

УДК 621.396.2:625.316.2

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ПРОГНОЗИРОВАНИЯ ТРАФИКА СЕТИ МОБИЛЬНОЙ СВЯЗИ

В. М. Безрук

Доктор технических наук, профессор, заведующий
кафедрой
Кафедра «Сети связи»**
Контактный тел.: 067-722-31-18
E-mail: bezruk@kture.kharkov.ua

И. В. Корсун

Фирма «Эрикссон»
г. Киев
Контактный тел.: 067-294-83-69
E-mail: imrir@ukr.net

В. А. Тихонов

Доктор физико-математических наук, профессор*
Контактный тел.: 067-575-70-36
E-mail: slavatihonoff@mail.ru

Н. В. Кудрявцева*

Контактный тел.: 093-829-00-25
*Кафедра «Радиоэлектронные системы»**
E-mail: kimberly86@list.ru
**Харьковский национальный университет
радиоэлектроники
пр. Ленина, 14, г. Харьков, Украина, 61166

Представлені теоретичні основи побудови моделей лінійного передбачення часових рядів, що характеризують функціонування системи стільникового зв'язку. Наведені результати досліджень ефективності прогнозування для часових рядів, отриманих у процесі реальної роботи мережі мобільного зв'язку стандарту WCDMA

Ключові слова: нестационарний процес, прогнозування

Представлены теоретические основы построения моделей линейного предсказания временных рядов, характеризующих функционирование системы сотовой связи. Приведены результаты исследований эффективности прогнозирования для временных рядов, полученных в процессе реальной работы сети мобильной связи стандарта WCDMA

Ключевые слова: нестационарный процесс, прогнозирование

Theoretical bases are presented of construction of models of linear prediction of time series, characterizing functioning of cellular communication network. Results of researches efficiency of forecasting are represented for time series. This results are obtained in real work of mobile communication of standard of WCDMA network

Keywords: nonstationary process, forecasting

Введение

При решении задач планирования сетевых ресурсов, а также эффективного управления и оптимизации сети мобильной связи зачастую необходимо прогнозировать возможные сценарии развития поведения сети при изменении параметров эксплуатации. К таковым могут относиться: введение в эксплуатацию нового оборудования каналов связи, изменение маршрутизации трафика зональности (кластеризации) контроллеров базовых станций, введение в сервис новых мультимедийных услуг и дополнительных сервисов. Поэтому актуальным является применение математических методов, позволяющих на основе статистики, собранной автоматизированными измерительными комплексами, выполнить прогнозирование изменения характеристик сети.

В данной статье представлены теоретические и практические аспекты построения моделей линейного прогнозирования временных рядов, характеризующих функционирование сети сотовой связи. Адекватность построенных моделей проверялась по эффективности прогнозирования временных рядов, накопленных в процессе реальной работы сети сотовой связи стандарта WCDMA. Представлены некоторые результаты исследований эффективности прогнозирования на примере трафика передачи данных абонентов вниз.

1. Анализ исходных данных о функционировании сети сотовой связи

Анализировались следующие данные о функционировании сети сотовой связи: запросы на установление

высокоскоростной сессии передачи данных, успешно установленные высокоскоростные сессии передачи данных, попытки установления вызова, ответные вызовы, проключенные вызовы, трафик передачи данных абонентов вверх и вниз, статистика передачи данных по радиоподсистеме, на одной сектор-несущей частоте для одной базовой станции. В настоящей статье представлены результаты исследования и прогнозирования лишь для некоторых данных - трафика передачи данных абонентов вниз.

Данные о функционировании сети сотовой связи стандарта WCDMA за 222 суток, полученные для трафика передачи данных абонентов вниз через каждые 30 минут, приведены на рис. 1. Предварительный анализ данных показывает, что они имеют нестационарность в виде тренда и две сезонные составляющие. В данных ярко выражена сезонная составляющая, связанная с суточной периодичностью интенсивности переговоров. Наблюдается также устойчивая сезонная составляющая с недельным периодом. Растущий тренд интенсивности трафика хорошо заметен на графике просуммированных за каждые сутки трафика передачи данных (рис. 2). Однако в целом тренд неоднороден, так как в нем присутствуют участки с разной скоростью роста и даже спада.

