

УДК 519.246.8+004.(891.2+7)

## О прототипе экспертной системы для поддержки автоматизации прогнозирования характеристик телетрафика

Тарек Юсеф Бади Биштави, Г.Н. Жолткевич, Ю.В. Соляник  
*Харьковский национальный университет имени В.Н. Каразина, Украина*

В работе описан основной набор функций прототипа, программной системы прогнозирования телетрафика TAPES (teletraffic analysis and prediction expert system).

**Ключевые слова:** прогнозирование, временные ряды, случайные процессы, телетрафик, экспертные системы, нейронные сети, boosting, анализ данных, компьютерные сети.

У роботі описаний основний набір функцій прототипу програмної системи прогнозування телетрафіка TAPES (teletraffic analysis and prediction expert system).

**Ключові слова:** прогнозування, часові ряди, випадкові процеси, телетрафік, експертні системи, нейронні мережі, boosting, аналіз даних, комп'ютерні мережі.

The paper describes the main features of teletraffic analysis and prediction software system prototype TAPES (teletraffic analysis and prediction system).

**Key words:** prediction, forecasting, time series, stochastic processes, teletraffic, expert systems, neural networks, boosting, data analysis, computer networks.

### 1. Введение

Моделирование трафика сетей передачи данных является актуальной задачей [1]. Одним из важных направлений исследований в этой предметной области является автоматизация процедур анализа и прогнозирования временных рядов значений характеристик телетрафика. В работе описан основной набор функций прототипа программной системы прогнозирования телетрафика TAPES (teletraffic analysis and prediction expert system).

### 2. Описание проблемы

Прогнозирование телетрафика играет значительную роль в проектировании и управлении компьютерными сетями. Однако свойства нестационарности, самоподобия и нелинейности его характеристик ограничивают применение эффективных методов стохастического прогнозирования, разработанных для стационарных или специальных случаев нестационарных временных рядов, и делают эту задачу трудновыполнимой [1].

Авторам этой работы неизвестно о существовании распространяемого программного продукта, обеспечивающего решение этих задач и пригодного к использованию в индустрии. Далее мы перечислим основные проблемы, стоящие на пути проектирования и реализации таких программных систем.

Автоматизированная система лишена одного из важнейших видов анализа зависимостей – визуального анализа графиков. На сегодняшний день эти методы по-прежнему практически незаменимы. Решение аналитика по выбору модели временного ряда зачастую следует за анализом графиков исходного ряда, его первых разностей и их коррелограмм. Визуальный анализ – неотъемлемая часть базовых технологий прогнозирования нестационарных временных рядов, таких как ARIMA (Autoregressive integrated moving average, технология Бокса-

Дженкинса). Поэтому эффективность работы автоматизированной системы, использующей эти технологии, будет значительно уступать эффективности работы квалифицированного инженера.

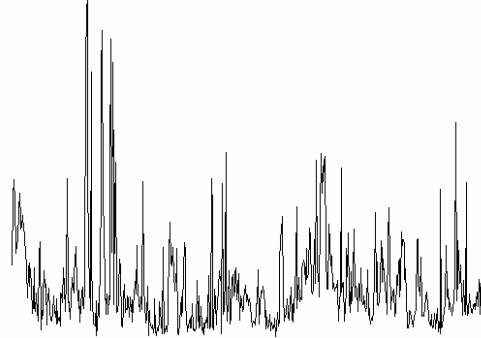


Рис.1. Фрагмент значений телетрафика LBL-TCP-3 (Internet Traffic Archive [2]), агрегированных по 10 с.

Ведущие технологии прогнозирования временных рядов, такие как ARIMA, FARIMA (fractional autoregressive integrated moving average), GARCH (general autoregressive conditional heteroskedasticity models), State Space Models, нейросетевые, - предполагают, в процессе построения модели, неоднократные шаги по принятию решений человеком для выбора различных вариантов их использования. Например, инженер должен выбрать класс моделей - линейная / нелинейная, принять решение по способу предварительной обработки данных, на основе визуального анализа графиков функций автокорреляции и частной корреляции принять решение о типе модели, значениях некоторых параметров, степени жесткости требований к ее адекватности и пр. Поэтому в автоматизированной системе должны быть заложены возможности решать или обходить эту проблему.

