

УДК 517.977.1, 004.855.5

# ВИРТУАЛЬНЫЕ АНАЛИЗАТОРЫ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ ПРИ ПЕРЕМЕННЫХ РЕЖИМАХ

**Н.Н. Бахтадзе**

*Институт проблем управления им. В.А. Трапезникова РАН*  
Россия, 117997, Москва, Профсоюзная ул., 65  
E-mail: [sung7@yandex.ru](mailto:sung7@yandex.ru)

**А.А. Черешко**

*Институт проблем управления им. В.А. Трапезникова РАН*  
Россия, 117997, Москва, Профсоюзная ул., 65  
E-mail: [chereshkoalex@gmail.com](mailto:chereshkoalex@gmail.com)

**Д.В. Елпашев**

*Институт проблем управления им. В.А. Трапезникова РАН*  
Россия, 117997, Москва, Профсоюзная ул., 65  
E-mail: [den.elpshv@gmail.ru](mailto:den.elpshv@gmail.ru)

**В.Н. Кушнарев**

*Институт проблем управления им. В.А. Трапезникова РАН*  
Россия, 117997, Москва, Профсоюзная ул., 65  
E-mail: [grand\\_yarl@mail.ru](mailto:grand_yarl@mail.ru)

**Ключевые слова:** виртуальный анализатор, идентификация, технологический процесс, ассоциативный поиск.

**Аннотация:** Приведены результаты построения виртуальных анализаторов (ВА) технологических процессов (ТП) в условиях перемены режима функционирования на примере процесса синтеза метил-трет-бутилового эфира (МТБЭ). Данный ТП характеризуется нелинейной динамикой, в связи с чем стандартные методы виртуального анализа показателей качества являются недостаточно точными, что влечет за собой значительные ошибки в прогнозах этих показателей. Исследованы возможности применения метода многомерной регрессии на основе ассоциативного поиска.

## 1. Введение

Установка синтеза метил-трет-бутилового эфира (МТБЭ) предназначена для синтеза МТБЭ из метанола и изобутилена, содержащегося в изобутан-изобутиленовой фракции с последующим выделением товарного продукта. Сырьём для получения МТБЭ служат метанол, изобутилен и изобутан-изобутиленовая фракция. Синтез МТБЭ проводится на катализаторах различных марок. Смена марок катализатора является одной из причин нелинейности такого типа ТП. Для прогноза показателей качества ТП предлагается использовать модели, построенные при помощи алгоритмов ассоциативного поиска, которые используют базу индуктивных знаний ТП для построения на каждом временном такте лучшей в смысле среднеквадратического критерия модели ТП.

В работе приведены результаты численного моделирования и сравнительный анализ двух типов моделей ВА [1] – стандартной линейной регрессии и линейной регрессии на основе ассоциативного поиска.

## 2. Многомерный регрессионный анализ

При построении ВА проанализирована информация о функционировании технологического объекта за три года. Данные ТП прошли предварительную статистическую обработку, включающую валидацию и фильтрацию выбросов. Для отбора информативных параметров, которые присутствуют в модели ВА в качестве входных переменных, производится корреляционный анализ данных ТП. Вычисляется коэффициент корреляции, в зависимости от различных временных сдвигов для всех параметров ТП. Затем производится расчет корреляционной матрицы, что позволяет выбрать наиболее информативные входы для построения модели.

Структура и параметры модели линейной регрессии представлены в Таблице 1.

**Таблица 1.** Модель ВА для определения концентрации метанола в МТБЭ на основе линейной регрессии

ВА на основе линейной регрессии концентрации метанола в МТБЭ, % масс.		
Наименование входа модели	Временной сдвиг	Параметр модели
Загрузка установки с учетом концентрации изобутилена	0 мин.	$b_1 = 0.0069$
Разница температур между 5 и 17 тарелками ректификационной колонны	0 мин.	$b_2 = -0.105$
Расход флегмы в реактор	0 мин.	$b_3 = 0.036$
Расход флегмы в ректификационную колонну	0 мин.	$b_4 = 0.013$
Температура верха реактора	0 мин	$b_5 = 0.238$
Давление реакционной смеси	215 мин	$b_6 = -1.0$
Свободный член	-	$bias = -0.083$

