

АЛГОРИТМЫ DATA MINING ПРИ ПОИСКЕ ЭФФЕКТИВНЫХ УСЛОВИЙ ПРОВЕДЕНИЯ ХИМИЧЕСКИХ РЕАКЦИЙ

Биллиг В.А., Звягинцев Н.В.

Тверской государственной технической университет, г. Тверь

Поступила в редакцию 25.11.2021, после переработки 20.12.2021.

В настоящее время накоплено значительное количество экспериментальных данных, фиксирующих процесс протекания химических реакций. Анализ этих данных комплексом алгоритмов Data Mining дает важную практическую информацию для поиска эффективных условий проведения реакций, при которых получается максимальное количество целевого продукта при минимальных затратах. В данной работе на примере работы с базой, содержащей данные о протекании реакции карбонилирования различных олефинов, показано, как разработанный нами программный комплекс позволяет извлечь полезные знания, способствующие повышению эффективности химических реакций.

Ключевые слова: Data Mining, ассоциативные правила, деревья решений, деревья регрессий, химические реакции.

Вестник ТвГУ. Серия: Прикладная математика. 2021. № 4. С. 29–42.
<https://doi.org/10.26456/vtprm625>

Введение

Данная статья является расширением доклада, сделанного на Всероссийской конференции «Математические основы информатики и информационно коммуникационных систем» (2021). Химические реакции (ХР) являются сложными процессами, на протекание которых оказывает влияние масса условий: давление, температура, состав и природа взаимодействующих веществ и катализаторов. Вместе с тем, важной задачей является поиск наиболее эффективных условий, когда при минимальных затратах получается максимальное количество целевого продукта. Для решения этой задачи необходимо проанализировать влияние всех условий на протекание ХР. Решение можно осуществить несколькими путями — моделированием механизма и кинетики самой ХР [1], а также применением комплекса алгоритмов Data Mining к экспериментальным данным о протекании ХР. Второй подход чаще всего является менее затратным с вычислительной точки зрения, однако требует существенного объема экспериментальных данных и разработки подходов к кодированию информации о факторах протекания химических реакций [2].

1. Формирование БД

Важным этапом второго подхода является формирование БД с результатами экспериментов при различных условиях. Данные об условиях протекания химических реакций являются слабосвязанными данными. В различных публикациях рассматривают влияние различного набора факторов, данные могут быть указаны в различных единицах измерения. В рамках данной работы исследуется процесс протекания реакции карбонилирования различных олефинов. Данная ХР хорошо экспериментально изучена, поэтому удалось собрать достаточно представительную выборку [3–8], содержащую данные о влиянии различных условий на получение целевого продукта. В состав результирующей выборки включены:

1. конверсия исходного вещества и селективность (для формирования выходных данных — целевые параметры),
2. давление, температура, состав исходного вещества и состав катализатора (для входных параметров).

Данные о давлении были пересчитаны в атмосферах, данные о температуре приведены к градусам Цельсия, а на основе информации о химическом составе были сформированы атрибуты булева типа:

1. *l_has_p* — в состав лиганда входит фосфор,
2. *substr_n_c* — количество атомов углерода в исходном веществе,
3. *substr_ol* — исходное вещество является спиртом,
4. *prec_cl* — катализатор содержит хлор,
5. *acid_type* — тип кислоты (органическая или не органическая).

Созданная выборка, содержащая 183 записи, является частью БД, содержащей больше входных параметров. Мы оставили наиболее информативные параметры, определенные по данным предварительных исследований.

2. Комплексный анализ данных

Эксперт, занимающийся анализом данных, обычно хочет получить ответы на вопросы: «Каковы ожидаемые значения целевых параметров при заданном наборе входных параметров?», «Каковы наиболее информативные параметры, влияющие на результат, и как управлять ими для достижения нужного выхода?». Зачастую для эксперта важны не столько конкретные значения параметров, сколько качественные оценки. Ему необходимы правила, позволяющие предсказать класс, которому принадлежит целевой параметр при управлении входными данными. Понятно, что эксперту нужны не только правила, но и характеристики этих правил, такие как достоверность правила, частота применения правила, возможно и другие характеристики. По нашему глубокому убеждению, ответы на эти вопросы можно получить, применяя широкий арсенал методов извлечения знаний из данных, известных как методы Data Mining.

