У роботі запропоновано аналітичний опис потоку даних у комп'ютерній мережі, отриманий по реальним статистичним даним. Аналітична модель може бути використана при аналізі ефективності комп'ютерної мережі

Ключові слова: комп'ютерні мережі, математичне моделювання, мережевий трафік, аналітична модель

В работе предлагается аналитическое описание потока данных в компьютерной сети, полученное на основе реальных статистических данных. Аналитическая модель может быть использована при анализе эффективности узлов компьютерной сети

Ключевые слова: компьютерные сети, математическое моделирование, сетевой трафик, аналитическая модель

It was suggested the analytic description of network flowing, based on real statistic data. Analytic model may be used in network hub effectiveness detection

Key words: networks, mathematical modeling, network traffic, analytic model

УДК 681.325

СТОХАСТИЧЕСКАЯ МОДЕЛЬ СЕТЕВОГО ТРАФИКА

П.Е. Пустовойтов
Кандидат технических наук, доцент
Кафедра «Системы информации»
Национальный технический университет «Харьковский политехнический институт»
ул. Фрунзе, 22, г. Харьков, Украина, 61002
Контактный тел.: 097-737-44-91
E-mail: pustovoitov@kpi.kharkov.ua

1. Введение, постановка задачи

По мере глобального роста и интенсивности использования Internet и корпоративных объединенных компьютерных сетей на передний план выходит ряд требований к качеству функционирования сетей, среди которых одним из самых важных является эффективная борьба с перегрузкой. При этом для удовлетворения этих новых требований недостаточно просто увеличить пропускную способность сети. Необходимы результативные методы управления трафиком и дисциплиной обслуживания потоков в узлах сети. Дело в том, что при возникновении перегрузки реальной становится опасность потери пакетов. Эти потерянные пакеты необходимо передавать повторно, что заметно увеличивает нагрузку на сеть и является причиной значительных задержек [1]. Более серьезный негативный феномен, называемый «глобальной синхронизацией» [2], состоит в следующем. Как правило, эффект переполнения очередей затрагивает несколько очередей одновременно. При этом в результате принятия мер для снятия перегрузки суммарный объем трафика резко падает и в течение определенного времени сеть используется не в полной мере. Это приводит к увеличению объемов трафика и возникает новый цикл «перегрузка-недогрузка».

При анализе возникающих здесь ситуаций для выбора мер противодействия перегрузкам необходимо иметь в виду, что поток сообщений для каждого узла сети представляет собой суперпозицию некоторого числа неоднородных потоков разной интенсивности и содержащих пакеты разной, в среднем, длины. Известные методы борьбы с перегрузкой [3] учитывают эти обстоятельства в недостаточной мере.

В связи с этим возникает необходимость рассмотрения методов борьбы с перегрузкой, основанных на управлении дисциплиной обслуживания с учетом различий в интенсивностях потоков, суммарное воздействие которых формирует перегрузку. Понятно, что рассмотрение вопросов управления дисциплиной обслуживания невозможно без предварительного описания потоков сообщений, поступающих на вход узла. Поставим задачу построения аналитической модели реального сетевого трафика (использованы статистические данные, полученные с прокси-сервера НТУ «ХПИ» за различные промежутки времени).

