

Эволюционно-статистический подход к самоорганизации прогностических моделей управления ТП

А.А. Мусаев (СПИИ РАН)

Предложен подход к решению задачи прогнозирования параметров состояния и выходных параметров ТП, основанный на сочетании алгоритмов эволюционного моделирования с традиционными вычислительными схемами многомерного статистического анализа. Разработан алгоритм и показаны особенности функционирования программного модуля, осуществляющего эволюционно-статистическое прогнозирование.

Введение. Особенности эволюционного прогноза

В основе технологии эволюционного моделирования (ЭМ), предложенной в [1], лежит идея замены процесса прямой оптимизации управления сложной динамической системой процессом самоорганизации управляющих решений в соответствии с принципами дарвиновской эволюционной теории – изменчивостью, селекцией и отбором. При этом процесс изменения структуры каждой модели осуществляется случайным образом. Новое поколение моделей проверяется по экзогенному критерию выживаемости, определяемому качеством решения функциональных задач. В соответствии с этим критерием осуществляется отбор группы наиболее эффективных моделей-родителей следующего поколения. Далее, путем случайных изменений, вносимых в структуру или параметры моделей-родителей, формируется множество моделей-потомков, образующих новое поколение. Таким образом, образуется последовательность поколений модифицированных моделей с неубывающим показателем качества решения исходной функциональной задачи.

По существу речь идет о стохастической самоорганизации математической модели. При этом задается лишь направление эволюции, процесс модификации моделей остается неконтролируемым и формируется самим компьютером в соответствии с принятой технологией генерации случайных последовательностей. По мнению авторов метода эволюционного моделирования [1], такой подход позволяет по-новому подойти к задаче создания искусственного интеллекта, когда роль внешнего дополнения в теореме Геделя о неполноте будет выполнять компьютер, обладающий возможностью свободы выбора.

В отечественной литературе советского периода данный подход был рассмотрен в [2, 3 и др.].

Важно заметить, что в процессе отбора выживают и формируют новое поколение не только наилучшая (среди предыдущего поколения) модель, а группа моделей, наиболее полным образом удовлетворяющие критерию селекции. Такой подход, названный принципом незаконченных решений, позволяет реализовать постулат системного анализа, в соответствии с которым наилучшее решение формируется из последовательности шагов или элементов, не являющихся оптимальными.

Очевидно, что решение задачи изменчивости можно решать различными методами. Предложен один из возможных подходов к этому вопросу, основанный на математической технологии генетических

алгоритмов (ГА). При этом соответствующие изменения в моделях-родителях осуществляются по аналогии с селекционными методами, принятыми в биологии и сельском хозяйстве [4]. По существу ГА представляют собой регуляризованные алгоритмы случайного поиска.

В рамках поставленной задачи, эволюционный подход с элементами ГА используется для модификации структуры модели, т.е. выбора набора предикатов и образованных ими локальных функционалов (последние – для нелинейных моделей). Параметрическая подгонка осуществляется классическими методами многомерного статистического анализа [5, 6, 7].

Математические технологии эволюционно-статистического моделирования

Имеется некоторая априорная нелинейная гиперповерхность отклика m -го порядка

$$y = y(x_1, x_2, \dots, x_m), \quad (1)$$

описывающая связь выходных (прогнозируемых) характеристик ТП с параметрами управления и состояния технологической установки. Совокупность обучающих данных представлена набором из N векторных наблюдений, отражающих значения управлений, состояний и результатов управлений (выходных характеристик) $Z_{\langle N; m \rangle}$. Для снижения значимости нестационарности контролируемых процессов организуется обучающая скользящая выборка наблюдений, на основе которой осуществляется подгонка (обычно методом наименьших квадратов (МНК)) гиперповерхности (1) полиномиальной моделью не выше 2-го порядка [4, 7]. При этом

$$\tilde{y} = f(\tilde{x}_1, \dots, \tilde{x}_m, \tilde{x}_1^2, \dots, \tilde{x}_m^2, \tilde{x}_1\tilde{x}_2, \dots, \tilde{x}_1\tilde{x}_m, \dots, \tilde{x}_{m-1}\tilde{x}_m). \quad (2)$$

Иными словами, формируется наилучшая аппроксимация, для которой

$$Q_N = \sum_{i=1}^N (y_i - \tilde{y}_i)^2 = \min. \quad (3)$$

Состав предикатов модели (2) может быть любым. Таким образом, речь идет о выборе наилучшей структуры модели, используемой для прогноза на каждом шаге оценивания. Очевидно, что такой выбор можно сделать полным перебором переменных, входящих в (1). В случае использования модели второго порядка, в правую часть (1) входят m исходных предикатов, m квадратов этих же предикатов и $(m \times m - m)/2$ различных парных произведений независимых переменных

У модели есть право занять свое место в живой жизни, и есть право ожить.

