

Comparison and Application of Various Parameter Adaptive Control Algorithms // Automatica. 1980. Vol. 16, № 2. P.117-133. 5. *Саридис Дж.* Самоорганизующиеся стохастические системы управления. М.: Наука, 1980. 400 с. 6. *Васильцова Н.В., Васильцова С.В., Шамша Б.В.* Алгоритм статистического анализа данных в АСУТП производства обесфторенных фосфатов, 1986. 18 с. Деп. УкрНИИТИ.

Поступила в редколлегию 18.05.98

УДК 681.51.015

СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДЛЯ ОЦЕНКИ КАЧЕСТВА ПРИКЛАДНЫХ ПРОГРАММ ИДЕНТИФИКАЦИИ

ВАСИЛЬЦОВА Н. В., ШАМША Б. В.

Рассматривается процесс автоматизированного проектирования программного обеспечения идентификации. Описываются характеристики качества алгоритмов и программ и влияющие на них статистические свойства объектов. Предлагается обобщенный подход к оценке качественных показателей алгоритмов с помощью имитационного моделирования.

Для решения практических задач идентификации и управления сложными технологическими объектами (процессами) существует достаточно большое число алгоритмов. Однако, несмотря на то, что арсенал алгоритмов велик, возникает немало препятствий их эффективному использованию. Каждый алгоритм ориентирован, как правило, на решение конкретного класса задач и обеспечивает эффективность решения лишь при выполнении некоторых теоретических предположений. Выбор того или иного алгоритма для конкретного объекта представляет определенную трудность, связанную с анализом и оценкой их эффективности в зависимости от класса объекта и его статистических характеристик. В связи с нелинейным характером алгоритмов идентификации и управления их аналитическое исследование представляет достаточно сложную задачу. Наиболее доступным методом исследования качественных и количественных показателей эффективности использования алгоритмов в автоматизированных системах управления технологическими процессами (АСУТП) является метод имитационного моделирования [1,2].

Анализ литературных источников [1,2] показал, что в настоящее время разработано большое количество подходов к синтезу систем имитационного моделирования.

В настоящей работе делается попытка обобщить накопленный опыт разработки имитационных моделей и предложить обобщен-

ный подход к их синтезу, основанный на предварительном анализе эффективности моделей с учетом статистических характеристик исследуемого объекта.

Использование имитационного моделирования при выборе алгоритма идентификации и управления, а также проведение эксперимента с разработанными моделями даст глубокое понимание исследуемого процесса (объекта) и системы управления им. Применение методов имитационного моделирования значительно сократит сроки выполнения работ по проектированию математического и программного обеспечения систем управления (МО и ПО) и значительно сэкономит материальные средства, выделяемые на разработку систем.

Рассмотрим имитационный подход к проектированию МО и ПО АСУТП, основанный на анализе эффективности использования алгоритмов идентификации и управления. Базой для использования такого подхода служат наборы алгоритмических и программных модулей решения отдельных задач управления.

Представим процесс проектирования модульной системы как разработку такой формализованной процедуры F выбора методов M , алгоритмов A , программ P решения K задач в АСУТП, которая

приводит к выбору системы $S = \bigcup_{k=1}^K S_k$, состоящей из подсистем S_k ,

$k = \overline{1, K}$. Выбор каждой S_k сводится к выбору такого подмножества M^k, A^k, P^k решения задачи, которое удовлетворяло бы заданному критерию качества и статистическим характеристикам объекта. Представим синтез S_k следующим образом:

$$\{M, A, P, H, E\} \xrightarrow{F} S_k \{M^k \in M; A^k \in A; P^k \in P\};$$

$$E^* = \text{extr}_{M^i \in M} \text{extr}_{A^j \in A} \text{extr}_{P^{ij} \in P} E(M^i, A^j, P^{ij}, H), \quad (1)$$

где M – класс методов; A – банк алгоритмов; P – библиотека программ решения отдельных задач; H – характеристики задачи; E – показатели качества подсистемы S .