2. Основные результаты

Реальные входящие потоки пакетов в узле компьютерной сети являются нестационарными. Непосредственный анализ наблюдений за трафиком позволяет выявить наличие суточных, недельных и сезонных колебаний интенсивности потоков. При этом, для шести рабочих дней недели с двенадцатичасовой продолжительностью работы, практически в каждом потоке амплитуда сезонных колебаний заметно меньше амплитуды недельных и особенно суточных колебаний, которая является самой высокой. С учетом этого обстоятельства выберем интервал наблюдений меньшим четверти периода сезонных колебаний и займемся отысканием амплитуды суточных и недельных колебаний. Разобьем интервал наблюдений на совокупность подынтервалов длиной Δ (например, $\Delta = \frac{1}{2}$ часа). Зафиксируем количество пакетов, поступающих в пределах каждого подынтервала. Обработаем теперь совместно случайные значения числа пакетов, приходящих в одно и то же время суток и в одноименные дни недели (например, интервал с $10^{\underline{30}}$ до $11^{\underline{00}}$ вторника). Вычислим среднее значение и дисперсию числа пакетов, поступающих в каждый из подынтервалов. При этом получим совокупность средних значений $m_1, m_2, ..., m_n$ случайного числа пакетов и их дисперсии $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, $n = 24 \cdot 6 = 144$. Пусть b_0 - среднее число пакетов, поступающих в систему в течение рабочего дня, b_1 - амплитуда суточных колебаний числа пакетов, поступающих в систему, b_2 - амплитуда недельной модуляции амплитуды суточных колебаний числа поступающих пакетов.

Введем теперь модель, описывающую зависимость средней интенсивности входящего потока от времени

$$\lambda(t) = \left(b_0 + b_1 \sin\frac{2\pi}{24}t\right)\left(1 + b_2 \sin\frac{2\pi}{144}t\right) =$$
$$= b_0 + b_1 \sin\frac{2\pi}{24}t + b_2 b_0 \sin\frac{2\pi}{144}t + b_1 b_2 \sin\frac{2\pi}{24}t \sin\frac{2\pi}{144}t =$$
$$= a_0 + a_2 \sin\frac{2\pi}{24}t + a_2 \sin\frac{2\pi}{144}t + a_3 \sin\frac{2\pi}{24}t \sin\frac{2\pi}{144}t. \qquad (1)$$

Параметры a_0, a_1, a_2 модели (1) найдем, используя данные о реальных наблюдениях, по методу наименьших квадратов.

Введем функционал

$$J = \sum_{k=1}^{n} \left[m_k - \lambda(t_k)\right]^2 =$$
$$= \sum_{k=1}^{n} \left[m_k - \left(a_0 + a_2 \sin\frac{2\pi}{24}t_k + a_2 \sin\frac{2\pi}{144}t_k + a_3 \sin\frac{2\pi}{24}t_k \sin\frac{2\pi}{144}t_k\right)^2\right]. \qquad (2)$$

Адекватность модели может быть проверена по критерию Фишера [4] с учетом анализа автокорреляции остатков [5].

Запишем функционал (2) в матричной форме. Введем матрицу H и векторы A, M:

$$H = \begin{pmatrix} 1 & \sin\frac{2\pi}{24}t_1 & \sin\frac{2\pi}{144}t_1 & \sin\frac{2\pi}{24}t_1 \sin\frac{2\pi}{144}t_1 \\ 1 & \sin\frac{2\pi}{24}t_2 & \sin\frac{2\pi}{144}t_2 & \sin\frac{2\pi}{24}t_2 \sin\frac{2\pi}{144}t_2 \\ ... & ... & ... & ... \\ 1 & \sin\frac{2\pi}{24}t_n & \sin\frac{2\pi}{144}t_n & \sin\frac{2\pi}{24}t_n \sin\frac{2\pi}{144}t_n \end{pmatrix},$$

$$A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}, \quad M = \begin{pmatrix} m_1 \\ m_2 \\ ... \\ m_n \end{pmatrix}.$$

