

РЕГУЛЯРИЗАЦИЯ РЕШЕНИЙ ПРИ СТАТИСТИЧЕСКОЙ ОБРАБОТКЕ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ КОГНИТИВНЫХ МЕТОДОВ МОДЕЛИРОВАНИЯ

Веревкин А.П., Муртазин Т.М.

Уфимский государственный нефтяной университет, Уфа

Ключевые слова: регрессионная модель, временной ряд, взаимная корреляция, когнитивный подход, нормализация.

Аннотация. Системы усовершенствованного управления (АРС-системы) базируются на использовании моделей технологических процессов, позволяющих оперативно прогнозировать изменение технологических параметров (ТП) и показателей качества продуктов (ПКП) производства. Для идентификации структуры и параметров моделей обычно используется статистическая информация, которая представляет собой результаты пассивных экспериментов. Данные режимных параметров в системах управления архивируются в базе данных в виде временных последовательностей без обеспечения их достоверности и однородности. В условиях не стационарности временных рядов и неоднородности данных резко снижается качество моделей расчета ТП и ПКП, поэтому они не могут, как правило, быть применены для разработки моделей без предварительной подготовки. Подготовка направлена на обеспечение условий стационарности рядов, адекватности моделей и включает в себя этапы выделения кластеров (фрагментов) данных и их фильтрацию, для которых в дальнейшем будут формироваться ситуационные модели. В статье рассматриваются методы приведения данных из архивов АСУТП в общедоступный формат, их фильтрации и кластеризации. Приведено описание методов фильтрации и кластеризации на основе использования так называемых контрольных когнитивных моделей и коэффициентов кросс-корреляции ТП и ПКП.

SOLUTIONS REGULARIZATION IN STATISTICAL PROCESSING TIME SERIES BASED ON COGNITIVE MODELING METHODS

Verevkin A.P., Murtazin T.M.

Ufa State Petroleum Technological University, Ufa

Keywords: regression model, time series, cross-correlation, cognitive approach, normalization.

Abstract. Advanced control systems (APC-systems) are based on the use of process models that allow you to quickly predict changes in technological parameters (TP) and product quality indicators (QI) of production. Statistical information, which is the results of passive experiments, is usually used to identify the structure and parameters of models. Data of operating parameters in control systems are archived in the database in the form of time sequences without ensuring their reliability and uniformity. In conditions of non-stationarity of time series and heterogeneity of data, the quality of TP and control panel calculation models sharply decreases, therefore, as a rule, they cannot be used for model development without preliminary preparation. The training is aimed at ensuring the conditions for the stationarity of the series, the adequacy of the models and includes the stages of selecting clusters (fragments) of data and filtering them, for which situational models will be formed in the future. The article discusses methods of bringing data from the archives of the automated control system into a common format, filtering and clustering them. The methods of filtering and clustering based on the use of so-called control cognitive models and cross-correlation coefficients of TP and QI are described.

При разработке систем оперативного управления технологических процессов и диагностики широко используются модели, для идентификации

структуры и параметров которых обычно используется статистическая информация [1-3]. Статистические данные имеют вид временных последовательностей, для которых, обычно, условия однородности и стационарности не выполняются, что резко снижает качество получаемых по ним моделей [4].

Проблема не стационарности объекта управления может быть связана как с влиянием не измеряемых факторов, так и изменением характеристик объекта (системы) во времени, например, из-за наличия динамики процессов управления. Применение традиционных подходов к получению моделей в форме регрессионного уравнения в условиях не стационарности объекта может привести к оценкам коэффициентов уравнений регрессии, которые противоречат закономерностям процесса (не физичность модели) как в отношении знаков коэффициентов, так и их значений (неустойчивость модели). Поэтому при проведении подготовки данных и моделировании, предлагается комплексно использовать приемы статистического анализа, когнитивную информацию и когнитивные модели [5-7].

Обобщение этого подхода состоит в использовании когнитивной модели (КМ) процесса, параметры которой определяются с точностью до интервальных значений, назначаемых экспертно или на основе «строгих» моделей. При этом когнитивная модель может быть не сбалансирована и иметь низкую адекватность по отношению к реальным данным, но качественно верно отображать связи переменных. Задачу моделирования предлагается решать в несколько этапов.

На первом этапе экспертным путем (когнитивно) определяется набор параметров (концептов) модели, которые, по мнению эксперта, должны входить в когнитивную модель, и формируется когнитивная модель $M(R, T, Q, A)$, где R – множество управлений, T – множество технологических параметров (параметров состояния), Q – множество зависимых переменных, например, показателей качества продуктов, A – матрица коэффициентов модели. Дополнительно, на основе результатов кросс-корреляционного анализа в состав концептов могут включаться переменные, имеющие высокий коэффициент корреляции с параметрами множеств Q и R или T , если это не противоречит закономерностям процесса. Таким образом, обобщенная структура когнитивной модели процесса будет учитывать как эвристическую, так и эмпирическую информацию. Заметим, что при проведении кросс-корреляционного анализа временных рядов должны учитываться транспортные задержки (лаги) между временными рядами, определяемые известными методами [4].

На втором этапе выполняется кластеризация временных рядов (множеств) $R(t)$, $T(t)$, $Q(t)$ как функций времени t . Фрагменты временных рядов выбираются так, чтобы средние значения переменных для разных фрагментов отличались заметно.

Предварительно для каждого фрагмента должны быть изъяты промахи (аномальные наблюдения) [8]. Число наблюдений должно быть не менее 10-15 на каждом интервале дискретности.

