

ДИАГНОСТИКА ОШИБКИ ОЦЕНКИ ВИРТУАЛЬНОГО АНАЛИЗАТОРА¹**С.А. Шевлягина***Институт автоматизации и процессов управления ДВО РАН, Россия, Владивосток, samotylova@dvo.ru*

Аннотация. При управлении технологическими процессами некоторые показатели качества продукции не измеряются в режиме реального времени из-за технических или экономических ограничений. В качестве эффективного решения задачи их оценки в режиме реального времени используются виртуальные анализаторы, которые обеспечивают стабильную и надежную оценку этих показателей на основе исторических данных об измеряемых переменных. При функционировании виртуальных анализаторов их прогнозы могут указывать на значительные изменения в технологическом процессе. Для диагностики предсказаний виртуального анализатора предлагается использовать числа Шепли при анализе причинно-следственных связей технологических переменных. Используемый подход продемонстрирован на реальных промышленных данных.

Ключевые слова: диагностика ошибки, виртуальный анализатор, идентификатор Хампеля, числа Шепли, производство пропилен.

DIAGNOSIS OF ESTIMATION ERROR OF A SOFT SENSOR**S.A. Shevlyagina***Institute of automation and control process FEB RAS, Russia, Vladivostok, samotylova@dvo.ru*

Abstract. In process control, some product quality indicators are not measured in real time due to technical or economic constraints. As an effective solution for their real-time assessment, soft sensors are used to provide a stable and reliable assessment of these variables based on historical data of readily measurable process variables. During the operation of soft sensors, their predictions can indicate significant changes in the process. To diagnose soft sensor predictions, Shapley's numbers are used to diagnosis the causal relationships of process variables. The approach used is demonstrated on real industrial data.

Keywords: error diagnosis, soft sensor, Hampel identifier, Shapley's number, propylene production.

Shapley's number

Для цитирования: Шевлягина С.А. Диагностика ошибки оценки виртуального анализатора // Математические методы в технологиях и технике. 2024. № 1. С. 24-28.

На многих промышленных предприятиях важные переменные, обуславливающие качество производимой продукции, например, концентрация, могут измеряться только периодически, например, каждые 4 часа, и со значительной временной задержкой. Своевременное получение этой информации может снизить эффективность технологических процессов. Одним из возможных подходов к решению этой проблемы является использование виртуальных анализаторов (ВА).

В последнее время ВА получили широкое применение в промышленности. Как правило, они представляют прогностические модели, полученные на основе доступных данных о параметрах технологического процесса. В основном они обеспечивают онлайн-прогнозирование некоторых переменных, которые имеют важное значение для контроля и управления качеством и безопасностью производства, т.к. аппаратные измерительные приборы не всегда возможны [1]. В целом, виртуальные анализаторы можно разделить на два типа: первопринципные модели (модели белого ящика) и модели, построенные на основе данных (модели черного ящика). Использование первопринципных моделей затруднительно, т.к. они слишком сложны в построении. ВА, основанные на промышленных данных, из-за практической пользы и

¹ Работа выполнена в рамках гос. задания Минобрнауки (FWFW-2021-0003).

независимости от априорных знаний, являются популярными и достаточно эффективными [2].

Определенные прогнозы виртуального анализатора при оценке заданного показателя качества могут указывать на значительные изменения в технологическом процессе. Диагностика возникновения изменения в прогнозе виртуального анализатора играет важную роль в мониторинге технологического процесса. Быстрое и точное определение изменений имеет важное значение для своевременного восстановления эффективности технологического процесса. В работе предлагается простой способ обнаружения изменений на основе идентификатора Хампеля с последующей диагностикой причины их возникновения, а именно какие именно технологические переменные изменяют оценку ВА. Вклад каждой технологической переменной в предсказания ВА оценивается числами Шепли.

В качестве объекта исследования используется промышленный объект – производство пропилена (рис. 1). Для заданного объекта необходимо осуществлять оценку потерь целевого продукта (пропилен) в пропановой фракции.

