Ю.Д. АРИНЕНКОВ, ***к.т.н., доцент,***
Донецкий национальный технический университет

ЭКОНОМИКО-ТЕХНОЛОГИЧЕСКИЙ ПРОГНОЗ ПО МАССИВАМ ДАННЫХ

Постановка проблемы

Вопрос построения производственных функций является центральным в экономико-математическом моделировании и принятии решений о планировании и управлении технологиями. В угольной промышленности его решение усложняется изменчивостью содержания полезного компонента в рудном теле при перемещении или смене добычного забоя, при транспортировке и при переработке на обогатительной фабрике. Подобная ситуация с дефицитом информации о свойствах поступающего сырья и ожидаемом качестве и количестве получаемой продукции имеет место и на других производствах, разрешается же она в теории на высоком уровне – применением генетических алгоритмов, а на практике – методами экстраполяции, зависящими от производственной функции, или экспертными методами. Но аналитическое описание производственной функции часто не известно, а опыт экспертов может быть недостаточен или даже ошибочен для оценки конкретной ситуации. Поэтому обоснование предпочтительного метода описания производственных функций и прогноза по такому описанию за область определения исходных данных является актуальной задачей для практического решения проблемы экономико-математического моделирования и оптимизации технологии в промышленных условиях действующего предприятия.

Анализ исследований по проблеме

В литературе по экономико-математическому моделированию важное место занимают «планово-экономические расчёты» [1], но термин «прогноз» как вероятностное суждение о будущем на основе специального научного исследования не употребляется. Более того, при анализе экономической деятельности предприятия даже для нужд конкурентной разведки, когда цель исследования явно прогностическая, считается приемлемым ограничиться использованием уравнения регрессии [2].

Вместе с тем известно, что регрессионное описание массива данных следует использовать в пределах области определения этого массива [3]. В вычислительной математике нахождение значения функции за пределами её определения называют экстраполированием назад или вперёд, в зависимости от того, в каком направлении, от начала или от конца интервала определения функции, выполняется экстраполяция (по Ньютону, по Ричардсону) [3]. Таким образом, экстраполяция увязана с интерполяцией по точкам от заданной и точно определённой в точках массива функции.

Характеристики экономических и технологических исследований обычно представляют в форме таблиц с достаточно точно измеренными данными, однако, найти сколько-нибудь удовлетворительное их описание в виде явной функции как закономерности обычно не удаётся. Тем не менее, для описания и таких характеристик предложены прогностические методы рекурсивного knot- и area-продолжения массивов данных, которые (продолженные массивы данных) затем интерполируются кусочно-линейными функциями [4], но методы экстраполяции для таких массивов не разработаны. В таком случае определённо нужно ставить задачу о выборе предпочтительного метода прогноза непосредственно по массиву данных изучаемого процесса или явления.

Постановка задачи

Пусть имеем массивы данных $X=\{x_0, x_1, x_2, \ldots, x_n\}$, $Y=\{y_0, y_1, y_2, \ldots, y_n\}$, где x – аргумент (независимая переменная), y – функция (зависимая переменная), n – размер массивов, причём, X и Y в общем случае имеют различные размерности и поэтому находятся в неевклидовом пространстве; упорядоченные пары этих массивов образуют множество (x, (y)), задаются простым перечислением по результатам соответствующего исследования и являются отношением (x, (y))∈ C, где C⊂X×Y.

Отношение C имеет отображения типа биекции и толерантности. Для описания такого массива разработан метод рекурсивного knot-продолжения. Этот метод продолжает (увеличивает) массив только в пределах его первичных границ, т.е. добавляет новые точки между первичными точками, не выходя за первую и последнюю точки исходного массива, однако, для нахождения новых опорных точек в крайних интервалах всё же требуется прогноз направления гипотетической функции за пределы указанного массива. Исходя из этой необходимости ставится задача обосновать предпочтительный для указанных начальных условий метод прогностического продолжения исходного массива C.

