УДК 519.216

АНАЛИЗ АЛГОРИТМА ОБНАРУЖЕНИЯ НЕОДНОРОДНОСТИ ТЕЛЕТРАФИКА

Галактионова О.В.
Кафедра математической статистики и системного анализа

Поступила в редакцию 20.02.2011, после переработки 10.03.2011.

Анализируется алгоритм, позволяющий разделить случаи однородного и неоднородного трафика телекоммуникационных систем.

The algorithm allowing to separate the cases of the homogeneous and nonhomogeneous traffic of telecommunication systems is analyzed.

Ключевые слова: долговременная зависимость, дробное броуновское движение, оценка параметров, телекоммуникационные сети.
Keywords: long-range dependence, fractional brownian motion, parameter estimation, telecommunications networks.

Введение

При проектировании телекоммуникационных систем необходимо учитывать все компоненты трафика. Обычно предполагается, что трафик представляет из себя сумму нескольких независимых компонент с различными свойствами. Каждая компонента описывается самоподобным процессом, наиболее важной характеристикой которого является параметр Херста. Поэтому важной практической задачей для неоднородного трафика является оценка параметров Херста каждой аддитивной компоненты в отдельности. В данной работе в качестве модели неоднородного трафика рассматривается сумма двух независимых дробных броуновских движений (ДБД) с различными показателями Херста. Такие модели есть частный случай моделей, предложенных в работе [1].

В терминах спектральной плотности свойство долговременной зависимости (ДВЗ) записывается в виде:

$$f_X(\nu) \sim c_f |\nu|^{-\alpha}, \nu \to 0, \tag{1}$$

при $\alpha \in (0,1)$.

В этом определении присутствуют два параметра (α, c_f). Заметим, что параметр Херста связан с параметром ДВЗ α соотношением $H = (1+\alpha)/2$. Поэтому, оценив один из них, мы можем найти оценку и для второго параметра.

Далее мы рассматриваем случай, когда результирующий процесс состоит из двух независимых стационарных компонент с нулевыми средними и параметрами Херста H_1 и H_2 соответственно. В силу аддитивности преобразования Фурье рассматриваемую модель неоднородного трафика разумно определить следующим образом: при $|\nu| > 0$ спектральная плотность случайного процесса удовлетворяет соотношению:

$$f_X(\nu) \sim c_1|\nu|^{-\alpha_1} + c_2|\nu|^{-\alpha_2}, \nu \to 0,\ 0 < \alpha_2 < \alpha_1 < 1. \tag{2}$$

Наша задача состоит в том, чтобы оценить параметры ДВЗ α_1, α_2 или параметры Херста H_1, H_2. Решение такой задачи предложено в работе [2]. Однако, прежде чем приступить к оценке параметров, необходимо уточнить ситуацию. А именно, выяснить какой процесс исследуется: однородный или неоднородный.

Целью статьи является анализ алгоритма, позволяющего разделить случаи однородного и неоднородного трафика телекоммуникационных систем.

1. Ограничения алгоритма

Для оценок параметра ДВЗ α и параметра Херста H Patrice Abry и Darryl Veitch в 1996 году предложили использовать метод, основная идея которого заключается в использовании последовательности вейвлет-коэффициентов процесса. Подробное описание предложенной методики, а также разработанного на ее основе алгоритма оценки параметров неоднородного трафика можно найти в работе [2]. Кратко перечислим основные ограничения, при которых работает алгоритм, имеющие существенное значение при численном моделировании.

Пусть $\psi_0(t) \in L^2(\mathbb{R})$ есть так называемый материнский вейвлет.