Журнал "Автоматизация в промышленности"

ных) – всего N_{ALL} переменных. Тогда полное число возможных моделей составит:

$$N_{models} = C_{N_{all}}^1 C_{N_{all}}^2 \dots C_{N_{all}}^{N_{all}-1}$$

Нетрудно убедиться, что при наличии большого числа параметров ТП (обычно $m > 30$), получаемых в процессе мониторинга и используемых в качестве предикатов прогностической модели, объем проводимых вычислений окажется недопустимо большим даже для современных компьютеров.

Эволюционное моделирование

Альтернативу к полному перебору составляет регуляризованный случайный поиск. В качестве методики регуляризации (в рамках настоящей статьи) выступает технология *эволюционного моделирования*, имитирующая процесс последовательной биологической оптимизации, описываемой дарвиновской теорией эволюции видов. В основе ЭМ положены принципы рандомизации, наследственности, селекции, отбора, подробно рассмотренные, например, в [1].

В соответствии с указанной технологией поиск оптимальной структуры модели управляемого ТП (2) состоит в следующем. На основе ранее полученных результатов мониторинга осуществляется классическая параметрическая идентификация (обычно методом линейной регрессии [5, 6]) некоторой базовой модели, структура которой выбрана случайно или из каких-то экзогенных соображений, связанных со знанием предметной области.

Далее, исходя из структуры базовой модели-предка, случайным образом формируется M_M моделей-потомков. Выбор изменений осуществляется с помощью генератора случайных чисел с равномерным распределением из интервала [0, 1]. Вероятности тех или иных изменений задаются в виде параметров программы.

Для полиномиальной модели 1...2 порядков на каждом шаге смены поколений возможны следующие изменения структуры модели:

- добавление с вероятностью $\alpha 1$ одного предиката первого порядка, неиспользованного в модели-предке на предыдущей итерации, или вычитание с вероятностью $\alpha 2$ из модели-предка одного предиката первого порядка;

- добавление с вероятностью $\beta 1$ одного предиката второго порядка, неиспользованного в модели-предке на предыдущей итерации, или вычитание с вероятностью $\beta 2$ одного предиката второго порядка из модели-предка;

- добавление с вероятностью $\gamma 1$ одного слагаемого, образованного перемножением двух предикатов (и не содержащегося в модели-предке), или вычитание с вероятностью $\gamma 2$ из модели-предка одного слагаемого, образованного перемножением двух предикатов;

- мутации или одновременное изменение (с вероятностью δ) группы параметров (добавление или вычитание).

Модели-потомки в совокупности с моделями-предками образуют поколение моделей. Каждая из моделей поколения проходит процесс параметрической идентификации на основе метода множественной регрессии или канонических корреляций. При этом оценка параметров моделей обычно осуществляется по выборке обучающих данных (т.е. данных с заранее определенными входами/выходами), сформированной на скользящем окне наблюдения из общего массива ретроспективных данных. После этого каждая из моделей поколения проходит верификацию (тестируется) на множестве тестовых данных (также имеющих заранее определенные входы/выходы). По результатам тестирования осуществляется их ранжирование в соответствии со значениями выбранных показателей качества. Обычно в качестве таких показателей выбираются меры точности или достоверности прогноза.

Далее, в соответствии с выбранным критерием отбора осуществляется селекция наилучшего подмножества моделей, образующих совокупность моделей-предков нового поколения. Формирование очередного поколения означает начало следующей итерации процесса эволюции, включающей генерацию структурно измененных моделей-потомков, оценку качества всех моделей поколения, их ранжирование и селекцию.

