ПРОГНОЗИРОВАНИЕ ПОТОКА ДАННЫХ В СЕТЯХ МОБИЛЬНОЙ СВЯЗИ

Тихонов В.А.[1], Чеботарьова Д.В.[2]

Кафедра ИМИ, Харьковский национальный университет радиоэлектроники, Харьков, Украина,

E-mail: vyacheslav.tykhonov@nure.ua[1], daria.chebotarova@nure.ua[2]

Аннотация* — *В работе решена задача планирования сетевых ресурсов, а также эффективного управления и оптимизации. Для этого прогнозируются возможные сценарии развития поведения сети при изменении параметров эксплуатации. Прогнозирование трафика важно при введении в эксплуатацию нового оборудования каналов связи, изменении маршрутизации трафика зональности контроллеров базовых станций, введении в эксплуатацию новых мультимедийных услуг и дополнительных сервисов. Нестационарный трафик предложено описывать моделями авторегрессии - проинтегрированного скользящего среднего (АРПСС).

Ключевые слова – прогнозирование, анализ данных, трафик, случайный процесс, модель, линейное предсказание, тренд, сезонная составляющая.

I. Введение

При решении задач планирования сетевых ресурсов, а также эффективного управления и оптимизации зачастую необходимо прогнозировать возможные сценарии развития поведения сети при изменении параметров эксплуатации. К таковым могут относится: введение в эксплуатацию нового оборудования каналов связи, изменение маршрутизации трафика зональности (кластеризации) контроллеров базовых станций, введение в сервис новых мультимедийных услуг и дополнительных сервисов. В некоторых плохо прогнозируемых случаях изменение телекоммуника-ционного оборудования сети приводит к частичной либо полной остановке какого-либо сервиса, а введение новых услуг вызывает лавинообразный рост трафика либо абонентской базы оператора.

Для прогнозирования стационарных и нестационарных случайных процессов успешно применяется модель линейного предсказания [1, 2]. Нестационарность таких процессов проявляется в виде тренда определенной формы. Этот класс нестационарных процессов можно определить как стационарные однородные процессы, не имеющие постоянного среднего значения. В [1] рассматриваются нестационарные процессы, тренд которых можно удалить с помощью взятия d-кратной разности отсчетов процесса. Операция нахождения d-кратной разности эквивалентна взятию d-й производной от значений случайного процесса. Таким образом, эта операция заменяет случайный процесс его соответствующей производной, устраняя при этом тренд некоторого порядка. После удаления тренда случайного процесса его d-я производная становится стационарной и ее можно анализировать методами, применяемыми для стационарных моделей линейного предсказания. Для таких нестационарных процессов, описываемых моделями авторегрессии - проинтегрированного скользящего среднего (АРПСС) [1], можно определить разностное уравнение, системную функцию, найти параметры модели, импульсную характеристику, общее решение разностного уравнения.

В работе рассматриваются нестационарные процессы с сильной квазипериодической сезонной составляющей, которая влияет на параметры рассчитываемой стационарной модели процесса.

II. Прогнозирование методом линейного предсказания

Предварительный анализ данных о функционировании системы за 356 дней, измеренных ежедневно, показывает, что они имеют нестационарность в виде тренда и сезонной составляющей (рис.1). В данных наблюдается растущий тренд трафика. Однако в целом тренд неоднороден, т.к. в нем присутствуют участки роста трафика с разной скоростью. В данных наблюдается сильная сезонная составляющая, связанная с недельной периодичностью трафика.

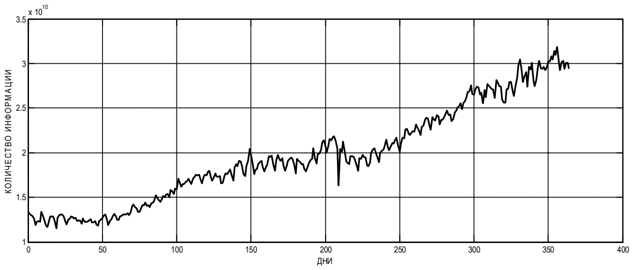

Рис. 1. График передачи данных.

Для синтеза модели линейного предсказания интерес представляют статистические характеристики коррелированных стационарных данных характеризующих функционирование системы. Анализ этих данных затруднен наличием тренда и сезонных колебаний процесса. Внутринедельные колебания объема данных приводят к тому, что корреляционная функция процесса характеризует в основном сезонные колебания и тренд. Для получения модели стационарного процесса необходимо вначале устранить тренд, а затем сезонные колебания. Для преобразованного таким способом процесса, строится стационарная модель линейного предсказания. Полная модель нестационарного процесса, используемая для прогнозирования, будет мультипликативно включать модели стационарной, трендовой и сезонной составляющей.

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

$$\omega_1[t] = \nabla x[t] = x[t] - x[t-1]. \quad (1)$$

Безтрендовый процесс имеет корреляционную функцию с выраженной незатухающей цикличностью. Анализ данных и корреляционной функции показал, что период цикла скорее всего равен 7 дней.