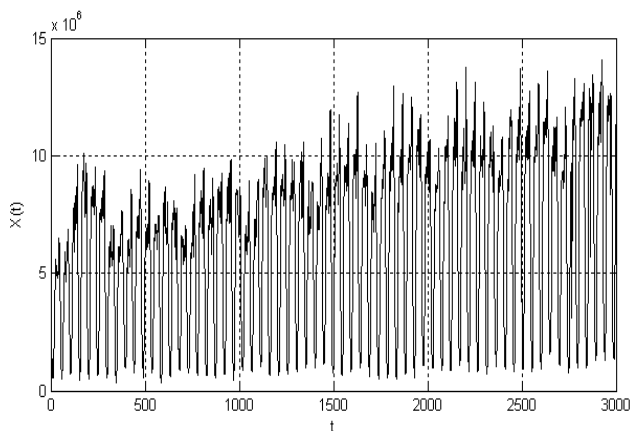


Рис. 1. Трафик передачи данных

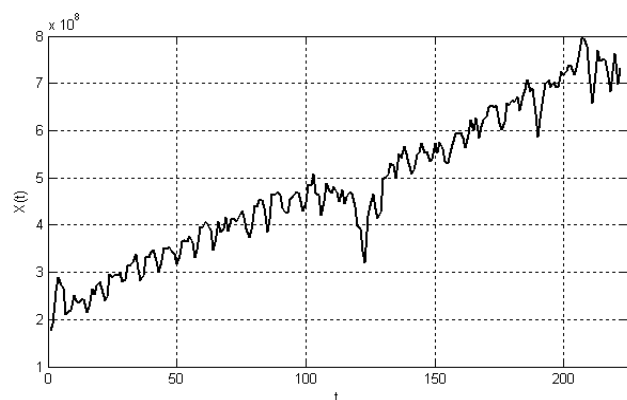


Рис. 2. Трафик передачи данных, просуммированных за каждые сутки

Другие данные о функционировании сети имеют подобные характеристики, за исключением запросов

и успешно установленных высокоскоростных сессий передачи данных. У этих близких по характеру изменения данных циклы менее устойчивы как по периоду, так и по амплитуде колебаний. Кроме того, их суточные тренды менее однородны, чем у других данных, что существенно влияет на точность прогноза.

2. Выбор и обоснование математической модели данных

Анализ прогнозируемых данных и их статистических характеристик показал, что на первом этапе исследований рационально использовать модели линейного предсказания. Для синтеза модели линейного предсказания интерес представляют статистические характеристики коррелированных стационарных данных, характеризующих функционирование системы. Анализ этих данных затруднен наличием сезонных колебаний процесса и трендом. Внутрисуточные колебания интенсивности приводят к тому, что корреляционная функция процесса описывает в основном сезонные колебания и тренд. Для получения модели стационарного процесса необходимо вначале устранить тренд и сезонные колебания.

Полная модель нестационарного процесса, используемая для прогнозирования, будет мультипликативно включать модели стационарной, трендовой и сезонной составляющих.

Нестационарные данные, характеризующие функционирование системы сотовой связи, рационально моделировать процессами авторегрессии-проинтегрированного скользящего среднего (АРПСС) [1]. Для исключения сезонной составляющей применяется упрощающий оператор взятия разности $\nabla_s = 1 - z^{-s}$, где z^{-s} - оператор сдвига, действие которого определяется выражением $z^{-s}x[t] = x[t-s]$. Так как прогнозируемые данные представляют собой значения интенсивности за каждые 30 минут, то для исключения внутрисуточных колебаний нужно выбрать $s=48$, т.е. сдвиг равный количеству отсчетов за одни сутки. После действия оператором ∇_s на исходный процесс

$$\omega_1[t] = \nabla_s x[t] = x[t] - x[t-48], \quad (1)$$

получаем новый случайный процесс $\omega_1[t]$ без внутрисуточных колебаний.

Для устранения тренда используется метод, применяемый для синтеза нестационарных моделей АРПСС [1], основанный на применении к случайным процессам с трендом оператора взятия разности

$$\omega[t] = \nabla^d x[t] = (1-z)^d x[t], \quad (2)$$

где d - порядок стационарных приращений процесса.

Для линейного тренда полагаем $d=1$, а для квадратичного необходимо брать $d=2$. Заметим, что для нестационарных процессов с произвольным законом изменения тренда применение этого метода дает существенные погрешности прогноза. Для линейных

участков тренда (рис. 2) из (1) с учетом (2), получаем при $d=1$

$$\omega_2[t] = \nabla \omega_1[t] = \omega_1[t] - \omega_1[t-1]. \tag{3}$$

Исключение из нестационарного процесса сезонной составляющей и тренда позволяет синтезировать модели линейного предсказания стационарного процесса. На графике корреляционной функции процесса трафика передачи данных абонентов вниз, представленной на рис. 3а, видно, что в процессе присутствуют значимые корреляции для малых сдвигов времени. Кроме этого, вблизи сдвига 48 отсчетов наблюдаются отрицательные значения корреляции. Анализ корреляционной функции при больших сдвигах указывает на цикличность корреляционной функции, которая наблюдается для сдвига равного недельному интервалу $T_n = 48 \times 7$ дней = 336 отсчетов. При долгосрочном прогнозе для устранения этой сезонной составляющей можно также использовать упрощающий оператор ∇_s , где $s=336$.

