О ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ АЛГОРИТМА «ГУСЕНИЦА» К ЗАДАЧЕ КОНТРОЛЯ СЕТЕВОГО ТРАФИКА

Д.С. Безрукавный
Московский государственный университет леса, г. Мытищи
Тел.: (495) 674-41-21, e-mail: d3dimonbe@rambler.ru

В задаче мониторинга сетевого оборудования одним из важных аспектов является наблюдение сетевого трафика с целью выявления возможных перегрузок каналов и определения сетевых атак. Информация о сетевом трафике имеет статистический характер и представляет собой временные последовательности. Эти последовательности исследуются методами статистического анализа. Большинство временных рядов загрузки каналов состоит из трёх компонент: трендовой, периодической и случайной. Основной интерес при исследовании представляют трендовая и периодическая компоненты. Поэтому для их выделения необходима предварительная фильтрация шумов – случайной компоненты. Предлагается применить для решения этой задачи алгоритм «Гусеница», [1] основанный на методах линейной алгебры. Его можно разбить на 4 этапа.

Первый этап – развертка одномерного ряда в многомерный. Выберем некоторое число $M<N$ (где N – длина анализируемого ряда), называемое *длиной гусеницы*, и представим первые M значений последовательности в качестве первой строки матрицы $\boldsymbol{X}$. В качестве второй строки матрицы берем значения последовательности с x_2 по x_{M+1}. Последней строкой с номером $k = N - M + 1$ будут последние M элементов последовательности.

На втором этапе производится анализ главных компонент (АГК). Сначала вычисляется матрица $\boldsymbol{V} = (1/k)\boldsymbol{X}^T\boldsymbol{X}$. Несмотря на то, что ее элементы не центрированы, мы будем называть ее ковариационной матрицей, иногда добавляя слово "нецентральная". Следующий шаг, как обычно в АГК, состоит в вычислении собственных чисел и собственных векторов матрицы $\boldsymbol{V}$, т.е. разложение ее $\boldsymbol{V} = \boldsymbol{P}\mathbf{L}\boldsymbol{P}^T$, где **L** - диагональная матрица, на диагонали которой стоят упорядоченные по убыванию собственные числа, а $\boldsymbol{P}$ - ортогональная матрица собственных векторов матрицы $\boldsymbol{V}$.

Если изучается выборка из случайной совокупности, то собственные числа матрицы $\boldsymbol{V}$ являются выборочными дисперсиями соответствующих главных компонент, а квадратные корни из них – выборочными стандартными отклонениями. Графическое представление собственных чисел и некоторых функций от них в АГК традиционно используется для выявления структуры исследуемой совокупности и отбора и интерпретации главных компонент.

Третий этап – отбор главных компонент. В силу свойств матрицы $\boldsymbol{P}$ мы можем представить матрицу ряда $\boldsymbol{X}$ как $\boldsymbol{X} = \boldsymbol{Y}\,\boldsymbol{P}^T$. Таким образом, мы получаем разложение матрицы ряда по ортогональным составляющим (главным компонентам).

В то же время преобразование $y_j = \boldsymbol{X}\,p_j$ является линейным преобразованием исходного процесса с помощью дискретного оператора свертки, т.е.

$$y_j[l] = \mathrm{S}^{M}_{q=1} x_{lq}\,p_{jq} = \mathrm{S}^{M}_{q=1} x_{l+q-1}\,p_{jq}.$$

Таким образом, процедура "Гусеница" порождает набор линейных фильтров, настроенных на составляющие исходного процесса. При этом собственные векторы матрицы $\boldsymbol{V}$ выступают в роли переходных функций соответствующих фильтров.

Четвёртый этап – восстановление одномерного ряда. Эта процедура основана на разложении $\boldsymbol{X} = \boldsymbol{Y}\,\boldsymbol{P}^T$. Будем говорить, что восстановление проводится по данному набору главным компонентам, если при применении формулы восстановления $\boldsymbol{X} = \boldsymbol{Y}^*\,\boldsymbol{P}^T$ матрица $\boldsymbol{Y}^*$ получена из матрицы $\boldsymbol{Y}$ обнулением всех не входящих в набор главных компонент. Таким образом, мы можем получить интересующее нас приближение матрицы ряда, очищенное от случайной шумовой компоненты.

Разработана программа «Анализатор трафика 2.0», созданная на основании полученных результатов. Она служит для облегчения задачи сетевого администрирования с целью статистического анализа и мониторинга телекоммуникационной системы (сети). В функции этой программы входят:

- непрерывный мониторинг трафика сети;
- сравнение текущих параметров с нормальными;
- выдача системному администратору предупреждений и рекомендаций, в случае возникновения отклонений.

Следует отметить, что применение этого алгоритма требует в общем случае стационарности исходного ряда. Поэтому его использование следует проводить с осторожностью.

Работа выполняется при поддержке гранта РФФИ № 05-07-90360.

Литература

1. Данилов Д.Л., Жиглявский А.А. Главные компоненты временных рядов: метод "Гусеница". Санкт-Петербургский университет, 1997.

2. Кендалл М. Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976.