УДК 681.513.6

С.В. ПОПОВ

ОБ ОДНОМ АДАПТИВНОМ АЛГОРИТМЕ ОЦЕНИВАНИЯ ПАРАМЕТРОВ 2-D МОДЕЛЕЙ

Рассматривается задача адаптивной обработки стохастических полей наблюдений на основе 2-D подхода и показывается, что широко применяемый для оценивания параметров 2-D моделей рекуррентный метод наименьших квадратов недостаточно эффективен при обработке нестационарных данных. Предложен адаптивный алгоритм оценивания параметров со скалярным коэффициентом усиления, обладающий выраженными следящими свойствами, доказана его сходимость. Результаты имитационного моделирования подтверждают высокую эффективность предложенного алгоритма.

В различных областях прикладных исследований, связанных с обработкой данных экологического мониторинга, экономических показателей, физических исследований и т.д., часто возникают задачи обработки двумерных полей наблюдений. В силу специфики организации данных обработка такой информации стандартными методами анализа оказывается малоэффективной. Это выражается в том, что в модель включается или слишком много данных, чтобы охватить все релевантные факторы, что неизбежно приводит к резкому увеличению количества настраиваемых параметров и ухудшению сходимости алгоритма настройки, или, во избежание этих трудностей, количество параметров уменьшается, но при этом теряется и полезная информация. В данном случае эффективными являются методы, основанные на 2-D подходе [1-3]. В них естественным образом учитывается влияние данных, расположенных в непосредственной близости к анализируемой точке, и количество настраиваемых параметров оказывается небольшим.

Для представления полей наблюдений в 2-D подходе используется двумерный эквивалент одномерных моделей авторегрессии и скользящего среднего (АРСС). В частности, если поле доступно в режиме растрового сканирования, то значение сигнала в пикселе с координатами (m, n) – s(m, n) может быть представлено как авторегрессия со скользящим средним от множества доступных данных в текущей и всех предыдущих строках. Если предположить, что сигнал s(m, n) генерируется процессом типа двумерного белого шума w(m, n) с нулевым математическим ожиданием и дисперсией σ_w^2, то можно предложить двумерную АРСС модель в виде

$$\sum_{(i,j)\in\beta} a(i,j)s(m-i,n-j) = \sum_{(i,j)\in\alpha} c(i,j)w(m-i,n-j). \qquad (1)$$

Двумерная последовательность белого шума имеет следующую корреляционную структуру:

$$M\{w(m,n)w(m-i,n-j)\} = \begin{cases} \sigma_w^2, \text{ при } i=j=0; \\ 0, \text{ в противном случае.} \end{cases}$$

При использовании нотации с оператором сдвига назад [4] соотношение (1) может быть переписано в другом виде:

$$A\left(p^{-1},q^{-1}\right)s(m,n) = C\left(p^{-1},q^{-1}\right)w(m,n),$$

где $p^{-i}x(m,n) = x(m-i,n)$, $q^{-j}x(m,n) = x(m,n-j)$, а полиномы A и C определены следующим образом:

$$\begin{cases} A\left(p^{-1},q^{-1}\right) = \sum_{(i,j)\in\beta} a(i,j)p^{-i}q^{-j}, \\ C\left(p^{-1},q^{-1}\right) = \sum_{(i,j)\in\alpha} c(i,j)p^{-i}q^{-j}. \end{cases}$$

Коэффициенты a(i, j) относятся к авторегрессионной части модели и имеют ненулевые значения в области b, которая называется опорной областью множества коэффициентов a(i, j). В терминах двумерной обработки сигналов опорная область b является аналогом порядка полинома авторегрессии. Коэффициенты c(i, j) относятся к части скользящего среднего. Опорной областью для c(i, j) является a.