Проблемы определения приемлимых значений некоторых параметров являются общими для многих моделей нестационарных временных рядов. Такими параметрами, в частности, являются:

- общее количество точек ряда, используемых при его анализе,
- количество точек ряда, составляющих «горизонт» прогноза (на сколько шагов вперед он строится),
- размер «окна» прогноза (или иначе – порядок авторегрессии, размерность «обучающего примера»).

### 3. Существующие решения

Известны разработки интеллектуальных систем поддержки принятия решений, целью которых являлось повышение качества обслуживания клиентов компьютерных сетей. Есть работы, описывающие попытки построения экспертных систем, ориентированных на консультирование анализа и прогнозирования временных рядов.

В [3] описана экспертная система с нечетким выводом, ориентированная на управление высокоскоростным телетрафиком в реальном времени в рамках технологии ATM (asynchronous transfer mode). В [4] описана нейро-нечеткая система, разработанная для моделирования транспортного трафика. В [5]

говорится об автоматизации анализа финансовых временных рядов с использованием эвристик. Проект SFILTEX [6] представляет собой реализацию экспертной системы поддержки статистического анализа временных рядов. В [7] описывается программное обеспечение, обеспечивающее взаимодействие модуля прогнозирования временных рядов, реализованного на языке Java, и экспертной системы, реализованной на базе CLIPS [8]. В последней работе в качестве моделей временных рядов используются нейронные сети, параметры архитектур которых настраиваются автоматически с помощью эвристик; также реализованы элементы технологии ARIMA. После публикаций [6] (1991 г.) и [7] (2010 г.) пока не появилась информация о распространяемом программном обеспечении или о поддержке этих проектов.

#### **4. Цели работы**

Основной целью этой работы являлась разработка прототипа полуавтоматизированной системы прогнозирования характеристик телетрафика. Программная реализация этой системы имела цель исследования пригодности выбора тех или иных средств (парадигм подходов к решению задач, математических моделей, технологических решений).

Как известно из опыта разработок систем поддержки принятия решений, проекты с узкой специализацией предметной области имеют больше шансов на успех. Круг задач TAPES ограничивается прогнозированием пульсирующего [9] телетрафика, представленного значениями одномерного временного ряда.

#### **5. Выбор средств для решения задач проекта**

В связи с тем, что причиной основных проблем автоматизации анализа временных рядов является необходимость в знаниях специалиста (эксперта в предметной области), естественно было рассмотреть разработку модуля, инкапсулирующего функциональность экспертной системы производственного типа. Так как разработка экспертной системы вообще представляет собой задачу масштаба крупного проекта, было решено ограничить риски за счет выбора одного из наиболее надежных из доступных средств разработки подобных систем (были выбраны среда и язык CLIPS).

Выбор CLIPS обусловлен следующими причинами:

1. первая версия разработана NASA;
2. более 25 лет непрерывной поддержки;
3. прекрасно документирована;
4. мультипарадигменный язык – возможности
  - a. декларативного стиля создания правил (продукций),
  - b. процедурного стиля кодирования вспомогательных задач,
  - c. объектно-ориентированного подхода к проектированию базы знаний.
5. возможность программного использования всех функций из внешних систем;
6. возможности вызова внешних процессов;
7. открытый код (C/C++);
8. распространяется бесплатно;

9. CLIPS - возможно, наиболее широко используемая среда разработки экспертных систем в последние 20 лет.

Используемые в TAPES математические модели и технологии моделирования телеграфика (нейронные сети, ARIMA, FARIMA, State Space Time Series Models, GARCH, алгоритмы Data Mining) представляют собой стандартный минимальный набор средств, реализуемых в большинстве ведущих в области анализа данных программных средах, таких как STATISTICA, SPSS, MATLAB, R, Greta, Tisean, продукты OхMetrics и др. Как показано в [11], модели ARIMA, FARIMA, GARCH можно выразить с помощью State Space Time Series Models (далее SSM). Практически, базовую функциональность модуля прогнозирования TAPES можно было бы реализовать с помощью линейных моделей SSM и нескольких видов архитектур нейронных сетей (нелинейные модели). Однако мы предпочли дублирование возможностей различных подходов в целях исследования степени их пригодности в различных смыслах. Нелинейные модели SSM [11] не использовались.