3. Программная платформа

Для анализа данных мы используем программный комплекс, содержащий три модуля:

1. Предварительная подготовка данных.
2. Комплекс алгоритмов анализа данных.
3. Визуализация и объяснение результатов анализа.

Для части алгоритмов анализа данных разработана собственная реализация, для части используются алгоритмы, входящие в пакет `sklearn` языка Python. Предварительная подготовка данных является важной частью работы с данными. На этом этапе решаются такие проблемы как восстановление пропусков в записях базы данных, а, главное, преобразование данных к виду, требуемому тем или иным алгоритмом. Одни алгоритмы могут работать только с транзакционными данными, другие – с категориальными данными, третьи – с непрерывными данными, возможно, приведенными к одному масштабу. Комплекс алгоритмов включает алгоритмы:

1. Построения ассоциативных правил.
2. Различные вариации алгоритма кластеризации `k-means` для работы с непрерывными данными и алгоритма `CLOPE` для работы с категориальными (транзакционными) данными.
3. Построения деревьев решений для задачи классификации.
4. Построения деревьев решений для задачи регрессии. Построения ассоциативных правил.
5. Различные вариации алгоритма кластеризации `k-means` для работы с непрерывными данными и алгоритма `CLOPE` для работы с категориальными (транзакционными) данными.
6. Построения деревьев решений для задачи классификации.
7. Построения деревьев решений для задачи регрессии.

Остановимся на некоторых алгоритмах и результатах их работы.

4. Алгоритм `ConApriori` построения ассоциативных правил

У разработанного нами алгоритма `ConApriori` есть два существенных отличия в сравнении с классическим алгоритмом `Apriori`:

1. способ представления данных,
2. способ построения достоверных ассоциативных правил.

Каждая запись базы данных представлена одним числом. Числовое представление записей базы данных не зависит от размера записи и позволяет эффективно вычислять значение базовой функции Support за время $O(N)$ с минимальной константой, где N – число записей базы данных. В отличие от классического алгоритма, в котором строятся частые правила на основе ранее построенных частых правил, в данном алгоритме строятся достоверные правила на основе ранее построенных достоверных правил, что повышает эффективность алгоритма. Подробное описание алгоритма приведено в работе [9]. На Рис. 1 показан начальный этап процесса построения ассоциативных правил. На этом этапе можно

1. выбрать базу данных и связаться с ней,
2. показать параметры базы данных с их возможным содержательным описанием,
3. показать базу данных,
4. задать требуемые характеристики правил,
5. построить правила.

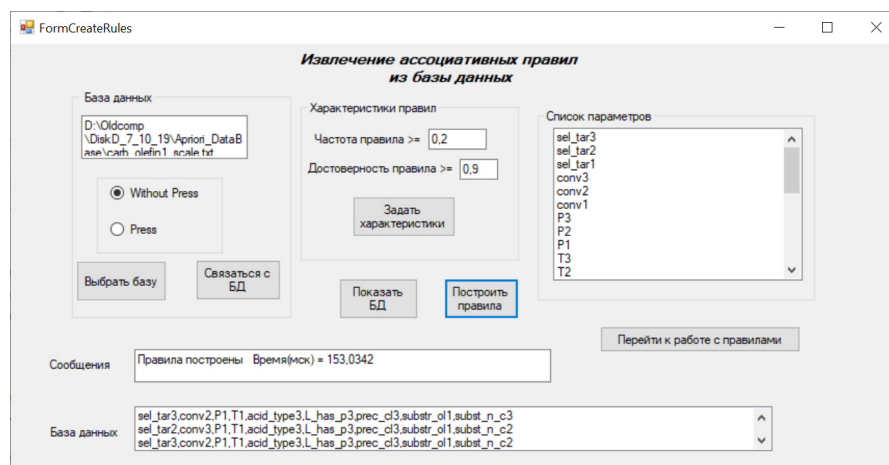


Рис. 1: Начальный интерфейс проекта построения ассоциативных правил

Алгоритм построения правил позволяет находить правила с заданной частотой и достоверностью. Кроме этого, для каждого правила вычисляется характеристика, называемая lift, определяющая степень корреляции между посылкой правила и его заключением. Анализ полученных правил для рассматриваемой нами базы данных показывает, что с высокой достоверностью, большей 0,9, и с достаточно высокой частотой появления правила, большей 0,2, можно определить лишь один класс выходного продукта – *sel_tar1*. Таких правил найдено 16. Приведем первые три из этих правил:

