

Методы интеллектуального анализа
данных в управлении химико-
технологическими процессами

Звягинцев Н.В.

к.т.н., доцент, с.н.с. Биллиг В.А.

Тверской государственный технический университет

Актуальность

- Имеется значительное количество экспериментальных данных о протекании различных химических процессов
- Активное развитие методов машинного обучения, в том числе методов Data Mining
- Важность повышения эффективности проведения химических процессов, выбор условий, снижающих влияние на окружающую среду и т.д.

Особенности химико- технологических процессов

Необходимо учитывать:

- факторы различной физической природы
- химическую структуру всех участников химического процесса
- время протекания процесса

Data Mining и «традиционные» методы моделирования

Под традиционными методами подразумеваются методы химической кинетики и квантово-химические методы. Возможна ли комбинация традиционных методов с методами Data Mining? Конечно, например:

- Формирование выборок на основе кинетических моделей
- Применение дескрипторов, рассчитанных в том числе квантово-химическими методами

Подготовка данных

Подготовка данных включает в себя следующие этапы:

- Преобразование сведений о химической структуре в набор целочисленных векторов с информацией о количестве определенных атомов, фрагментов или химических связей в соединении
- Масштабирование данных
- Перерасчет данных в целевые единицы измерения

Формирование выборок

В данной работе использовались две выборки:

- Выборка, полностью сформированная на основе экспериментальных данных (183 записи)
- Модельная выборка, сформированная на основе кинетической модели, путем интегрирования системы дифференциальных уравнений, описывающих кинетику процесса (30420 записей).

Какие инструменты применялись?

- Поиск ассоциативных правил, в том числе алгоритм **Apriori**.
- Построение деревьев на основе подготовленных выборок (например, деревья классификации и деревья регрессии).
- Использовались собственные реализации алгоритма **Apriori**, а также реализации из библиотеки *sklearn* для Python.

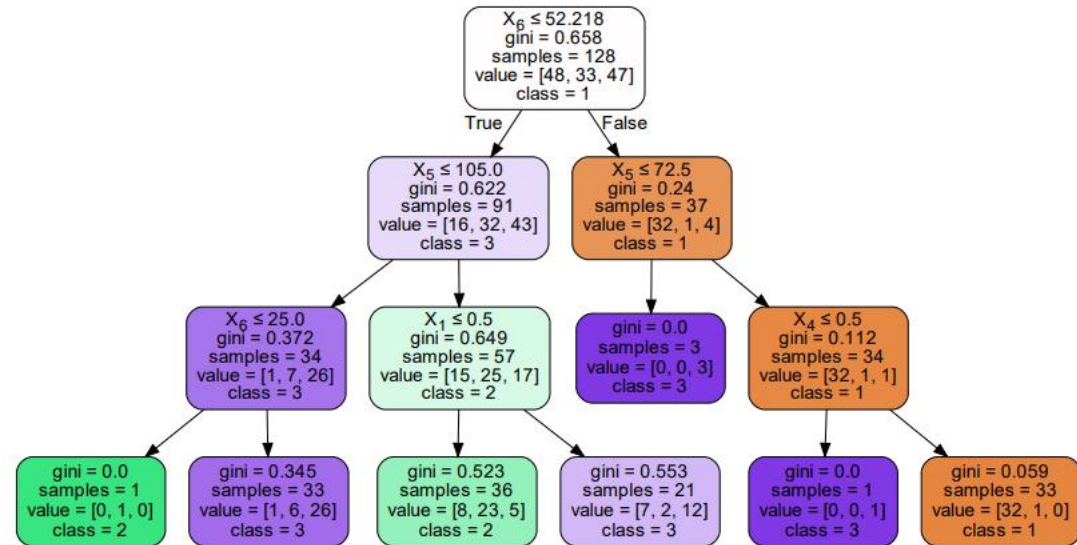
Пример: построение дерева классификации

Использовался алгоритм *DecisionTreeClassifier* (примесь Джини в качестве критерия классификации).

Построение дерева на основе экспериментальных данных.