Тогда

$$J = (HA - M)^T (HA - M). \qquad (3)$$

Минимизируя (3) по вектору A, получим искомый вектор оценок уравнения регрессии

$$A = (H^T H)^{-1} H^T M. \qquad (4)$$

При этом, так как

$$H^T H = \begin{pmatrix} 1 & 1 & ... & 1 \\ \sin\frac{2\pi}{24}t_1 & \sin\frac{2\pi}{24}t_2 & ... & \sin\frac{2\pi}{24}t_2 \\ \sin\frac{2\pi}{144}t_1 & \sin\frac{2\pi}{144}t_2 & ... & \sin\frac{2\pi}{144}t_n \\ \sin\frac{2\pi}{24}t_1 \sin\frac{2\pi}{24 \cdot 14}t_1 & \sin\frac{2\pi}{24}t_2 \sin\frac{2\pi}{24 \cdot 14}t_2 & ... & \sin\frac{2\pi}{24}t_n \sin\frac{2\pi}{144}t_n \end{pmatrix} \times$$

$$\times \begin{pmatrix} 1 & \sin\frac{2\pi}{24}t_1 & \sin\frac{2\pi}{144}t_1 & \sin\frac{2\pi}{24}t_1 \sin\frac{2\pi}{144}t_1 \\ 1 & \sin\frac{2\pi}{24}t_2 & \sin\frac{2\pi}{144}t_2 & \sin\frac{2\pi}{24}t_2 \sin\frac{2\pi}{144}t_2 \\ ... & ... & ... & ... \\ 1 & \sin\frac{2\pi}{24}t_n & \sin\frac{2\pi}{144}t_n & \sin\frac{2\pi}{24}t_n \sin\frac{2\pi}{144}t_n \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum_{k=1}^{n} \sin\frac{2\pi}{24}t_k & \sum_{k=1}^{n} \sin\frac{2\pi}{144}t_k & \sum_{k=1}^{n} \sin\frac{2\pi}{24}t_k \sin\frac{2\pi}{144}t_k \\ \sum_{k=1}^{n} \sin\frac{2\pi}{24}t_k & \sum_{k=1}^{n} \sin^2\frac{2\pi}{24}t_k & \sum_{k=1}^{n} \sin\frac{2\pi}{24}t_k \sin\frac{2\pi}{144}t_k & \sum_{k=1}^{n} \sin^2\frac{2\pi}{24}t_k \sin\frac{2\pi}{144}t_k \\ \sum_{k=1}^{n} \sin\frac{2\pi}{144}t_k & \sum_{k=1}^{n} \sin\frac{2\pi}{24}t_k \sin\frac{2\pi}{144}t_k & \sum_{k=1}^{n} \sin^2\frac{2\pi}{144}t_k & \sum_{k=1}^{n} \sin\frac{2\pi}{24}t_k \sin^2\frac{2\pi}{144}t_k \\ \sum_{k=1}^{n} \sin\frac{2\pi}{24}t_k \sin\frac{2\pi}{144}t_k & \sum_{k=1}^{n} \sin^2\frac{2\pi}{24}t_k \sin\frac{2\pi}{144}t_k & \sum_{k=1}^{n} \sin\frac{2\pi}{24}t_k \sin^2\frac{2\pi}{144}t_k & \sum_{k=1}^{n} \sin^2\frac{2\pi}{24}t_k \sin^2\frac{2\pi}{144}t_k \end{pmatrix} =$$

$$= \begin{pmatrix} n & & & 0 \\ & \sum_{k=1}^{n} \sin^2\frac{2\pi}{24}t_k & & \\ & & \sum_{k=1}^{n} \sin^2\frac{2\pi}{144}t_k & \\ 0 & & & \sum_{k=1}^{n} \sin^2\frac{2\pi}{24}t_k \sin^2\frac{2\pi}{144}t_k \end{pmatrix},$$

то

$$\left(H^T H\right)^{-1} = \begin{pmatrix} n^{-1} & & & 0 \\ & \left(\sum_{k=1}^{n} \sin^2 \frac{2\pi}{24} t_k\right)^{-1} & & \\ & & \left(\sum_{k=1}^{n} \sin^2 \frac{2\pi}{144} t_k\right)^{-1} & \\ 0 & & & \left(\sum_{k=1}^{n} \sin^2 \frac{2\pi}{24} t_k \sin^2 \frac{2\pi}{144} t_k\right)^{-1} \end{pmatrix} =$$

$$= \begin{pmatrix} \eta_{11} & & & 0 \\ & \eta_{22} & & \\ & & \eta_{33} & \\ 0 & & & \eta_{44} \end{pmatrix}.$$