P2 => sel_tar1: частота = 0,21; достоверность = 0,93; лифт = 2,54

P2, T1 => sel_tar1: частота = 0,21; достоверность = 0,93; лифт = 2,54
 P2, acid_type3 => sel_tar1: частота = 0,21; достоверность = 0,95;
 лифт = 2,60

Определяющим фактором, влияющим на результат, является давление. Параметр P2 присутствует в посылке всех правил, заключение которых содержит выходной параметр *sel_tar1*. Правила, позволяющие определить два других класса выходного продукта, существуют, но их достоверность ниже достоверности определения класса *sel_tar1*.

5. Алгоритм построения дерева классификации DecisionTreeClassifier

Для построения дерева классификации мы использовали алгоритм, реализованный в Python пакете *tree*, входящем в состав пакета *sclearn*. При построении дерева классификации можно использовать два критерия – примесь Джини и энтропию. Примесь Джини считается по следующей формуле:

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

Для критерия энтропии применяется известная формула Шеннона:

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

В обеих формулах используются p_i – вероятности появления классов, рассчитываемые как частоты появления класса в выборке. Для крайних случаев, когда в выборке присутствует ровно один класс, или когда частота появления всех N классов одинакова, обе формулы принимают минимально и максимально возможные значения. Минимальное значение для обеих формул, когда существует p_k , равное 1, а остальные p_i , ($i \neq k$) равны 0:

$$\min(gini) = \min(entropy) = 0.$$

Максимальное значение *gini*, когда все p_k равны $1/N$:

$$\max(gini) = 1 - 1/N.$$

Максимальное значение *entropy*, когда все p_k равны $1/N$:

$$\max(entropy) = \log_2 N.$$

Качественно оба критерия отвечают требованиям, предъявляемым к подобным критериям. Чем сильнее перемешаны классы в записях базы данных, чем выше неопределенность в предсказании класса, тем большие значения имеют оба критерия. Для наших данных оба критерия строят качественно похожие деревья. Для построения дерева классификации используется та же БД, что и при построении ассоциативных правил. Для построения ассоциативных правил все параметры классифицируются, так что каждая запись базы данных содержит значение

параметра только одного класса. При построении дерева классификации предварительная классификация выполняется только для целевого параметра, представленного тремя классами, такими же, как и при построении ассоциативных правил. Для входных параметров сохраняются исходные данные непрерывного типа. Как это обычно делается, исходная база данных разбивается на два подмножества – обучающую выборку и контрольную выборку, позволяющую оценить качество классификации. Дерево решений для классификации целевого параметра *sel_tar* строится по обучающей выборке, содержащей 128 записей, в которых целевой параметр представлен тремя классами примерно равной мощности. На Рис. 2 показано построенное дерево классификации, ограниченное четырьмя уровнями, практически содержащими всю полезную информацию.

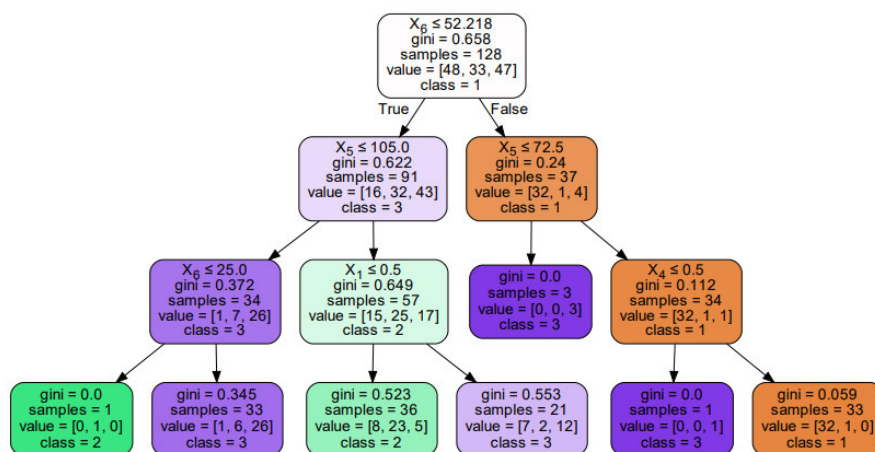


Рис. 2: Дерево классификации

В результате анализа этого дерева можно построить следующие правила:

```

P > 52 & T > 72 => sel_tar1
P < 52 & T > 105 & substr_ol = "спирт" => sel_tar2
P > 25 & P < 52 & T < 105 => sel_tar3
  