Далее предполагается, что выполнены следующие ограничения:

Предположение 1. *Базисная система функций строится с использованием оператора масштабного преобразования:*

$$\psi_{j,0} = 2^{-j/2}\psi_0(2^{-j}t).$$

Предположение 2. *ψ_0 имеет N нулевых моментов:*

$$\int t^k \psi_0(t) dt = 0,$$

$k = 0, ..., N-1$ *(но не для $k \geq N$!).*

В реальных задачах вместо непрерывного используется так называемое дискретное вейвлет-преобразование. Если исходные данные имеют длину n, то выходные данные $c_{j,k}$ разбиты на $J = \lfloor \log_2 n \rfloor$ октав. Число доступных вейвлет-коэффициентов для каждой октавы j равно $n_j = \lfloor 2^{-j} n \rfloor$. Коэффициенты вейвлет-преобразования определяются по правилу:

$$d_{j,k} = \int \psi_{j,k}(t) X(t) dt, \tag{3}$$

где интеграл от случайной функции понимается в среднеквадратическом смысле. Набор этих коэффициентов обладает следующими свойствами:

Свойство 1. *В силу Предположения 1, имеет место точная масштабная инвариантность (степенное поведение):*

$$E(d_{j,k})^2 = 2^{j\alpha} c_f C\ , \tag{4}$$

где $C = \int |\nu|^{-\alpha} |\Psi_0(\nu)|^2 d\nu$, Ψ_0 - преобразование Фурье функции ψ_0.

Свойство 2. *В силу Предположений 1 и 2, $d_{j,k}$ есть совокупность случайных величин, которые являются квази-декоррелированными. В частности, долговременная зависимость, присутствующая в области временного представления, полностью отсутствует в плоскости (j,k) вейвлетных коэффициентов.*

Кроме этого для процесса $(X(t), t \geq 0)$ выполнены следующие дополнительные условия:

Предположение 3. *Процесс $(X(t), t \geq 0)$ и, следовательно, $d_{j,k}$ являются гауссовскими.*

Предположение 4. *Для фиксированного j с.в. $\{d_{j,k}\}_{k\in\mathbb{Z}}$ независимы и одинаково распределены.*

Предположение 5. *Последовательности $\{d_{j,k}\}_{k\in\mathbb{Z}}$ при разных j независимы.*

2. Алгоритм оценки параметров неоднородного трафика

Алгоритм оценки параметров неоднородного трафика подробно рассмотрен в статье [2]. Коротко напомним основные его этапы:

- Применяя дискретное вейвлет-преобразование к исследуемому процессу $X(t)$, получаем вейвлет-коэффициенты $d_{j,k}$.

- Находим величины μ_j и y_j по формулам

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} d_{j,k}^2, \tag{5}$$

$$y_j = log_2 \mu_j - g_j, \tag{6}$$

  где n_j - число коэффициентов, доступных для анализа в октаве j, $g_j = \psi(n_j/2)/\ln 2 - \log_2(n_j/2)$.

- Далее, используя МНК, рассчитываем оценки $(\widehat{a_1}, \widehat{b_1})$ параметров α_1 и $\log_2 c_1$:

$$\widehat{a_1} = \frac{\sum y_j (S x_j - S_x)/\sigma_j^2}{S S_{xx} - S_x^2} = \sum w_j y_j, \tag{7}$$

$$\widehat{b_1} = \frac{\sum y_j (S_{xx} - S_x x_j)/\sigma_j^2}{S S_{xx} - S_x^2} = \sum v_j y_j, \tag{8}$$

  где $S = \sum 1/\sigma_j^2$, $S_x = \sum x_j/\sigma_j^2$, $S_{xx} = \sum x_j^2/\sigma_j^2$, в качестве фактора x_j выступает номер октавы $x_j = j$. Расчеты выполняются по второй половине октав исходного диапазона $[J/2, \ldots, J]$.

- Находим величины μ_j' по формулам:

$$\mu_j' = \mu_j - Q_j(\widehat{b_1}, \widehat{a_1}), \tag{9}$$

где $Q_j = (\widehat{b}, \widehat{a}) = q 2^{j\widehat{b} + \widehat{a}}$, $q = \prod \frac{exp[lw_j + v_j)\psi(n_j/2)]\Gamma(n_j/2)}{\Gamma(lw_j + v_j + n_j/2)}$, и Q_l определено только для тех j для которых $lw_j + v_j + n_j/2 > 0$.