Метод решения задачи

Для выполнения рекурсии по методу knot-продолжения массива вначале индексы всех элементов массива i=0, 1, 2, …, n умножаются на коэффициент 2. В результате первой рекурсии массив будет продлён до n_1=2n интервалов, все старые точки теперь будут иметь чётные индексы, а нечётные индексы будут заняты новыми опорными точками. В работе [4] показано, что для начала рекурсии, кроме проведения хорд, стягивающих каждую пару смежных точек массива, от точки (x_0, y_0) необходимо построить луч влево как прогноз назад на несуществующую точку (x_{-2}, y_{-2}) и продолжение вправо которого необходимо для формирования доверительного треугольника, своим центром тяжести определяющего новую опорную точку на интевале (x_0, y_0) - (x_2, y_2). Для предсказания коэффициента наклона A этого луча там применена формула

$$A = B^2 \cdot C, \qquad (1)$$

где B, C – коэффициенты наклона стягивающих хорд первого и второго интервалов,

$$B = \frac{y_2 - y_0}{x_2 - x_0}, \qquad (2)$$

$$C = \frac{y_4 - y_2}{x_4 - x_2}. \qquad (3)$$

Понятно, что формула (1) получена на основании условия равенства отношений коэффициентов наклонов стягивающих хорд смежных интервалов

$$A/B=B/C, \qquad (4)$$

поэтому коэффициент наклона прогнозирующего луча будет равен

$$A = \left(\frac{y_2 - y_0}{x_2 - x_0}\right)^2 \cdot \frac{x_4 - x_2}{y_4 - y_2}. \qquad (1,a)$$

Для прогноза вперёд аналогичным образом получена формула

$$C = \left(\frac{y_n - y_{n-2}}{x_n - x_{n-2}}\right)^2 \cdot \frac{x_{n-2} - x_{n-4}}{y_{n-2} - y_{n-4}}. \qquad (5)$$

Эти зависимости вполне удовлетворительно работают на изучавшихся в [4] монотонных характеристиках при прогнозе максимальной крупности по ситовой характеристике.

Легко проверить, что условие (4) даёт обратимое решение, т.е. A(y)=A^{-1}(x). Однако, на немонотонной последовательности формулы (1) и (5) дают неправдоподобные направления.

В качестве альтернативы для (4) рассмотрим условие равенства приращений коэффициентов наклона в начале последнего интервала и в конце его как

$$C-B=k(B-A), \qquad (6)$$

где k – коэффициент значимости (веса) первого приращения наклона, пусть 0 < k < 1, а коэффициенты наклона двух последних стягивающих хорд определяются по формулам

$$A = \frac{y_{n-2} - y_{n-4}}{x_{n-2} - x_{n-4}}, \qquad (7)$$

$$B = \frac{y_n - y_{n-2}}{x_n - x_{n-2}}. \qquad (8)$$

Из (6)

C=B+k(B-A). (9)

Теперь из (9) при k=0 получим условие осторожного прогноза вперёд

C=B, (10)

а при k=1 – смелого прогноза вперёд

C=2B-A. (11)

Формулу для прогноза назад получим из (9), поменяв в ней местами A и B, как

A=B+k(B-C), (12)

но коэффициенты B и C для (12) теперь будут определяться через (2) и (3) соответственно. Коэффициент k в формуле (12) имеет точно такое же значение, как и в формуле (9), и при k=0 получим условие осторожного прогноза назад

A=B, (13)

а при k=1 – условие смелого прогноза назад

A=2B-C. (14)

Проверим функцию (9) на обратимость. Обратимой называем функцию, которая при $y_i=f_1(x_i)$ даёт $x_i=f_2(y_i)$ для всех i в интервале её определения. Отметим левую часть (9) звёздочкой как копию и в полученной копии функции заменим коэффициенты наклона A и B (но не коэффициент k) на их обратные величины. После несложных преобразований установим, что $C^{-1} \neq (C^*)^{-1}$, т.е.функция с вычислением коэффициента наклона по формуле (9) не обратима, она не будет выполнять принятое условие «$y_i=f_1(x_i)$ есть то же самое $x_i=f_2(y_i)$ для всех i в интервале её определения», так что результат прогноза будет зависеть от выбора системы координат. Этот же вывод справедлив и в отношении структуры (12). Зависимость решений (9) и (12) от выбора системы координат воспринимается как существенный критерий их несовершенства и побуждает искать очередную альтернативу.