```

В корне дерева имеет место следующая ситуация: Параметр *gini* = 0.66, всего примеров 128, примеров каждого класса целевого параметра соответственно (48, 33, 47). При построении дерева находятся два наиболее информативных параметра — давление (X_6) и температура (X_5). Первое из приведенных выше правил позволяет распознать, что целевой параметр относится к первому классу, с вероятностью 0.93 и частотой появления правила 0.25. Эти результаты хорошо согласуются с результатами построения ассоциативных правил. Второе и третье правило позволяют распознавать второй и третий класс целевого параметра. Частота применения этих правил также находится на уровне 0.25, но достоверность распознавания ниже достоверности распознавания первого класса. Для второго класса достоверность равна 0.63, для третьего — 0.77. Рассматривать правила с меньшей частотой особого смысла не имеет, скорее всего, они носят случайный

характер, поскольку примеров таких правил немного. Представляет интерес построение оценки суммарной достоверности правильного распознавания классов. Учитывая, что у нас имеется контрольная выборка, можно построить эту оценку, определив в соответствии с построенными правилами классы по входным данным контрольной выборки и сравнив полученные результаты с известными результатами, имеющимися в базе данных для этой выборки. На Рис. 3 показана гистограмма ошибок предсказания:

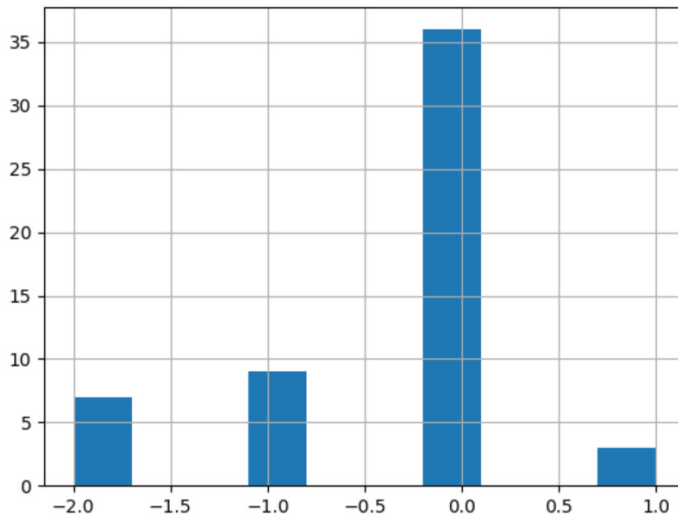


Рис. 3: Гистограмма ошибок предсказания классов

Из 55 примеров контрольной выборки в 36 случаях классы распознаются без ошибок, так что средняя достоверность правильного распознавания находится на уровне 0.65. Еще в 12 случаях ошибка предсказания не более одного класса. Классы распознаются с разной достоверностью.

6. Алгоритм построения дерева регрессии `DecisionTreeRegressor`

Если дерево классификации позволяет предсказать появление того или иного класса для заданных входных данных, то дерево регрессии позволяет предсказать конкретное значение выходного параметра. Понятно, что корректное предсказание конкретного значения — это более трудная задача, требующая больше информации, требующая существенного увеличения объема исходной базы данных. К сожалению, собранные нами экспериментальные данные позволяют проводить лишь качественный анализ на уровне классов, а не на уровне конкретных значений. Как было показано с достаточной частотой и достоверностью удается построить правила, позволяющие предсказать лишь некоторые классы выходного параметра `sel_tar`, отвечающего за селективность. Классы второго выходного параметра — определяются с достоверностью, не превышающей значения 0.5. Для построения дерева регрессии мы использовали созданную нами модельную базу данных. Входными данными модели являются:

1. Концентрация стирола в качестве основного компонента химической реакции.
2. Концентрации воды и катализатора.
3. Давление и температура, как факторы, обеспечивающие протекание реакции.

Выходным параметром, как и в случае экспериментальных данных, является параметр conv , отвечающий за конверсию. Модельная база данных включает 30420 записей, в которых рассмотрены различные комбинации входных данных.