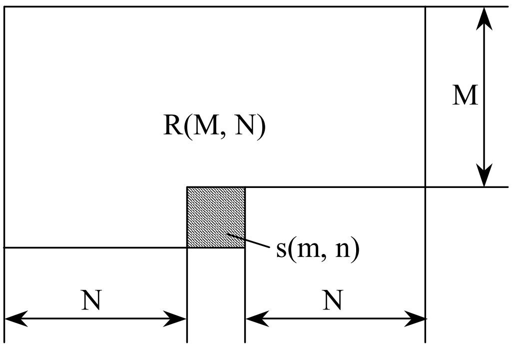
Коэффициент a(0, 0) в авторегрессионной части принимается равным единице. Таким образом, модель сигнала может быть записана в виде

$$s(m,n) = -\sum_{(i,j)\in\beta-(0,0)} a(i,j)s(m-i,n-j) + \sum_{(i,j)\in\alpha} c(i,j)w(m-i,n-j), \qquad (2)$$

где запись b – (0, 0) означает область b, за исключением пикселя (0, 0).

Заметим, что уравнение (2) сходно с хорошо известным разностным представлением сигнала, в котором текущее значение представляется в виде взвешенной суммы предыдущих значений сигнала и шума. Размеры и форма опорных областей b и a определяют, какие из прошлых значений сигнала и шума входят в эти суммы. В данном контексте в качестве опорной области часто используется [1] несимметричная полуплоскость (рисунке), обозначаемая R(M, N). Такая опорная область включает M предыдущих строк и по N столбцов с обеих сторон от текущего пикселя.

При такой форме опорной области полиномы $A(p^{-1}, q^{-1})$ и $C(p^{-1}, q^{-1})$ принимают вид



Опорная область R(M, N)

$$\begin{cases} A(p^{-1},q^{-1}) = \sum\limits_{i=0}^{M}\sum\limits_{j=0}^{N} a(i,j)p^{-i}q^{-j} + \sum\limits_{i=1}^{M}\sum\limits_{j=1}^{N} a(i,-j)p^{-i}q^{j}, \\ C(p^{-1},q^{-1}) = \sum\limits_{i=0}^{M}\sum\limits_{j=0}^{N} c(i,j)p^{-i}q^{-j} + \sum\limits_{i=1}^{M}\sum\limits_{j=1}^{N} c(i,-j)p^{-i}q^{j}. \end{cases}$$

Будем предполагать, что рассматривается устойчивый минимально-фазовый двумерный случайный процесс. Это предположение налагает ограничения на коэффициенты АРСС-модели (1), которые превращаются (как и в одномерном случае) в условия устойчивости полиномов. Устойчивость двумерных полиномов является сложной проблемой, которая на сегодня окончательно не решена.

Для оценивания параметров описанной модели чаще всего [1] применяются различные модификации рекуррентного метода наименьших квадратов (РМНК), обладающие хорошими сглаживающими свойствами. Однако в случае нестационарности параметров модели, характерной для многих практических случаев обработки полей наблюдений, данный алгоритм оказывается недостаточно эффективным. Поэтому возникает задача синтеза адаптивного алгоритма оценивания параметров модели (1), обладающего выраженными следящими свойствами.

Рекуррентное уточнение 2-D модели, в принципе, не отличается от одномерного случая в том, что данные и неизвестные параметры могут быть представлены в векторной форме:

$$\begin{cases} y(m,n) = \theta^{T}x(m,n) + w(m,n), \\ \hat{y}(m,n) = \hat{\theta}^{T}x(m,n), \end{cases}$$

где q – вектор неизвестных параметров; x(m, n) – вектор наблюдений; $\hat{\theta}$ – оценка вектора параметров модели.