Для сезонных процессов с трендом существует проблема выбора подходящей модели из множества возможных. Как показали исследования, для анализируемых данных при краткосрочном прогнозе нестационарная модель, соответствующая корреляционной функции (рис. 3), имеет вид

$$(1 - z^{48})(1 - \Phi z)x[t] = a[t], \tag{4}$$

где Φ - коэффициент модели авторегрессии (АР) первого порядка стационарного процесса $\omega_2[t]$. Тогда сезонная модель описывается разностным уравнением

$$x[t] = \Phi x[t-1] + x[t-48] - \Phi x[t-49] + a[t]. \tag{5}$$

Выражение для расчета Φ получим, умножив (5) на $x[t-1]$ и усреднив. В результате преобразований получим

$$\Phi = \frac{R[47] - R[1]}{R[48] - R[0]}. \tag{6}$$

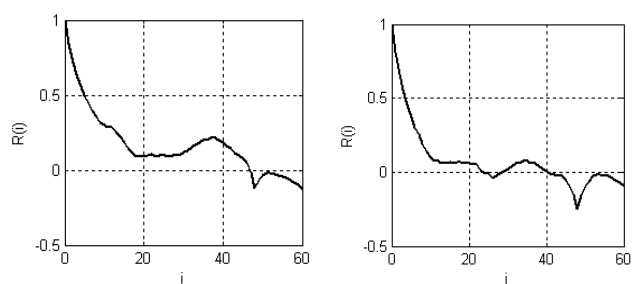


Рис. 3. Корреляционная функция процесса $\omega_2[t]$ трафика передачи данных: а - для данных без предварительного логарифмирования, б - для логарифмов данных

В ряде случаев для анализа временных рядов, а также при решении задачи прогнозирования, применяется предварительное нелинейное преобразование данных. В данной работе прогнозирование проводилось как по исходным данным, так и по их нелинейно преобразованным значениям. Заме-

тим, что применение нелинейного логарифмического преобразования уменьшает разброс данных для больших интенсивностей трафика, но увеличивает разброс данных при низких интенсивностях. На графике корреляционной функции (рис. 3б) прологарифмированного процесса $x[t]$ видно, что уменьшились колебания корреляционной функции относительно нулевой линии, а при сдвиге 48 ее значения стали более значимы. Для такой корреляционной функции можно применить более точную модель АР первого порядка.

3. Методы прогнозирования процессов функционирования сети

Рассмотрим теоретические и практические особенности прогнозирования процессов функционирования сети мобильной связи. Прогнозы с использованием разностного уравнения получаются с минимальной среднеквадратической ошибкой на момент t с упреждением l . Известно [1], что такой прогноз есть условное математическое ожидание $x[t+l]$ в момент t , при условии, что все $x[t]$ до момента t известны. Условные математические ожидания случайных величин, входящих в (5), равны

$$\{x[t+i]\} = \begin{cases} x[t+i], & i \leq 0, \\ \hat{x}[t+i], & i > 0, \end{cases} \quad \{a[t+i]\} = \begin{cases} a[t+i], & i \leq 0, \\ 0, & i > 0, \end{cases} \tag{7}$$

где $\hat{x}[t+i]$ - прогноз $x[t]$ в момент t на i шагов. Разностное уравнение, применяемое для прогноза на момент t на l шагов, получим из (4)

$$x[t+l] = \Phi x[t+l-1] + x[t+l-48] - \Phi x[t+l-49] + a[t] \tag{8}$$

Краткосрочный прогноз на упреждение l получаем как условное математическое ожидание от (8). Выражения для прогноза для разных упреждений l с учетом (7) принимают вид:

$$\hat{x}[t+1] = \Phi x[t] + x[t-47] - \Phi x[t-48], \quad l=1,$$

$$\hat{x}[t+l] = \Phi \hat{x}[t+l-1] + x[t+l-48] - \Phi x[t+l-49], \quad 2 \leq l \leq 48, \tag{9}$$

$$\hat{x}[t+49] = \Phi \hat{x}[t+48] + \hat{x}[t+1] - \Phi x[t], \quad l=49,$$

$$\hat{x}[t+l] = \Phi \hat{x}[t+l-1] + \hat{x}[t+1-48] - \Phi \hat{x}[t+1-49], \quad l > 49.$$