Алгоритм построения модельной базы данных приведен в следующем разделе, а сейчас рассмотрим результаты построения дерева регрессии по данным модельной базы данных. Для построения дерева регрессии мы использовали алгоритм `DecisionTreeRegressor` из пакета `tree`, входящего в состав пакета `sklearn Python`. Алгоритм строит бинарное дерево и по сути отличается от алгоритма построения дерева классификации лишь критерием разбиения множества примеров на два подмножества. Если в алгоритме классификации используются критерии энтропии и Джини, характеризующие степень перемешивания классов в примерах, то при построении дерева решений используется критерий mse (mean square error), - минимизирующий среднеквадратичную ошибку, вычисляемую для выходного параметра на множестве примеров, получаемых в результате разбиения. Чем меньше разброс значений выходного параметра, тем лучше разбиение, тем точнее можно предсказать результат. Как обычно, модельная база данных разбивается на два подмножества – обучающую выборку и контрольную выборку, позволяющую оценить качество регрессии. Дерево решений для целевого параметра строится по обучающей выборке, содержащей 21293 записи. На Рис. 4 показано построенное дерево регрессии, ограниченное четырьмя уровнями.

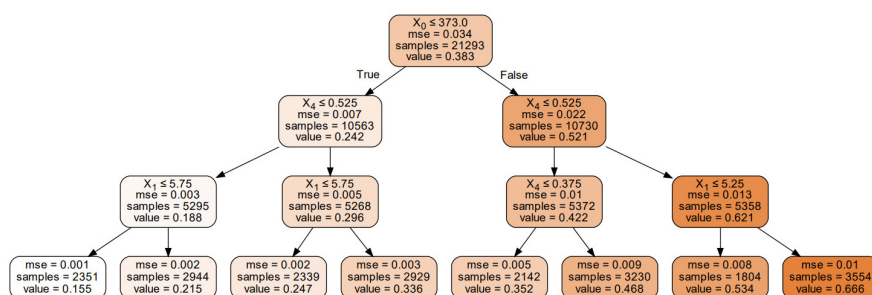


Рис. 4: Дерево регрессии модельной базы данных

Анализ построенного дерева позволяет построить ряд правил. Приведем два из этих правил:

1. Температура (X_0) ? 373 & Концентрация катализатора (X_4) \leq 0.525 & Давление ? 5.75 => Значение целевого параметра минимально ($\text{conv} = 0.155$)
2. Температура (X_0) ?373 & Концентрация катализатора (X_4) \geq 0.525 & Давление ?5.25 => Значение целевого параметра максимально ($\text{conv} = 0.688$)

Мы ограничили высоту дерева четырьмя уровнями. Это позволяет строить правила, задающие значение выходного параметра с высокой частотой. Число примеров для каждого листа построенного дерева составляет примерно 2000–3000, то есть примерно 10% от общего числа записей. Вместе с тем критерий mse в каждом листе достаточно мал, примерно в 10 раз меньше в сравнении со значением этого параметра в корне дерева. Конечно, разрешив строить дерево большей глубины, можно добиться существенного уменьшения значения этого критерия. Естественно, как следствие, уменьшится частота появления соответствующих правил. Так, ограничив дерево решений шестью уровнями, удалось снизить среднеквадратичную ошибку более чем в два раза, сохранив приемлемое число примеров в листьях дерева (500 и более примеров). Мы не приводим эти результаты, поскольку графическое отображение такого дерева вызывает естественные затруднения ввиду размера дерева. Контрольная выборка позволяет сравнить предсказанные значения со значениями выходного параметра, хранящимися в базе данных. На Рис. 5 показана гистограмма ошибок предсказания.