На третьем этапе для каждого фрагмента последовательно оценивают линейные уравнения связи зависимой переменной от каждого регрессора и

получают частные коэффициенты регрессии $a_{i(r)}$, $a_{i(t)}$ где $i(r)$, $i(t)$ – номер фрагмента для переменной управления r из множества R и технологического параметра t из множества T , соответственно. Количество коэффициентов $a_{i(r)}$ составит $n_r \times \sum_{i=1}^{n_r} n_{i(r)}$, n_r – количество параметров множества R , $n_{i(r)}$ – количество выделенных фрагментов для переменной r . Количество коэффициентов $a_{i(t)}$ составит $n_t \times \sum_{i=1}^{n_t} n_{i(t)}$, n_t – количество параметров множества T , $n_{i(t)}$ – количество выделенных фрагментов для переменной t .

На четвертом этапе выполняют процесс фильтрации частных коэффициентов регрессии и выделяют коэффициенты регрессии $\bar{a}_{i(r)}$, $\bar{a}_{i(t)}$, удовлетворяющие следующим трем критериям:

1. Знак коэффициентов кросс-корреляции, определяемых для пар Q и T или R , должен совпадать со знаком «весов» дуг (или последовательности дуг) когнитивной модели;
2. Значения оценок частных коэффициентов регрессии при факторах должны лежать в диапазоне изменения коэффициентов связей концептов в когнитивной модели;
3. Значения коэффициентов кросс-корреляции по модулю должны превышать некоторые граничные значения, которые позволяют считать связь достаточно сильной.

На пятом этапе из отобранных на четвертом этапе частных коэффициентов регрессии рассчитываются коэффициенты регрессии для пар Q и T или R , которые «взвешивают» для каждого параметра с учетом нормировки:

$$a(R) = \sum_{i=1}^{\bar{n}_{i(r)}} \bar{a}_{i(r)} \times c_{i(r)}, \quad a(T) = \sum_{i=1}^{\bar{n}_{i(t)}} \bar{a}_{i(t)} \times c_{i(t)},$$

где $\bar{n}_{i(r)}$, $\bar{n}_{i(t)}$ – количество частных коэффициентов регрессии для параметров множеств R и T , прошедших фильтрацию; $c_{i(r)}$, $c_{i(t)}$ – коэффициенты взвешивания, определяемые как

$$c_{i(r)} = \left| \frac{\overline{Cov_{i(r)}(R, Q)}}{\sum_{i=1}^k \overline{Cov_{i(r)}(R, Q)}} \right|, \quad c_{i(t)} = \left| \frac{\overline{Cov_{i(t)}(T, Q)}}{\sum_{i=1}^l \overline{Cov_{i(t)}(T, Q)}} \right|,$$

$\overline{Cov_{i(r)}(R, Q)}$, $\overline{Cov_{i(t)}(T, Q)}$ – нормируемые коэффициенты ковариации, определимые как

$$\overline{Cov(R, Q)} = \frac{Cov(R, Q)}{M(R) \times M(Q)}, \quad \overline{Cov(T, Q)} = \frac{Cov(T, Q)}{M(T) \times M(Q)},$$

$M(R)$, $M(T)$, $M(Q)$ – математические ожидания параметров множеств R , T , Q на выделенных временных фрагментах.

Таким образом, комплексирование когнитивного и статистического моделирования способно обеспечить теоретическую обоснованность получаемых

моделей, а механизмы оперативной адаптации [9] повысить их вычислительную точность под изменяющиеся характеристики процесса.

Отметим, что даже при относительно невысоком коэффициенте детерминации полученные модели могут успешно использоваться для целей прогноза в задачах оперативного управления и оптимизации технологических режимов.

Список литературы

1. Веревкин А.П., Кирюшин О.В. Автоматизация технологических процессов и производств в нефтепереработке и нефтехимии. – Уфа: Изд-во УГНТУ, 2005. – 71 с.
2. Terrence Blevins, Willy K. Wojsznis, Mark Nixon. Advanced Control Foundation: Tools, Techniques and Applications. – ISA, 2012. – 556 p.
3. Kadlec P., Gabrys B. and Strandt S. Data-driven soft sensors in the process industry // Computers and Chemical Engineering. 2009. Vol. 33. P. 795-814.
4. Орлова Е.В. Эконометрическая методология исследования систем: учебник. – М.: Инфра, 2019. – 216 с.
5. Васильев В.И., Ильясов Б.Г. Интеллектуальные системы управления. Теория и практика: учебное пособие. – М.: Радиотехника, 2009. – 329 с.
6. Веревкин А.П. Когнитивные модели в системах искусственного интеллекта: цели и методы построения // Интеграция науки и образования в вузах нефтегазового профиля-2016 / Материалы Международной научно-методической конференции, посвященной 60-летию филиала УГНТУ. – Салават, 2016. – С. 167-170.
7. Веревкин А.П., Калашник Д.В., Хуснияров М.Х. Моделирование оперативного определения индекса расплава для управления процессом производства полиэтилена // Башкирский химический журнал. – 2013. – Т. 20, № 1. – С. 69-74.
8. Веревкин А.П., Муртазин Т.М., Денисов С.В., Устюжанин К.Ю. Подготовка данных для построения виртуальных анализаторов в задачах усовершенствованного управления // Автоматизация в промышленности. – 2019. – №3. – С. 12-17.
9. Веревкин А.П., Муртазин Т.М., Григорьева Ю.Л. Когнитивное моделирование процессов нефтепереработки с упрощенной процедурой адаптации динамических моделей // Территория Нефтегаз. – 2018. – № 7-8. – С. 14-18.

Сведения об авторах:

Веревкин Александр Павлович – д.т.н., профессор, профессор кафедры «Автоматизация, телекоммуникация и метрология»;

Муртазин Тимур Мансурович – к.т.н., доцент, доцент кафедры «Автоматизация, телекоммуникация и метрология».