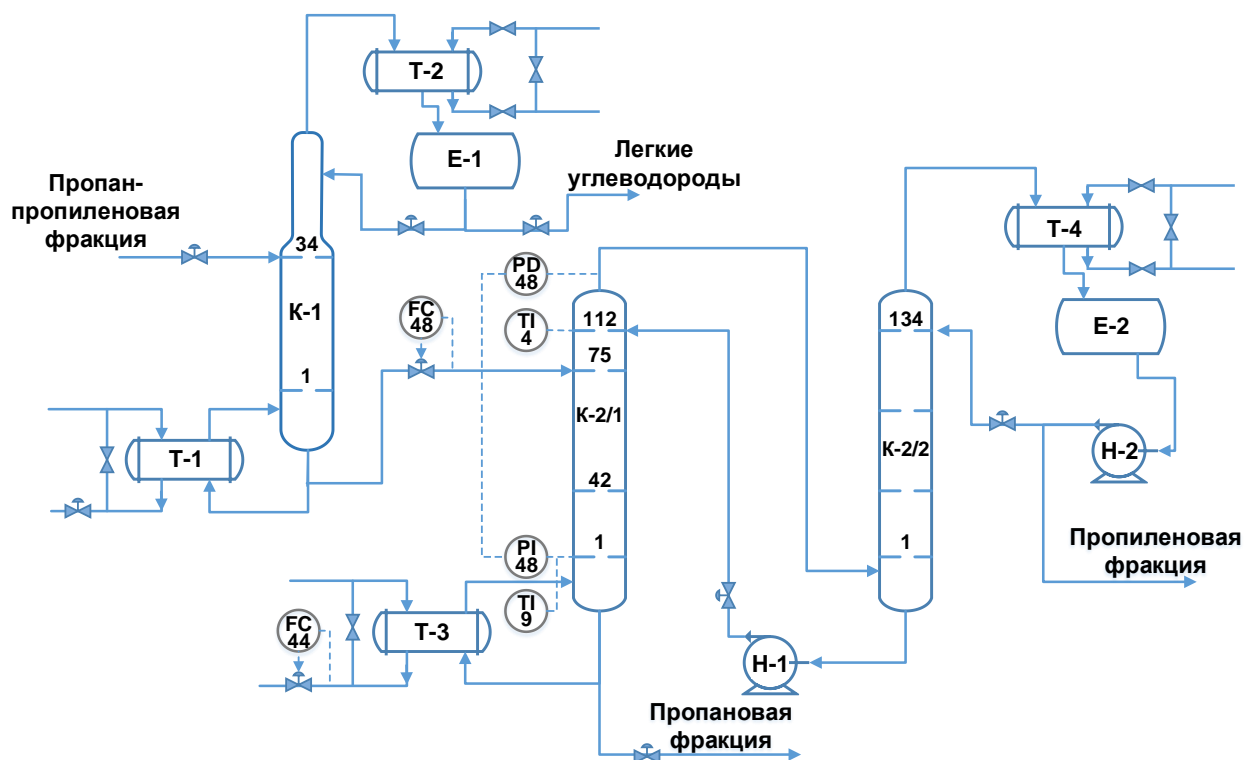


Рис. 1. Технологическая схема фрагмента производства пропилена

Жидкая пропан-пропиленовая фракция (ППФ) подается в колонну К-1, которая предназначена для отпарки из ППФ легких углеводородных фракций: C_2 , метана, водорода. Нагрев кубовой жидкости в колонне К-1 осуществляется в межтрубном пространстве кипятильника Т-1, на основе тепла циркуляционной горячей воды, расход которой регулируется клапаном. Пары с верха колонны К-1 поступают в межтрубное пространство дефлегматора Т-2, где охлаждаются и частично конденсируются обратной водой. Жидкие углеводороды из емкости Е-1 через регулирующий клапан, подаются на верхнюю тарелку колонны К-1 в качестве орошения. Кубовый продукт колонны К-1 (жидкая ППФ) через регулирующий клапан поз. FC-48 направляется для разделения в колонну К-2. Процесс разделения ППФ из-за близких температур кипения углеводородов, входящих в ее состав, требует колонную аппаратуру с большим количеством тарелок. Колонна К-2 конструктивно выполнена из двух последовательно работающих колонн К-2/1 и К-2/2. Нагрев кубовой

жидкости колонны К-2/1 осуществляется в межтрубном пространстве кипятильника Т-3 горячей циркуляционной водой. Расход циркуляционной горячей воды регулируется клапаном поз. FC-44. Кубовый продукт колонны К-2/1 (пропановая фракция) выводится на печи пиролиза. Пары с верха колонны К-2/1 направляются в нижнюю кубовую часть колонны К-2/2. Жидкость из куба колонны К-2/2 насосом Н-1 через регулирующий клапан, подается на верхнюю 112-ю тарелку колонны К-2/1 в качестве орошения. Пары с верха колонны К-2/2 охлаждаются и частично конденсируются в межтрубном пространстве конденсатора Т-4 оборотной водой. Газожидкостная смесь из конденсатора Т-4 поступает в флегмовую емкость Е-2. Жидкая пропиленовая фракция из емкости Е-2, насосом Н-2 подается на верхнюю 134-ю тарелку колонны К-2/2 в качестве орошения. Балансовый избыток пропилена по уровню в емкости Е-2 выводится от насоса Н-2 на дальнейшее производство.