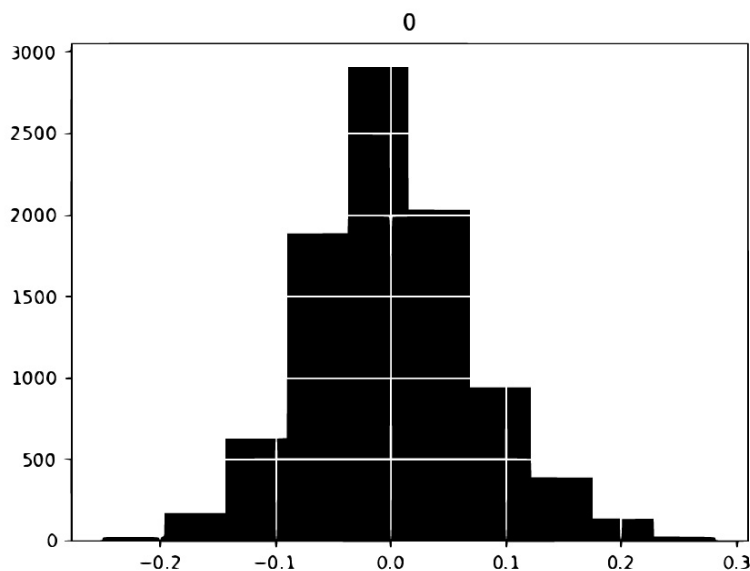


Рис. 5: Гистограмма ошибок предсказания для дерева регрессии

По виду гистограммы можно сделать вывод, что ошибки предсказания согласуются с нормальным законом распределения.

7. Алгоритм построения модельной базы данных

Аппарат Data Mining можно применять не только к экспериментальным данным, но и к результатам, полученным при моделировании реакций. База данных для построения дерева регрессии была построена нами на основе кинетической

модели, сводящейся к интегрированию системы дифференциальных уравнений. Опуская детали, рассмотрим схему построения кинетической модели химической реакции. Химическая реакция, преобразующая исходные вещества в некоторый продукт, может быть представлена в общем виде:



где:

ν_i — стехиометрические коэффициенты,

A_i — исходные вещества,

B — продукт химической реакции.

Химическая реакция разворачивается во времени. С течением времени меняется концентрация исходных веществ и целевого продукта. Скорость протекания протой химической реакции $w(t)$ определяется законом действующих масс:

$$w(t) = k \prod_{i=1}^N C_{A_i}^{\nu_i}, \quad (4)$$

где:

ν_i — стехиометрические коэффициенты,

C_{A_i} — исходные вещества,

k — константа химической реакции, определяемая уравнением Аррениуса:

$$k(T) = A \exp(-E_{act} / RT). \quad (5)$$

В уравнении (5): A — экспоненциальный множитель,

E_{act} — энергия активации элементарной стадии,

R — универсальная газовая постоянная,

T — температура.

В большинстве случаев химическая реакция является сложной, - ее протекание определяется совокупностью параллельно или последовательно протекающих стадий, каждая из которых описывается стехиометрическим уравнением (3). Скорости изменения концентрации каждого из веществ, участвующих в реакции, различны и независимы. Кинетическая модель сложной химической реакции задается системой уравнений, каждое из которых описывает скорость реакции соответствующего вещества:

$$w_{A_i}(t) = \sum_{k=1}^M d_k w_k(t), \quad (6)$$

где: w_{A_i} — скорость изменения концентрации компонента реакции,

w_k — скорость стадии химической реакции для данного компонента, d_k — весовой коэффициент. Зная уравнения w_{A_i} в соотношении (6), можно интегрировать выражения вида и найти изменения концентраций:

$$dC_{A_i} = w_{A_i}(t) dt. \quad (7)$$

Для данной математической модели химической реакции, имея массив исходных концентраций всех веществ, в том числе промежуточных, а также зная константы

химических реакций, можно рассчитать эволюцию системы. При интегрировании использовался метод Эйлера. При формировании кинетических уравнений для создания нашей модельной БД были использованы результаты работы [10]. Задав сетку исходных значений концентрации, температуры, времени, давления, и решив соответствующую систему уравнений, мы сгенерировали модельную БД.

Заключение

1. Рассматриваемые нами базы данных, как экспериментальная, так и модельная, позволяют извлечь знания, которые могут помочь эксперту, занимающемуся исследованием эффективности протекания процесса карбонилирования олефинов.

2. Комплексное применение алгоритмов Data Mining повышает надежность и способствует улучшению качества анализа данных.

3. Исследование непосредственно реакции карбонилирования показало, что существенное влияние на эффективность реакции оказывает давление.

