



Методы оперативного определения не измеряемых поточными приборами качественных показателей химико-технологических процессов

Е.А. Гребенюк, Э.Л. Ицкович (ИПУ РАН)

Рассматриваются методы виртуальных измерений качественных показателей химико-технологических процессов, которые целесообразно использовать в системах автоматизации производственных установок. Выделяются и анализируются особенности отдельных методов виртуальных измерений, реализующих их алгоритмов, областей их применения.

Ключевые слова: виртуальные измерения, автоматизация технологических процессов, применение нейросети, контроль качества уравнением регрессии.

Введение

Все виды автоматического определения текущих значений любой величины на производстве можно подразделить на три категории:

- непосредственное измерение значения величины поточным прибором, когда чувствительный элемент прибора реагирует на изменение самой искомой величины (например, измерение температуры жидкости вставленной в нее термопарой);
- косвенное измерение значения величины поточным прибором, определяемое известной точной зависимостью по непосредственным автоматическим приборным измерениям значений других величин, с которыми искомая величина связана функционально (например, определение расхода газа по трубе точной расчетной формулой по измеренным значениям перепада давления на вставленной в трубу диафрагме, температуры и абсолютного давления газа);
- оценка текущего значения величины статистическими методами по измеренным значениям других величин и/или по редким лабораторным замерам искомой величины.

Первые две категории измерений часто объединяются наименованием "прямые измерения". Третья категория измерений под наименованием "виртуальные измерения" приобретает все большее значение при автоматизации химико-технологических процессов. Этому способствует ряд обстоятельств. С одной стороны, современные алгоритмы точного регулирования процессов требуют знания текущих качественных показателей их входных/выходных материальных потоков; с другой — прямое определение текущих качественных показателей в большинстве случаев на предприятиях отсутствует по следующим причинам:

- отсутствуют прямые автоматические способы измерения многих качественных, химических величин;
- слишком большая цена на ряд поточных анализаторов, и нехватка у предприятий финансовых средств на их приобретение;
- сложность и высокая стоимость обслуживания многих поточных анализаторов из-за требуемой частоты и дороговизны их калибровки в производственных ус-

ловиях, которую должна проводить посторонняя организация;

- большое транспортное запаздывание поточных анализаторов, составляющее десятки минут, что препятствует их непосредственному применению в системах регулирования ТП.

В то же время, наблюдаемое увеличение вычислительной мощности микропроцессорных средств автоматизации и развитие статистических методов создают практические условия для оценки текущих изменений значений качественных величин без их прямого измерения. Приближенная оценка их оперативных значений может производиться различными методами в зависимости от свойств самих величин и характера их статистических связей с измеряемыми величинами.

Виртуальные измерения на базе статистического анализа

Построение линейного уравнения регрессии

При оценке текущего значения неизмеряемой искомой величины ТП по имеющимся статистическим зависимостям между ней и рядом других величин этого процесса, значения которых измеряются, во многих случаях можно применить метод регрессионного анализа [1], заключающийся в построении линейного уравнения регрессии, аргументами которого являются значения отобранных измеряемых величин (так называемых регрессоров), а определяемой функцией — значение искомой величины.

При этом следует учесть временные соответствия между искомой величиной и каждым регрессором — измеряемой величиной. Поскольку регрессоры, в общем случае, отделены от искомой величины различными динамическими каналами, то сопоставление значений любого регрессора и искомой величины в единый момент времени не определяет их истинную статистическую связь: значение величины на выходе динамического канала зависит от значения величины на входе не только в данный момент времени, но и во все предшествующие моменты, которые сохраняет память канала. Улучшить оценку можно было бы за счет учета и компенсации динамической связи между искомой величиной и регрессором, но это достаточ-

но сложно и в подавляющем большинстве случаев представляется мало реальным. На практике применяется метод сдвига во времени входящих в уравнение регрессии значений регрессоров и искомой величины, который частично компенсирует имеющуюся между ними динамическую связь. Для этого в уравнение регрессии, определяющее значение искомой величины в текущий момент времени, подставляют значение каждого регрессора в момент времени, сдвинутый относительно текущего момента на интервал времени, соответствующий максимальному значению модуля взаимной корреляционной функции данного регрессора и искомой величины.

В общем случае виртуальная оценка искомой величины Y в момент времени t линейным уравнением регрессии, в котором аргументами являются p измеряемых величин (регрессоров) $X_1, \dots, X_k, \dots, X_p$, имеет вид:

$$y(t) = \alpha_0 + \alpha_1[x_1(t - t_{\max 1}) - m_{x1}] + \alpha_2[x_2(t - t_{\max 2}) - m_{x2}] + \dots + \alpha_p[x_p(t - t_{\max p}) - m_{xp}], \quad (1)$$

где $y(t)$ — оценка значения величины Y в момент времени t ; $\alpha_0, \alpha_1, \alpha_p$ — коэффициенты уравнения регрессии; $m_{x1}, m_{x2}, \dots, m_{xp}$ — оценки математических ожиданий величин $X_1, X_2, \dots, X_p$; $x_1(t - t_{\max 1})$ — значение величины X_1 в момент времени, сдвинутый относительно текущего времени t на значение $t_{\max 1}$, соответствующее максимальному значению модуля взаимной корреляционной функции между величинами Y и X_1 ; $x_2(t - t_{\max 2})$ — значение величины X_2 в момент времени, сдвинутый относительно текущего времени t на значение $t_{\max 2}$, соответствующее максимальному значению модуля взаимной корреляционной функции между величинами Y и X_2 и т. д.

Расчет коэффициентов уравнения регрессии выполняется методом наименьших квадратов. В соответствии с этим методом определяются такие значения коэффициентов, при которых сумма квадратов отклонений истинных значений искомой величины от значений, рассчитанных по уравнению регрессии, будет минимальной.