Для исключения сезонной составляющей применяется упрощающий оператор взятия разности

$$\nabla_s = 1 - z^{-s}, \quad (2)$$

где z^{-s} - оператор сдвига, действие которого определяется выражением:

$$z^{-s} x[t] = x[t-s]. \quad (3)$$

Так как прогнозируемый процесс представляет собой значения переданного объема данных за каждый день, то для исключения внутрисуточных колебаний нужно выбрать $s = 7$ После действия оператором ∇_s при $s = 7$ на процесс без тренда

$$\omega_2[t] = \nabla_7 \omega_1[t] = \omega_1[t] - \omega_1[t-7], \tag{4}$$

получаем новый стационарный случайный процесс $\omega_2[t]$ без внутринедельных колебаний, который используется для расчета коэффициентов АР Φ_i.

В ряде случаев для анализа временных рядов, а также при решении задачи прогнозирования, применяется предварительное нелинейное преобразование данных. Применяемое нелинейное преобразование уменьшает разброс данных для больших интенсивностей, но увеличивает разброс данных при низких интенсивностях. В данной работе прогнозирование проводилось по исходным данным, без предварительного нелинейного преобразо-вания.

Прогнозы способом разностного уравнения получают с минимальной среднеквадратической ошибкой на момент t с упреждением l. Показано [1], что такой прогноз есть условное математическое ожидание $x[t+l]$ в момент t, при условии, что все $x[t]$ до момента t известны.

Для прогноза на момент t на l шагов используем разностное уравнение

$$\begin{aligned} x[t+l] &= (\Phi_1+1)x[t-1+l] - (\Phi_1-\Phi_2)x[t-2+l] - \\ &- (\Phi_2-\Phi_3)x[t-3+l] - \Phi_3 x[t-4+l] + x[t-7+l] - \\ &- (\Phi_1+1)x[t-8+l] + (\Phi_1-\Phi_2)x[t-9+l] + \\ &+ (\Phi_2-\Phi_3)x[t-10+l] + \Phi_3 x[t-11+l] + a[t+l]. \end{aligned} \tag{5}$$

Прогнозы зависят от того, насколько модель, построенная по историческим данным, будет справедлива и для будущих данных. Следовательно, точность прогноза зависит от однородности данных, т.е. одинакова ли скорость роста тренда, есть хорошая повторяемость периода цикла, выполняется ли условие стационарности для стационарной составляющей процесса. Как видно из графика на рис. 1 скорость роста тренда не всегда одинакова для данных. Поэтому для прогноза выбирались однородные участки данных. Анализ графика корреляционной функции стационарной составляющей данных указывает, что период цикла недостаточно точно соответствует семи дням. Удаление циклической составляющей не полностью подавляет долгосрочные колебания корреляционной функции. Все это влияет на точность прогноза.

На рис. 2 представлены графики прогнозов передачи данных трафика. Для прогноза использовалась модель АР третьего порядка с коэффициентами АР:

$$\Phi_1 = -0{,}450,\ \Phi_2 = -0{,}254,\ \Phi_3 = -0{,}197 \cdot$$

Коэффициенты АР рассчитывались по историческим данным до точки времени начала прогноза. Начало и интервал прогноза выбирались равными $t = 270$ и $l = 30$ дней соответственно.

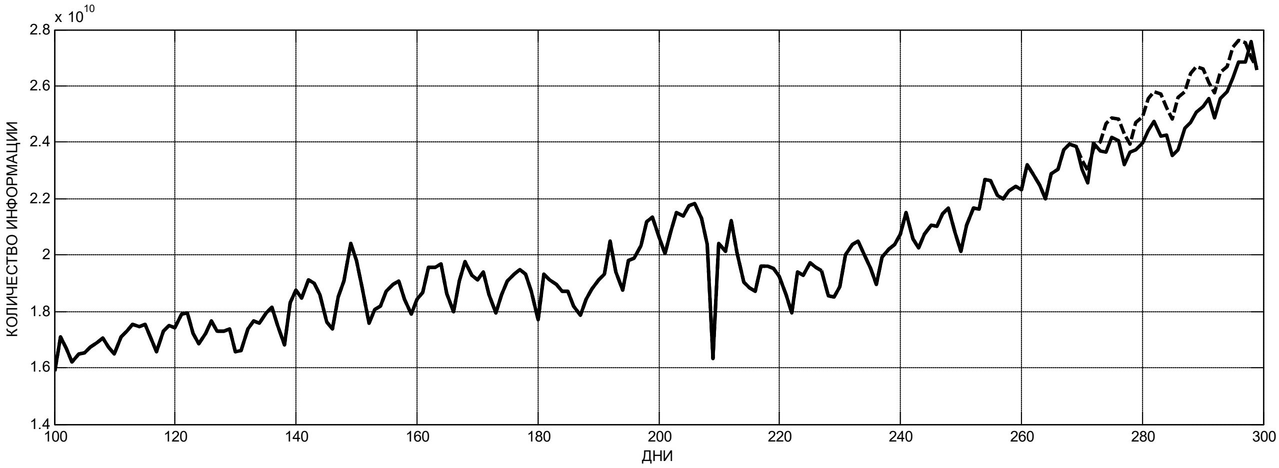

Рис. 2. Прогнозирование выборки передачи данных трафика при $t = 270$ и $l = 30$ дней.

III. Выводы

Используемый метод позволяет простым способом восстановить в прогнозе тренд, сезонную составляющую и модель стационарного случайного процесса. Анализ графиков данных и прогноза показывает некоторое отставание данных от прогноза. Однако в конце прогноза это отличие незначительно. Таким образом используемый метод может быть использован при прогнозировании широкого класса процессов, анализируемых оператором мобильной связи.

IV. Список литературы

[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 STATISTICA в среде Windows: Учебн. пособие. – 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.