## 3. Многомерный регрессионный анализ на основе ассоциативного поиска

Как показали результаты моделирования, линейная регрессионная модель ВА предсказывает концентрацию остаточного метанола для определенного режима ТП с приемлемой точностью. Недостатком такой модели является неадекватность прогнозов в условиях нелинейности ТП, обусловленной работой на разных технологических режимах. Попытка построить единую линейную модель для разных технологических режимов показала, что достаточно точно описать ТП в рамках классической линейной регрессии крайне затруднительно. Необходимо «вручную» переключать между различными линейными моделями. Альтернативным решением является основанный на применении индуктивных знаний алгоритм ассоциативного поиска [2, 3] для построения лучшей по среднеквадратическому критерию модели, новой на каждом такте прогнозирования.

Для построения модели выбираются только те точки, которые являются близкими к текущей. В качестве критерия близости принимается Евклидово расстояние  $(d(\bar{x}_1, \bar{x}_2) = \sqrt{\sum_{i=1}^S (x_{1i} - x_{2i})^2})$  между точками.

В результате ассоциативного поиска находится множество наборов ассоциативных данных  $\{x_{N-j}(n), y_{N-j}(n)\}$ , которые используются для вычисления коэффициентов регрессионной модели по МНК.

Таким образом, общая аппроксимирующая гиперплоскость строится не на всей области возможных значений входного вектора, а лишь для окрестности текущего входного вектора  $\bar{x}_N$ .

Для повышения быстродействия алгоритма ассоциативного поиска можно заранее применить алгоритмы кластеризации (классификация без учителя) входных векторов из базы данных по критерию ассоциативного поиска. Целью кластеризации является выделение определенного числа групп объектов, имеющих схожие признаки, и отделение их от более отличающихся объектов. После завершения кластеризации каждому объекту входного вектора присваивается кластерная метка.

На каждом рабочем такте происходит определение, к какому кластеру относится текущий входной вектор. Затем из базы знаний кластера отбираются входные векторы, ближайšie к текущему, и соответствующие реальные значения выходов. Формируется и решается система линейных уравнений относительно неизвестных коэффициентов модели и прогнозируемого выхода. После нескольких тактов индуктивную базу знаний требуется пополнить и повторить алгоритм кластеризации для уточнения кластеров и их моделей.

Модель ВА на основе ассоциативного поиска позволяет достаточно просто автоматически формировать модели, соответствующие определенным режимам. Модель определяет, какой наиболее вероятный текущий режим работы и строит ту модель, что наиболее адекватно его описывает. Важным условием работы является наличие достаточного количества данных для каждого режима, что позволит модели более корректно строить точечные модели.

#### **4. Модели ассоциативного поиска с переменной структурой для случая резких изменений режима**

Выбор структуры модели (1) технологического процесса определяется на основе предварительного анализа на этапе обучения. В частности, применяются методы корреляционного анализа. На практике встречаются ситуации, когда выборочный коэффициент корреляции весьма чувствителен к наличию в выборке резко выпадающих наблюдений при наличии выбросов (outlier) в выборке. Также не всегда имеет место нормальное распределение исследуемых процессов, что искажает результат корреляционного анализа.

Достаточно распространенным для практики случаем является «кусочно-стационарный» характер исследуемого процесса, обусловленный, в частности, резкими изменениями режима процесса. Примерами тому могут служить: существенные изменения характеристик входных показателей процесса нефтепереработки при поступлении на вход сырья из иного источника; перемена режима либо отключение подачи катализаторов в процессах химического производства; поступление комплектующих от нового производителя в машиностроении, и т.п. Представляется целесообразным осуществлять на этапе обучения «двухуровневую» кластеризацию.

Прежде всего, для всех входных факторов (компонент вектора входов) осуществляется определение ранговых коэффициентов корреляции Спирмена [4]:

$$(1) \quad \rho_s = 1 - \frac{6}{N(N-1)(N+1)} \sum_{i=1}^N (R_{is} - P_i)^2,$$

где:  $N$  – объем выборок  $x_s = (x_{1,s}, \dots, x_{N,s})$ ,  $s = 1, \dots, S$ ;  $y = (y_1, \dots, y_N)$ ;  $S$  – размерность вектора входов;  $P_i$  – ранг наблюдения  $x_{i,s}$  в ряду  $x_s$ ;  $R_{is}$  – ранг наблюдения  $y_i$  в ряду  $y$ .