Для реализации этого метода производится набор обучающей выборки значений искомой величины, измеряемой лабораторным анализатором в отдельные моменты t_i и соответствующих им по времени значений регрессоров (непрерывно измеряемых величин), и по этой выборке объемом n указанным методом вычисляют коэффициенты уравнения регрессии. Ниже представлен математический вид минимизируемой функции:

$$\sum_{i=1}^n [y(t_i) - y_{\text{изм}}(t_i)]^2 = \sum_{i=1}^n \left\{ \alpha_0 + \alpha_1[x_1(t_i - t_{i\max 1}) - m_{x1}] + \dots + \alpha_p[x_p(t_i - t_{i\max p}) - m_{xp}] - y_{\text{изм}}(t_i) \right\}^2 \rightarrow \min, \quad (2)$$

где $y_{\text{изм}}(t_i)$ — измеренное лабораторным анализатором

значение искомой величины Y в момент t_i , n — объем обучающей выборки, состоящий из измеряемых датчиками в соответствующие моменты времени значений регрессоров и определяемых лабораторным анализатором значений искомой величины.

Оценка значения искомой величины при наличии двух регрессоров. Распространенным частным случаем виртуального измерения искомой величины Y является ее оценка линейным уравнением регрессии по двум измеряемым величинам X_1 и X_2 . Первая из них зависит от ряда факторов, от которых изменяется и величина Y , а вторая в основном определяется другими факторами, которые также оказывают существенное влияние на величину Y , но мало влияют на первую измеряемую величину. Оценка величины Y в момент времени t определяется следующим уравнением:

$$y(t) = \alpha_0 + \alpha_1[x_1(t - t_{\max 1}) - m_{x1}] + \alpha_2[x_2(t - t_{\max 2}) - m_{x2}]. \quad (3)$$

Формулы для расчета коэффициентов этого уравнения регрессии в результате минимизации функции (2) при двух регрессорах принимают вид:

$$\alpha_0 = m_y - \alpha_1 m_{x1} - \alpha_2 m_{x2}, \quad (4)$$

$$\alpha_1 = \frac{S_y}{S_{X_1}} \cdot \frac{r_{YX_1} - r_{YX_2} \cdot r_{X_1X_2}}{1 - r_{X_1X_2}^2}, \quad (5)$$

$$\alpha_2 = \frac{S_y}{S_{X_2}} \cdot \frac{r_{YX_2} - r_{YX_1} \cdot r_{X_1X_2}}{1 - r_{X_1X_2}^2}, \quad (6)$$

где m_y — оценка среднего значения искомой величины Y ; S_y — оценка среднеквадратичного отклонения искомой величины Y от ее среднего значения; S_{X_1}, S_{X_2} — оценки среднеквадратичных отклонений регрессоров X_1 и X_2 от их средних значений; r_{YX_1}, r_{YX_2} — оценки коэффициентов корреляции между значениями величин Y и X_1 и Y и X_2 , сдвинутыми во времени на интервалы $t - t_{\max 1}$ и $t - t_{\max 2}$ соответственно; $r_{X_1X_2}$ — оценка коэффициента корреляции между значениями величин X_1 и X_2 , сдвинутыми между собой во времени на интервал $t_{\max 1} - t_{\max 2}$.

Оценка значения искомой величины при наличии p регрессоров. В общем случае, виртуальное измерение искомой величины Y реализуется линейным уравнением регрессии, в котором аргументами являются p измеряемых величин $X_1, \dots, X_k, \dots, X_p$. При этом коэффициенты уравнения регрессии, вычисляемые методом наименьших квадратов, имеют следующий вид:

$$\alpha_0 = m_y - \alpha_1 m_{x1} - \alpha_2 m_{x2} - \dots - \alpha_p m_{xp}, \quad (7)$$

$$a_j = \frac{S_y}{S_{X_j}} \frac{L_j}{L}, \quad j = 1, 2, \dots, p, \quad (8)$$

$$\text{где } L = \begin{vmatrix} 1 & r_{X_1X_2} & r_{X_1X_3} & \dots & r_{X_1X_p} \\ r_{X_1X_2} & 1 & r_{X_2X_3} & \dots & r_{X_2X_p} \\ r_{X_1X_3} & r_{X_2X_3} & 1 & \dots & r_{X_3X_p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{X_1X_p} & r_{X_2X_p} & r_{X_3X_p} & \dots & 1 \end{vmatrix}, \quad (9)$$

L_j получен из L заменой его j -го столбца на столбец из элементов $r_{yx_1}, r_{yx_2}, \dots, r_{yx_p}$.

Область применения рассмотренного линейного уравнения регрессии определяется следующими факторами:

- наличие в технологическом объекте различных измеряемых величин входных компонентов, режимных характеристик, выходных продуктов, которые статистически взаимосвязаны с искомой неизмеряемой величиной;

- близость к линейности статистических связей измеряемых и искомой величин (последнее существенно повышает точность оценки);

- стабильность во времени ТП, когда все изменения их входных, режимных и выходных показателей можно считать на достаточно длительных интервалах времени стационарными случайными процессами.

Возникающие время от времени значительные изменения параметров статистических связей, вызванные сменой производительности агрегата или подачей входных компонентов другого качества, или переменной требований к выходным продуктам, то есть переход ТП на другой стабильный режим требует пересчета коэффициентов построенного уравнения регрессии.

Принципиально любые измеряемые величины, имеющие значительные связи с искомой величиной, могут быть использованы для ее оценки, но на практике следует ограничиваться теми величинами, статистические взаимосвязи которых между собой минимальны (что позволяет охватить влияние на искомую величину разных независимых факторов и тем самым повысить точность ее оценки), а между ними и искомой величиной — максимальны.

Конкретное построение уравнения регрессии, определяющего зависимость оценки искомой величины от значений измеряемых величин, предусматривает наличие обучающей выборки, состоящей из временных рядов измеряемых и искомой величин. Временные ряды измеряемых регрессоров и лабораторных анализов искомой величины обычно имеются в исторической БД рассматриваемого ТП, в противном случае они должны быть получены набором значений измеряемых величин и периодических лабораторных анализов (или специально организованных ручных замеров) искомой величины.