В связи с этим выбор эффективного программного модуля может быть представлен как последовательный выбор метода, алгоритма и программы (1). Тогда для проектирования системы S необходимо решить следующие задачи:

– сформировать перечень K задач, решаемых в системе S , т.е. поставить задачу проектирования подсистем S_k ;

– сформировать класс методов M , банк алгоритмов A , библиотеку программ P решения k -й задачи;

– выявить характеристики и показатели E качества отдельных модулей, по которым можно производить сравнение между собой альтернативных модулей;

– выявить специфические свойства объекта H , которые влияют на качество решения задач в условиях функционирования системы;

– установить математическую зависимость между характеристиками качества модулей E и свойствами объекта H ;

– разработать формализованную процедуру F автоматизированного проектирования прикладного ПО, т.е. разработать критерии и алгоритмы выбора наилучшего модуля.

Для решения задачи выявления характеристик H и показателей E качества отдельных модулей известно достаточно много методов и алгоритмов, среди которых: методы максимального правдоподобия, минимизации байесовского риска и группового учета аргументов, обычный и взвешенный МНК, гребневые и робастные процедуры, рекуррентные, адаптивные и другие алгоритмы [3,4].

Качество методов идентификации может быть определено по следующим показателям:

– среднеквадратичной ошибке оценки коэффициентов модели

$$E_1 = M\{(\hat{B} - B)'(\hat{B} - B)\},$$

где $M\{\}$ – символ математического ожидания; B – вектор истинных коэффициентов модели; \hat{B} – вектор оценки коэффициентов модели;

– среднеквадратичной ошибке оценки коэффициентов, зависящей от плана

$$E_2 = M\{(\hat{B} - B)'(X'X)(\hat{B} - B)\};$$

– максимальной координатной ошибке оценки коэффициента модели

$$E_3 = \max(\hat{b}_j - b_j),$$

где $j = \overline{1, m^*}$; m^* – количество оцениваемых коэффициентов;

– относительной ошибке оценки коэффициента модели

$$E_4 = 1 / m^* \sum_{j=1}^{m^*} (\hat{b}_j - b_j) / b_j^2;$$

– общей дисперсии

$$E_6 = \text{Tr} \left\{ M \left\{ \left[\hat{B} - M \{ \hat{B} \} \right] \left[\hat{B} - M \{ \hat{B} \} \right]' \right\} \right\};$$

– максимальной координатной относительной оценке коэффициента модели

$$E_7 = \max \left| \left(\hat{b}_j - b_j \right) / b_j \right|, \quad j = \overline{1, m^*};$$

– смещению оценки коэффициента модели

$$E_8 = \left(M \{ \hat{B} \} - B \right) \left(M \{ \hat{B} \} - B \right)'$$

Такой набор показателей может использоваться для любых методов идентификации, являясь их исчерпывающей характеристикой.

Выбор методов идентификации для их последующей реализации определяется целью исследования, классом объекта, постановкой задачи, объемом и способом представления априорной информации. Состав банка алгоритмов определяется прежде всего спецификой объекта, а именно, временными и статистическими характеристиками протекающих в объекте процессов. Класс алгоритмов выбирается проектировщиком по результатам статистического анализа экспериментальных данных. Выбор программ для включения в библиотеку зависит от наличия готовых программных модулей и возможности программной реализации других алгоритмов.

Основными показателями качества программы являются: общее время, затраченное на поиск решения; общий объем требуемой оперативной памяти ЭВМ; достижимая точность решения.

Реализация процедур автоматизированного выбора модулей предполагает возможность их сравнительного анализа по некоторому критерию качества. Обязательным условием выбора показателя качества для сравнения модулей является возможность количественной оценки критерия. Более подробно характеристики качества алгоритмов и программ проанализированы в [5].

Ограничим класс объектов квазистационарными на интервале объектами, которые описываются линейными по параметрам моделями. Это ограничение не вносит каких-либо принципиальных изменений в модель и не снижает общности исследований. Если его необходимо снять, то множество H необходимо дополнить характеристиками тренда, а линейный оператор в модели заменить нелинейным.

Для рассматриваемого класса объектов множество H формируется из следующих параметров [1]: $h_1 = n$ – объем выборки; $h_2 = m$ – количество независимых переменных; $h_3 = \lambda_{\min}/\lambda_{\max}$ – показатель обусловленности матрицы плана; $h_4 = \sigma^2$ – дисперсия ошибки в зависимой переменной; $h_5 = \alpha'\alpha$ – длина вектора истинных коэффициентов в модели; $h_6 = \sigma_{\varepsilon}^2/\sigma_y^2$ – соотношение дисперсий "шум/сигнал" для зависимой переменной; $h_7 = \eta$ – уровень засоренности распределения, если последнее является смесью $f(\varepsilon) = (1-\eta)f_1(\varepsilon) + \eta f_2(\varepsilon)$; $h_8 = \tau_3$ – время затухания автокорреляционной функции $R(\tau_3)$; $h_9 = \sigma_{\xi}^2/\sigma_x^2$ – соотношение дисперсии "шум/сигнал" для независимых переменных; $h_{10} = \max r_{\xi,x}$ – максимальная корреляция ошибок с переменными x .