Тогда

$$A = \begin{pmatrix} \eta_{11} & & & 0 \\ & \eta_{22} & & \\ & & \eta_{33} & \\ 0 & & & \eta_{44} \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ \sin\frac{2\pi}{24} t_1 & \sin\frac{2\pi}{24} t_2 & \dots & \sin\frac{2\pi}{24} t_n \\ \sin\frac{2\pi}{144} t_1 & \sin\frac{2\pi}{144} t_2 & \dots & \sin\frac{2\pi}{144} t_n \\ \sin\frac{2\pi}{24} t_1 \sin\frac{2\pi}{24\cdot 14} t_1 & \sin\frac{2\pi}{24} t_2 \sin\frac{2\pi}{24\cdot 14} t_2 & \dots & \sin\frac{2\pi}{24} t_n \sin\frac{2\pi}{144} t_n \end{pmatrix} \times$$

$$\times \begin{pmatrix} m_1 \\ m_2 \\ \dots \\ m_4 \end{pmatrix} = \begin{pmatrix} \eta_{11} & & & 0 \\ & \eta_{22} & & \\ & & \eta_{33} & \\ 0 & & & \eta_{44} \end{pmatrix} \begin{pmatrix} \sum_{k=1}^{n} m_k \\ \sum_{k=1}^{n} m_k \sin\frac{2\pi}{24} t_n \\ \sum_{k=1}^{n} m_k \sin\frac{2\pi}{144} t_n \\ \sum_{k=1}^{n} m_k \sin\frac{2\pi}{24} t_n \sin\frac{2\pi}{144} t_n \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{1}{n}\sum_{k=1}^{n} m_k \\ \frac{1}{\sum_{k=1}^{n} \sin^2 \frac{2\pi}{24} t_n} \sum_{k=1}^{n} m_k \sin\frac{2\pi}{24} t_n \\ \frac{1}{\sum_{k=1}^{n} \sin^2 \frac{2\pi}{144} t_n} \sum_{k=1}^{n} m_k \sin\frac{2\pi}{144} t_n \\ \frac{1}{\sum_{k=1}^{n} \sin^2 \frac{2\pi}{24} t_n \sin^2 \frac{2\pi}{144} t_n} \sum_{k=1}^{n} m_k \sin\frac{2\pi}{24} t_n \sin\frac{2\pi}{144} t_n \end{pmatrix} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}. \quad (5)$$

В результате проведения расчетов по формуле (5) с учетом реальных значений m_k, $k = 1,2,\dots,n$, получим

$$A = \begin{pmatrix} 3.16\cdot 10^5 \\ 1.28\cdot 10^5 \\ 0.46\cdot 10^5 \\ 0.21\cdot 10^5 \end{pmatrix}.$$

Проведем статистический анализ результатов обработки наблюдений. Эта процедура содержит несколько этапов.

Проверка однородности дисперсий. Так как в каждой точке факторного пространства (ФП) возможных значений переменной t_k, $k = 1,2,\dots,m$, проводилась серия наблюдений, содержащая 6 повторных опытов, то результаты экспериментов имеют вид:

$$\% = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{61} & x_{62} & \dots & x_{6m} \end{pmatrix}.$$

Здесь

x_{ik} - случайное значение наблюдаемой переменой в k-й точке проведения эксперимента, полученное в i-ой неделе, $i = 1,2,\dots,6$, $k = 1,2,\dots,m$.