Число итераций эволюции (число поколений) может задаваться заранее или ограничиваться некоторым максимальным значением при обеспечении возможности остановки итерационного цикла (эволюции) после достижения требуемого уровня достоверности прогноза.

Генетические алгоритмы

В процессе реализации механизма структурной изменчивости моделей оказалось удобно использовать технологию ГА. Общая структурная схема ГА представлена на рис. 1. В соответствии с методологией ГА, каждой из моделей сопоставляется двоичное число в виде последовательности единиц и нулей (хромосома). Общее количество двоичных разрядов этого числа соответствует общему числу возможных предикатов, включающему число параметров управления и состояния, а также, в случае использования нелинейных моделей, соответствующих функционалов от исходных параметров.

Наличие единицы означает введение в модель предиката, номеру которого в хромосоме соответствует его номер в общем фиксированном списке всех возможных предикатов. Таким образом, инициализация популяции означает формирование множества хромосом со случайными комбинациями нулей и единиц в их двоичных кодах. При этом каждой хромосоме однозначно соответствует модель с заданной структурой предикатов.

В отличие от чисто стохастической модификации моделей, характерной для "чистой" эволюции, технология ГА предполагает формирование новых моделей путем комбинации хромосом родительских пар, выбор которых производится в соответствии с некоторыми регуляризирующими правилами, характерными для методов биологической селекции. В частности, функция репродукции популяции хромосом может осуществляться в соответствии с правилами выбора родительских пар. Обычно родителей выбирают исходя из результатов тестовой оценки их эффективности: вероятность того, что данная хромосома станет родителем, пропорциональна полученной оценке. Репродукция осуществляется индивидуально для одного родителя путем изменения его хромосомы либо для двух родителей путем кроссовера генов. В результате описанных операций на каждом этапе эволюции получаются популяции с новыми индивидами.

В базовом варианте эволюционно-статистического прогнозатора (ES-прогнозатора) ГА используются в основном в интересах модификации структуры модели в процессе эволюции. При этом каждая модель определяется соответствующей хромосомой. Локализация соответствующего изменения разыгрывается с помощью датчика равномерно распределенных случайных чисел, масштабированного длиной хромосомы (максимальным числом возможных предикатов).

Многомерная регрессия

Параметрическое оценивание для каждой версии структуры модели ТП осуществлялось на основе вычислительной схемы многомерной регрессии $Y = AX + V$, где Y – вектор наблюдений параметров выходного процесса, X – матрица предикатов, получаемая в процессе мониторинга состояния ТП, V – вектор аддитивной погрешности наблюдений. При этом наилучшей параметрической оценкой в классе несмещенных оценок (при выполнении известного набора ограничений и условий [4, 5, 7]) является оценка по МНК $a = (X'X)^{-1} X'Y$. Данный подход хорошо известен, и на нем останавливаться не будем.

В качестве развития данного подхода использовалась вычислительная схема канонических корреляций, позволяющая учесть корреляционные связи между параметрами выходных потоков ТП [7].

Другое обобщение данного подхода связано с вычислительными схемами робастного оценивания, позволяющие повысить устойчивость параметрического оценивания к отклонениям статистических характеристик наблюдений от априорной гауссовской модели [6]. Как правило, техника робастной идентификации приводит к нелинейным моделям и соответствующим итерационным вычислительным схемам. Определенной альтернативой к робастным М-оценкам может служить сочетание традиционной вычислительной схемы МНК с предварительной обработкой данных, включающей обнаружение и обработку аномальных наблюдений и последовательную

