

Теоретико-информационные методы прогнозирования и статистического анализа временных рядов

АВТОРЫ: д.т.н. Рябко Б.Я.

Методы прогнозирования и статистического анализа временных рядов находят самое широкое практическое применение и по этой причине находятся в центре внимания многих исследователей как у нас в стране, так и за рубежом.

В 2016 г. автором были опубликованы результаты многолетних исследований, посвященные применению идей и методов теории информации для построения асимптотически оптимальных методов прогнозирования стационарных эргодических процессов и показано, как так называемые универсальные коды можно применять для проверки статистических гипотез о законе распределения и памяти, или связности, процессов.

Основные результаты можно сформулировать в виде следующих утверждений:

Теорема 1. Пусть $x_1 x_2 \dots x_t$ порождается стационарным эргодическим источником и пусть F – универсальный код. Тогда погрешность прогноза по мере $M(F) = \exp(-|F(x_1 x_2 \dots x_t)|)$ стремится к нулю при возрастании длины последовательности t . (Здесь $|u|$ длина слова u .)

Теорема 2. Пусть M – некоторый стационарный эргодический источник, порождающий символы из алфавита A , и $x_1 x_2 \dots x_t$ – последовательность символов из этого алфавита. Рассматриваются две гипотезы $H_0 = \{x_1 x_2 \dots x_t \text{ порождена } M\}$ и $H_1 = \text{не } H_0$ и тест для их проверки:

$$\text{Если } -\log M(x_1 x_2 \dots x_t) - |F(x_1 x_2 \dots x_t)| < -\log a, \text{ то } H_0, \text{ иначе } H_1,$$

где a – уровень значимости теста и F – универсальный код. Тогда ошибка первого рода этого теста не превосходит a , а ошибка второго рода стремится к 0 при увеличении t .

Подобные критерии сформулированы для проверки гипотезы о величине памяти марковских цепей и о покомпонентной независимости многомерных процессов.

Описанный метод прогнозирования применялся для реальных процессов (солнечная активность, измеряемая в числах Вольфа, колебания уровня моря в нескольких регионах, индекс безработицы и др.) и во всех случаях показал высокую точность относительно ранее известных методов.

Предложенные тесты применялись для проверки генераторов случайных чисел, используемых в системах защиты информации. При их помощи были обнаружены отклонения от равномерного распределения («полной случайности») для широко известного генератора псевдослучайных чисел RC4 и ряда других.

ПУБЛИКАЦИИ:

1. B. Ryabko, J. Astola, M. Malyutov. Compression-Based Methods of Statistical Analysis and Prediction of Time Series // Springer International Publishing Switzerland, 2016.