Для каждого столбца матрицы X проведем усреднение результатов повторных опытов для наблюдаемой переменной и рассчитаем дисперсии этих результатов

$$m_k = \frac{1}{q}\sum_{i=1}^{6} x_{ik}, \quad \sigma_k^2 = \frac{1}{q-1}\sum_{i=1}^{6}(x_{ik} - M_k)^2,$$

$$k = 1,2,\dots,m, \quad q = 6.$$

Проверка однородности дисперсий выполняется по критерию Фишера. С этой целью найдем

$$\sigma_{max}^2 = \max_k\{\sigma_k^2\}, \quad \sigma_{min}^2 = \left\{\min_k \sigma_k^2\right\}$$

и вычислим

$$F_p = \frac{\sigma_{max}^2}{\sigma_{min}^2}.$$

Полученное значение F_p сравним с критическим F_{kp}, извлеченным из таблицы распределения Фишера для заданного уровня значимости $\alpha = 0.05$ и числа степеней свободы ν_1 и ν_2 числителя и знаменателя, равных числу повторных опытов минус число оцениваемых параметров (в данном случае $\nu_1 = \nu_1 = q - 1$).

При этом, если $F_p < F_{kp}$, то гипотеза об однородности принимается, в противном случае её следует отклонить и принять меры для улучшения однородности (например, провести дополнительные опыты в точке с максимальной дисперсией).

В рассматриваемой задаче устранение неоднородности дисперсий путем проведения дополнительных экспериментов невозможно. При этом для оценивания параметров уравнений регрессии в случае неоднородности дисперсий использование МНК в форме (4) некорректно. С целью учета различий в дисперсиях опытов введем дисперсионную матрицу

$$D = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & \ddots & \\ & & & \\ 0 & & & \sigma_n^2 \end{pmatrix},$$

с использованием которой функционал наименьших квадратов (3) примет вид

$$J = (HA - M)^T D^{-1}(HA - M).$$

Минимизация этого функционала по вектору параметров модели A приведет к соотношению

$$\hat{A} = (H^T D^{-1} H)^{-1} H^T D^{-1} M. \quad (6)$$

Если гипотеза об однородности принимается, то дисперсии опытов усредняют по формуле

$$s_0^2 = \frac{1}{n}\sum_{k=1}^{n}\sigma_k^2 \quad (7)$$

и, таким образом, получают дисперсию воспроизводимости эксперимента. При этом $f_0 = n(q-1)$ - число степеней свободы.

Итак, если в ходе расчета дисперсии воспроизводимости эксперимента установлена однородность дисперсий в разных точках ФП, то, используя соотношения (4), осуществляется вычисление коэффициентов уравнения регрессии (1). Если же выявлена неоднородность дисперсий, то для оценки компонентов вектора A необходимо использовать формулу (6).

В рассматриваемой задаче имеем $\sigma_{\max}^2 = 4.7 \cdot 10^5$, $\sigma_{\min}^2 = 3.1 \cdot 10^5$. При этом $F_p = 1{,}516$. В соответствии с таблицей критических точек распределения Стьюдента для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $q - 1 = 5$ получим $F_{кр} = 2{,}57$. Поскольку $F_p < F_{кр}$ считаем, что проведенный эксперимент однороден. После усреднения дисперсий получим дисперсию воспроизводимости $s_0^2 = 3.74 \cdot 10^5$.

Затем, после вычисления коэффициентов уравнения регрессии переходят непосредственно к статистическому анализу полученных результатов, который проводится в два этапа: 1) оценка значимости коэффициентов уравнения регрессии, 2) оценка адекватности модели. Перейдем к их рассмотрению.