В связи с тем, что SSM уделяется пока сравнительно мало внимания в русскоязычной литературе, посвященной прогнозированию временных рядов, мы приведем краткое описание базовой математической модели и способа прогнозирования. Общий случай для линейной гауссовой модели одномерного временного ряда можно представить с помощью уравнений:

$$\begin{aligned} y_t &= Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H_t), \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t), \\ \alpha_1 &\sim N(0, P_1), \end{aligned}$$

где  $y_t$  - «наблюдаемые» (observed) значения временного ряда;  $\alpha_t$  - вектор состояния («ненаблюдаемые» (unobserved) переменные); матрицы  $Z, H, T, R, Q, \alpha_1, P_1$  предполагаются известными или определяются по данным с помощью алгоритмов обучения (таким образом, возможна автоматизация построения моделей временного ряда);  $t = 1 \dots n$ .

Прогноз строится с помощью фильтра Калмана [11]:

$$\begin{aligned} \alpha_{t+1} &= T_t \alpha_t + K_t v_t, \\ K_t &= T_t P_t Z_t^T F_t^{-1}, \\ P_{t+1} &= T_t P_t L_t^T + R_t Q_t R_t^T, \\ F_t &= Z_t P_t Z_t^T + H_t, \\ v_t &= y_t - Z_t \alpha_t, \\ L_t &= T_t - K_t Z_t. \end{aligned}$$

Для определения параметров моделей SSM разработаны эффективные по скорости работы алгоритмы. Изложение теории и приложений SSM дано в [11].

Средства реализации задач прогнозирования определялись целями и классом проекта (исследовательский работающий прототип программной системы). Поэтому мы позволили себе использовать возможности вызова функций других программных продуктов (MATLAB, STATISTICA, SPSS, Tisean, SsfPack, Greta, Stata), а платформу .NET корпорации Microsoft использовали как

інтегруючу середу, включаючу функції взаємодіяння з користувачем (использовался язык программирования C#).

#### **6. Основные возможности системы**

Взаємодіяння користувача з системою забезпечено з допомогою спеціально розробаного языка выполнимых спецификаций *TapesSpec*. Иначе можно специфицировать задание системе с помощью средств графического интерфейса, но использование языка дает больше возможностей, например, при задании ограничений на значения параметров модели. Интерпретатор *TapesSpec* был разработан на языке Prolog (использовалась среда SWI-Prolog).

Значительную часть правил базы знаний экспертной системы составляют эвристики, используемые при идентификации моделей ARIMA, FARIMA, SSM, GARCH, нейронных сетей и оценивании их параметров. Приведем примеры эвристик, источниками которых являются рекомендации разработчиков математических моделей и экспертов-практиков:

- *если требуется выразить сезонность (периодичность) временного ряда с помощью аддитивной модели, то имеет смысл применить логарифмическое преобразование к исходным данным;*
- *если значения оценок автокорреляционной функции на первых лагах значимы и экспоненциально убывают, а значения оценок частной корреляционной функции имеют выброс на лаге 1, и нет значимых оценок значений на других лагах, то это говорит в пользу модели AR(1) (процесс авторегрессии первого порядка);*
- *если размерность скрытого слоя перцептрона сделать избыточно большой (например, больше размера входного слоя), затем усреднить результаты обучения ансамбля  $k$  (параметр метода) таких предикторов, то высока вероятность того, что эффективность прогноза ансамбля будет близка к эффективности прогноза нейронной сети, обученной в тех же условиях, с такой же размерностью входного слоя и суб-оптимальной размерностью скрытого слоя.*

Решение по определению некоторых «проблемных параметров», таких как количество используемых в анализе точек временного ряда, может быть принято на основе нескольких подходов, среди которых - методы, нелинейного анализа временных рядов, используемые для определения достаточной размерности пространства состояний восстанавливаемой по временному ряду реконструкции динамической системы (используется программный пакет Tisean).

В TAPES есть возможность использования технологии «усиления» (boosting) [13]. Ее применение позволяет повысить точность прогноза за счет получения группы предикторов со слабо коррелированными остатками. В [14] было показано экспериментально, что этот эффект может быть достигнут с помощью многокритериального обучения нейронной сети прямого распространения.