В теории и практике адаптивной идентификации как наиболее простые широко распространены алгоритмы, основанные на стохастической аппроксимации [5]. Вместе с тем эти процедуры пригодны лишь для идентификации стационарных процессов. С тем, чтобы придать подобного рода алгоритмам следящие свойства, введем в рассмотрение процедуру вида

$$\begin{cases} \hat{\theta}(m,n) = \hat{\theta}(m,n-1) + \dfrac{a}{p(m,n)}\left(y(m,n) - \hat{\theta}^T(m,n-1)x(m,n)\right)x(m,n), \\ p(m,n) = \lambda(m,n)p(m,n-1) + \|x(m,n)\|^2, \\ 0 \le \lambda(m,n) \le 1, \quad 0 < a < 2, \quad p(1,1) = 1, \end{cases} \tag{3}$$

являющуюся расширением алгоритма, предложенного в [6], на двумерный случай. Заметим, что при l(m, n) = 0 (3) превращается в двумерный аналог алгоритма Качмажа [7], а при l(m, n) = 1 – в алгоритм Гудвина-Рэмеджа-Кейнеса [8].

Принципиальным отличием алгоритма (3) от процедуры стохастической аппроксимации Гудвина-Рэмеджа-Кейнеса является использование в нем параметра дисконтирования l(m, n), что обеспечивает процедуре требуемые следящие свойства и в то же время исключает возможность "взрыва параметров", характерного для взвешенного метода наименьших квадратов.

Запишем (3) относительно ошибок оценивания параметров:

$$\tilde{\theta}(m,n) = \theta - \hat{\theta}(m,n) = \tilde{\theta}(m,n-1) + \frac{a}{p(m,n)}x(m,n)\varepsilon(m,n),$$

где $\varepsilon(m, n)$ – обновляющая последовательность, и введем функцию Ляпунова:

$$\begin{gathered} V(m,n) = \tilde{\theta}^T(m,n)\tilde{\theta}(m,n) = V(m,n-1) - \frac{2a}{p(m,n)}\tilde{\theta}^T(m,n-1)x(m,n)\left(\varepsilon(m,n) - w(m,n)\right) - \\ - \frac{2a}{p(m,n)}\tilde{\theta}^T(m,n-1)x(m,n)w(m,n) + \frac{a^2}{p^2(m,n)}\left(\varepsilon(m,n) - w(m,n)\right)^2 x^T(m,n)x(m,n) + \\ + 2w(m,n)x^T(m,n)x(m,n)\left(\varepsilon(m,n) - w(m,n)\right) + w^2(m,n)x^T(m,n)x(m,n). \end{gathered} \tag{4}$$

Усредняя (4) по $w(m, n)$, получаем

$$\begin{gathered} M\{V(m,n)\} = V(m,n-1) - \frac{2a}{p(m,n)}\tilde{\theta}^T(m,n-1)x(m,n)\left(\varepsilon(m,n) - w(m,n)\right) + \\ + \frac{a^2}{p^2(m,n)}\left(\varepsilon(m,n) - w(m,n)\right)^2\|x(m,n)\|^2 + \|x(m,n)\|^2\sigma_w^2. \end{gathered}$$

Исследование сходимости алгоритма, базирующееся на рассмотрении последовательности (4) как супермартингала, совпадает с подходом Гудвина-Рэмеджа-Кейнеса, при этом принципиальным моментом является выполнение на каждой итерации условия

$$p^2(m,n) > p(m,n)p(m,n-1),$$

откуда следует, что параметр l(m, n) должен находиться в интервале

$$1 - \frac{\|x(m,n)\|^2}{p(m,n-1)} < \lambda(m,n) \le 1,$$

или $\left(1 - \lambda(m,n)\right)\sum_{j=1}^{n}\lambda^{n-j-1}(m,j)\|x(m,j)\|^2 \le \|x(m,n)\|^2$.

Таким образом, сходимость алгоритма (3) может обеспечиваться либо постоянным ростом величины $\|x(m,n)\|^2$, что может быть сделано искусственно путем увеличения амплитуды входных сигналов (режим активного эксперимента), либо, что более реально, соответствующим изменением параметра l(m, n), который должен увеличиваться от 0 до 1.