Для построения ВА используются следующие технологические параметры в качестве входных переменных: FC-44 – расход оборотной воды; ТИ-4 – температура верха колонны К-2/1; ТИ-9 – температура низа колонны К-2/1; PD-48 – перепад давления в колонне К-2/1; FC-48 – расход из К-1 в К-2/1. В качестве выходной переменной являются потери пропилена (оценка пропилена в пропановой фракции). На рис. 2 представлены промышленные данные, нормированные с Z-оценкой.

Для построения виртуального анализатора используется Байесовская линейная регрессия [3, 4].

Для диагностики причинно-следственных связей входных переменных используются числа Шепли. Для каждой входной переменной i числа Шепли определяются как:

$$\phi_i(f, x) = \sum_{S' \subseteq M \setminus i} w_x(S') [f(S' \cup \{i\}) - f(S')], \quad w_x(S') = \frac{|S'|!(M - |S'| - 1)!}{M!}, \quad (1)$$

где f – прогнозирующая модель; x – входная переменная; $\phi_i(\bullet)$ – значение Шепли входной переменной i в модели f ; M – количество входных переменных; S' – подмножество входных переменных.

Для $w_x(S')$ знаменатель $M!$ представляет все возможные комбинации входных переменных; числитель $|S'|!(M - |S'| - 1)!$ означает время появления $S' \cup \{i\}$ во всех $M!$ комбинациях; $f(S' \cup \{i\}) - f(S')$ указывает ожидаемый предельный вклад входной переменной i в одной комбинации. Значение числа Шепли входной переменной i равно средневзвешенному значению по всем комбинациям входных переменных.

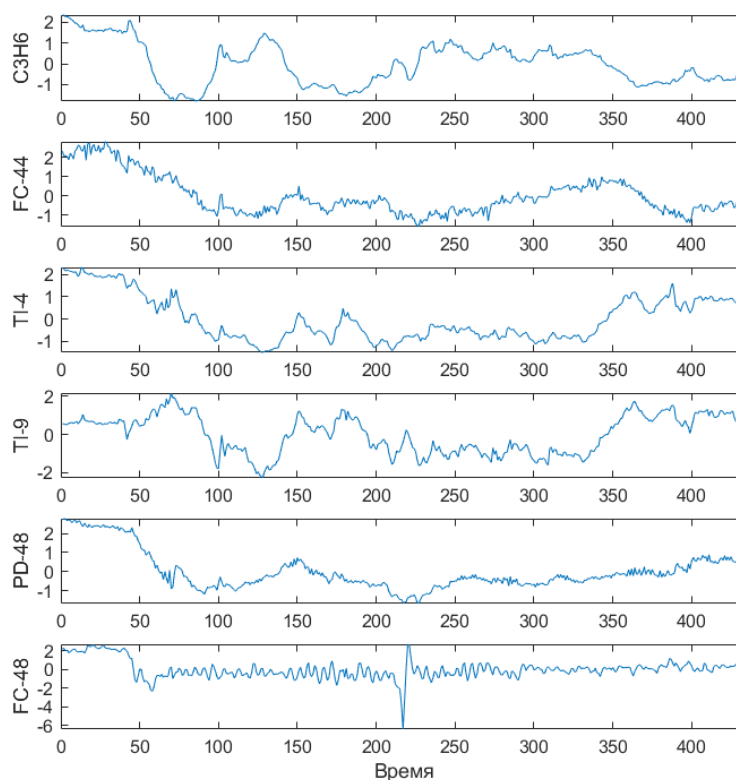


Рис. 2. Графическое представление реальных данных процесса

Для одной входной переменной j её вклад в прогнозирующую модель $\hat{f}(x)$ можно вычислить следующим образом:

$$\phi_j(\hat{f}) = \beta_j x_j - E(\beta_j X_j), \quad (2)$$

где β_j – коэффициент входной переменной j ; $E(\beta_j X_j)$ – предполагаемый усредненный эффект этой переменной.