Оценка значимости коэффициентов уравнения регрессии. Вначале рассчитывается дисперсия ошибок оценок коэффициентов регрессии. С этой целью вычислим ковариационную матрицу ошибок оценок параметров уравнения регрессии по формуле:

$$\Psi = s_0^2 \left(H^T H\right)^{-1} =$$

$$= s_0^2 \begin{pmatrix} \sigma_{a_0}^2 & \mathrm{cov}(a_0, a_1) & \ldots & \mathrm{cov}(a_0, a_m) \\ \mathrm{cov}(a_1, a_0) & \sigma_{a_1}^2 & \ldots & \mathrm{cov}(a_1, a_m) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{cov}(a_m, a_0) & \mathrm{cov}(a_m, a_1) & \ldots & \sigma_{a_m}^2 \end{pmatrix}.$$

В рассматриваемой задаче

$$(H^T H)^{-1} = \begin{pmatrix} \eta_{11} & & & 0 \\ & \eta_{22} & & \\ & & \eta_{33} & \\ 0 & & & \eta_{44} \end{pmatrix}.$$

Поэтому

$$\Psi = \begin{pmatrix} s_0^2\eta_{11} & & & 0 \\ & s_0^2\eta_{22} & & \\ & & s_0^2\eta_{33} & \\ 0 & & & s_0^2\eta_{44} \end{pmatrix}.$$

Пусть

a_i -истинное значение i -го коэффициента уравнения регрессии,

$\hat{a}_i$ - оценка этого значения по результатам эксперимента.

Тогда случайная величина

$$T_i = \frac{\left|\hat{a}_i - a_i\right|}{\sigma_{a_i}} \quad (8)$$

распределена по закону Стьюдента со $\sum_{j=1}^{n}(q_j - 1)$ степенями свободы. Построим теперь доверительный интервал для a_i. Пусть ε_{a_i} - половина этого интервала. Величину ε_{a_i} выберем так, чтобы, $P\left(\left|\hat{a}_i - a_i\right| < \varepsilon_{a_i}\right) = \gamma$, где γ - доверительная вероятность. Перейдем в левой части этого равенства к случайной величине T_i:

$$P\left(\frac{\left|\hat{a}_i - a_i\right|}{\sigma_{a_i}} < \frac{\varepsilon_{a_i}}{\sigma_{a_i}}\right) = \gamma$$

или

$$P\left(T_i < \frac{\varepsilon_{a_i}}{\sigma_{a_i}}\right) = \gamma.$$

Используя таблицы T-распределения, найдем такое число t_{a_i}, что

$$P\left(T_i < t_{a_i}\right) = \gamma.$$

Тогда

$$\frac{\varepsilon_{a_i}}{\sigma_{a_i}} = t_{a_i}.$$

При этом $\varepsilon_{a_i} = \sigma_{a_i} t_{a_i}$ и

$$\left|\hat{a}_i - a_i\right| < \sigma_{a_i} t_{a_i}$$

или

$$\hat{a}_i - \sigma_{a_i} t_{a_i} < a_i < \hat{a}_i + \sigma_{a_i} t_{a_i}. \quad (9)$$

Таким образом, найден доверительный интервал $\left[\hat{a}_i - \sigma_{a_i} t_{a_i}, \hat{a}_i + \sigma_{a_i} t_{a_i}\right]$, накрывающий истинное значение коэффициента a_i с вероятностью не ниже γ. Далее, считают, что коэффициент a_i значим с надежностью γ, если соответствующий доверительный интервал не накрывает нуль. В противном случае этот коэффициент следует признать незначимым и приравнять к нулю. Соответствующий фактор должен быть из уравнения регрессии исключен.

В рассматриваемой задаче имеем

$$\left(H^T H\right)^{-1} = \begin{pmatrix} 6{,}9\cdot 10^{-3} & & & 0 \\ & 2{,}1\cdot 10^{-3} & & \\ & & 1{,}4\cdot 10^{-3} & \\ 0 & & & 0{,}72\cdot 10^{-3} \end{pmatrix}.$$

Тогда

$$\Psi = \begin{pmatrix} 2{,}59\cdot 10^{3} & & & 0 \\ & 7{,}88\cdot 10^{2} & & \\ & & 5{,}25\cdot 10^{2} & \\ 0 & & & 2{,}62\cdot 10^{2} \end{pmatrix}.$$

