УДК 51.7, 519.12

С. С. Замай, В. А. Охонин, А. Л. Щемель

Институт вычислительного моделирования СО РАН
Академгородок, 50, стр. 44, Красноярск, 660036, Россия
E-mail: zamay@ats.krsnkr.ru; vokhonin@yorku.ca;
andong.china@gmail.com

ПРОБЛЕМА НАДЕЖНОСТИ МНОГОПАРАМЕТРИЧЕСКИХ РЕГРЕССИОННО-ИМИТАЦИОННЫХ МОДЕЛЕЙ СЛОЖНЫХ СИСТЕМ *

Рассматривается возможность синтеза имитационного и адаптивно-регрессионного подхода, на основе использовании в качестве адаптивных величин отклонений параметров модели от априорных, в пределах доверительных интервалов этих априорных значений. Показано, что в рамках данного подхода получает естественное решение проблема плохой обусловленности многопараметрических моделей.

Ключевые слова: имитационные модели, регрессионные модели, многопараметрические модели, сложные системы.

Введение

К настоящему времени назрела необходимость, и существует возможность, перестройки систем управления:

– в принципе решены проблемы информационных коммуникаций, вычислительные мощности стали доступны;

– разрозненные, изолированные фрагменты экономическо-социальных структур, интегрируются в единую метаструктуру;

– заканчиваются попытки свести экономическо-социальные взаимоотношения к одной жесткой, фиксированной схеме, апеллирующие к гомогенному представлению о системе;

– резко повышается спрос на точные схемы управления, обеспеченные математическими моделями, реализованными с использованием современных информационных технологий.

Для обеспечения потребности развивающейся структуры управления необходимо обеспечить возможность гибкого трансферта данных, соответствующих различным уровням описания систем. Традиционно для этих целей развивался имитационный подход, обеспечивающий трансферт данных с более детального уровня описания на менее детальный. При этом отсутствовала возможность обратного трансферта. Вместе с тем существует регрессионный подход, обеспечивающий моделирование макросистемы без обращения к детальному описанию. Возможна и задача обратного трансферта данных: с уровня макросистемы на уровень детального описания. Здесь мы рассмотрим возможность решения данной задачи, вместе с задачей улучшения точности прогноза, на основе сочетания имитационнного и регрессионного подхода.

При попытке синтеза имитационного моделирования [1] и регрессионного подхода, основанного на «подгонке» модели под наблюдаемые, одна из сложностей состоит в том, что при построении имитационной модели количество параметров модели зачастую определяется не требованиями, вытекающими из матстатистики, а структурой описываемого объекта. В связи с этим проблема недоопределенности коэффициентов, в рамках наличных данных о поведении системы, не допускает решения в виде выбрасывания «статистически неопределенных» коэффициентов. Идеологически родственное «замораживание статистически недостоверных поправок» к параметрам выглядит логически малообоснованным. Предлагаемый подход состоит в минимизации, вместо расстояния от предикторных данных до фактических, этого

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 09-07-98002-р_сибирь_а).

ISSN 1818-7900. Вестник НГУ. Серия: Информационные технологии. 2010. Том 8, выпуск 2

расстояния, дополненного взвешенной среднеквадратичной оценкой, расстояния между априорными и фактическими значениями параметров модели, взятой с некоторым положительным коэффициентом.

Данный подход применим также к известной проблеме снижения качества прогноза [2], при увеличении точности аппроксимации, за счет увеличения числа параметров аппроксимации, при использовании адаптивных моделей [3; 4].

Проиллюстрируем данный подход на ситуации, когда имеется множество параметров модели, и совокупность n функций оценки h^i. Пусть шкала измерения параметров выбрана таким образом, что все доверительные интервалы априорного оценивания параметров модели равны. Для верификации надежности предсказаний используем аппроксимацию метода бутстрепа. Обозначим x_s – компоненты вектора поправок к априорным величинам параметров, H – совокупность всех функций оценки:

$$H = \sum_{i=1}^{n} h^i, \;\; H' = H + \varepsilon \times \bar{x}^2 / 2, \tag{1}$$

где введена модифицированная вышеуказанным образом функция оценки H'. Для приближенной, суммарной, бутстреп-оценки можно получить приближенное соотношение, рассматривая исключение одной из функций оценки как малое возмущение. На этом пути можно теоретически оценить оптимальное значение веса дополнительной оценки ε, минимизирующее ожидаемую бутстреп-оценку. Также можно обнаружить аналогию с методами регуляризации некорректных задач.