Формулы для коррекции и дисперсии ошибки прогноза выражаются через импульсную характеристику линейной системы, описываемой разностным уравнением (5). Процесс на выходе линейной системы с импульсной характеристикой $h[i]$ может быть представлен в виде свертки

$$x[t] = h(z)a[t] = \sum_{i=0}^{\infty} h[i]a[t-i]. \tag{10}$$

Подставляя (10) в (4) и приравнявая коэффициенты при одинаковых z^{-1} , можно получить выражения для вычисления импульсной характеристики линейной системы.

Ошибка прогноза для упреждения l равна [1]

$$e[t+l] = a[t+l] + h[1]a[t+l-1] + \dots + h[l-1]a[t+1]. \quad (11)$$

Прогноз будет несмещенный, так как условное математическое ожидание ошибки прогноза удовлетворяет равенству $\hat{e}[t+l]=0$. Дисперсия ошибки прогнозирования D_e находится из (11)

$$D_e(l) = (1+h^2[1]+h^2[2]+\dots+h^2[l-1])D_a.$$

При доверительной вероятности $1-\epsilon$, интервальная оценка прогноза $\hat{x}[t+l]$ определяется соотношением [2]

$$\hat{x}[t+l](\pm) = \hat{x}[t+l] \pm u_{1-\epsilon/2} \left(1 + \sum_{i=1}^{l-1} h_i^2 \right)^{1/2} \sigma_a,$$

где σ_a - среднеквадратическая ошибка прогнозирования, $u_{1-\epsilon/2}$ - квантиль нормального распределения уровня значимости ϵ .

4. Результаты исследований возможностей прогнозирования трафика

Ниже представлены результаты краткосрочных прогнозов трафика передачи данных сети сотовой связи с учетом сезонной составляющей с периодом 48 отсчетов. На рис. 4 представлены графики прогнозов логарифмов данных трафика передачи данных абонентов вниз. Для прогноза использовались модели АР первого порядка с коэффициентом $\Phi = 0,7728$. Прогнозы с использованием предварительного нелинейного предсказания более точны в наиболее интересной области – при высокой интенсивности трафика. Неточности прогноза в области низкой интенсивности также могут дать существенный вклад в дисперсию погрешности прогноза.

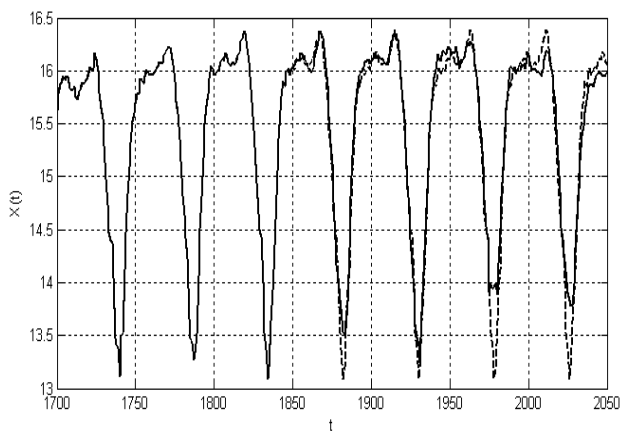


Рис. 4. Прогнозирование прологарифмированной выборки данных трафика передачи данных абонентов вниз

Анализ показал, что точность прогноза существенно не зависит от коэффициентов АР. Это связано с тем, что в выражении (5) первое слагаемое компенсируется третьим слагаемым, при достаточно близких значениях соответствующих данных в

предыдущих циклах, которое имеет обратный знак. Поэтому точность прогноза зависела от положения цикла, с которого начинался прогноз и от длины прогноза.

Исследовалась зависимость ошибки прогноза от положения начального цикла прогноза. Интервал прогноза составлял 336 отсчетов, а корреляционная функция рассчитывалась по предшествующим 200 отсчетам.

На рис. 5 показана зависимость дисперсии ошибки прогноза при сдвиге начала прогноза на сутки (48 отсчетов). Дисперсия ошибки прогноза вычислялась по формуле

$$D(t) = \frac{1}{l} \sum_{i=t}^{i=t+l} [(x(i) - \hat{x}(i)) - \overline{(x(i) - \hat{x})}]^2.$$

Из данных трафика следует, что в субботу и воскресенье трафик существенно меньше, чем в будние дни. Поэтому если прогноз на будние дни строится по данным субботы или воскресенья, то это существенно увеличивает дисперсию прогноза. На точность прогноза заметное влияние оказывают аномальные значения вблизи верхних (в дневное время суток) и нижних (в предутреннее время суток) значений трафика. Поэтому значения дисперсии погрешности прогноза на графике рис. 5 не имеют строгой недельной цикличности.

