

МЕТОД НАИМЕНЬШИХ МОДУЛЕЙ В ЗАДАЧЕ ФИЛЬТРАЦИИ ВРЕМЕННЫХ РЯДОВ

Н.Ю. Жбанова

Липецкий государственный технический университет
Россия, 398600, Липецк, Московская ул., 30
E-mail: zbanoid@gmail.com

С.А. Жбанов

Липецкий государственный технический университет
Россия, 398600, Липецк, Московская ул., 30
E-mail: pm.lisky@gmail.com

Ключевые слова: временные ряды, трендовая фильтрация, метод наименьших модулей

Аннотация: В статье рассматриваются перспективные алгоритмы трендовой фильтрации временных рядов, позволяющие создавать эффективные комплексы программ для расчета и оптимизации структурно сложных стохастических систем с очередями, в том числе многоканальных систем с самоподобным входящим потоком требований и не экспоненциальных сетей массового обслуживания.

1. Введение

В настоящее время для изучения свойств сложных систем используется подход, основанный на анализе сигналов, произведенных системой. Это актуально в тех случаях, когда математически описать изучаемый процесс невозможно, но в распоряжении исследователя имеется некоторая характерная наблюдаемая величина. Поэтому анализ систем, особенно при экспериментальных исследованиях, часто реализуется посредством обработки регистрируемых сигналов.

Как и большинство других видов анализа, анализ временных рядов предполагает, что данные содержат систематическую составляющую (обычно включающую несколько компонент) и случайную составляющую, которая затрудняет обнаружение регулярных компонент. На сегодняшний день известно множество способов восстановления систематической составляющей, в частности, методы параметрической аппроксимации, методы непараметрического оценивания, авторегрессии и другие. В последнее время среди исследователей наблюдается повышение интереса к методам трендовой фильтрации временных рядов, в основе которых лежит метод наименьших квадратов (МНК).

Одним из наиболее известных, простых и часто используемых алгоритмов фильтрации считается фильтр Ходрика-Прескотта (*HP*). Он используется для сглаживания временных рядов, выделения сезонных колебаний и трендов [1]. Впервые *HP*-фильтр был использован для анализа бизнес-циклов послевоенной экономики США, и активно применяется до сих пор в анализе экономических и технологических процессов и систем. Альтернативой *HP*-фильтру является фильтр, основанный на применении нормы l_1 . Подход был разработан в 2009 году С. Бойдом и соавторами [2] как модификация

фильтра Ходрика-Прескотта. Различные аспекты применения данного подхода при решении прикладных задач моделирования и управления можно найти, например, в [3-5].

Однако, при анализе ряда стохастических систем с очередями [6], в частности, информационно-вычислительных, телекоммуникационных и транспортных сетей во входящем потоке требований часто обнаруживается долгосрочная зависимость или самоподобие, поэтому его моделирование с помощью марковского процесса, который приводит к недооценке нагрузки, зачастую неоправданно. С учетом того, что самоподобные процессы более точно характеризуют поведение сетевого трафика, чем марковские, важной задачей представляется разработка и развитие существующего математического обеспечения для фильтрации входящего трафика с самоподобием.

Перспективным направлением для решения данной задачи является разработка алгоритмов трендовой фильтрации с различными метриками и их сравнительный анализ. В данной работе предлагается модификация метода Бойда, основанная на применении метода наименьших модулей (МНМ) [7]. В серии экспериментов показано преимущество МНМ-фильтра для анализа временных рядов с ярко выраженным самоподобием.