Q_j - несмещенная оценка $2^{j\alpha_1}C_1$ (см. (4)), следовательно

$$E\mu'_j = 2^{j\alpha_2}C_2.$$

- Определим
$$y'_j = \log_2 \mu'_j - g_j. \tag{10}$$

- Пусть $\mu'_0 > 0$ - некоторое фиксированное число. Будем считать y_j определенным только для тех j, для которых $\mu'_j > \mu'_0$. Обозначим это множество за J_2.

- Далее, используя МНК, рассчитываем оценки $(\widehat{a_2}, \widehat{b_2})$ параметров α_2 и $\log_2 c_2$ при условии, что величины y'_j определены только для $j \in J_2 = \{j : \mu'_j > \mu'_0, \mu'_0 > 0\}$.

Известно, что полученные оценки являются несмещенными и состоятельными.

3. Численное моделирование

Для анализа точности полученных теоретических результатов было использовано численное моделирование. Отметим, что предложенный алгоритм разработан с учетом выполнения некоторых ограничений. В частности, согласно Предположению 3, исходный процесс должен быть гауссовским. Поэтому далее для анализа свойств алгоритма мы используем модельный процесс, состоящий из двух независимых дробных броуновских шумов с различными параметрами Херста (т.е. приращений дробных броуновских движений, которые образуют стационарную последовательность).

Для моделирования процесса ДБД мы использовали метод, основанный на алгоритме случайного перемещения средней точки (RMA). Данный метод генерирует приблизительно самоподобную последовательность с относительной ошибкой менее 6%, в области $0.5 \le H \le 0.95$. По сравнению с генератором, основанным на БПФ, данный метод работает намного быстрее. Поэтому для практического моделирования трафика рекомендуют использовать RMA-метод. Теоретическая алгоритмическая сложность метода - $O(n)$.

При расчете вейвлет преобразования одним из важных вопросов является выбор материнского вейвлета. В работе [3] показано, что для оценки параметра Херста единственным необходимым условием является число исчезающих моментов. Другие свойства вейвлетов (симметричность, ортонормированность) не имеют никакого теоретического и существенного практического значения. В качестве конкретного вейвлет-базиса можно использовать вейвлеты Добеши, так как для них заранее известно число исчезающих моментов. Заметим, что это не единственный вариант. Могут использоваться и другие вейвлеты.

Обратимся к вопросу о стационарности исследуемого процесса. Хорошо известно, что в условиях эффекта долговременной зависимости трудно проверять гипотезу о стационарности данных. В работе [4] показано, что за счет правильного выбора вейвлета можно выявить тенденции данных, а также уменьшить их влияние на полученные оценки. В данном случае особую роль играет число исчезающих моментов N выбранного вейвлета. Этот параметр также важен при оценивании дисперсии оценки параметра самоподобия. В той же работе показано, что чем больше N, тем лучше оценка. Однако, при увеличении числа N увеличивается и число вейвлет-коэффициентов, которые исключаются при выполнении расчетов. Это связано с тем, что при проведении эксперимента мы имеем дело с данными конечного размера и, как следствие, проявляются так называемые «краевые эффекты». В качестве компромиссного решения предлагается выбирать число исчезающих моментов $N \simeq H + 1$ ([5]).

Хорошо известно, что для вычисления коэффициентов $d_{j,k}$ вейвлет преобразования требуется задание начальной последовательности $a_k(0,k)$. Такой начальный ряд вычисляется непосредственно из исходного процесса X_t, а затем используется для расчетов вейвлет-коэффициентов с помощью быстрого рекурсивного пирамидального алгоритма. Часто упрощение $a_k(0,k) = x(k)$ приводит к ошибкам в $d_{j,k}$. Однако в работах [6], [7] показано, что ошибки инициализации являются существенными на первых октавах и быстро уменьшаются с увеличением j. В случае анализа ДВЗ первые октавы для рассчетов не берутся. Поэтому в качестве начальной последовательности будем использовать исходный ряд.