Выборка регрессоров из всей совокупности измеряемых на технологической установке величин может быть проведена разными путями, различающимися исходными данными, простотой реализации и точностью оценки искомой величины, получаемой по уравнению регрессии.

Метод формирования набора регрессоров с помощью парных коэффициентов корреляции величин

Следует сразу отметить, что данный метод, несмотря на свою относительную простоту и практическую распространенность, имеет ряд достаточно существенных недостатков, сужающих рациональную область его применения и снижающих точность оценивания.

Метод включает ряд последовательно реализуемых этапов.

Первый этап. Экспертное выделение из всего множества измеряемых величин в рассматриваемом ТП той совокупности измеряемых величин, которые являются возможными регрессорами. На основе качественного анализа физических и химических свойств ТП и опыта работы по его управлению выделяются измеряемые величины, претендующие на роль регрессоров.

Второй этап. Вычисление парных коэффициентов корреляции между искомой величиной и каждым предполагаемым регрессором и выделение группы значимых регрессоров, то есть регрессоров, парные коэффициенты корреляции которых с искомой величиной превосходят некоторое граничное, задаваемое пользователем значение, например 0,4 или 0,5. Для выделенных регрессоров вычисляются также корреляционные отношения между каждым из них и искомой величиной, и для каждого регрессора по сопоставлению полученных значений парного коэффициента корреляции с корреляционным отношением отбираются те из выделенных регрессоров, которые имеют статистическую связь с искомой величиной, близкую к линейной зависимости. Линейность статистической связи между двумя величинами характеризуется близостью значений их парного коэффициента корреляции и корреляционного отношения, поскольку корреляционное отношение характеризует степень связи независимо от ее формы, а парный коэффициент корреляции характеризует степень линейной связи. Приблизительно можно также судить о линейности связи между любой измеряемой и искомой величинами путем построения графика, по осям которого откладываются отдельные их значения.

Третий этап. Вычисление парных коэффициентов корреляции между отдельными регрессорами из полученной на предыдущем этапе группы. Удаление из выделенной группы регрессоров тех, которые имеют значимую связь с другими регрессорами этой группы. Степень связи двух регрессоров определяется значением коэффициента парной корреляции между ними. Если этот коэффициент превосходит некоторое задаваемое пользователем значение, например 0,7 или 0,8, то из этой пары удаляется регрессор, у которого меньший парный коэффициент корреляции с искомой величиной.

Четвертый этап. По оставшимся в группе регрессорам методом, описанным выше, вычисляются коэффициенты уравнения регрессии. Для этого используется обучающаяся выборка измеряемых величин (регрессоров) и соответствующих им по времени лабораторных анализов искомой величины. Построенное уравнение линейной регрессии действительно либо для одного режима работы установки, либо для некоторого множества близких режимов в зависимости от того, в каких условиях работы установки получена данная выборка. Практически она должна со-

стоять из ряда десятков лабораторных анализов искомой величины и соответствующих им измерений регрессоров, то есть должна охватывать достаточно длительный интервал времени работы установки.

Приведенный способ отбора регрессоров, хотя и используется на практике, является недостаточно корректным и может иметь повышенную погрешность оценки искомой величины. Это произойдет, если среди выбранных регрессоров двое (коллинеарность) или большее число (мультиколлинеарность) величин окажутся связанными между собой близкой к линейной функциональной зависимостью. В этом случае одна из зависимых величин является линейной комбинацией остальных. При указанной взаимозависимости регрессоров влияние отдельного регрессора на оценку искомой величины может быть искажено (как в большую, так и в меньшую сторону) воздействиями связанных с ним регрессоров. В рассмотренном способе нельзя проранжировать регрессоры по силе их независимого влияния на искомую величину, так как используемые в нем коэффициенты парной корреляции характеризуют тесноту связи между искомой величиной и каждым регрессором с учетом взаимосвязей его с другими регрессорами. Полученные оценки коэффициентов уравнения регрессии, при наличии существенно взаимосвязанных регрессоров, становятся ненадежными: они имеют сравнительно большие ошибки и изменяются с изменением объема обучающей выборки. Поэтому, в общем случае, целесообразно использовать другую процедуру отбора регрессоров, которая исключает присутствие в уравнении регрессии зависимых регрессоров.

Метод формирования набора регрессоров с помощью частных коэффициентов корреляции величин

Метод состоит из ряда последовательно реализуемых этапов.

Первый этап. Экспертное выделение из всего множества измеряемых величин в рассматриваемом ТП той совокупности измеряемых величин, которые являются возможными регрессорами. Этап не отличается от первого этапа метода, указанного в предыдущем разделе.

Второй этап. По выделенным на предыдущем этапе регрессорам строится уравнение линейной регрессии (1) (общее уравнение регрессии), и определяются его коэффициенты. Для этого, как и в предыдущем методе, используется обучающаяся выборка измеряемых величин (регрессоров) и соответствующих им по времени лабораторных анализов искомой величины.

Кроме этого, аналогичным способом строятся еще p других, частных уравнений регрессии, в каждом из которых исключен один из регрессоров общего уравнения регрессии. Так, при исключении X_j регрессора j -е частное уравнение регрессии примет следующий вид:

$$y_j(t) = \alpha_0 + \alpha_1[x_1(t - t_{\max 1}) - m_{x1}] + \dots + \alpha_{j-1}[x_{j-1}(t - t_{\max j-1}) - m_{xj-1}] + \alpha_{j+1}[x_{j+1}(t - t_{\max j+1}) - m_{xj+1}] + \dots + \alpha_p[x_p(t - t_{\max p}) - m_{xp}]. \quad (10)$$

Третий этап. Производится вычисление коэффициентов детерминации для общего и каждого частного уравнения регрессии, показывающих какая часть дисперсии искомой величины Y объясняется влиянием остальных измеряемых величин, не используемых в данном уравнении регрессии.