Задачи

- кодирования процедур предварительного анализа данных,
- преобразований временного ряда,

- формирования множеств обучающих примеров - обучающего множества, множества для кросс-валидации, множества для тестирования модели (back-testing),
  - построения моделей и численное оценивание их параметров;
  - расчета точечных и интервальных оценок прогнозов (в случае задачи регрессии),
  - проверки адекватности моделей (диагностические тесты, применяемые к остаткам),
  - метадиагностики различных методов (сравнительный анализ),
  - расчета значения функции потерь
- были решены с помощью средств систем MATLAB и STATISTICA (использовались их встроенные языки программирования).

Система TAPES может работать в следующих режимах:

1. *«автоматический»* - пакетное задание выполняется в соответствии со спецификацией (без вмешательства пользователя);
2. *«полуавтоматический»* – система может обращаться к пользователю по поводу выбора вариантов решения;
3. *«среда разработки экспертной системы»* - в этом режиме можно модифицировать базу знаний;
4. *«вычислительный пакет»* - TAPES можно использовать для выполнения расчетов, вывода графиков и т. п.

Это деление условно, - например, возможности режима 4 можно использовать в любом другом режиме. Вопросы разграничения доступа к функциям системы мы здесь не затрагиваем, однако перечислим возможных пользователей TAPES:

1. проектировщики и администраторы компьютерных сетей, инженеры смежных специальностей, (режимы 1 и 2, требуется знание языка спецификации задания или навыки работы с графическим интерфейсом системы);
2. специалисты по прогнозированию временных рядов (все режимы, кроме 3);

инженеры по знаниям (специалисты по разработке и сопровождению баз знаний интеллектуальных систем - режим 3).

### **7. Перспективы дальнейшей работы**

Целями нашей дальнейшей работы являются развитие базы знаний экспертной системы, а также проектирование и реализация программного продукта на основе разработанного прототипа.

### **ЛИТЕРАТУРА**

1. Шелухин О.И., Тенякшев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях. – Монография. / Под ред. Шелухина О.И. – М.:Радиотехника, 2003. – 480 с.
2. The Internet Traffic Archive [Электронный ресурс]. - Режим доступа: <http://http://ita.ee.lbl.gov>.

3. Ascia G., Catania V., Ficili G., Palazzo S., Panno D. A VLSI fuzzy expert system for real-time traffic control in ATM networks //IEEE Transactions on Fuzzy Systems. – 1997. 5(1). - 20–31.
4. Quek C., Pasquier M., Lim B. A novel self-organizing fuzzy rule-based system for modelling traffic flow behaviour // Expert Syst. Appl. - 2009. 36(10). – 12167-2178.
5. Castillo O., Melin P. An intelligent system for financial time series prediction combining dynamical systems theory, fractal theory and statistical methods // Computational Intelligence for Financial Engineering. – 1995. 151-155.
6. Krtolica R., Obradovic I., Bozinovic R. SFILTEX: toward an expert system for time-series filter design //Journal of Applied Statistics. – 1991. 18(1). - 121 – 137.
7. Nikolov V., Bogdanov V. Integration of Neural Networks and Expert Systems for Time Series Prediction [Электронный ресурс]. - Режим доступа: <http://doi.acm.org/10.1145/1839379.1839475>. - 2010. – PDF формат, версия 1.5.
8. CLIPS: A Tool for Building Expert Systems [Электронный ресурс]. - Режим доступа: <http://clipsrules.sourceforge.net>.
9. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. 3-е изд. - СПб.: Питер, 2006. – 958 с.
10. Уотермен Д. Руководство по экспертным системам. - М.: Мир, 1989. – 338 с.
11. Durbin J., Koopman S.J. Time Series Analysis by State Space Methods. - Oxford: Oxford University Press, 2001. – pp. 272.
12. Chatfield, C. The Analysis of Time Series. An Introduction (6th edn.). -London: Chapman & Hall, 1996. – pp. 283.
13. Freund Y, Schapire R.E. A decision-theoretic generalization of on-line learning and an application to boosting //Journal of Computer and System Sciences. – 1997. 55(1). - 119--139.
14. Биштави Тарек Юсеф Бади, Жолткевич Г.Н., Соляник Ю.В. Многокритериальное обучение предиктора в задаче прогнозирования телеграфика // Системи обробки інформації.– 2010. т. 2, №83. – 163-165.12.