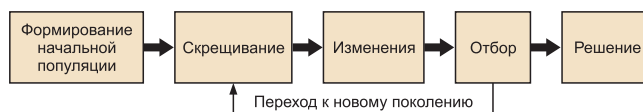


Рис. 1. Структура ГА

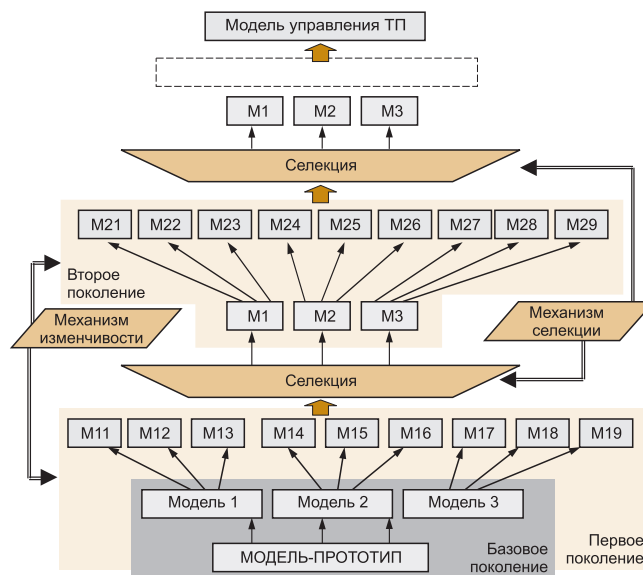


Рис. 2. Общая функциональная структура алгоритма эволюционного формирования модели управления ТП

сглаживающую фильтрацию (например, с помощью калмановского или экспоненциального фильтров).

Таким образом, разработанная схема эволюционно-статистического прогнозатора основывается на симбиозе трех различных математических методологий: эволюционном моделировании, ГА и многомерном статистическом оценивании.

Функциональная структура эволюции модели прогноза

Общая функциональная структура алгоритма эволюционного формирования модели прогноза приведена на рис.2. Последовательность эволюции модели прогноза представлена схемой процесса, развивающегося снизу вверх. Первоначально имеется некоторая базовая модель (прототип), структура которой выбирается случайно или исходя из имеющегося априорного опыта по управлению ТП. С помощью генератора изменчивости модель-прототип видоизменяется, порождая три (в общем случае – M) новых модели.

В свою очередь каждая из трех новых моделей является родителем (прототипом), случайные изменения которого порождают еще по три новых модели. Вся совокупность исходных моделей образует первое поколение моделей, подлежащих оптимальной параметрической идентификации. Оценка параметров моделей производится в соответствии с традиционными алгоритмами множественной линейной регрессии или канонических корреляций.

Далее каждая из моделей первого поколения проходит процедуру тестирования путем реализации построенного на их основе прогноза на множестве тес-

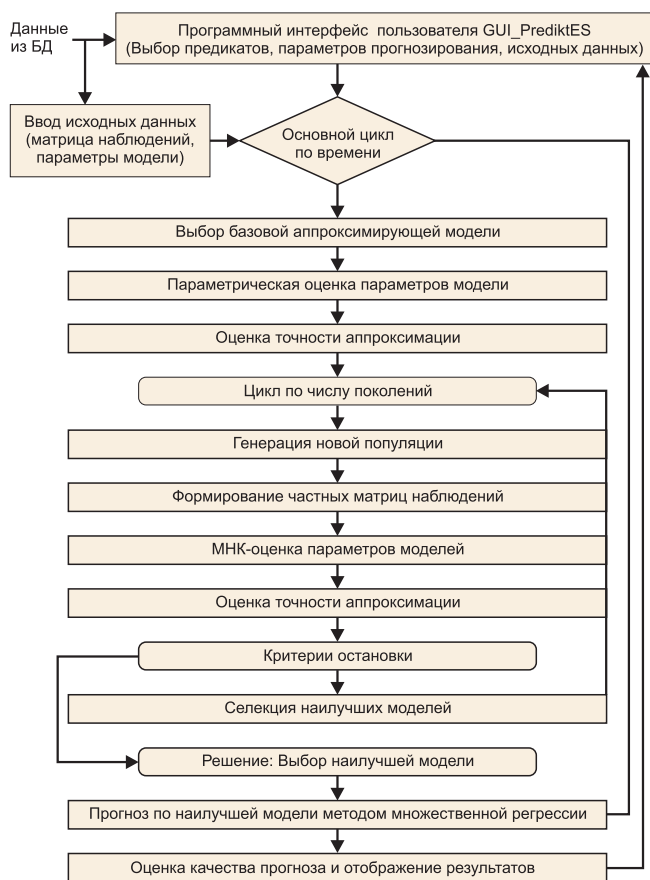


Рис. 3. Блок-схема программы прогнозирования на основе эволюционного моделирования

товых данных. Сравнение точностей получаемых прогнозов позволяет выявить N_s наилучших моделей (число "выживших" после отбора моделей определяется либо заранее фиксированным числом N_{s0} , либо порогом точности прогноза).