Для общего уравнения регрессии вида (1) коэффициент детерминации R вычисляется по формуле:

$$R^2 = 1 - \frac{\sum_{i=1}^n [y_{изм}(t_i) - y(t_i)]^2}{\sum_{i=1}^n [y_{изм}(t_i) - m_y]^2}, \quad (11)$$

где m_y — оценка среднего значения искомой величины, $y_{изм}(t_i)$ — значение искомой величины, измеренное лабораторным анализатором в момент t_i , n — объем обучающей выборки.

Для каждого j -го частного уравнения регрессии вида (10) коэффициент детерминации вычисляется по формуле:

$$R_j^2 = 1 - \frac{\sum_{i=1}^n [y_{изм}(t_i) - y_j(t_i)]^2}{\sum_{i=1}^n [y_{изм}(t_i) - m_y]^2}. \quad (12)$$

Четвертый этап. Вычисляются частные коэффициенты корреляции между каждым регрессором и искомой величиной. Частный коэффициент корреляции r_{yx} при $X_1 \dots X_{j-1}, X_{j+1} \dots X_p = \text{const}$, оценивающий влияние регрессора X_j на искомую величину Y при неизменном уровне $p-1$ других регрессоров $X_1 \dots X_{j-1}, X_{j+1} \dots X_p$, включенных в уравнение регрессии, вычисляется по формуле:

$$r_{yx} \text{ при } x_1 \dots x_{j-1}, x_{j+1} \dots x_p = \text{const} = \sqrt{1 - \frac{1 - R_j^2}{1 - R^2}}. \quad (13)$$

Так, частный коэффициент корреляции первого порядка (то есть, когда в общем уравнении регрессии имеется два регрессора), оценивающий влияние регрессора X_1 при постоянстве регрессора X_2 может быть вычислен по формуле:

$$r_{yx_1} \text{ при } x_2 = \text{const} = \sqrt{1 - \frac{1 - R_1^2}{1 - R^2}}. \quad (14)$$

Если значение частного коэффициента корреляции между регрессором X_j и искомой величиной Y равно нулю, то этот регрессор не влияет на значение искомой величины. Для оценки близости к нулю оценки частного коэффициента корреляции, полученной по выборке объема n с p регрессорами, используется неравенство (15). Не нулевыми считаются частные коэффициенты корреляции, превышающие по абсолютной величине следующий порог [1]:

$$r_{yx_j} \text{ при } x_1 \dots x_{j-1}, x_{j+1} \dots x_p = \text{const} > \frac{2}{\sqrt{n-p}}. \quad (15)$$

Ввиду этого, из уравнения регрессии должны быть исключены все регрессоры, не удовлетворяющие этому неравенству.

Пятый этап. Выполняется процедура устранения регрессоров, связанных функциональной линейной зависимостью (устранение мультиколлинеарности). Для выполнения этой процедуры по каждому из отобранных на предыдущих этапах регрессоров строятся уравнения регрессии по остальным регрессорам. Так, регрессор X_j , обозначаемый на данном этапе искомой величиной, вычисляется по уравнению регрессии:

$$X_j(t) = a_0 + a_1 X_1(t) + \dots + a_{j-1} X_{j-1}(t) + a_{j+1} X_{j+1}(t) + \dots + a_p X_p(t), \quad (16)$$

где $a_0, a_1, \dots, a_p$ — коэффициенты уравнения регрессии, вычисляемые по формулам (7-9).

Затем для каждого регрессора, входящего в уравнение (16), вычисляются коэффициент детерминации и частный коэффициент корреляции по отношению к искомой величине X_j . Для этого используются формулы (11-13) применительно к искомой величине X_j , определяемой уравнением (16) по регрессорам $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p$.

Если значение коэффициента детерминации k -го регрессора X_k больше значения его частного коэффициента корреляции, то регрессор X_k ответствен за мультиколлинеарность, и он исключается из числа регрессоров. Если аналогичное условие имеет место для нескольких регрессоров, то исключается из числа регрессоров только один из них, имеющий наименьшее значение частного коэффициента корреляции.

Оставшееся число регрессоров подвергается аналогичной процедуре:

- по каждому из них строятся уравнения регрессии по остальным регрессорам, также вычисляются коэффициент детерминации и частный коэффициент корреляции каждого оставшегося регрессора по отношению к регрессору, играющему роль искомой величины, и удаляется регрессор, у которого значение коэффициента детерминации больше значения его частного коэффициента корреляции и т. д. до тех пор, пока все построенные коэффициенты детерминации не окажутся меньше частных коэффициентов корреляции оставшихся регрессоров.

Шестой этап. По оставшимся регрессорам строится необходимое уравнение регрессии и по обучающей выборке рассмотренным методом наименьших квадратов определяются его коэффициенты.

Статистический прогноз значения искомой величины

Для искомого качественного показателя, который является случайным стационарным процессом с настолько низкочастотными изменениями относительно его среднего значения, что между некоторым числом соседних по времени его лабораторных замеров существует значимая корреляционная связь, можно реализовать виртуальную оценку текущего значения искомой величины путем статистической экстраполяции ряда его последних по времени лабораторных замеров — построением линейного разностного уравнения [2].

Пусть лабораторные измерения случайного стационарного процесса Y представляют собой последовательность замеров $y_t, y_{t-1}, \dots$ в дискретные моменты

времени $t, t-1, \dots$. Тогда, после перехода к новым переменным $z_t = y_t - m_y$, где m_y — оценка среднего значения величины Y , связь между последовательными значениями лабораторных измерений описывается линейным разностным уравнением вида:

$$\hat{z}_t = \sum_{i=1}^p \beta_i z_{t-i}, \quad (17)$$

где $\hat{z}_t$ — оценка преобразованного значения последнего лабораторного замера $z_t = y_{\text{изм}}(t) - m_y$ в момент времени t ; β_i — коэффициенты уравнения; p — порядок уравнения.

Переход к новым переменным z_t выполняется для удаления из уравнения постоянного члена, что необходимо для получения приведенных ниже оценок коэффициентов.

Используя полученное уравнение, построим прогноз следующего значения z_{t+1} на момент времени $t+1$:

$$\overline{z}_{t+1} = \sum_{i=1}^p \beta_i z_{t-i+1}. \quad (18)$$

Определим коэффициенты уравнения (18) [2].