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

- конверсия исходного вещества
- селективность целевые параметры),
- давление,
- температура,
- состав исходного олефина,
- состав катализатора.



Пример: построение дерева регрессии

Использовался алгоритм *DecisionTreeRegressor* (энтропия в качестве критерия классификации). Модельная выборка

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

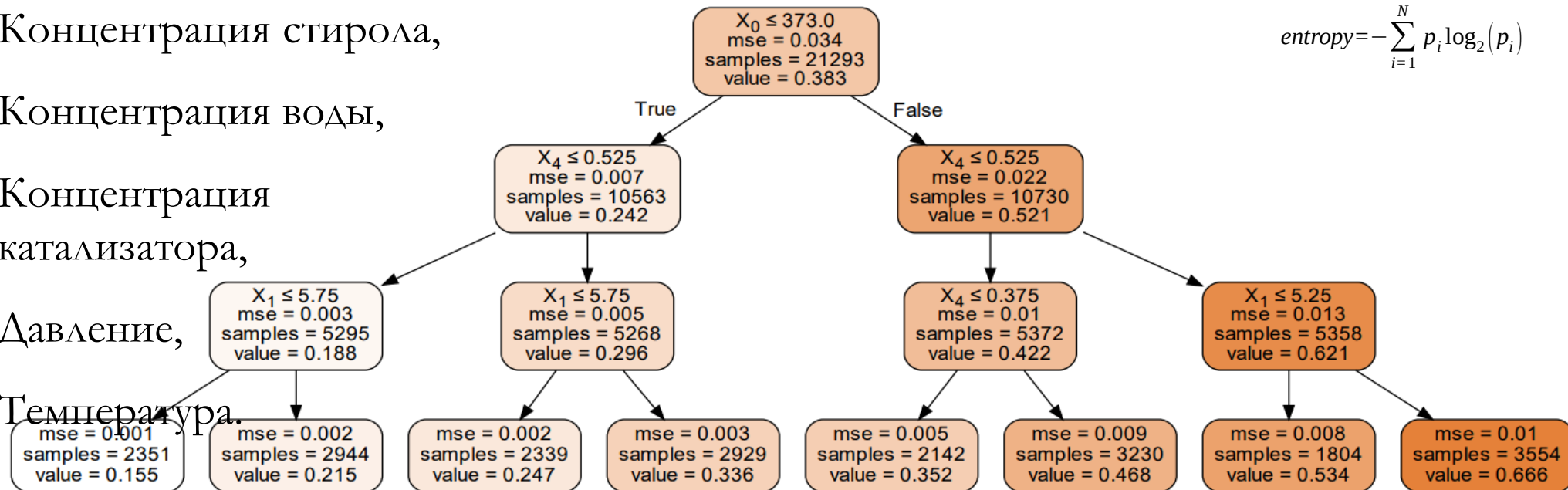
Концентрация стирола,

Концентрация воды,

Концентрация катализатора,

Давление,

Температура.



Управление процессом карбонилирования олефинов

$$S_T = \begin{cases} 0, \text{если } T < T_{\min} \text{ или } T > T_{\max} \\ (T - T_E) / (T_{\min} - T_E), \text{если } T \in [T_{\min}, T_E] \\ (T_{\max} - T) / (T_{\max} - T_E), \text{если } T \in [T_E, T_{\max}] \end{cases}$$

$$S_P = \begin{cases} 0, \text{если } P \notin [P_{\min}, P_{\max}] \\ (P - P_{\min}) / (P_{\max} - P_{\min}), \text{если } P \in [P_{\min}, P_{\max}] \end{cases}$$

$$\epsilon_{tot} = \begin{cases} \epsilon = S_T + S_P, \text{если } \epsilon < 1 \\ 1, \text{если } \epsilon \geq 1 \end{cases}$$

№	кон v	T, °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.

Выводы

- Методы интеллектуального анализа данных являются удобным инструментом, позволяющим выбирать наиболее эффективные условия проведения химических процессов.
- Предложенные методы апробированы на процессе карбонилирования олефинов, имеющем важное практическое значение, в том числе в фармацевтической промышленности при производстве препаратов группы ибупрофена.