Характеристики h_j могут изменяться в широких пределах, но для рассматриваемого класса объектов интервалы их варьирования оказываются более узкими. На основе авторских исследований реальных процессов данные интервалы определены по каждой h_j как пара:

$$\begin{aligned} & (h_{j\max} - h_{j\min}) : h_1 = 20 - 200 ; h_2 = 5 - 50 ; h_3 = 10^{-1} - 10^{-3} \\ & (\text{или } h_3 = 10 - 1000) ; \\ & h_4 = 0 - 10 ; h_5 = 1 - 100 ; \\ & h_6 = 0 - 60 ; h_7 = 0 - 20 ; h_8 = 0 - 50 ; h_9 = 0 - 60 ; \\ & h_{10} = 0 - 0.5 . \end{aligned}$$

Интервалы исследования необходимо принять более широкими, чем они обычно встречаются на практике. Это вызвано стремлением изучить поведение алгоритма в экстремальных ситуациях на границах области изменения параметров объекта Ω_H . Кроме указанных $h_j, j = \overline{1,10}$, параметрами модели являются математические ожидания, дисперсии и другие характеристики случайных величин и их распределений.

Рассмотрим подробно метод исследования алгоритмов идентификации в целях количественной оценки показателей их качества.

Для принятия решения о выборе наиболее предпочтительной альтернативы информацию могут дать лишь экспериментальные исследования модуля на тестовых задачах с известными характеристиками [5].

Рассмотрим постановку и решение задачи машинной имитации при исследовании программных реализаций алгоритмов идентификации. Пусть $E_l = \{e_s\}_{s=1}^S$ – совокупность показателей качества l -го модуля, определенных ранее. На каждый показатель e_s влияет один или несколько параметров $h_j \in H$. Для получения экспериментальной характеристики качества l -го модуля необходимо построить регрессионные зависимости $e_s = \psi_s(H) \forall s$, где $\psi(\cdot)$ – некоторая детерминированная функция.

Чтобы получить зависимости e_s , проводится серия машинных экспериментов по $h_j \in H$, где независимыми переменными являются h_j , зависимыми – каждый из e_s . Эта задача распадается на подзадачи: выбор плана эксперимента, синтез имитационных моделей с параметрами h_j .

Учитывая, что основным требованием к синтезируемым моделям является минимум остаточной дисперсии и дисперсий оценок коэффициентов, в силу малой априорной информации для проведения машинных экспериментов следует обратиться к методам оптимального планирования эксперимента. Для исследования эффективности алгоритмов идентификации более предпочтительными являются D-оптимальные планы [6].

Синтез имитационной модели гипотетического объекта с параметрами H базируется на методах статистического имитационного моделирования [2], которые позволяют генерировать выборки псевдослучайных чисел с заданными характеристиками. Для получения выборки с заданными характеристиками необходимо решить следующие задачи.

Моделирование одномерной случайной выборки. Вследствие стохастической природы исследуемых процессов используются генераторы псевдослучайных чисел (ГПСЧ). Подробный анализ существующих ГПСЧ приведен в [1]. В результате сравнительного анализа оказалось, что лучшими характеристиками обладает ГПСЧ [7], использованный в разработанной программной имитационной модели.

Имитация выборки с заданной автокорреляцией. Метод имитации выборки с заданной автокорреляционной функцией подробно рассмотрен в [8].

Имитация ошибок в переменных. Ошибка в зависимой и независимой переменных может носить импульсный и флуктуационный характер. Первый вид ошибки характеризуется как выброс, второй – как постоянно действующая случайная ошибка (ошибка измерений).

Имитация выбросов производится по схеме $X_j = X_j + d\tilde{\varepsilon}_{\tilde{k}}$, где $X_j = \{x_{ij}\}_{i=1}^h$ – вектор j -й переменной; d – величина выброса; $\tilde{\varepsilon}_{\tilde{k}}$ – нулевой вектор с единицей на месте \tilde{k} – й компоненты.