4. Алгоритм обнаружения неоднородности трафика

В работе [2] описан численный эксперимент, на основании которого выявлены закономерности, позволяющие определить однородность-неоднородность исследуемого процесса. Сравнив результаты рассчитанных оценок по всему диапазону октав и по второй половине октав, мы можем однозначно ответить на вопрос об однородности процесса. Если оценка, рассчитанная по исходной методике выше или равна оценке, рассчитанной по второй половине октав, то это говорит о том, что процесс однороден. В противном случае мы имеем дело с неоднородным процессом.

Этот алгоритм хорошо работает в случае, когда веса разных компонент различаются не сильно. Хотелось бы понять как работает этот алгоритм в случаях, когда компоненты процесса имеют разные веса. В частности выявить границы применимости этого метода. Провести аналитическое исследование в такой задаче не удается. Поэтому далее проводится численное исследование свойств предложенного алгоритма выявления неоднородности трафика и оценки параметров трафика.

Проведем эксперимент с различными весами компонент неоднородного процесса. Для анализа выбран процесс (a_1·ДБШ(0.9)+a_2·ДБШ(0.6)). Проведено по 500 независимых измерений каждого процесса. Отдельная реализация смоделированного процесса содержала 2^{16} точек. Заметим, что при оценке неоднородного трафика используется следующее обозначение: $\widehat{H_1}$ - оценка для параметра Херста $H_1 = 0.9$, $\widehat{H_2}$ - оценка для параметра Херста $H_2 = 0.6$.

Результаты эксперимента представлены в таблице 1.

Таблица 1: Оценка параметров неоднородного процесса с разными весами компонент

Веса компонент (a_1/a_2)	Оценка параметра Херста (метод Abry и Veitch) $\widehat{H}$; $[H_{min}, H_{max}]$	Оценки параметров Херста компонент процесса $\widehat{H_1}$; $[H_{1,min}, H_{1,max}]$ $\widehat{H_2}$; $[H_{2,min}, H_{2,max}]$
0.9/0.1	$\widehat{H}$ = 0.886; [0.865,0.907]	$\widehat{H_1}$ = 0.885; [0.850,0.910]
0.8/0.2	$\widehat{H}$ = 0.872; [0.850,0.894]	$\widehat{H_1}$ = 0.873; [0.848,0.898]
0.7/0.3	$\widehat{H}$ = 0.841; [0.819,0.863]	$\widehat{H_1}$ = 0.911; [0.886,0.936] $\widehat{H_2}$ = 0.597; [0.574,0.620]
0.6/0.4	$\widehat{H}$ = 0.810; [0.789,0.831]	$\widehat{H_1}$ = 0.907; [0.882,0.932] $\widehat{H_2}$ = 0.611; [0.587,0.635]
0.5/0.5	$\widehat{H}$ = 0.752; [0.732,0.784]	$\widehat{H_1}$ = 0.897; [0.870,0.924] $\widehat{H_2}$ = 0.621; [0.594,0.648]
0.4/0.6	$\widehat{H}$ = 0.706; [0.684,0.728]	$\widehat{H_1}$ = 0.891; [0.866,0.916] $\widehat{H_2}$ = 0.609; [0.586,0.632]
0.3/0.7	$\widehat{H}$ = 0.674; [0.652,0.692]	$\widehat{H_1}$ = 0.887; [0.861,0.913] $\widehat{H_2}$ = 0.597; [0.574,0.620]
0.2/0.8	$\widehat{H}$ = 0.660; [0.639,0.681]	$\widehat{H_1}$ = 0.662; [0.636,0.714]
0.1/0.9	$\widehat{H}$ = 0.631; [0.610,0.652]	$\widehat{H_1}$ = 0.629; [0.604,0.654]