Математический формализм

Пусть при аппроксимации всей совокупности n задач наилучшее значение поправок к параметрам модели, минимизирующее оценку H, равно x'_s. Введем сдвинутые на эту величину переменные y_s:

$$y_s = x_s - x'_s. \tag{2}$$

В новых переменных наилучшее аппроксимирующее значение равно нулю. Однако практический интерес представляет не наилучшая аппроксимация известных данных, а наилучший прогноз неизвестных данных. Такой прогноз может достигаться при иных значениях параметров. Рассмотрим для поиска именно наилучших прогнозных параметров следующую процедуру:

– поочередно из обучающей выборки изымается одна из всей совокупности задач;

– по оставшейся выборке, на основе функции H', находятся минимизирующие эту функцию переменные, по ним рассчитывается оценка для изъятой задачи, сумма оценок по всем изъятым задачам, H_{test}, является оценкой прогноза.

Результат такой процедуры исследуем методами теории возмущений, полагая, что изъятие, а равно и модификация функции оценки, согласно (2), сдвигают значения новых параметров от нулевых на малую величину, и функции оценки можно раскладывать в ряд по этому сдвигу параметров. Это приводит нас к совокупности соотношений:

$$\begin{gathered}
\bar{h}^i = \left(\partial h^i / \partial y_s \big|_{\bar{y}=0} \right), \quad \hat{h}^i = \left\| \partial^2 h^i / \partial y_s \partial y_r \big|_{\bar{y}=0} \right\|, \quad \hat{H} = \frac{n-1}{n} \sum_{i=1}^{n} \hat{h}^i; \\
\hat{p}^i = \hat{h}^i - \frac{1}{n} \sum_{i=1}^{n} \hat{h}^i = \hat{h}^i - \frac{1}{n-1} \hat{H}, \quad \sum_{i=1}^{n} \bar{h}^i = 0, \quad \sum_{i=1}^{n} \hat{p}^i = 0; \\
\bar{y}^i = \arg\min \left(\frac{1}{2} \bar{y}^i (\hat{H} + \varepsilon - \hat{p}^i) \bar{y}^i + (\varepsilon \bar{x}' - \bar{h}^i) \bar{y}^i \right); \\
(\hat{H} + \varepsilon - \hat{p}^i) \bar{y}^i = \bar{h}^i - \varepsilon \bar{x}', \quad h^i \approx \frac{1}{2} \bar{y}^i \hat{h}^i \bar{y}^i + \bar{h}^i \bar{y}^i, \quad H_{test} = \sum_{i=1}^{n} h^i,
\end{gathered} \tag{3}$$

где введены вектор первых производных функций оценки по параметрам и, положительно определенная, в силу того, что описывает окрестности минимума, матрица вторых производных, которые используются для приближенного вычисления функции оценки прогноза. Из (3) следует:

$$\begin{aligned} &H_{test} \approx \sum_{i=1}^{n} \bar{h}^{i}(\hat{H}+\varepsilon-\hat{p}^{i})^{-1}(\bar{h}^{i}-\varepsilon\bar{x}')+\frac{1}{2}\sum_{i=1}^{n}(\bar{h}^{i}-\varepsilon\bar{x}')g; \\ &g(\hat{H}+\varepsilon-\hat{p}^{i})^{-1}(\hat{p}^{i}+\frac{1}{n-1}\hat{H})(\hat{H}+\varepsilon-\hat{p}^{i})^{-1}(\bar{h}^{i}-\varepsilon\bar{x}'). \end{aligned} \tag{4}$$

Используя (3), можно показать, что при большом числе примеров n в первом неисчезающем приближении можно пренебречь в (4) операторами $\hat{p}^{i}$. При этом (4) перейдет в