Список литературы

- [1] Звягинцев Н.В., Елисеев О.Л., Кондратьев Л.Т., Лapidус А.Л. Исследование влияния природы лигандов на региоселективность реакции карбонилирования стирола в присутствии комплексов палладия (II) // Доклады Академии наук. 2010. Т. 434, № 2. С. 189–195.
- [2] Биллиг В.А., Звягинцев Н.В. Информационная система обработки и хранения данных о кинетике химических реакций // Программные продукты и системы. 2018. Т. 31, № 4. С. 808–813. <http://dx.doi.org/10.15827/0236-235X.124.808-813>
- [3] Jang E.J., Lee K.H., Lee J.S., Kim Y. Regioselective synthesis of ibuprofen via the palladium complex catalyzed hydrocarboxylation of 1-(4-isobutylphenyl) ethanol // Journal of Molecular Catalysis A-chemical. 1999. Vol. 138. Pp. 25–36. [https://doi.org/10.1016/S1381-1169\(98\)00142-3](https://doi.org/10.1016/S1381-1169(98)00142-3)
- [4] Ionescu A., Laurency G., Wendt O.F. Regioselectivity in aqueous palladium catalysed hydroxycarbonylation of styrene: a catalytic and mechanistic study // Dalton Transactions. 2006. № 32. Pp. 3934–3940. <https://doi.org/10.1039/B607331J>
- [5] Kruis D., Ruiz N., Janssen M.D., Boersma J., Claver C., Koten G. Highly selective hydrocarboxylation of styrene with oxalic acid or water using palladium ortho-amino arenethiolates with intramolecular coordinating nitrogen Lewis bases // Inorganic Chemistry Communications. 1998. Vol. 1, № 8. Pp. 295–298. [https://doi.org/10.1016/S1387-7003\(98\)00079-3](https://doi.org/10.1016/S1387-7003(98)00079-3)
- [6] Bianchini C., Lee H.M., Mantovani G., Meli A., Oberhauser W. Bisalkoxycarbonylation of styrene by pyridinimine palladium catalysts // New Journal of Chemistry. 2002. № 4. Pp. 387–397. <https://doi.org/10.1039/B108804C>

- [7] Rio I., Ruiz N., Claver C., Veen L., Leeuwen P. Hydroxycarbonylation of styrene with palladium catalysts: The influence of the mono- and bidentate phosphorus ligand // *Journal of Molecular Catalysis A: Chemical*. 2000. Vol. 161, № 1-2. Pp. 39–48. [https://doi.org/10.1016/S1381-1169\(00\)00359-9](https://doi.org/10.1016/S1381-1169(00)00359-9)
- [8] Jayasree S., Seayad A., Chaudhari R.V. Highly active supported palladium catalyst for the regioselective synthesis of 2-arylpropionic acids by carbonylation // *Chemical Communications*. 1999. № 12. Pp. 1067–1068. <https://doi.org/10.1039/A902541C>
- [9] Billig V.A. Effective algorithm for constructing associative rules // *Software and Systems*. 2017. Vol. 30, № 2. Pp. 196–206. <http://dx.doi.org/10.15827/0236-235X.118.196-206>
- [10] Li Y. Studies in hydrocarboxylation of styrene and derivatives using palladium complex catalysts: PhD Thesis. 2010.

Образец цитирования

Биллиг В.А., Звягинцев Н.В. Алгоритмы Data Mining при поиске эффективных условий проведения химических реакций // *Вестник ТвГУ. Серия: Прикладная математика*. 2021. № 4. С. 29–42. <https://doi.org/10.26456/vtprmk625>

Сведения об авторах

1. **Биллиг Владимир Арнольдович**

профессор кафедры ПО факультета ИТ Тверского государственного технического университета.

Россия, 170026, г. Тверь, наб. Афанасия Никитина, д. 22.

E-mail: Vladimir-Billig@yandex.ru

2. **Звягинцев Николай Васильевич**

аспирант кафедры ПО факультета ИТ Тверского государственного технического университета.

Россия, 170026, г. Тверь, наб. Афанасия Никитина, д. 22.

E-mail: n.zvyagintsev@gmail.ru

DATA MINING ALGORITHMS IN SEARCH OF EFFECTIVE CONDITIONS FOR CONDUCTING CHEMICAL REACTIONS

Billig Vladimir Arnoldovich

Professor at IT department,
Tver State Technical University
Russia, 170026, Tver, Afanasy Nikitin embankment, 22.
E-mail: Vladimir-Billig@yandex.ru

Zvyagintsev Nikolay Vasilevich

Postgraduate student at IT department,
Tver State Technical University
Russia, 170026, Tver, Afanasy Nikitin embankment, 22.
E-mail: n.zvyagintsev@gmail.ru

Received 25.11.2021, revised 20.12.2021.