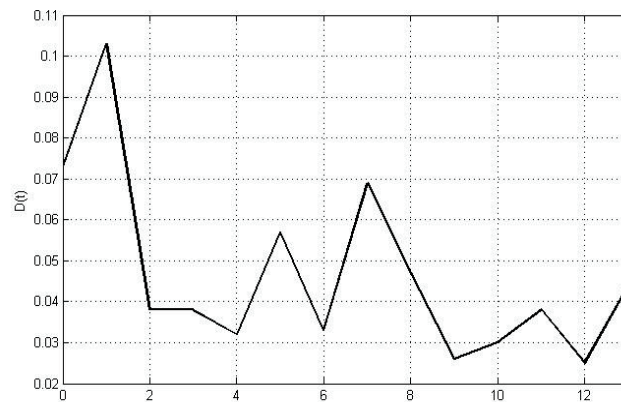


Рис. 5. Изменение дисперсии прогноза на неделю вперед от начала прогноза

Как отмечалось выше, кроме суточного цикла, в данных присутствует недельный цикл. Поэтому рационально построить модель линейного предсказания недельных данных и получить среднесрочный прогноз на несколько недель вперед. Если на этом интервале трендом данных можно пренебречь, то можно использовать модель

$$(1-z^{336})(1-\Phi z)x[t] = a[t].$$

Из этого уравнения следует, что разностное уравнение, применяемое для прогноза на момент t на l шагов имеет вид

$$x[t+l] = \Phi x[t+l-1] + x[t+l-336] - \Phi x[t+l-337] + a[t] \quad (12)$$

Выражение для расчета коэффициента AP получается из (12)

$$\Phi = \frac{R[0] - R[335]}{R[0] - R[336]}$$

Исследовалась точность прогноза от положения начального цикла прогноза. Интервал прогноза составлял 672 отсчета (2 недели), а корреляционная функция рассчитывалась по предшествующим 200 отсчетам. На рис. 6 показана зависимость дисперсии ошибки прогноза при сдвиге начала прогноза на сутки (48 отсчетов). На дисперсию ошибки среднесрочного прогноза оказывает влияние наличие неустойчивого тренда. Поэтому дисперсия ошибки вначале растет, а потом падает.

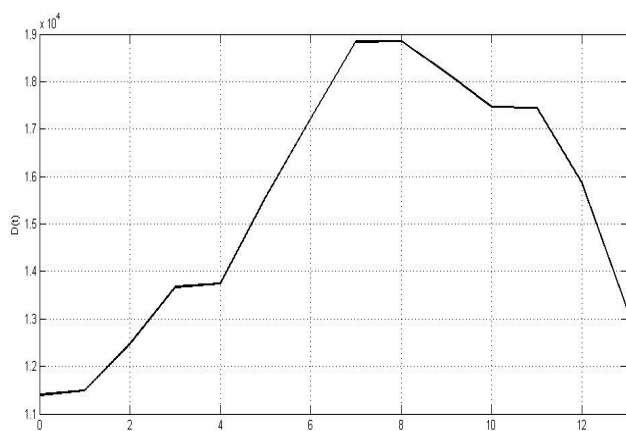


Рис. 6. Изменение дисперсии прогноза данных на 2 недели вперед от начала прогноза

Выводы

1. Проведен статистический анализ реальных данных о функционировании сети мобильной связи стандарта WCDMA.
2. На основе результатов анализа данных проведено обоснование и выбор математической модели линейного предсказания для адекватного описания реальных данных и прогнозирования процессов функционирования сети мобильной связи.
3. С использованием накопленных данных о функционировании сети мобильной связи исследованы практические возможности и особенности прогнозирования трафика при различных исходных условиях.
4. Получены положительные результаты прогнозирования трафика, которые могут быть использованы для эффективного управления и оптимизации ресурсов в процессе функционирования сети мобильной связи.

Литература

1. Бокс Дж., Дженкинс Г. Анализ временных рядов: Пер. с англ. – М.: Мир, 1974. – Вып.1. – 406с.
2. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTIKA в среде Windows. – М.: Финансы и статистика, 1999. – 384 с.
3. Montgomery D.C., Johnson L. A., Gardiner J. S. Forecasting & Time Series Analysis. – Mc.Graw-Hill Inc., 1990. – P. 384.