Так как критическое значение $t_a = 1.98$ (при доверительной вероятности $\gamma = 0.95$), то

$$\varepsilon_{a_0} = \sigma_{a_0} \cdot t_a = 100.8,$$
$$\varepsilon_{a_1} = \sigma_{a_1} \cdot t_a = 55.59,$$
$$\varepsilon_{a_2} = \sigma_{a_2} \cdot t_a = 45.37,$$
$$\varepsilon_{a_3} = \sigma_{a_3} \cdot t_a = 32.05.$$

При этом во всех случаях доверительный интервал не накрывает нуль и, следовательно, все коэффициенты уравнения регрессии (1) значимы.

Проверка адекватности уравнения регрессии. Для проверки адекватности уравнения регрессии, полученного после отбрасывания незначимых факторов, необходимо кроме дисперсии воспроизводимости рассчитать дисперсию адекватности, характеризующую степень отклонения линии регрессии от значения функции отклика в точках ФП, соответствующих проведенным экспериментам. Дисперсия адекватности вычисляется по формуле

$$s_{ад}^2 = \frac{1}{n-\ell}\sum_{j=1}^{n}\left(\bar{y}_j - \sum_{i=0}^{\ell-1}\hat{a}_i x_{ji}\right)^2 = \frac{1}{n-\ell}\sum_{j=1}^{n}\left(m_k - \lambda(t_k)\right)^2. \quad (10)$$

Для расчета $s_{ад}^2$ используются усредненные по результатам повторных опытов значения функции отклика $\bar{y}_j$, $j = 1,\dots,n$. Поэтому дисперсия адекватности не зависит от дисперсии ошибки измерений. В соотношении (10) ℓ - число членов в уравнении регрессии, оставшихся после оценки значимости. Таким образом, дисперсия адекватности характеризует остаточную сумму квадратов отклонении результатов реальных экспериментов от построенного уравнения регрессии, приходящуюся на один свободный опыт (то есть на одну степень свободы).

Теперь адекватность полученного уравнения регрессии проверяют по критерию Фишера, сравнивая отношение

$$F_p = \frac{s_{ад}^2}{s_0^2}$$

с критическим значением

$$F_p < F_{кр} \quad (11)$$

для числа степеней свободы $f_{ад} = n - \ell$ и $f_0 = \sum_{j=1}^{n}(q_j - 1)$ и заданного уровня значимости α .

Если неравенство (11) выполняется, то уравнение регрессии считается адекватным, в противном случае - нет.

Принятие гипотезы об адекватности эквивалентно принятию гипотезы о равенстве дисперсий адекватности и воспроизводимости. Адекватность уравнения регрессии означает, что рассеяние экспериментальных данных относительно уравнения регрессии имеет тот же порядок, что и рассеяние, связанное с ошибками опыта. При этом отклонение экспериментальных точек относительно построенного уравнения регрессии объясняется именно этими ошибками, а не ошибочной гипотезой о структуре модели. В рассматриваемой задаче непосредственный подсчет дисперсии адекватности по формуле (10) дал следующий результат $\sigma_{ад}^2 = 4{,}17\cdot 10^5$. . При этом расчетное значение критерия Фишера

$$F_p = \frac{4.17\cdot 10^5}{3.75\cdot 10^5} = 1{,}083$$

Поскольку критическое значение критерия Фишера для $\alpha = 0{,}05$ равно $F_{кр} = 2{,}38,$ то

$$F_p < F_{кр}$$

и, следовательно, модель адекватна.

Анализ автокорреляции остатков. Обнаружение автокорреляции остатков в отклонениях от тренда осуществляется с использованием критерия Дарбина-Уотсона, вычисляемого по формуле

$$d = \frac{\sum_{j=1}^{n}(\xi_{j+1} - \xi_1)^2}{\sum_{j=1}^{n}\xi_j^{\ 2}}, \quad (12)$$

где

ξ_j - отклонение от тренда в j -м наблюдении.