2. Фильтрация самоподобных временных рядов

2.1. Самоподобные процессы

Наличие во временных рядах свойства самоподобия, которое характерно для современных телекоммуникационных сетей связи, означает, что число требований, поступающих в систему в текущий интервал времени, может зависеть от количества требований, поступивших в предшествующие (и даже значительно отдаленные) временные интервалы. Важнейшим параметром, характеризующим такую долговременную память в случайном потоке требований, является показатель Хёрста H . В процессе наблюдения самоподобие (или персистентность) при $0,5 < H < 1$, когда увеличение нагрузки на систему в прошлом наиболее вероятно приведёт к ещё большей нагрузке в будущем, и наоборот. При $H = 0,5$ процесс становится марковским, а при $H < 0,5$ – антиперсистентным, т. е. нарушающим тенденцию предыдущих наблюдений. Одними из простейших самоподобных процессов являются фрактальное броуновское движение и фрактальный гауссовский шум.

Анализ долгосрочной статистики нагрузки на систему, решение задачи фильтрации данных и оценка показателя Хёрста являются важными предварительными этапами для синтеза самоподобного входящего трафика.

2.2. МНМ в задаче фильтрации

Рассмотрим входящий поток z_t как сумму систематической составляющей x_t и случайной составляющей v_t . Задачей трендовой фильтрации является нахождение по наблюдаемым значениям z_t трендовой составляющей x_t , которая, с одной стороны, не сильно отличается от исходной последовательности, а, с другой, должна обладать заранее известными, поддающимся математическому описанию, свойствами. Данной постановке задачи удовлетворяет следующий критерий

$$(1) \quad \min_x \|z - x\|_p^p + \mu \|Dx\|_l^l,$$

где $p \in \mathbf{N}$ и $l \in \mathbf{N}$ – нормы векторов, μ – параметр регуляризации, подбираемый экспериментально, а матрица D имеет вид:

$$D = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & 1 & -2 & 1 & 0 \\ 0 & \dots & 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

В критерии (1) возможно использование различных норм. Например, в случае евклидовой нормы $p = l = 2$ для x_t имеем сглаженную оценку Ходрика-Прескотта. Если же выбрать $p = 2$, $l = 1$, то для x_t уже строится кусочно-линейная оценка Бойда, которая уже относится к задаче выпуклого программирования, имеющей более высокую вычислительную сложность.

Фильтры Ходрика-Прескотта и Бойда хорошо себя зарекомендовали для восстановления систематической составляющей x_t марковских процессов [4], однако, для самоподобных процессов перспективным является рассмотрение аналога критерия (1), в котором $p = l = 1$ и используется метод наименьших модулей:

$$(2) \quad \min_x \|z - x\|_1 + \mu \|Dx\|_1.$$

2.3. Вычислительный эксперимент

Для решения задачи МНМ-фильтрации (2) специальные алгоритмы не разрабатывались и были применены стандартные методы оптимизации пакета Scilab. Была проведена серия экспериментов, целью которой являлся сравнительный анализ результатов МНМ-подхода и алгоритма l_1 , а также алгоритма Ходрика-Прескотта, на исходных данных с различными коэффициентами самоподобия. Во-первых, в качестве исходного ряда использовались искусственно сгенерированные данные с показателями Херста $0,6 \leq H \leq 0,9$, которые можно отнести к самоподобным процессам (рис. 1). Данные генерировались путем наложения фрактального гауссовского шума на кусочно-линейную функцию z_t , которая считалась истинным сигналом в рамках данного исследования. Для самоподобных процессов было проведено 40 опытов, для марковского процесса (использовался стандартный белый шум) – 15. В качестве меры расхождения для каждого опыта была вычислена абсолютная ошибка E .