Воспользуемся известным в математике определением угла между двумя прямыми, заданными коэффициентами наклона [3], с помощью которого сопоставим углы поворота $\Delta\varphi_{B-A}$ между хордами с коэффициентами наклонов A и B и $\Delta\varphi_{C-B}$ между хордами с коэффициентами наклонов B и C как

$$\Delta j_{B-A} = arctg\left(\frac{B-A}{1+A\cdot B}\right), \qquad (15)$$

$$\Delta j_{C-B} = arctg\left(\frac{C-B}{1+B\cdot C}\right). \qquad (16)$$

Устанавливая равенство левых частей (15) и (16), этим самым устанавливаем равенство между выражениями под знаками arctg() правых частей этих же выражений, т.е.

$$\frac{B-A}{1+A\cdot B} = \frac{C-B}{1+B\cdot C}. \qquad (17)$$

Оставляя искомый коэффициент наклона в левой части (17), получим

$$A = \frac{C-2B-B^2C}{B^2-2BC-1}, \qquad (18)$$

где коэффициенты наклона B и C определяются через (2) и (3), и

$$C = \frac{A-2B-B^2A}{B^2-2BA-1}, \qquad (19)$$

где коэффициенты наклона A и B

определяются через (7) и (8) соответственно.

Уже изложенным ранее приёмом можно проверить, что формулы и (18), и (19) являются обратимыми. Однако, по структуре формул (18) и (19) не видно, как ввести в них коэффициент k для управления весом удалённого интервала, который был введен в (6), (9), (12). Поэтому решим предыдущую задачу, не обращаясь к формулам (15)-(17), выразив углы наклона стягивающих хорд через коэффициенты их наклона:

$$\Delta\varphi_{B-A} = \mathrm{arctg}B - \mathrm{arctg}A, \qquad (20)$$

$$\Delta\varphi_{C-B} = \mathrm{arctg}C - \mathrm{arctg}B. \qquad (21)$$

Приравняв левые части (20) и (21), получим условие равенства обеих углов поворота

$$\mathrm{arctg}B - \mathrm{arctg}A = \mathrm{arctg}C - \mathrm{arctg}B. \qquad (22)$$

Из этого условия формулируем условия прогноза назад и вперёд соответственно

$$\mathrm{arctg}A = 2\mathrm{arctg}B - \mathrm{arctg}C, \qquad (23)$$

$$\mathrm{arctg}C = 2\mathrm{arctg}B - \mathrm{arctg}A. \qquad (24)$$

Далее, переходя от углов поворота к коэффициентам наклона, получаем:

$$A = \mathrm{tg}(2\mathrm{arctg}B - \mathrm{arctg}C), \qquad (25)$$

$$C = \mathrm{tg}(2\mathrm{arctg}B - \mathrm{arctg}A). \qquad (26)$$

Численным методом можно проверить, что формулы (25) и (26) дают такие же результаты, как и формулы (18) и (19), следовательно, они также дают обратимые решения. Но теперь из структуры формул (25) и (26) становится очевидной возможность ввести в них коэффициент влияния дальнего интервала k следующим образом:

$$A = \mathrm{tg}[(1+k)\mathrm{arctg}B - k\cdot\mathrm{arctg}C], \qquad (27)$$

$$C = \mathrm{tg}[(1+k)\mathrm{arctg}B - k\cdot\mathrm{arctg}A]. \qquad (28)$$

Таким образом, получены четыре варианта функций для выполнения прогноза назад по первым трём точкам массива и вперёд по трём последним точкам массива. При этом метод может применяться на всех рекурсиях в пределах, соответственно определяемых размером первого и последнего интервала на каждой рекурсии и поэтому даст прогноз от самого далёкого положения до ближайшего, примыкающего к границам определения массива. Однако, критерии выбора альтернативных формул для прогноза не являются критериями отбора предпочтительных для прогноза формул.

Для выявления предпочтительного варианта прогноза из рассмотренных выполнено имитационное исследование, результаты которого обсуждаются ниже.

Основные результаты и применения

В результате проведенного анализа формулы (5), (9), (19) и (26) предложены как альтернативные для решения задачи прогноза вперёд по трём крайним точкам исходного массива. Для выбора предпочтительной формулы был поставлен вычислительный эксперимент. План эксперимента и результаты вычислений коэффициента C по четырём указанным формулам сведены в табл.1.