$$\begin{aligned} &H_{test} \approx \sum_{i=1}^{n} \bar{h}^{i}(\hat{H}+\varepsilon)^{-1}(\bar{h}^{i}-\varepsilon\bar{x}')+\frac{1}{2n-2}\sum_{i=1}^{n}(\bar{h}^{i}-\varepsilon\bar{x}')g; \\ &g(\hat{H}+\varepsilon)^{-1}\hat{H}(\hat{H}+\varepsilon)^{-1}(\bar{h}^{i}-\varepsilon\bar{x}')=\sum_{i=1}^{n}\bar{h}^{i}(\hat{H}+\varepsilon)^{-1}\bar{h}^{i}+ \\ &+\frac{1}{2n-2}\sum_{i=1}^{n}\bar{h}^{i}(\hat{H}+\varepsilon)^{-1}\hat{H}(\hat{H}+\varepsilon)^{-1}\bar{h}^{i}+ \\ &+\frac{\varepsilon^{2}}{2n-2}\sum_{i=1}^{n}\bar{x}'(\hat{H}+\varepsilon)^{-1}\hat{H}(\hat{H}+\varepsilon)^{-1}\bar{x}' \approx \sum_{i=1}^{n}\bar{h}^{i}(\hat{H}+\varepsilon)^{-1}\bar{h}^{i}+ \\ &+\frac{n\varepsilon^{2}}{2n-2}\bar{x}'(\hat{H}+\varepsilon)^{-1}\hat{H}(\hat{H}+\varepsilon)^{-1}\bar{x}'. \end{aligned} \tag{5}$$

Проведем разложение (5) по собственным векторам матрицы $\hat{H}$, соответствующим собственным значениям q_s, большим нуля, в силу положительной определенности соответствующей матрицы

$$\begin{aligned} &\nu_{s}=\sum_{i}(\bar{h}_{s}^{i})^{2}, \quad \mu_{s}=\sum_{s}\bar{x}_{s}^{2}; \\ &H_{test} \approx \sum_{s}\frac{\nu_{s}}{q_{s}+\varepsilon}+\sum_{s}\frac{\varepsilon^{2}\mu_{s}q_{s}}{(2-2/n)(q_{s}+\varepsilon)^{2}}. \end{aligned} \tag{6}$$

Согласно (6), при малых q_s и нулевом ε, тестовая ошибка может значительно возрастать из-за малых знаменателей в членах первой суммы в (6). По сути, в этом случае проявляется плохая обусловленность марицы $\hat{H}$. Содержательно наличие у этой матрицы малых собственных значений ведет к возможности чрезвычайно сильного отклика на сдвиг параметров системы и ухудшает прогнозные качества модели. Ненулевой параметр ε демпфирует роль всех собственных значений, меньших этого параметра, и снимает проблему плохой обусловленности. Можно усмотреть в данном подходе прямые аналоги со стандартными методами регуляризации некорректных задач.

Все члены первой суммы в (6) представляют собой монотонно убывающие функции параметра ε, обращающиеся в нуль в бесконечности. Все члены второй суммы представляют собой монотонно убывающие функции этого параметра, обращающиеся в нуль в нуле. Так может существовать оптимальное, большее нуля и меньшее бесконечности, значение параметра ε, доставляющее минимум суммарной оценке ошибки тестирования. Также возможен вариант, когда оптимальное значение равно бесконечности. Вариант, когда оптимальное значение равно нулю, очевидно исключен: первая производная (6) по ε в нуле отрицательна.

Содержательно нулевые значения параметра ε соответствуют чисто аппроксимационной модели. Бесконечно большие значения параметра ε соответствуют чисто имитационной модели, когда параметры модели вообще не подгоняются под реальные данные.

В качестве грубой оценки величины оптимального соотношения весов рассмотрим следствие предположения, что основную роль в (6) играют некоторые члены с некими «типичными» значениями параметров, соотносимыми у различных собственных значений. В этом случае (6) переходит в выражение

$$H_{test} \sim \frac{\nu'}{q'+\varepsilon} + \frac{\varepsilon^2 \mu' q'}{(2-2/n)(q'+\varepsilon)^2}, \tag{7}$$

где ν', μ' и q' соответствуют неким «типичным» значениям. Минимум величины (7), соответствующий обращению в нуль первой производной, достигается при

$$\varepsilon = (1-1/n)\nu' q' / \left(\mu' q'^2 - \nu'(1-1/n)\right). \tag{8}$$

Данный минимум соответствует физически допустимым положительным значениям, если $\mu' q' > \nu'(1-1/n)$. С ростом размера выборки матрица $\hat{H}$ растет, ориентировочно пропорционально числу членов выборки минус единица, а также растет, ориентировочно пропорционально числу членов выборки, параметр ν'. Полагая для оценки

$$q' \sim \gamma(n-1), \quad \nu' \sim \lambda n, \tag{9}$$

получим для (8):

$$\varepsilon = \lambda\gamma / \left(\mu'\gamma^2 - \lambda/(n-1)\right).$$

Согласно (9), при небольших размерах выборки оптимум может соответствовать бесконечному значению ε, но при больших размерах выборки оптимальное значение стремится к некоему постоянному положительному значению. Из (6) можно получить и строгий достаточный критерий того, что смешанный адаптивно-имитационный подход будет точнее чисто имитационного. Для этого достаточно, чтобы при больших величинах параметра ε ошибка нарастала. В этом случае (6) переходит в

$$H_{test} \approx \sum_s \frac{n\mu_s q_s}{(n-1)} + \left(\sum_s \nu_s - \frac{n}{(n-1)}\sum_s \mu_s q_s^2\right) / \varepsilon. \tag{10}$$