Для вычисления коэффициента β_1 рассмотрим уравнение:

$$\begin{aligned} \hat{z}_t &= \beta_1 z_{t-1}, \\ \overline{z}_{t+1} &= \sum_{i=1}^p \beta_i z_{t-i+1}, \end{aligned} \quad (19)$$

Для вычисления коэффициента β_2 рассматривается уравнение второго порядка:

$$\begin{aligned} \hat{z}_t &= \beta_1 z_{t-1} + \beta_2 z_{t-2}; \\ \beta_2 &= \frac{\begin{vmatrix} 1 & r(1) \\ r(1) & r(2) \end{vmatrix}}{\begin{vmatrix} 1 & r(1) \\ r(1) & 1 \end{vmatrix}}. \end{aligned} \quad (20)$$

Для вычисления коэффициента β_j , где j последовательно принимает значения $j = 3, 4, \dots, p$, рассматривается уравнение

$$\begin{aligned} \overline{z}_{t+1} &= \sum_{i=1}^j \beta_i z_{t-i+1}; \\ \beta_j &= \frac{L_{jj}}{L_j}, \end{aligned} \quad (21)$$

$$\text{где } L_j = \begin{vmatrix} 1 & r(1) & r(2) & \dots & r(j-1) \\ r(1) & 1 & r(1) & \dots & r(j-2) \\ r(2) & r(1) & 1 & \dots & r(j-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(j-1) & r(j-2) & r(j-3) & \dots & 1 \end{vmatrix}$$

$$L_{jj} = \begin{vmatrix} 1 & r(1) & r(2) & \dots & r(1) \\ r(1) & 1 & r(1) & \dots & r(2) \\ r(2) & r(1) & 1 & \dots & r(3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(j-1) & r(j-2) & r(j-3) & \dots & r(j) \end{vmatrix}.$$

*Наука - это великая красота.
Ученый у себя в лаборатории не просто техник:
это ребенок лицом к лицу с явлениями природы,
действующими на него, как волшебная сказка.*

Мария Кюри - Складовская

Здесь $r(j)$, $j = 1, 2, \dots, p$ — оценки коэффициентов автокорреляции случайного процесса Y , которые вычисляются по формуле:

$$r(j) = \frac{\sum_{t=1}^n z_t z_{t-j}}{S_Y^2}, \quad (22)$$

где n — число сопоставимых по разности времени $t - j$ замеренных лабораторных анализов величины Y в обучающей выборке, по которой строится уравнение (18), S_Y — оценка среднеквадратичного отклонения искомой величины Y от ее среднего значения.

Прогнозная оценка процесса Y в момент его следующего лабораторного замера $t + 1$ определяется как

$$\overrightarrow{y_{t+1}} = \overrightarrow{z_{t+1}} + m_y,$$

где $\overrightarrow{z_{t+1}}$ вычисляется уравнением (18). Использование линейного разностного уравнения (18) для оценки значения искомого процесса в интервале между последним по времени лабораторным анализом в момент $t(y_t)$ и будущим лабораторным замером в момент $t+1(y_{t+1})$ может базироваться на линейной интерполяции между этими значениями. В этом случае значение $\overrightarrow{y(t^*)}$, где $t^* = (t) + \Delta < t + 1$ для любого Δ в диапазоне $0 < \Delta < 1$ вычисляется по формуле:

$$\overrightarrow{y(t^*)} = (1 - \Delta)y_t + \Delta \overrightarrow{y_{t+1}}. \quad (23)$$

Последовательное вычисление коэффициентов уравнения (18) по формулам (19-21) позволяет получить прогноз, имеющий минимальную среднеквадратичную ошибку и использующий минимально необходимое число членов уравнения (18), поскольку при описанном выше способе вычисления значений коэффициентов уравнения (18) порядок уравнения (18) определяется числом ненулевых коэффициентов β_i .

Коэффициенты уравнения считаются ненулевыми (значимыми), если их значения, рассчитываемые по имеющейся обучающей выборке, превышают по абсолютной величине указанный ниже порог:

$$\beta_i > \frac{2}{\sqrt{n-i}}. \quad (24)$$

Если, начиная с некоторого k , не выполняется условие (24) для всех $i > k$, то порядок модели выбирается равным k .

Общие замечания

по использованию статистического анализа

1. Полученные любым из описанных способов статистические уравнения действительны для виртуального измерения величины Y в тех режимах работы агрегата, в которых по обучающей выборке определены коэффициенты уравнений. При других режимах следует за-

ново набирать статистику измеряемых автоматически величин X_k и измеряемой лабораторными замерами величины Y и корректировать коэффициенты уравнений.

На практике существенные изменения режима работы агрегата, свойств перерабатываемого сырья, характеристик оборудования происходят зачастую непредсказуемо для оператора процесса и иногда непрерывно во времени, поэтому важно при использовании уравнений предусмотреть:

- коррекцию оценки искомой величины по ряду ее последовательных лабораторных замеров путем изменения среднего значения искомой величины m_y , чтобы в последующие моменты лабораторных замеров их значения были достаточно близки (коррекция оценки по единичному лабораторному замеру может быть ошибочной, например, из-за несовпадения момента времени взятия пробы для лабораторного анализа, принятому моменту оценки величины);

- возможную постепенную автоматическую подстройку всех коэффициентов уравнений на основе совокупности полученных за значительный временной интервал функционирования виртуального анализатора значений регрессоров и лабораторных замеров искомой величины Y .

2. Значения всех приведенных статистических характеристик: средних значений и среднеквадратичных отклонений; авто- и взаимных корреляционных функций; коэффициентов автокорреляции и корреляционных отношений; коэффициентов детерминации, парной и частной корреляции; коэффициентов уравнений могут автоматически рассчитываться в любом статистическом пакете программ (например, Statistica, SPSS, MESOSAUR), поэтому получение всех используемых при построении статистических уравнений показателей и параметров не требует ручного труда.