В таблице правдоподобные результаты вычислений выделены полужирным шрифтом. Как видим, только формула (9) во всех случаях дала правдоподобные значения коэффициента наклона C, что даёт основание именно её рекомендовать для применения в прогнозных задачах.

Таблица 1

План и результаты численного эксперимента по прогнозу вперёд

Точки i плана заданных коэффициентов			Прогнозируемый по различным формулам коэффициент наклона C			
Точка плана i	Коэффициент A	Коэффициент B	Формула (5)	Формула (9)	Формула (19)	Формула (26)
1	2	0	0	-2	-2	-2
2	2	1	0,5	0	0,5	0,5
3	0	1	$+\infty$	2	$+\infty$	0
4	-2	1	-0,5	4	-0,5	-0,5
5	-2	0	0	2	2	2
6	-2	-1	-0,5	0	-0,5	-0,5
7	0	-1	$+\infty$	-2	$+\infty$	0
8	2	-1	0,5	-4	0,5	0,5
9	0	0	0	0	0	0

Эта формула проста в применении. Полагая, что величина интервалов массива данных соответствует свойствам изучаемого явления и методу измерения его параметров, принимаем величину интервала прогнозирования равной величине последнего интервала (для прогноза вперёд)

$$\Delta x_+ = x_n - x_{n-2}, \qquad (29)$$

Прогнозирующие лучи выходят из точки (x_n, y_n) и пересекаются с абсциссой

$$x_+ = x_n + \Delta x_+ \qquad (30)$$

при ординатах

$$y_B = y_n - B \cdot \Delta x_+, \qquad (31)$$

$$y_C = y_n - C \cdot \Delta x_+. \qquad (32)$$

Полученные три точки (x_n, y_n), (x_+, y_B) и (x_+, y_C) образуют доверительный треугольник, через центр тяжести которого по предположению с наибольшей вероятностью должна пройти прогнозирующая кривая. Координаты этой точки определяются по известным формулам:

$$x = \frac{x_n + 2x_+}{3}, \qquad (33)$$

$$y = \frac{y_n + y_B + y_C}{3}. \qquad (34)$$

В качестве примера для прогноза взяты данные, считанные с графика из [2] (таблица 2).

Таблица 2

Значения прибыли от поставок слябов на американский рынок из [2]

i	x, тыс.т	y, млн.$ США
0	105	69,6
2	123	68,4
4	131	52,2
6	143	55,2
8	162	61,2
10	168	46,2
12	182	45
Прогноз по регрессии	196	42

По данным этой таблицы выполним прогноз вперёд. Для этого вначале по формулам (7) и (8) вычислим коэффициенты

$$A = \frac{46,2-61,2}{168-162} = -2,5\,\frac{\text{млн.долларов}}{\text{тыс.тонн}},$$

$$B = \frac{45-46,2}{182-168} = -0,0857\,\frac{\text{млн.долларов}}{\text{тыс.тонн}}.$$

Далее по формуле (9) при k=0,5 получим искомую величину

C=-0,0857+0,5·(-0,0857+2,5)=1,12145.

Затем по формуле (29) находим величину интервала прогноза

Δx_+=182-168=14 [тыс.т]

По формуле (30) вычисляем абсциссу прогноза

x_+=182+14=196 [тыс.т]

По (31) и (32) находим ординаты прогнозируемых точек

y_B=45+(-0,0857)·14=43,800[млн.долл. США]

y_C=45+1,12145·14=60,700[млн.долл. США]

В результате имеем три точки для построения доверительного треугольника прогноза, по которым с помощью (33) и (34) находим координаты центра тяжести доверительного треугольника прогноза

$$x = \frac{182+196+196}{3} = 191,33\ [\text{тыс.т}]$$

$$y = \frac{45+46,200+29,300}{3} = 40,167[\text{млн.долл.США}]$$

Как видим, по регрессии точка далёкого прогноза вперёд была (196; 42), см.табл.2, а по разработанному методу это (196; 60,7), т.е. на 18,7 млн.долларов США выше, чем по регрессии. Но главный результат следует видеть в том, что примененный метод позволяет рассматривать экономический процесс [2] как существенно колеблющийся, чтобы затем вскрыть причины этих колебаний и, возможно, использовать их в интересах предприятия.