Имитация ошибок измерений в зависимой и независимых переменных осуществляется в соответствии с выражениями $y^{(0)} = y^{(0)} + e$; $x_j^{(0)} = x_j^{(0)} + x$, где $e = \{e_t\}_{t=1}^h$, $x = \{x_t\}_{t=1}^h$ – векторы ошибок. Векторы имитируются как равномерно или нормально распределенные псевдослучайные числа. Однако реальные распределения редк удается описать идеальной зависимостью.

Наибольший интерес представляют η – засоренные распределения вида $f(\xi) = (1 - \eta)f_1 + \eta f_2(\xi)$, где $f_1(\cdot)$ и $f_2(\cdot)$ – известные плотности распределений (базовая и засоряющая). Реальные распределения близки к нормальным, поэтому в качестве базового $f_1(\cdot)$ принимается нормальное распределение с нулевым средним и известной дисперсией. В качестве $f_2(\cdot)$ принимается распределение с большей, чем у $f_1(\cdot)$, дисперсией. Это позволяет удлинить “хвосты” распределения. Если параметр расположения распределения $f_2(\cdot)$ принять равным нулю, то имеет место случай симметричного засорения. Если центр распределения смещен относительно нуля, то имитируется асимметричное засорение. Дисперсии ошибки выбираются из учета требуемого соотношения “сигнал/шум” $q_x = \sigma_\xi^2 / \sigma_x^2$, $q_y = \sigma_e^2 / \sigma_y^2$.

Моделирование многомерной выборки предполагает имитацию требуемой корреляционной структуры исходных данных или, иначе, моделирование m одномерных выборок с заданными взаимно корреляционными свойствами, которые определяют степень обусловленности исходных данных. Получение матрицы X с заданной

обусловленностью сводится к моделированию X с заданной корреляционной матрицей \tilde{R} , а последнее – к моделированию \tilde{R} с заданными собственными числами $\Lambda = \{\lambda_j\}$, $j = \overline{1, m}$. В общем случае имитация X осуществляется по схеме

$$Z = U\Lambda^{1/2} \in N(O, \Lambda) \Rightarrow X = ZC' \in N(O, \tilde{R}),$$

где $U \sim N(O; I)$; $\tilde{R} = C\Lambda C'$; $N(\cdot)$ – нормальный закон распределения; C – матрица собственных векторов.

В [9] проведен анализ существующих методов моделирования \tilde{R} с заданными λ_j . В результате исследования авторами разработан алгоритм генерации \tilde{R} , отличающийся от существующих использованием модифицированного алгоритма ортогонализации Грама – Шмидта вместо обычно используемого классического метода [10]. Вычислительная схема моделирования X с заданной R с использованием модифицированного алгоритма Грама – Шмидта имеет вид:

- первый этап: генерация $U \sim N(O; I)$;
- второй этап: моделирование $\Lambda = \text{diag}\{\lambda_j\}$ с учетом требуемых соотношений;
- третий этап: моделирование C – матрицы собственных векторов с использованием модифицированного алгоритма Грама – Шмидта;
- четвертый этап: вычисление $\tilde{R} = C\Lambda C'$;
- пятый этап: получение $X: X = U$.

Имитация структуры и параметров объекта. Матрица X , полученная на предыдущих этапах моделирования, используется для получения вектора зависимой (выходной) переменной y . Структура модели, согласно принятому классу, выбирается линейной по параметрам и произвольной по входным переменным $y = \hat{F}(x)B + \varepsilon$.

Вид детерминированного преобразования \hat{F} задается в одном из следующих классов: тождественных $\varphi(x_j) = x_j$; полиномиальных $\varphi(x_j) = p_k(x_j)$, где k – порядок полинома; обратных $\varphi(j) = x_j$; экспоненциальных $\varphi(x_j) = e^{x_j}$; индексации по времени $\varphi(x_j) = t$; задержек на j тактов $\varphi(x_j) = x_j(t - i)$; суперпозиций $\varphi(x_j) = \varphi_1[\varphi_2(x_j)]$;

произведений $\varphi(x_j) = \varphi_1(x_j) \cdot \varphi_2(x_j)$; индикаторных $\varphi(x_j) = 0$, если $x_j \in X$, и $\varphi(x_j) = 1$, если $x_j \notin X$.