Примеры построения статистических уравнений

В качестве примеров описанных подходов рассмотрим виртуальные оценки качества продуктов, полученных на отдельных установках химико-технологического производства.

Технологическая установка № 1.

Не измеряемое поточным прибором качество продукта Y раз в смену определяется лабораторным анализом. Оператор ведет ТП по совокупности семи измеряемых режимных величин: X_k , где $k = 1, 2, 3, 4, 5, 6, 7$. Получена обучающая выборка, состоящая из 60 совместных измерений, из них 39 измерений используются непосредственно для обучения (вычисления параметров), а 21 измерение — для тестирования полученных оценок.

Построим корреляционную матрицу (совокупность парных коэффициентов корреляции) всех измеряемых величин и показателя Y (табл. 1).

Если при построении уравнения регрессии формировать набор регрессоров с помощью парных коэффициентов корреляции и считать граничным значением

Таблица 1

	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
Y	1,000	0,017	-0,289	0,018	-0,303	-0,407	-0,077	0,193
X ₁	0,017	1,000	0,034	0,527	0,411	0,449	0,599	0,296
X ₂	-0,289	0,034	1,000	0,062	-0,247	0,514	0,393	0,188
X ₃	0,018	0,527	0,062	1,000	0,389	0,414	0,406	0,288
X ₄	-0,303	0,411	-0,247	0,389	1,000	0,449	0,342	0,083
X ₅	-0,407	0,449	0,514	0,414	0,449	1,000	0,773	0,510
X ₆	-0,077	0,599	0,393	0,406	0,342	0,773	1,000	0,547
X ₇	0,193	0,296	0,188	0,288	0,083	0,510	0,547	1,000

коэффициента корреляции величину 0,5, то, как видно из корреляционной матрицы регрессоров, имеющих парный коэффициент корреляции с искомой величиной Y, превышения этой границы не существует, то есть уравнение регрессии построить нельзя.

Для формирования набора регрессоров с помощью частных коэффициентов корреляции рассчитаем их по формулам (11-13):

$$\begin{aligned} r_{YX_1 \text{ при } X_2 \dots X_7} &= 0,06; r_{YX_2 \text{ при } X_1 X_3 \dots X_7} = 0,02; \\ r_{YX_3 \text{ при } X_1 X_2 X_4 \dots X_7} &= 0,104; r_{YX_4 \text{ при } X_1 X_2 X_3 X_5 \dots X_7} = 0,07; \\ r_{YX_5 \text{ при } X_1 \dots X_4 X_6 X_7} &= 0,399; r_{YX_6 \text{ при } X_1 \dots X_5 X_7} = 0,415; \\ r_{YX_7 \text{ при } X_1 \dots X_6} &= 0,475. \end{aligned}$$

Отберем в качестве регрессоров те, для которых частный коэффициент корреляции является значимым (15), то есть превышает порог

$$h = \frac{2}{\sqrt{39}} = 0,32.$$

Регрессоры X₅, X₆ и X₇ имеют значимые частные коэффициенты корреляции. Проверка их на мультиколлинеарность по формуле (16) и затем по формулам (11-13) показала, что коэффициент детерминации в регрессии X₆ на X₅ и X₇ больше значения частного коэффициента корреляции между X₆ и Y, поэтому его следует исключить из числа регрессоров. Окончательно уравнение регрессии строится по измеряемым величинам X₅ и X₇:

$$Y = 43,919 - 0,13X_5 + 0,086X_7.$$

Среднеквадратичное отклонение оценки величины Y по этому уравнению от фактического значения составляет 0,184.

Построим статистический прогноз величины Y.

Все коэффициенты уравнения (18), кроме первого, незначимы: $\beta_1 = 0,79$, $\beta_2 = 0,052$, $\beta_3 = 0,047$.

Рассчитанная среднеквадратическая погрешность такого прогноза равна 0,139.

Технологическая установка № 2.

Не измеряемое поточным прибором качество продукта Y раз в 4 часа определяется лабораторным анализом. Оператор ведет ТП по совокупности семи измеряемых режимных величин: $\beta_1 X_k$, где $k = 1, 2, 3, 4, 5, 6, 7$. Получена обучающая выборка, состоящая из 60 совместных измерений, из них 42 измерения используются непосредственно для обучения (вычисления параметров), а 18

Таблица 2

	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
Y	1,000	0,573	0,357	-0,568	-0,841	0,354	0,465	0,808
X ₁	0,573	1,000	0,347	-0,257	-0,485	0,295	0,510	0,491
X ₂	0,357	0,347	1,000	-0,126	-0,440	0,343	0,432	0,407
X ₃	-0,568	-0,257	-0,126	1,000	0,764	0,172	-0,030	-0,193
X ₄	-0,841	-0,485	-0,440	0,764	1,000	0,038	-0,193	-0,521
X ₅	0,354	0,295	0,343	0,172	0,038	1,000	0,776	0,668
X ₆	0,465	0,510	0,432	-0,030	-0,193	0,776	1,000	0,676
X ₇	0,808	0,491	0,407	-0,193	-0,521	0,668	0,676	1,000

измерений — для тестирования полученных оценок. Корреляционная матрица представлена в табл. 2.

Выделим регрессоры, имеющие парный коэффициент корреляции с искомой величиной Y больше граничного значения -0,5. Это измеряемые величины X₁, X₃, X₇.

Построим по ним уравнение регрессии:

$$Y = 10,004 + 0,048X_1 - 0,046X_3 + 0,102X_7.$$

Среднеквадратичная погрешность оценки Y равна 0,259.