Далее экспертным образом устанавливается пороговое значение для ранговых корреляционных коэффициентов каждой из компонент вектора входов, определяющее значимость компонент входных векторов, включая значения в прошлые моменты времени, для формирования выхода. Таким образом определяется, нужно ли эти значения входов включать в модель.

Сформируем кластеры для различных структур точечных моделей и пронумеруем их следующим образом. Пусть номер кластера  $C$  содержит  $K$  позиций,  $K = (S + 1)(N - 1)$ , соответствующих значениям ранговых коэффициентов корреляции для компонент входных векторов в моменты времени  $t \leq N$  и значений выходов в прошлые моменты времени  $t \leq N - 1$ . При этом в каждой позиции стоит 1, если соответствующий ранговый коэффициент корреляции превосходит пороговое значение:  $\rho(t) \geq L$ , и 0 – в противном случае, например:

$$(2) \quad C = [10010011010111].$$

Для каждого кластера, содержащего значения входов и прошлых выходов, могут быть построены ассоциативные модели. При этом требуется произвести кластеризацию «второго уровня» – для значений входов и предыдущих выходов моделей для каждого из кластеров «первого уровня», соответствующих всем сформированным структурам.

## 5. Анализ результатов моделирования

В работе для построения модели ассоциативного поиска исследуемого ТП используются те же статистические данные по тем же входам, что и у линейной модели. В качестве индуктивной базы знаний используются входные данные за 3 года и соответствующие им выходы. На каждом такте линейная модель строится на основе 30 ближайших точек к текущей, которые отобраны из базы знаний. Данные предварительно нормируются, что позволяет в равной степени учитывать все входы модели в евклидовой метрике.

В Таблице 2 приведен сравнительный анализ двух типов моделей – стандартной линейной регрессии и линейной регрессии на основе ассоциативного поиска. На рис. 1рис. 1 представлены графики прогнозов для этих моделей. Как видно из рис. 1, обе модели улавливают динамику изменений содержания метанола в МТБЭ. Однако, модель классической линейной регрессии описывает все возникающие изменения существенно менее точно.

**Таблица 2.** Сравнение метрик предсказания линейной модели и модели на основе ассоциативного поиска

Метрика	Линейная регрессия	Линейная регрессия на основе ассоциативного поиска
Средняя абсолютная ошибка (MAE)	0.081	0.045
Средняя абсолютная процентная ошибка (MAPE)	17%	8.3%
Среднеквадратическая ошибка (MSE)	0.013	0.0045
Коэффициент детерминации (R2)	0.521	0.836



**Рис. 1.** График сравнения прогноза содержания метанола в МТБЭ для линейной и ассоциативной моделей.

## 6. Заключение

В данной работе представлен метод синтеза ВА технологического процесса синтеза МТБЭ. Отмечено, что ввиду нелинейности такого типа ТП использование классической линейной регрессии может давать неадекватные оценки при прогнозе ключевых показателей качества. Исследована работа ВА на основе ассоциативного поиска, как альтернативный способ прогноза показателей качества ТП, который позволяет обойти проблемы с нелинейностью ТП и переменной режима функционирования.

Исследование выполнено в рамках научной программы Национального центра физики и математики, направление № 9 «Искусственный интеллект и большие данные в технических, промышленных, природных и социальных системах».

## Список литературы

1. Бахтадзе Н.Н. Виртуальные анализаторы (идентификационный подход) // Автоматика и телемеханика. 2004. № 11. С. 3-24.
2. Черешко А.А., Виртуальные анализаторы качества на основе цифровых моделей // Автоматизация в промышленности. 2022. № 7. С. 33-38.
3. Лотоцкий В.А., Чадеев В.М., Максимов Е.А., Бахтадзе Н.Н. Перспективы применения виртуальных анализаторов в системах управления производством // Автоматизация в промышленности. 2004. № 5. С. 23-29.
4. Stephanou M.; Varughese M. Sequential estimation of Spearman rank correlation using Hermite series estimators // Journal of Multivariate Analysis. 2021. Vol. 186. P. 104783. arXiv:2012.06287. doi:10.1016/j.jmva.2021.104783. S2CID 235742634.