Задание вектора V определяется, как правило, физическими соображениями. Однако для сокращения числа прогонов необходимо учесть следующий факт. Результаты исследований показали, что качество восстановленной модели в значительной мере зависит от степени корреляционной связи между вектором неизвестных параметров и собственными векторами корреляционной матрицы. В связи с этим при моделировании рассматривается лучший случай, когда $\alpha = C^{\max}$, где C^{\max} – собственный вектор, соответствующий максимальному собственному значению $\lambda_{\max}(X'X)$, и худший случай, когда $\alpha = C^{\min}$, а C^{\min} соответствует $\lambda_{\min}(X'X)$.

Рассмотрим укрупнённо процесс автоматизированного проектирования прикладного программного обеспечения идентификации.

В соответствии с перечнем задач идентификации проектировщику необходимо по каждой задаче выбрать метод решения и построить соответствующую цепочку базовых алгоритмов и пакет программ. С объекта в систему проектирования в реальном масштабе времени или по запросу поступает исходная информация, которая обрабатывается системой обработки данных (СОД) для определения характеристик объекта. На основании полученных от СОД рекомендаций по результатам ретроспективного анализа данных проектировщик выбирает класс алгоритмов решения задачи (или задач) идентификации. В распоряжении проектировщика имеется набор модулей решения отдельных задач идентификации. Используя генератор тестовых задач, проектировщик задает программу исследования модулей для конкретной ситуации, которая определяется характеристиками объекта и условиями проведения эксперимента. Система имитации формирует функции качества модулей, которые позволяют проектировщику с учетом целей исследования задать критерий выбора модуля, удовлетворяющего функциональным и базовым требованиям. В связи с тем, что задача формулируется как задача математического программирования, процедура F реализуется известными методами поиска экстремума. Если получены экспериментальные характеристики качества модулей E , то наиболее эффективным признается тот, который является решением задачи $P^* = \arg \min_{\gamma} \min_j e [P_j(\gamma), H]$, где γ – параметры алгоритма и программы, например, граничные значения,

масштабные коэффициенты, степени приближения и другие, которые заранее не определены, но требуют априорного задания.

Рассмотренная схема исследования может быть достаточно легко распространена на алгоритмы решения различных задач. Несмотря на то, что реализация автоматизированного проектирования алгоритмического обеспечения идентификации требует значительных вычислительных затрат, нередко этот подход является единственным, позволяющим обоснованно подойти к выбору алгоритма. На практике нередки случаи, когда для идентификации моделей объекта выбирается первый встретившийся алгоритм. Исследования авторов показывают, что для одних и тех же экспериментальных данных различные алгоритмы приводят к различным моделям, отличающимся не только порядком коэффициентов, но и их знаками. Этот и другие аналогичные факты требуют проведения серьезного предварительного исследования используемых алгоритмов на различных тестовых задачах.

Предлагаемый подход реализован авторами при автоматизации реальных производственных процессов.

Литература: 1. *Васильцова С.В., Шамша Б.В.* Использование имитационной модели при синтезе системы идентификации. 1982. 23 с. Деп. ВИНТИ, №3806. 2. *Клейнен Дж.* Статистические методы в имитационном моделировании. М.: Статистика, 1978. 221 с. 3. *Демиденко Е.З.* Линейная и нелинейная регрессия. М.: Финансы и статистика, 1981. 302 с. 4. *Алберт А.* Регрессия, псевдорегрессия и рекуррентное оценивание. М.: Наука, 1977. 224 с. 5. *Шамша Б.В., Васильцова С.В.* Оценка качества алгоритмического и программного обеспечения АСУТП. 1982. 23 с. Деп. ВИНТИ, № 3807. 6. *Круг Г.К., Сосулин Ю.А., Фатеев В.А.* Планирование эксперимента в задачах идентификации и экстраполяции. М.: Наука, 1977. 208 с. 7. *Greenwood I.A.* Portable generators for the random variables usual in reability simulatio //Recent Developments in Statistics, Amsterdam e.a., 1977. P. 677 -688. 8. *Поляк Ю.Г.* Вероятностное моделирование на ЭВМ. М.: Сов. радио, 1971. 400 с. 9. *Глазунова Н.А.* Исследование методов моделирования случайных процессов с заданными вероятностными свойствами. Автореф. канд. дисс. М., 1978. 21 с. 10. *Стренг Г.* Линейная алгебра и ее применения. М.: Мир, 1980. 454 с.

Поступила в редколлегию 28.05.98