"Выжившая" группа моделей вновь подвергается случайным модификациям в соответствии с технологиями ГА, в результате чего образуется множество модифицированных моделей-потомков, которые вместе с исходной группой образуют второе поколение. Далее вся совокупность моделей вновь подвергается параметрической идентификации, оценке качества и селекции и т.д.

Описанный процесс "размножения" и "селекции" моделей повторяется в течение заданного числа поколений. Остановка цикла генерации поколений может быть проведена и раньше, например, на основе критерия превышения порога точности или критерия сходимости результатов прогноза на тестовой совокупности данных. В конечном счете, итерационная процедура позволяет выявить наилучшую модель, на основе которой и будет проводиться текущий прогноз.

Программная реализация модуля эволюционно-статистического прогноза

Структура программной реализации модуля эволюционно-статистического прогноза, выполненной в рамках разработки представлена на рис. 3.

Формирование задания и выбор загружаемого массива данных осуществляется средствами программного интерфейса пользователя GUI_PrognosisES. Затем управление передается основной исполнительской процедуре модуля PrognMainAT6.m, обеспечивающей прогноз параметров выходных потоков. Далее организуется скользящее окно наблюдений, используемое для формирования динамической оценки корреляционной матрицы предикатов.

Функционирование программы связано с использованием двух вложенных циклов.

Внешний цикл осуществляет динамический отсчет последовательности поступления новых векторных наблюдений. На каждом шаге данного цикла осуществляется скольжение окна наблюдения заданного размера. Последнее (текущее) поступление данных представлено результатом мониторинга состояния ТП и управлений, на основе которых необходимо сформировать прогноз значений параметров выходных потоков. Собственно прогноз формируется методами многомерной статистики, опирающимися, при заданной модели связи входных/выходных параметров, на традиционную вычислительную схему МНК. При этом остается открытым вопрос о структуре указанной модели.

Последний вопрос решается на основе внутреннего цикла по числу эволюционирующих поколений математических моделей. Процесс формирования наилучшей модели осуществляется в соответствии с методологией эволюционного моделирования, структурная схема которого приведена выше на рис. 2. При этом, наряду с моделированием эволюции, в ходе генерации новых поколений моделей-потомков используется выше описанная технология ГА.

Выбранная в процессе эволюции наилучшая (для заданного числа поколений) модель служит основой для формирования текущего прогноза. На следующем шаге мониторинга состояния ТП данная вычислительная схема полностью повторяется.

Накопленные за определенный период времени результаты прогноза в совокупности с последующими данными мониторинга выходных процессов позволяют оценить качество формируемого прогноза. Как правило, для его оценки используются такие характеристики, как среднее квадратическое отклонение и среднее значение абсолютных значений относительных погрешностей прогноза.

Применение эволюционно-статистического прогнозатора

Результаты мониторинга состояния ТП (включая выходные данные on-line анализаторов и результаты лабораторных анализов) поступают в БД в виде временной последовательности многомерных данных. Накопленные данные образуют в БД двумерную таблицу размера $n \times m$, где n – число векторных наблюдений (строк), m – число контролируемых параметров (столбцов). Выбор загружаемого массива и параметров

прогностического модуля осуществляется средствами программного интерфейса пользователя.

Основная программа PredictorES выполнена в виде скрипт-файла, обращение к которому, как и вся работа с модулем, осуществляется через программный интерфейс GUI_Prognosator EG/AT6 (рис. 4)

Средствами панелей Prognosis parameters выбираются выходные параметры ТП, подлежащие прогнозированию. При этом число прогнозируемых параметров должно быть меньше или равно начальному числу предикатов модели. Настройка процесса ввода данных и скользящего окна наблюдения производится средствами панели Options. Здесь осуществляется выбор имени массива данных; число последних наблюдений, используемых при решении задачи эволюционного прогноза; общее число параметров управления состоянием и выхода; начальное число предикторов, используемых в модели; размер скользящего окна наблюдения; среднюю задержку реакции свойств выходных потоков на изменения управляющих воздействий.