Несмотря на кажущуюся простоту, изложенный метод следует применять в виде программного продукта на ЭВМ. Именно в такой форме он применен для прогноза максимальной крупности угля при вычислении гранулометрических характеристик на ЭВМ [5].

Выводы

Предложен вариант прогноза параметров экономико-технологической системы по трём последним точкам массива данных для прогноза вперёд. Для этой цели разработаны четыре альтернативных формулы. В результате численного эксперимента установлено, что предпочтительной является формула прогноза параметров через приращение коэффициента наклона прогнозирующей кривой по отношению к коэффициенту наклона стягивающей хорды последнего интервала массива соответственно приращению её коэффициента наклона по сравнению с коэффициентом наклона стягивающей хорды предыдущего к последнему интервалу массива данных. Дан численный пример решения.

Литература

1. Лотов А.В. Введение в экономико-математическое моделирование / Под ред. Н.И.Моисеева. – М.: Наука. Главная редакция физико-математической литературы, 1984. –392с.

2. Череватский Д.Ю., Роттер М.В. Использование математической статистики для нужд конкурентной разведки //Вісник Академії економічних наук України: Науковий щорічник -2003. -№2 (4). -С.143 - 145.

3. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров -М.: «Наука», Главная редакция физико-математической литературы, 1973. –632с.

4. Ариненков Ю.Д. Эффективность рекурсивного продолжения коротких массивов данных состава угля // ВІСТІ Донецького гірничого інституту. Всеукраїнський науково-технічний журнал гірничого профілю. Донецький національний технічний університет. –Донецьк, ДонНТУ, 2003. -№2. –С.88-92.

5. Arinenkov Yu. Universal model for research both optimization of technological processes and circuits of coal preparation factories on the COMPUTER/2nd Regional APCOM'97 Symposium on COMPUTER APPLICATIONS AND OPERATIONS RESEARCH IN THE MINERAL INDUSTRIES. Published by: The Moscow State Mining University Publishing Center. Moscow, Russia - 1997. 536p. - P.209-214.

Статья поступила в редакцию 14.05.2004

А.В. СМИРНОВ, ***к.т.н., доцент,***
Д.В. РЕВЕГА,
Донецкий национальный технический университет

СИНТЕТИЧЕСКИЕ СКОЛЬЗЯЩИЕ СРЕДНИЕ

Скользящее усреднение нашло самое широкое применение для выявления тенденций изменения во времени различных экономических характеристик и явлений [1,2]. Это произошло потому, что его алгоритмы просты, прозрачны и понятны исследователям временных рядов. Вместе с тем имеются и существенные недостатки рассматриваемых способов выявления трендов, которые вредны при реализации процессов оперативного управления экономическими системами на практике (реализация компьютерных биржевых торговых систем для формирования эффективных торговых входов/выходов в рынки; оперативное управление современным производством с помощью компьютерных систем поддержки принятия управленческих решений, корпоративных компьютерных систем и т.д.).

При этом основные и наиболее существенные недостатки многочисленных известных алгоритмов скользящего усреднения сводятся к следующему:

- объективно существующая и принципиально неустранимая (по мнению большинства исследователей временных рядов) временная задержка продуктов усреднения относительно исследуемого временного ряда на величину m/2, где: m – величина временного окна скользящего усреднения [1,2];

- высокая колеблемость выделяемых трендов, которая слабо уменьшается с ростом величины m [1,2];

- наличие линейных частотных искажений при выделении трендов (форма выделяемых трендов отличается от реально существующих тенденций временных рядов из-за специфических свойств амплитудно-частотных характеристик (АЧХ) – квадратов модулей комплексных коэффициентов передачи алгоритмов усреднения в полосе их прозрачности в частотной области [3]).

Отметим, что объективное существование названных двух первых недостатков алгоритмов усреднения фиксировалось в специальной литературе как принципиально неустранимое. Экономисты-исследователи вполне с ними мирились. И это было обусловлено тем, что в большинстве случаев ими ранее решались статистические экономические задачи без существенных временных ограничений на их решение.

Одной из первых работ, в которой автор обратил внимание на влияние АЧХ