Определим теперь регрессоры с помощью частных коэффициентов корреляции:

$$\begin{aligned} r_{YX_1 \text{ при } X_2 \dots X_7} &= 0,043; r_{YX_2 \text{ при } X_1 X_3 \dots X_7} = 0,309; \\ r_{YX_3 \text{ при } X_1 X_2 X_4 \dots X_7} &= 0,539; r_{YX_4 \text{ при } X_1 X_2 X_3 X_5 \dots X_7} = 0,006; \\ r_{YX_5 \text{ при } X_1 \dots X_4 X_6 X_7} &= 0,436; r_{YX_6 \text{ при } X_1 \dots X_5 X_7} = 0,167; \\ r_{YX_7 \text{ при } X_1 \dots X_6} &= 0,489. \end{aligned}$$

Выберем в качестве регрессоров те, для которых частный коэффициент корреляции является значимым, то есть превышает порог

$$h = \frac{2}{\sqrt{42}} = 0,31.$$

Таковыми регрессорами являются X₃, X₅, X₇. Проверка их на мультиколлинеарность показала, что коэффициент детерминации во всех построенных регрессиях меньше соответствующего частного коэффициента корреляции, поэтому строим уравнение регрессии с регрессорами X₃, X₅, X₇:

$$Y = 69,028 - 0,134X_3 - 0,115X_5 + 0,023X_7.$$

Среднеквадратичная погрешность оценки Y равна 0,201.

Построим статистический прогноз величины Y. Все коэффициенты уравнения (18) кроме первого незначимы: $\beta_1 = 0,64$, $\beta_2 = 0,077$.

Рассчитанная среднеквадратическая погрешность такого прогноза также равна 0,201.

Виртуальные измерения на базе нейронной сети

В общем случае наличия сугубо нелинейных связей между искомой величиной и статистически связанными с ней измеряемыми величинами целесообразно использовать в качестве виртуального измерителя нейронную сеть.

В типовых промышленных применениях используется трехслойная нейронная сеть, схема которой приведена на рисунке.

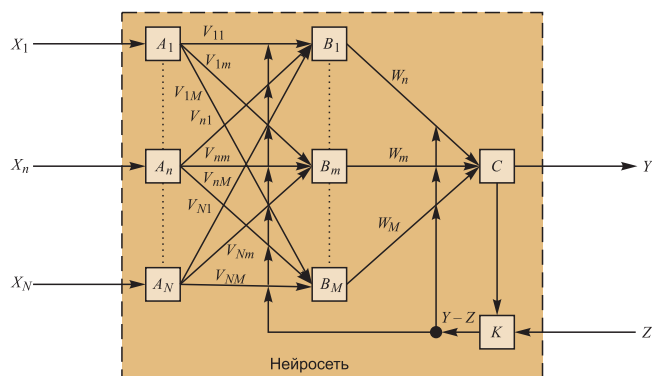


Рис. Схема нейронной сети – виртуального измерителя, где x_n – значение n -ой измеряемой величины ($n = 1...N$), Y – оценка значения искомой величины, Z – лабораторное измерение искомой величины, A_n – n -й элемент входного слоя сети ($n = 1...N$), B_m – m -й элемент промежуточного слоя сети ($m = 1...M$), C – элемент выходного слоя сети, V_{nm} – веса связей элементов входного и промежуточного слоев ($n = 1...N$, $m = 1...M$), W_m – веса связей элементов промежуточного слоя ($m = 1...M$) с элементом C , K – элемент сопоставления оценки искомой величины с ее лабораторным замером, $(Y - Z)$ – обучающее воздействие (корректировка весов связей)

Значения измеряемых величин поступают на элементы входного слоя (каждый его элемент соединен с определенным датчиком). Число входных элементов соответствует числу измеряемых величин, задействованных в системе оценки искомой величины. Каждый элемент промежуточного слоя взаимосвязан со всеми элементами входного и выходного слоев. Выходной слой, в котором в рассматриваемом случае находится один элемент, связан со всеми элементами промежуточного слоя и имеет выход, являющийся оценкой искомой величины. Взаимные связи эле-

ментов внутри каждого слоя отсутствуют. Имеющиеся значения связей любых элементов друг с другом определяются их весами (коэффициентами связи), изменения которых меняют общую зависимость между измеряемыми и искомой величиной. Настройка сети осуществляется процедурой обучения, которая заключается в подаче на сеть обучающего набора пар значений измеряемых величин и соответствующих им значений искомой величины (в качестве последних значений используются лабораторные замеры искомой величины), сделанные в один и тот же момент времени. Сам процесс настройки, автоматически реализуемый в нейросети, заключается в таком изменении весов связей между всеми элементами сети, которое уменьшает различие между имеющимся выходом сети в определенный момент времени и требуемым выходом – лабораторно измеренным значением искомой величины в этот же момент времени.

Практически обучающий набор должен иметь порядка сотни и более совпадающих во времени значений измеряемых величин и лабораторных замеров, из которых примерно 70% значений используются для обучения нейросети, а 30% – для тестирования результата настройки и оценки полученной точности виртуального измерения. Если в архиве системы автоматизации данного объекта присутствует необходимое число таких пар, полученных ранее при существующем режиме и свойствах ТП, то обучение нейросети может быть проведено достаточно быстро и точно; в противном случае обучающий набор должен быть получен во время текущей работы объекта, что может значительно удлинить процесс настройки нейросети. Уже включенная в

ОАО "Беллакт" совершенствует управление ремонтами

Компания АйТиЭм начала работу по проекту внедрения информационной системы управления техническим обслуживанием и ремонтами Волковысского ОАО "Беллакт" (г. Волковыск, Республика Беларусь) – современной, динамично развивающейся компании, лидирующей в Беларуси на рынке молочных продуктов для детского питания и другой молочной продукции. Значительное внимание компания Беллакт уделяет эффективности основных фондов. В рамках этих задач компания пригласила консультантов АйТиЭм для реализации информационной системы управления техническим обслуживанием и ремонтами (ИСУ ТОиР).

Компания АйТиЭм предложила Беллакт апробированное на практике комплексное решение, включающее не только ПО, но также наработки в области измерения эффективности процессов ТОиР, постановки современных практик ТОиР, внедрения информационных систем в технических службах предприятия.

К настоящему времени специалисты АйТиЭм провели аудит процессов ТОиР на предприятии, и на этой основе разработали проектную документацию – спецификации на внедрение ИСУ ТОиР, на создание БД сис-

темы, на инфраструктуру будущей информационной системы. В качестве программной платформы проекта будет использована EAM/MRO-система TRIM разработки НПП "СпецТек".