Свойства эволюционного процесса задаются тремя параметрами:

- коэффициентом репродукции, определяющим число моделей-потомков для каждой из моделей-предков;
- числом поколений процесса эволюции;
- числом наилучших моделей, успешно прошедших процедуру отбора (параметром селекции).

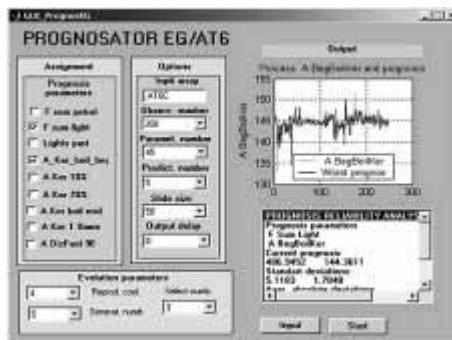


Рис. 4

Выводы результатов, включающих графические иллюстрации, значения прогнозируемых параметров и результаты статистического анализа качества прогноза, осуществляются в графические окна интерфейса.

Представленный программно-алгоритмический модуль эволюционно-статистического прогноза является практической разработкой и был апробирован на

данных, полученных в результате мониторинга процессов функционирования и управления технологической установки атмосферной трубчатки АТ6 (первичной перегонки нефти).

Список литературы

1. Фогель Л., Оуэнс А., Уолли М. Искусственный интеллект и эволюционное моделирование // Пер. с англ. под ред. А.Г. Ивахненко. М.: Мир. 1969.
2. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. Киев: Техника. 1975.
3. Ивахненко А.Г., Зайченко Ю.П., Димитров В.Д. Принятие решений на основе самоорганизации. М.: Сов. Радио. 1975.
4. Вороновский Г.А., Махотило К.В., Петрашев С.Н., Сергеев С.А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. Харьков: Основа. 1997. (<http://www.neuropower.de/rus/>).
5. Демиденко Е.З. Линейная и нелинейная регрессии. М.: Финансы и статистика. 1981.
6. Линник Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений М.: Физматгиз. 1958.
7. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей / Под ред. С.А. Айвазяна М.: Финансы и статистика. 1985.

Мусаев Александр Азерович – д-р техн. наук, проф.,

ведущий научный сотрудник Санкт-Петербургского института информатики и автоматизации РАН.

Контактный телефон (812) 350 58-85. E-mail: amusaev@szma.com

Omron совершенствует функциональность своей базовой серии регуляторов температуры

Компания Omron внесла ряд важных изменений в серию своих базовых регуляторов температуры, в то же время сохраняя неизменными простоту настройки и использования.

Новая серия регуляторов температуры E5CSV пришла на смену популярной серии E5CS, получившей всемирное признание как идеальный выбор для простого и эффективного регулирования температуры при низких затратах. Серия E5CSV унаследовала простоту настройку путем использования DIP и поворотных переключателей, большой светодиодный 7-сегментный индикатор, а также возможность выбора между дискретным (вкл/выкл) и пропорциональным интегрально-дифференциальным (PID) регулированием с самонастройкой. Эти регуляторы оснащены индикаторами выходного сигнала и аварийного состояния, а также индикатором отклонения от заданного значения.

Среди новых функций, появившихся в серии E5CSV – функция автонастройки, позволяющая делать процесс настройки и оптимизации еще более простым, чем это было ранее. Новые регуляторы снабжены функцией защиты заданного значения с целью уменьшения риска отказа прибора. Возможность выбора между типами входа (термопара/термометр сопротивления) обеспечивает простоту в работе с широким спектром датчиков температуры. Наличие нового 3,5-разрядного ЖК-дисплея означает, что E5CSV способен отобразить диапазон температур до 1999 °C.

Новая серия отвечает требованиям RoHS по применению опасных материалов при производстве, а также соответствует жестким нормам стандарта IP66. Глубина регуляторов также была уменьшена до всего лишь 78 мм.

[Http://www.omron-industrial.ru](http://www.omron-industrial.ru)

Уважаемые читатели! Продолжается подписка на журнал "Автоматизация в промышленности" на 2006 г. Оформить подписку Вы можете:

В любом почтовом отделении

Индексы в каталоге "Роспечать" – 81874,

в Объединенном каталоге "Пресса России" – 39206

В редакции и

Сети Интернет по адресу: www.avtprom.ru