сформулированной в [13], необходимо учитывать энергетические затраты и их влияние на окружающую среду и стоимость продукта. Синтез желательно проводить при температуре, близкой к температуре окружающей среды, и при атмосферном давлении.

В работе [14] предложены оценки стоимости температуры TS и давления TP при проведении ХТП. Исходя из необходимости уменьшить стоимость давления и температуры, повысив конверсию продукта, в [14] предложено несколько оценок ε_{tot} , в данной работе оценка эффективности предложена в виде:

Очевидно, наиболее эффективны физические условия с минимальным значением величины ε_{tot} . В табл. 1 представлены оценки эффективности условий протекания процесса карбонилирования стирола, рассчитанные по формуле (3) на основе экспериментальных данных из [8].

Таблица 1

Оценка эффективности условий проведения процесса карбонилирования стирола

№	$conv$	$T, ^\circ C$	$P, атм$	T_S	P_S	$1-conv$	ε_{tot}
1	0.73	40	30	0.18	0.26	0.27	0.71
2	0.81	60	30	0.08	0.26	0.19	0.53
3	0.89	80	30	0.02	0.26	0.11	0.39
4	0.98	100	30	0.00	0.26	0.02	0.28
5	0.93	100	20	0.00	0.51	0.07	0.58
6	0.75	100	40	0.00	0.69	0.25	0.94

Исходя из выбранной методики оценки, наиболее эффективны условия из эксперимента №4.

Заключение

В работе рассматриваются ряд вопросов управления химико-технологическими процессами. Важной задачей является выявление управляющих воздействий, для поиска которых применены методы и алгоритмы Data Mining. Поиск управляющих воздействий показан на примере процесса карбонилирования олефинов, имеющем важное практическое значение. В частности, показано большое влияние на протекания данного процесса температуры и давления. Показаны также способы оценки условной стоимости управляющих воздействий на протекание процесса.

Список литературы

1. Кафаров В. В., Перов В. Л., Мешалкин В. П. Принципы математического моделирования химико-технологических систем (Введение в системотехнику химических производств). М.: Химия, 1974. – 344 с.
2. Вильямс, Т. Дж. Проектирование химико-технологических процессов методами системотехники. Ленинград: Химия, 1967. – 187 с.

3. Указатель пакетов Python'a (Python Package Index — PyPI). – Режим доступа: <https://pypi.org/project/chemformula/>
4. Regioselective synthesis of ibuprofen via the palladium complex catalyzed hydrocarboxylation of 1-(4-isobutylphenyl) ethanol / Jang E. J. et al // *Journal of Molecular Catalysis A: Chemical*, 138 (1999), 25–36.
5. Ionescu A. Regioselectivity in aqueous palladium catalysed hydroxycarbonylation of styrene: a catalytic and mechanistic study / Ionescu A., Laurenczy G., Wendt O. F. // *Dalton Trans.*, 2006, 3934–3940.
6. Highly selective hydrocarboxylation of styrene with oxalic acid or water using pallaalum ortylo-ammo arenethiolates with intramolecular coordinating nitrogen Lewis bases / Kruis D. et al. // *Inorganic Chemistry Communications*, 1 (1998), 295–298.
7. Bis-alkoxycarbonylation of styrene by pyridinimine palladium catalysts / Bianchini C. et al // *New J. Chem.*, 2002, 26, 387–397.
8. Hydroxycarbonylation of styrene with palladium catalysts. The influence of the mono- and bidentate phosphorus ligand / del Rio I. et al // *Journal of Molecular Catalysis A: Chemical*, 161 (2000), 39–48.
9. Jayasree S. Highly active supported palladium catalyst for the regioselective synthesis of 2-arylpropionic acids by carbonylation / Jayasree S., Seayad A., Chaudhari R. V. // *Chem. Commun.*, 1999, 1067–1068.
10. Полак Л. С., Гольденберг М. Я., Левицкий А. А. Вычислительные методы в химической кинетике. М.: Наука, 1984. – 280 с.
11. Li Y. Studies in hydrocarboxylation of styrene and derivatives using palladium complex catalysts. PhD Thesis, 2010.
12. Биллиг В. А. Алгоритмы Data Mining при поиске эффективных условий проведения химических реакций / Биллиг В. А., Звягинцев Н. В. // *Вестник ТвГУ. Серия: Прикладная математика*, 2021, № 4, 29–42.
13. Anastas, P. T.; Warner, J. C. *Green Chemistry: Theory and Practice*, Oxford University Press: New York, 1998. – 135 p.
14. Звягинцев Н. В. Оценка эффективности условий проведения химических реакций / Звягинцев Н. В., Биллиг В. А. // Программные