Распределение значений критерия Дарбина-Уотсона протабулировано.

Вычисленное значение критерия d сравнивается с табличными d_1, d_2. Возможны следующие случаи:

- если $d < d_1$, гипотеза об отсутствии автокорреляции отвергается (имеется положительная корреляция);
- если $d_2 < d < 4 - d_2$, гипотеза об отсутствии автокорреляции принимается;
- если $d_1 \le d \le d_2$ или $4 - d_2 \le d \le 4 - d_1$, то необходимо дальнейшее исследование (например, по большему числу наблюдений);
- если $d > 4 - d_1$, то гипотеза об отсутствии корреляции отвергается (имеется отрицательная корреляция).

Соотношение (12) непосредственно следует из отношения, используемого для выявления наличия автокорреляции случайного процесса $\xi(t)$. Это соотношение имеет вид

$$r(t) = \frac{M[(\xi(t+1) - m(t+1)) - (\xi(t) - m(t))]^2}{M[(\xi(t) - m(t))^2]}. \quad (13)$$

Перепишем (13) следующим образом

$$r(t)=\frac{M[(\xi(t)-m(t))^2]+M[(\xi(t+1)-m(t+1))^2]}{M[(\xi(t)-m(t))^2]}-$$

$$-\frac{2M[(\xi(t)-m(t))(\xi(t+1)-m(t+1))]}{M[(\xi(t)-m(t))^2]}=$$

$$=\frac{D(t+1)+D(t)-2K(t,t+1)}{D(t)}. \qquad (14)$$

Здесь $D(t)$, $D(t+1)$ - дисперсии процесса в моменты времени t и $t+1$;

$K(t,t+1)$ - ковариация между случайными отсчетами процесса в моменты t и $t+1$.

Если процесс является стационарным (или нестационарность его такова, что она не слишком существенно проявляется на интервале $[t,t+1]$), то $D(t)\cong D(t+1)$. Тогда (14) приближенно можно записать так:

$$r(t)\cong 2-2k(t,t+1),$$

где $k(t,t+1)$ - коэффициент корреляции между случайными значениями процесса в моменты времени t и $t+1$.

В связи с этим ясно, что, если корреляция между этими случайными значениями отсутствует (или мала), то значение $r(t)$ приблизительно равно 2.

При увеличении значения коэффициента корреляции величина $r(t)$ уменьшается, приближаясь к нулю (для положительной корреляции), или к четырем (для отрицательной корреляции).

Понятно, что (12) является частным случаем (13), если процесс наблюдается в дискретные моменты времени $j=1,2,...,n$ и анализируется автокорреляция остатков, получающихся после исключения тренда.

В рассматриваемой задаче $d_1=1{,}41$, $d_2=1{,}72$, а значение критерия Дарбина-Уотсона $d=1{,}87$. Поэтому следует считать, что автокорреляция остатков отсутствует, подтверждая вывод об адекватности модели.

Выводы

Полученные описания динамики интенсивности потоков могут быть использованы при решении задач анализа эффективности обработки данных в узлах компьютерной сети, а также в задачах управления обслуживанием потоков в сети.

Литература

1. Столлингс В. Современные компьютерные сети. 2-е изд.: пер. с англ. / В . Столлингс. – СПб.: ПИТЕР, 2003. – 783с.
2. Karr P. Improving Round-Trip Estimates in Reliable Transport Protocols / P. Karr, C. Partridge // ACM Trans. On Coup. Sytems. – 1991. – №3. – p.88– 97
3. Goyal R. Improving Performance of TCP over the ATM Service / R. Goyal // Proc., ICC. – 1997. – №6.
4. Гмурман В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. – М.: Высш. шк., 1972. – 368с.
5. Старков И.И. Статистическая обработка наблюдений / И.И. Старков. – М.: БИНОМ, 2003 – 312с.