Согласно проектной документации НПП "СпецТек" поставит Волковысскому ОАО "Беллакт" 11 пользовательских лицензий на ПО TRIM. Далее, специалисты АйТиЭм выполнят работы по наполнению БД системы (паспортизация оборудования), по конфигурированию и установке ПО и БД, совместно с заказчиком разработают систему показателей процессов ТОиР и аналитические формы для их отображения и анализа, проведут обучение пользователей. Занятия с персоналом заказчика будут проводиться на основе регламента эксплуатации ИСУ ТОиР, который будет предусматривать распределение пользователей по функциям, цепочки выполнения действий пользователей по процессам при реализации передовых практик ТОиР, заложенных в функциональность TRIM. Важным этапом проекта станет также внедрение конвертера для обмена данными между ИСУ ТОиР и системой бухгалтерского учета на основе продуктов 1С.

[Http://www.trim.ru](http://www.trim.ru)

работу нейросеть непрерывно адаптируется: каждое лабораторное измерение поступает в нейросеть, где сравнивается с выходом сети, и полученная разность их значений используется для текущей корректировки весов связей элементов сети. Это позволяет сохранять достигнутую точность виртуального изменения при постепенных, достаточно плавных изменениях режима и/или свойств объекта. При резких и значительных изменениях в ТП нейросеть должна быть переведена в режим обучения и процесс ее обучения необходимо проводить заново.

Следует отметить, что первоначальная совокупность измеряемых величин, подключенных к входному слою нейросети, может быть уменьшена после окончания процесса обучения, если окажется, что некоторые измеряемые величины дают очень малый или даже нулевой вклад в оценку искомой величины (веса связей элементов, к которым подключены эти величины, с элементами промежуточного слоя сети получились близкими к нулю).

Такая нейросеть, сформированная в виде функционального блока, входит, например, в состав распределенной системы управления Delta V фирмы Emerson (www.metrans.ru/products/DeltaV). Она реализована в отдельном контроллере системы Delta V и допускает одновременное подключение к ее входному слою сигналов 16 различных датчиков. Хорошие результаты настройки нейросети дают обучающие выборки 200...300 измерений и соответствующих им лабораторных анализов. При переходе технологической установки на

другой режим управление по выходу нейросети останавливается и происходит корректировка (уточнение) модели, реализуемой в нейросети; после чего управление по выходу нейросети восстанавливается.

Заключение

Необходимость оценки многих, не измеряемых непосредственно производственных величин со временем все более увеличивается, поскольку для создания современных систем автоматизации производства, базирующихся на совершенных алгоритмах управления, требуется достаточно полная информация обо всех показателях производства, а она не может быть получена без использования различных методов виртуального измерения. Кроме того, в ряде случаев, даже когда для оценки значения величины можно выбрать установку поточного прибора или виртуального измерителя, целесообразно предпочесть второй вариант, поскольку он значительно дешевле и может удовлетворять по требованиям к точности оценки величины.

В связи с указанными обстоятельствами следует ожидать значительного возрастания интереса предприятий, проектных организаций, системных интеграторов к возможностям и существующим методам виртуального измерения.

Список литературы

1. Елисеева И. Эконометрика. Финансы и статистика. 2007.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып.1. Мир. 1974.

Гребенюк Елена Алексеевна — д-р техн. наук, ведущий научный сотрудник,

Ицкович Эммануил Львович — д-р техн. наук, проф., зав. лабораторией ИПУ им. В.А. Трапезникова РАН.

Контактный телефон (495) 334-90-21.

RTX 2011 – самая мощная версия расширения реального времени для Windows: поддержка до 32 процессоров!

Компания IntervalZero Inc. объявляет о выпуске новой версии расширения PB RTX для ОС семейства Microsoft Windows.

RTX 2011 позволяет отказаться от применения отдельных DSP-процессоров и вентильных матриц FPGA, что значительно сокращает стоимость разработки систем на базе ОС Windows, требующих поддержки жесткого PB и развитых функций пользовательского интерфейса. Новая версия поддерживает до 32 процессоров, работающих под RTX в полноценной SMP-конфигурации (Symmetric MultiProcessing – симметричное мультипроцессирование), что четырехкратно превышает возможности предыдущей версии RTX. Кроме того, RTX 2011 традиционно поддерживает современные версии продуктов компании Microsoft, такие как Windows 7, Windows Embedded Standard 7, Visual Studio 2010 SP1.

В отличие от традиционной архитектуры, которая предполагает использование двух сред разработки и двух независимых систем для одновременной поддержки PB и сложных пользовательских интерфейсов, RTX является единственным решением, позволяющим совместить все это в рамках одной интегрированной системы, реализующей как ЧМИ на основе ОС Windows, так и функциональность PB. При этом задействовано столько про-

цессорных ядер, сколько требует специфика конкретной прикладной задачи. Кроме того, расширение RTX является единственным продуктом, способным обеспечивать поддержку конфигураций SMP на одном компьютере исключительно на основе ОС Windows.

Компания IntervalZero также объявляет о доступности для новой версии RTX шести опций Runtime (сред исполнения), которая позволяет уже сегодня выбирать число поддерживаемых процессоров с возможностью масштабирования системы в будущем. Диапазон вариантов RTX 2011 Runtime простирается от Runtime Solo – поддержка PB на одном совместно используемом или выделенном для подсистемы RTSS (Real-Time SubSystem) ядре процессора до Runtime Ultimate – как одного совместно используемого ядра, так и с выделением для подсистемы RTSS до 31 процессорного ядра.

Новая версия расширения RTX является важным компонентом концепции Soft-Control Architecture (архитектура программного управления), с помощью которой многие OEM-клиенты компании IntervalZero экономят на разработке своих систем 25...50 % от прежнего уровня затрат, радикально увеличивают пропускную способность своих решений, а также повышают качество и функциональность конечных продуктов.

[Http:// www.rtsoft.ru](http://www.rtsoft.ru)