At present, a significant amount of experimental data has been accumulated, recording the process of the occurrence of chemical reactions. Analysis of these data by a set of Data Mining algorithms provides important practical information for finding effective conditions for carrying out reactions, at which the maximum amount of the target product is obtained at minimal cost. In this paper, using the example of working with a database containing data on the course of the carbonylation reaction of various olefins, it is shown how the software package developed by us allows us to extract useful knowledge that contributes to an increase in the efficiency of chemical reactions.

Keywords: Data Mining, association rules, decision trees, regression trees, chemical reactions.

Citation

Billig V.A., Zvyagintsev N.V., "Data Mining algorithms in search of effective conditions for conducting chemical reactions", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2021, № 4, 29–42 (in Russian). <https://doi.org/10.26456/vtpmk625>

References

- [1] Zvyagintsev N.V., Eliseev O.L., Kondratev L.T., Lapidus A.L., "Effect of the ligand nature on the regioselectivity of the styrene carbonylation in the presence of palladium(II) complexes", *Doklady Chemistry*, **434**:2 (2010), 189–195 (in Russian).
- [2] Billig V.A., Zvyagintsev N.V., "Information system for processing and storing information on chemical reaction kinetics", *Software and Systems*, **31**:4 (2018), 808–813 (in Russian), <http://dx.doi.org/10.15827/0236-235X.124.808-813>.

- [3] Jang E.J., Lee K.H., Lee J.S., Kim Y., “Regioselective synthesis of ibuprofen via the palladium complex catalyzed hydrocarboxylation of 1-(4-isobutylphenyl) ethanol”, *Journal of Molecular Catalysis A-chemical*, **138** (1999), 25–36, [https://doi.org/10.1016/S1381-1169\(98\)00142-3](https://doi.org/10.1016/S1381-1169(98)00142-3).
- [4] Ionescu A., Laurency G., Wendt O.F., “Regioselectivity in aqueous palladium catalysed hydroxycarbonylation of styrene: a catalytic and mechanistic study”, *Dalton Transactions*, 2006, № 32, 3934–3940, <https://doi.org/10.1039/B607331J>.
- [5] Kruis D., Ruiz N., Janssen M.D., Boersma J., Claver C., Koten G., “Highly selective hydrocarboxylation of styrene with oxalic acid or water using palladium ortho-amino arenethiolates with intramolecular coordinating nitrogen Lewis bases”, *Inorganic Chemistry Communications*, **1**:8 (1998), 295–298, [https://doi.org/10.1016/S1387-7003\(98\)00079-3](https://doi.org/10.1016/S1387-7003(98)00079-3).
- [6] Bianchini C., Lee H.M., Mantovani G., Meli A., Oberhauser W., “Bis-alkoxycarbonylation of styrene by pyridinimine palladium catalysts”, *New Journal of Chemistry*, 2002, № 4, 387–397, <https://doi.org/10.1039/B108804C>.
- [7] Rio I., Ruiz N., Claver C., Veen L., Leeuwen P., “Hydroxycarbonylation of styrene with palladium catalysts: The influence of the mono- and bidentate phosphorus ligand”, *Journal of Molecular Catalysis A: Chemical*, **161**:1-2 (2000), 39–48, [https://doi.org/10.1016/S1381-1169\(00\)00359-9](https://doi.org/10.1016/S1381-1169(00)00359-9).
- [8] Jayasree S., Seayad A., Chaudhari R.V., “Highly active supported palladium catalyst for the regioselective synthesis of 2-arylpropionic acids by carbonylation”, *Chemical Communications*, 1999, № 12, 1067–1068, <https://doi.org/10.1039/A902541C>.
- [9] Billig V.A., “Effective algorithm for constructing associative rules”, *Software and Systems*, **30**:2 (2017), 196–206, <http://dx.doi.org/10.15827/0236-235X.118.196-206>.
- [10] Li Y., *Studies in hydrocarboxylation of styrene and derivatives using palladium complex catalysts*, PhD Thesis, 2010.