Согласно (10), ошибка будет нарастать при увеличении параметра ε, в области его больших значений, если

$$(n-1)\sum_s \nu_s < n\sum_s \mu_s q_s^2. \tag{11}$$

Ввиду вышеизложенных соображений о характере роста матрицы $\hat{H}$ и ее собственных значений с ростом выборки из (11) также можно сделать вывод, что, при больших выборках, условие (11) будет выполнено, т. е. смешанный адаптивно-имитационный подход будет эффективнее чисто имитационного. Напомним, что данный подход всегда эффективнее аппроксимационного.

Отметим также аналогию между предлагаемым подходом и методами регуляризации некорректных задач [5], также модифиуцирующими малые собственные значения инвертируемых матриц.

Выводы

Согласно проведенному анализу, теоретически из трех подходов (аппроксимационного, имитационного и смешанного имитационно-аппроксимационного), если содержательно применимы все эти подходы, всегда наименее перспективным является аппроксимационный, основанный на минимизации отклонений модели от данных. При небольших выборках перспективнее имитационный, а при больших выборках преимущество теоретически всегда переходит к смешанному имитационно-аппроксимационному подходу.

Практическая реализация смешанного имитационно- аппроксимационного подхода состоит из следующих шагов:

– составляется параметрически зависимая имитационная модель;

– оцениваются вероятные доверительные интервалы параметров;

– по большой обучающей выборке строится функция (1) и процедура ее минимизации по параметрам модели;

– параметр функции, определяющий соотношение аппроксимационной части и веса отклонений параметров от их априорных оценок, находится посредством совокупности пробных бутстреп-процедур таким образом, чтобы минимизировать ошибку прогнозирования;

– найденные наилучшие прогнозирующие параметры модели используются для прогноза. Такой прогноз, на сегодняшнем уровне развития методов количественного прогнозирования, будет наиболее надежен.

Если предполагается дополнить задачу прогнозирования задачей оптимизации, подход будет нуждаться в еще некотором развитии, а именно: при поиске оптимума модель может стремиться уходить в области, далекие от тех, на которых параметры модели надежно верифицированы. Чтобы ограничить эту тенденцию, можно поступить следующим образом: построить бутстреп-методом ошибки прогнозирования для всех задач выборки, затем по этому дискретному набору точек поля ошибок построить модель поля ошибок. При поиске оптимума, в задаче оптимального управления, можно минимизировать или максимизировать не значение управляющей функции, а значение этой функции с учетом модели поля ошибок.

Список литературы

1. *Хемди А. Т.* Имитационное моделирование. Введение в исследование операций = Operations Research: An Introduction. 7-е изд. М.: Вильямс, 2007.

2. *Tetko I. V., Livingstone D. J., Luik A. I.* Neural network studies. 1. Comparison of overfitting and overtraining // Journal of Chemical Information and Computer Sciences. 1995. Vol. 35. No. 5. P. 826–833.

3. *Барцев С. И., Охонин В. А.* Адаптивные сети обработки информации. Красноярск, 1986. Препринт ИФ СО АН СССР № 59Б.

4. *Барцев С. И., Гилев С. Е., Охонин В. А.* Принцип двойственности в организации адаптивных систем обработки информации // Динамика химических и биологических систем. Новосибирск: Наука, 1989. С. 6–55.

5. *Тихонов А. Н., Арсенин В. Я.* Методы решения некорректных задач. 2-е изд., М.: Наука, 1979.

Материал поступил в редколлегию 16.03.2010

S. S. Zamay, V. A. Okhonin, A. L. Shemel

PROBLEM OF RELIABILITY OF MULTIPLEPARAMETER REGRESSION-PORTRAIT MODELS OF COMPLEX SYSTEMS

The opportunity of synthesis of imitating and regression approaches, based on use as adaptive values of deviations of parameters of model from aprioristic, within the limits of confidential intervals of these aprioristic values, is considered. In the frames the given approach, problem of bad conditionality of multi parametric models receives the natural decision.

Keywords: imitating models, regression models, multiparametric models, complex systems.