В результате анализа проведенного эксперимента можно сделать следующие выводы. Если компоненты процесса вносят равнозначный вклад, мы получаем верные оценки для параметров обоих компонент. Тоже самое справедливо и в случае, когда вес одной компоненты незначительно превышает вес другой. Но, начиная с определенного момента, в случае, когда вес одной компоненты значительно превышает вес другой, параметры компоненты с меньшим весом данным методом не удается оценить с достаточной точностью. На основании результатов численного моделирования выявлено, что граничным значением веса второй компоненты, когда алгоритм перестает работать, является 0.24.

На основе описанных численных экспериментов можно предложить следующий алгоритм обнаружения неоднородности телетрафика:

- Применяя дискретное вейвлет-преобразование к исследуемому процессу $X(t)$ получаем вейвлет-коэффициенты $d_{j,k}$.
- Используя метод, предложенный P.Abry и D.Veitch, рассчитываем оценку параметра Херста H.
- Используя алгоритм оценки параметров неоднородного трафика, описанный во втором разделе, рассчитываем оценки параметров Херста H_1 и H_2.
- Сравниваем полученные оценки:
  - если $H >= H_1$, то исследуемый процесс является однородным, но возможно присутствие второй компоненты с весом менее 24% от общей величины;
  - если $H < H_1$, то исследуемый процесс является неоднородным, причем

вес второй компоненты больше 24% от общей величины;

Заключение

В настоящей статье рассмотрен алгоритм обнаружения неоднородности телетрафика. Описан выбор материнского вейвлета для расчета вейвлет-преобразования в алгоритме. Проведен анализ весов компонент неоднородного процесса. Показано, что начиная с определенного момента, когда вес одной компоненты значительно превышает вес другой компоненты, данный алгоритм позволяет вычислить только параметры компоненты с большим весом.

Список литературы

[1] Галактионова О.В., Хохлов Ю.С. Модель телетрафика, объединяющая устойчивое движение Леви и дробное броуновское движение // Вестник Тверского государственного университета. Серия: Прикладная математика. — Вып. 3. — 2006. — С. 163-167.

[2] Галактионова О.В., Хохлов Ю.С. Оценка параметров неоднородного трафика // Вестник Тверского государственного университета. Серия: Прикладная математика. — Вып. 18. — 2010. — С. 87-102.

[3] Abry P. Ondelettes et Turbulence-Multiresolutions, Algorithmes de Decomposition, Invariance D'Echelle et Signaux de Pression // Paris, France: Diderot Editeur. — 1994. — Pp. 289.

[4] Abry P., Veitch D. Wavelet Analysis of Long Range Dependent Traffic // IEEE Transactions on Information Theory. — Vol. 44, № 1. — January 1998.

[5] Cohen A., Daubechies I., Vial P. Wavelet on the interval and fast wavelet transforms // Appl. Computat. Harmonic Anal. — 1993. — vol. 1, no. 1. — Pp. 54-81.

[6] Leland W.E. Taqqu M.S., Willinger W., Wilson D.V. On the self-similar nature of Ethernet traffic // Computer Communications Review. — 1993. — № 23. — P. 183-193. // Proceedings of the ACM/SIGCOMM93. — San Francisco, September 1993. — Reprinted in Trends in Networking - Internet, the conference book of the Spring 1995 Conference of the National Unix User Group of the Netherlands (NLUUG). // Also reprinted in Computer Communication Review. — 1995. — 25, № 1. — P. 202-212.

[7] Leland W.E. Taqqu M.S., Willinger W., Willson D.V. Statistical analysis and stochastic modeling of self-similar data traffic // In J. Labetoulle and J. W. Roberts, editors, The Funda mental Role of Teletraffic in the Evolution of Telecommunications Networks. — Amsterdam, 1994. — P. 319-328.