Для множества входных переменных p их общий вклад можно определить по формуле:

$$\sum_{j=1}^p \phi_j(\hat{f}) = \sum_{j=1}^p (\beta_j x_j - E(\beta_j X_j)) = \hat{f}(x) - E(\hat{f}(X)) = \hat{f}(x) - \phi_0. \quad (3)$$

На рис. 3а показано влияние входных данных на предсказание ВА. Каждое наблюдение представлено одной точкой в строке технологических параметров. На рис. 3б представлены средние абсолютные значения чисел Шепли по всем наблюдениям.

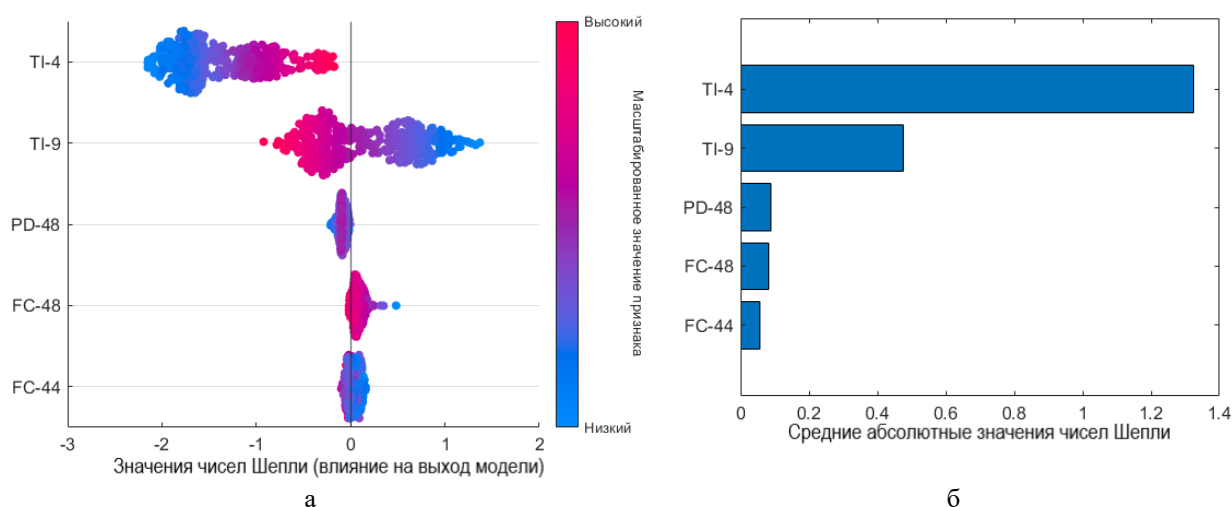


Рис. 3. Важность технологических параметров по всем наблюдениям:
а – итоговый график; б – средние абсолютные значения чисел Шепли

Итоговый график чисел Шепли по всем наблюдениям содержит только информацию о важности технологических параметров, которые не обязательно являются причинами изменений. Чтобы найти причину изменений, необходим более информативный график с локальными объяснениями входов.

Для обнаружения локальных изменений в прогнозе ВА используется идентификатор Хампеля, который считается надежным методом обнаружения выбросов [5]. В идентификаторе Хампеля в качестве опорного значения используется медиана выборки \tilde{x} , которая определяется как:

$$|x_i - \tilde{x}| > \alpha S, \quad (4)$$

$$S = \frac{1}{0,6745} \text{median}\{|x_i - \tilde{x}|\}, \quad (5)$$

где α – пороговый параметр ($\alpha = 3$), а S – оценка шкалы медианного абсолютного отклонения.

В результате обработки данных производства пропилена (рис. 1) определено 6 точек изменения прогноза ВА, которые соответствуют временным меткам: 129, 234, 235, 249, 251 и 276 (рис. 4). После определения местоположения точек изменения, необходимо определить, какие технологические параметры изменяют предсказания.

Для демонстрации процесса решения этой задачи возьмем две точки изменения, обнаруженные во временных метках 234 и 249. На рис. 5 показаны наборы чисел Шепли для временных меток 234 (рис. 5а) и 249 (рис. 5б).

Значения чисел Шепли для временных меток 234 и 249 показывают, что наибольший вклад оказывает температура верха колонны К-2/1 (поз. ТИ-4) как и на всем диапазоне наблюдений (рис. 3). Важность перепада давления (поз. РД-48) и температуры низа колонны К-2/1 (поз. ТИ-9) для рассматриваемых временных меток изменяется.

