

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ СЕТЕВОГО ТРАФИКА

Запорожец Д.Б., Скулиш М.А.

Институт телекоммуникационных систем

КПИ им. Игоря Сикорского, Украина

E-mail: zaporozhetsdb@gmail.com

Short-term Forecasting: Methods to Predict Network Traffic

We analyzed the ability of simple linear forecasting methods to predict short-term real network traffic behavior. To evaluate such ability we experimentally estimate forecasting accuracy of some basic methods operating time series based on a real network traffic.

Прогнозирование стохастического процесса является одной из наиболее общих прикладных научных проблем, поэтому в этой области были предложены многочисленные подходы и методы. Простые и в то же время надежные методы прогнозирования также имеют решающее значение для различных решений для сетевого трафика. Сегодня существуют два основных подхода к управлению сетью [1]: офлайн и онлайн. Оба этих метода нуждаются в прогнозировании поведения трафика. Разницу можно найти в длительности интервалов прогнозирования: автономные методы основаны на долгосрочном прогнозировании (часы и дни), в то время как онлайн-методы, основанные на быстрых изменениях в сети, работают с краткосрочными прогнозами (с секунд до десятков минут). Новизна этого исследования заключается в исследовании методов прогнозирования с использованием реальных временных рядов интенсивности сетевого трафика.

Примерами ранних подходов являются модели FARIMA [2] (авторегрессионная дробно интегрированная скользящая средняя) для долгосрочного прогнозирования интенсивности сетевого трафика. Еще одной примерной моделью является ARIMA [3] (авторегрессионная интегрированная скользящая средняя). Третий общий подход - использование нейронных сетей для прогнозирования [3].

Методы прогнозирования демонстрируют высокую точность даже для дальнего прогнозируемого горизонта. Но в то же время у них есть некоторые серьезные недостатки, когда применяются для управления онлайн-трафиком. Сначала для любой модели ARIMA / FARIMA требуется анализ временного ряда для оценки адекватных параметров модели. Нейронные сети свободны от таких ограничений, но им требуется сравнительно длительное время для обучения, в то время как реальное поведение трафика может быстро измениться, и такое изменение приведет к новому циклу обучения. И, наконец, такие методы довольно сложно реализовать в реальном сетевом оборудовании из-за значительной вычислительной сложности, особенно в случае нейронных сетей.

Описанные выше долгосрочные методы прогнозирования являются излишними для краткосрочного прогнозирования. Основными требованиями здесь являются сравнительная простота, полная автоматизация процессов и способность быстро реагировать на произошедшие изменения. Линейные методы, отвечают всем этим требованиям.

В целом, наблюдаемые временные ряды могут быть представлены суммой нескольких ненаблюдаемых компонентов [4].

$$X(t) = \mu(t) + S(t) + e(t)$$

Здесь $\mu(t)$ - модель тренда, $S(t)$ - сезонная составляющая, $e(t)$ - стационарный остаток. В случае краткосрочного прогнозирования можем пренебречь сезонной частью и сосредоточиться на плавном тренде и остаточной центрированной случайной составляющей.

Далее будет рассмотрено некоторые классические методы линейного прогнозирования.

Полиномиальная аппроксимация. Метод, основанный на представлении сегмента временных рядов в форме заданного степенного полинома.

$$P_k(x_t) = a_0 + \sum_{i=1}^k a_i x_{t-i}^i$$

Здесь a_0, \dots, a_n - коэффициенты, x_{t-1}, \dots, x_{t-n} - значения сегмента временных рядов, k - полиномиальная мощность. Коэффициенты могут быть получены методом наименьших квадратов, сводя к минимуму кумулятивную квадратичную ошибку предсказания

Полиномиальная экстраполяция. Основная разница между полиномиальной экстраполяцией и аппроксимацией является необходимость точно соответствовать всем значениям сегмента. Следствием такого условия является жесткое ограничение для размера сегмента: оно должно быть равно $n = k + 1$ балл.

Сглаживание с помощью формул Спенсера. Другими известными методами получения тренда по наблюдаемым данным являются формулы Спенсера [6]. В этой работе предоставлено формулы на основе 5- и 7-балльных формул, которые относятся к взвешенному движущемуся усредненному процессу. Исходные формулы Спенсера были модифицированы для экстраполяции значения тренда за пределами рассматриваемого сегмента временных рядов.

$$x_t[5] = \frac{1}{35} [-28x_{t-5} + 77x_{t-4} - 28x_{t-3} - 98x_{t-2} + 112x_{t-1}]$$

$$x_t[7] = \frac{1}{21} [-12x_{t-7} + 18x_{t-6} + 12x_{t-5} - 9x_{t-4} - 24x_{t-3} - 12x_{t-2} + 48x_{t-1}]$$

Здесь $x_t[5]$ - прогноз на основе 5-точечной формулы и $x_t[7]$ - следовательно, по 7-точечной формуле. Практически формулы Спенсера представляют собой низкочастотные фильтры, отбрасывающие высокочастотный шум стационарного остатка.

Линейное предсказание. Метод авторегрессии, основанный на представлении случайного процесса как временной инвариантной линейной системы [5].

$$\hat{y}(n) = - \sum_{k=1}^p a_k y(n-k)$$

Здесь $\hat{y}(n)$ - предсказание n-го значения временных рядов. Весовые параметры a_1, \dots, a_k могут быть получены путем минимизации кумулятивной ошибки предсказания квадрата на участке длины n.

$$\sum_n [y(n) - \hat{y}(n)]^2 \rightarrow \min_{\{a_k\}}$$

Экспоненциальное сглаживание. Экспоненциальное сглаживание можно рассматривать как частный случай взвешенного среднего значения с учетом всех предыдущих значений ряда с экспоненциально убывающими коэффициентами. Существует множество экспоненциальных сглаживающих типов моделей.

Наиболее оптимальным методом автономного краткосрочного сетевого прогнозирования является экспоненциальное сглаживание. Он сочетает в себе универсальность и в то же время надежность, точность среди всех рассмотренных методов. Экспоненциальное сглаживание с параметром адаптивного сглаживания было признано лучшим методом прогнозирования.

Были оценены методы краткосрочного прогнозирования. Полученные результаты помогут исследователям и инженерам разрабатывать решения для управления сетью, особенно системы балансировки нагрузки, которые нуждаются в таких методах, выбирать наиболее подходящие. Статистический анализ используемых временных рядов интенсивности реального сетевого трафика можно рассматривать как дополнительный вклад в работу, раскрывая внутреннюю структуру процессов и лучше понимая присущие ему свойства и характеристики.

Литература

1. Awduche D. et al. Overview and principles of Internet traffic engineering. RFC 3272, may, 2002.
2. Shu Y. et al. Traffic prediction using FARIMA models In: Communications, 1999. ICC'99. 1999 IEEE International Conference on. IEEE, 1999. pp. 891-895.
3. Rutka G. Network Traffic Prediction using ARIMA and Neural Networks Models In: Electronics and Electrical Engineering. 2008. pp. 47-52.
4. Koopman S. J., Shephard N. State space and unobserved component models. Cambridge University Press, 2004.
5. Makhoul J. Linear prediction: A tutorial review In: Proceedings of the IEEE. 1975. pp. 561-580.
6. Kendall, M. G., Stuart A.: The advanced theory of statistics. Vols. II and III. (1961).