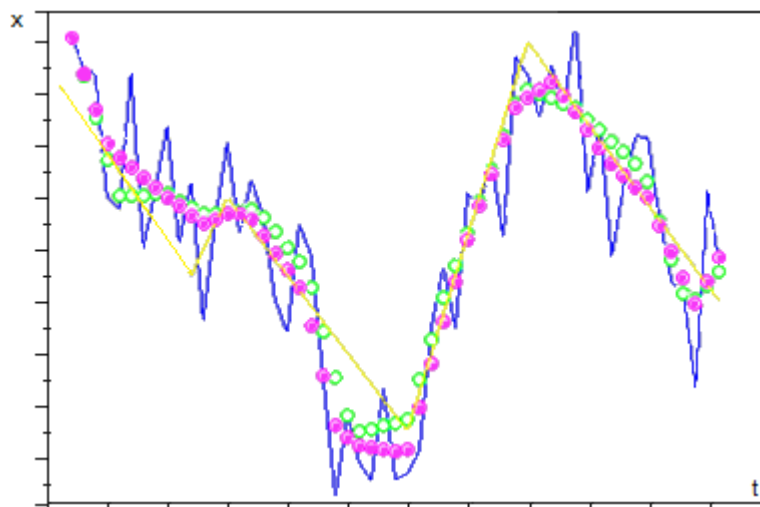


Рис. 1. Пример фильтрации самоподобного процесса ($H = 0,7$) с использованием МНМ-фильтра (зеленые маркеры) и l_1 -фильтра (фиолетовые маркеры).

Как и ожидалось, для марковского процесса l_1 -фильтр и фильтр Ходрика-Прескотта показали более высокие результаты, чем МНМ-фильтр. Средняя ошибка l_1 -фильтра по пятнадцати опытам составила 36,1, фильтра Ходрика-Прескотта – 42,5, МНМ-фильтра – 46,1. Параметр фильтрации μ подбирался вручную до достижения минимальной относительной ошибки.

Для процессов с самоподобием лучшие результаты показал МНМ-фильтр (хотя в ряде опытов у МНМ и l_1 -фильтра были близкие результаты, и в четырех опытах лучшие результаты были у фильтра l_1). Средняя ошибка l_1 -фильтра по сорока опытам составила 54,1, МНМ-фильтра – 46,9.

3. Заключение

На основании проведенных экспериментов можно сделать вывод, что предложенный алгоритм МНМ-фильтрации позволяет выявлять систематическую составляющую x_t в сильно зашумленных входящих данных с самоподобием, что значительно облегчает моделирование систем, характеризующихся самоподобием во входящих потоках. В дальнейшем планируется скомбинировать МНМ-фильтр и l_1 -фильтр с целью объединения их преимуществ.

Список литературы

1. Hodrick R.J., Prescott E.C. Postwar U.S. business cycles: An empirical investigation // Money, Credit, and Banking. 1997. Vol. 29, No. 1. P. 1-16.
2. Kim S.-J., Koh K., Boyd S., Gorinevsky D. l_1 Trend Filtering // Society for Industrial and Applied Mathematics. 2009. Vol. 51, No. 2. P. 339-360.
3. Поляк Б.Т. Методы l_1 -оптимизации в управлении и фильтрации // Материалы пленарных докладов III мультиконференции по проблемам управления. СПб.: ОАО «Концерн ЦНИИ «Электроприбор», 2010 г.
4. Стрельцов Г.А. Применение метода l_1 оптимизации к задачам выделения гармонических сигналов из шума // Труды 35-ой конференции «Информационные технологии и системы - 2012» ИТиС '12. Петрозаводск, 19-25 августа 2012 г. URL: <http://itas2012.iitp.ru/pdf/1569602847.pdf>
5. Сараев П.В., Блюмин С. Л., Галкин А.В. Некоторые современные подходы к моделированию и автоматизации сложных систем // Вести высших учебных заведений Черноземья. 2017. № 1 (47). С. 53-66.
6. Жбанов С. А. Разработка алгоритмов идемпотентной фильтрации временных рядов // Материалы XII Международной научно-практической конференции «Современные сложные системы управления». Т. 1. Липецк, 25-27 октября 2017 г. Липецк: Липецкий государственный технический университет, 2012. С. 83-85.
7. Мудров В.И., Кушко В. Л. Метод наименьших модулей. М.: «Знание». 1971. 64 с.