Оценим влияние l(m, n) на скорость сходимости алгоритма. Для этого запишем (4) в виде

$$V(m,n) = V(m,n-1) - \frac{2a}{p(m,n)}\tilde{\theta}^T(m,n-1)x(m,n)\varepsilon(m,n) + \frac{a^2}{p^2(m,n)}\varepsilon^2(m,n)\|x(m,n)\|^2.$$

Если в качестве характеристики скорости сходимости принять изменение функции Ляпунова на каждом шаге, то решение дифференциального уравнения

$$\frac{\partial(V(m,n-1) - V(m,n))}{\partial p(m,n)} = \frac{2a\tilde{\theta}^T(m,n-1)x(m,n)\varepsilon(m,n)}{p^2(m,n)} - \frac{2a^2\varepsilon^2(m,n)\|x(m,n)\|^2}{p^3(m,n)} = 0$$

позволяет получить значение p(m, n), обеспечивающее максимальную скорость сходимости, $p(m,n) = a\|x(m,n)\|^2$.

С другой стороны, p(m, n) определяется вторым соотношением из (3), откуда видно, что максимальное быстродействие алгоритма достигается при a = 1, l(m, n) = 0, что соответствует алгоритму Качмажа.

Как видно, "скоростные" свойства алгоритма (3) вступают в противоречие с его сглаживающими свойствами, поэтому в процессе настройки целесообразно начинать работу с малых значений l(m, n), обеспечивая тем самым высокую скорость сходимости, увеличивая далее этот параметр до уровня, обеспечивающего требуемый компромисс между сглаживающими и следящими свойствами.

Для подтверждения теоретических выкладок проведено имитационное моделирование, целью которого было сравнение предложенного алгоритма с РМНК. В качестве основы были взяты результаты и методика из [3]. Моделирование показало, что при решении различных классов задач двумерного прогнозирования и сглаживания алгоритм (3) обеспечивает выигрыш по точности в среднем около 2 раз по сравнению с РМНК, обладая при этом более высоким быстродействием, что позволяет эффективно использовать его в задачах обработки полей наблюдений в реальном времени.

Список литературы: 1. *Wellstead P.E., Wagner G.R., Caldas-Pinto J.R.* Two-dimensional adaptive prediction, smoothing and filtering // IEE Proceedings. 1987. Vol. 134. Pt. F. No. 3. P. 253-267. 2. *Wellstead P.E., Zarrop M.B.* Self-tuning systems. Control and signal processing. Chichester: John Wiley & Sons, 1991. 579 p. 3. *Caldas-Pinto J.R., Wellstead P.E.* Self-tuning filters and predictors for two-dimensional systems // Int. J. Cont. 1985. Vol. 42, No. 2. P. 457-515. 4. *Даджион Д., Мерсеро Р.* Цифровая обработка многомерных сигналов. М.:Мир, 1988. 488 с. 5. *Льюнг Л.* Идентификация систем: Теория для пользователя. М.: Наука, 1991. 432 с. 6. *Бодянский Е.В., Плисс И.П., Соловьева Т.В.* Многошаговые оптимальные упредители многомерных нестационарных стохастических управляемых процессов // Докл. АН УССР. Сер. А. 1986. № 12. С. 47-49. 7. *Kaczmarz S.* Angenaherte Auflosung von Systemen linearer Gleichungen // Bull. Int. Acad. Polon. Sci. Lett., Cl. Sci. Math. Nat., Ser. A. 1937. S. 355-357. 8. *Goodwin G.C., Ramadge P.J., Caines P.E.* Discrete time stochastic adaptive control // SIAM J. Contr. and Optim. 1981. 19. No. 6. P. 829-853.

Поступила в редколлегию 12.03.2001

Попов Сергей Витальевич, аспирант кафедры искусственного интеллекта, младший научный сотрудник проблемной НИЛ АСУ ХТУРЭ. Научные интересы: адаптивная обработка полей наблюдений, адаптивное прогнозирование временных последовательностей, нейро- и фаззи- технологии. Адрес: Украина, 61166, Харьков, пр. Ленина 14, тел. 40-98-90. E-mail: Serge.Popov@ieee.org.