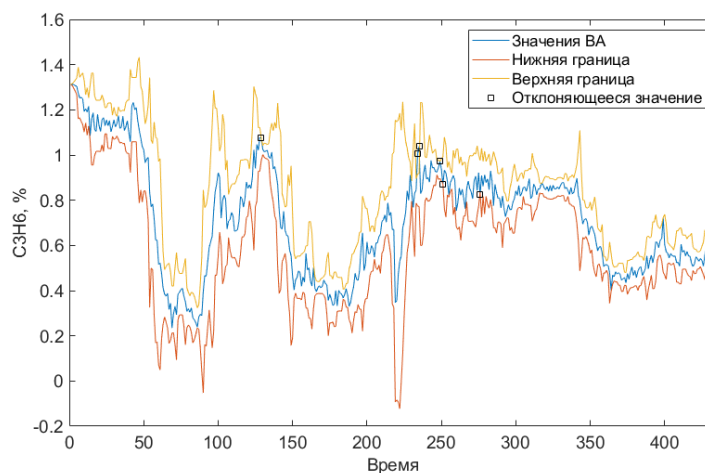


Рис. 5. Оценка потерь пропилена, полученная виртуальным анализатором

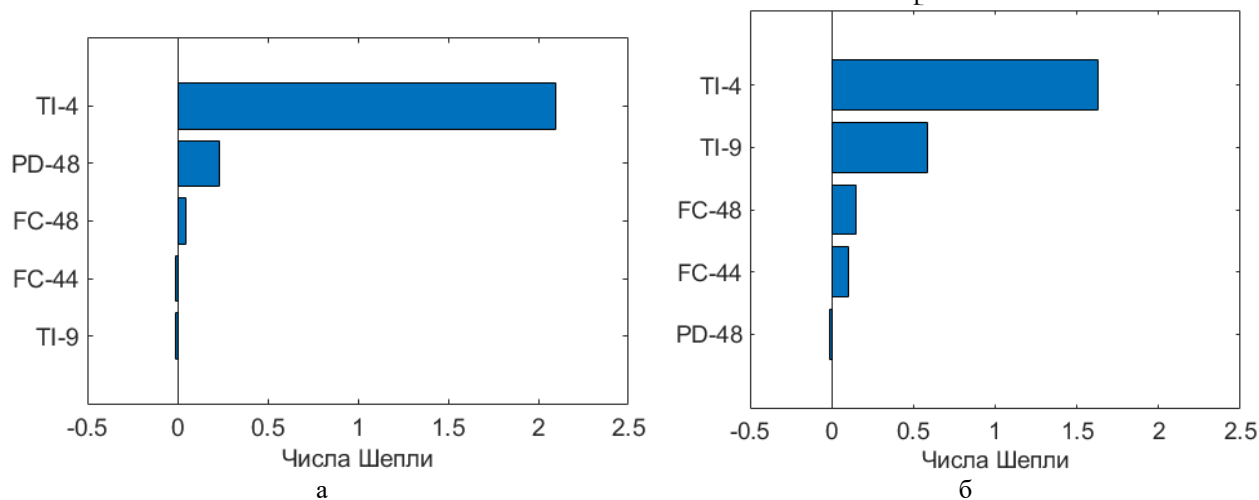


Рис. 5. Локальные значения чисел Шепли для:
а – временной метки 234; б – для временной метки 249

Наблюдается смена знака вклада в изменение предсказания ВА. Как видно из данного примера, в точках 234 и 249 значительно изменились входные характеристики ТИ-9, FC-44, и РД-48, однако определить, что именно вызвало изменение: один параметр или их комбинация сложно. Поэтому требуется провести дополнительную диагностику.

Важно отметить, что метод диагностики на основе чисел Шепли может распределить вклад каждого входа и обнаружить информативные связи между технологическими переменными и предсказанием.

Библиографический список

1. Shang C., Yang F., Huang D., Lyu W. Data-driven soft sensor development based on deep learning technique // Journal of Process Control. 2014. Vol. 24(3), pp. 223-233.
2. Ma L., Wang M., Peng K. A missing manufacturing process data imputation framework for nonlinear dynamic soft sensor modeling and its application // Expert Systems with Applications. 2024. Vol. 237, P. 121428.
3. Yang K., Jin H., Chen X., Dai J., Wang L., Zhang D. Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models // Chemometrics and Intelligent Laboratory Systems. 2016. Vol. 155, pp. 170-182.
4. Galeazzi A., de Fusco F., Prifti K., Gallo F., Biegler L., Manenti F. Predicting the performance of an industrial furnace using Gaussian process and linear regression: A comparison // Computers & Chemical Engineering. 2024. Vol. 181. P. 108513.
5. Davies P. L., Gather U. The identification of multiple outliers // Journal of the American Statistical Association. 1993. Vol. 88. № 423, pp. 782-792.