

Е. В. Пучков (Ростов-на-Дону, РГСУ). **Определение глубины погружения временного ряда с помощью методики box-counting.**

Финансовые временные ряды характеризуются большими колебаниями с неочевидной автокорреляцией. Сложность таких временных рядов делает их интерпретацию затруднительной. Остановимся на задаче текущего прогнозирования стохастической последовательности $x(t)$ по данным о ее предыстории $x(t-1), x(t-2), \dots$. Проблема сводится к нахождению оценки $\hat{x} = F(x(t-1), x(t-2), \dots, x(t-m))$ в реальном времени. В линейном случае эта задача хорошо исследована и успешно может быть решена с помощью адаптивных прогнозирующих авторегрессионных моделей [1]. Для построения же нелинейных прогнозов наиболее целесообразным представляется применение нейросетевого анализа. Для прогнозирования элементов временных рядов широко используют метод скользящих окон (windowing). Основная идея этого метода состоит в использовании двух окон, входного (W_{in}) и выходного (W_{out}), фиксированных размеров m и n , соответственно, перемещающихся (скользящих) вдоль оси абсцисс [2]. Входное окно формирует данные для входов нейронной сети, а выходное, соответственно, для выходов (оценок).

Задача обучения нейросети заключается в выборе размера входного окна и настройки весов нейронов. Остановимся на первой из этих задач. Размер входного окна (глубины погружения) влияет на качество прогноза временного ряда. В связи с этим возникает проблема нахождения оптимального размера входного окна. Выходное же окно определяется из условий задачи. Будем считать, что размер выходного окна равен 1, поскольку обобщение на случай других размеров выходного окна не составляет труда. Для определения размера входного окна применим методику box-counting [3]. Методика основана на подсчете числа заполненных ячеек, на которые разбивается пространство признаков. Признаками в данном случае являются значения временного ряда. Далее оценивается кросс-энтропия для значений отрезка временного ряда $X_t = (x_t, x_{t-1}, \dots, x_{t-m})$ по формуле

$$I(X_t, X_{t+1}) = \log \frac{N_{X_{t+1}} N_{X_t}}{N_{X_{t+1}, X_t}}, \quad (1)$$

где N_y — количество заполненных ячеек для отрезка временного ряда y ($y = X_{t+1}, X_t, X_{t+1}X_t$, соответственно).

Чем больше кросс-энтропия, тем больше определенности вносит знание значения X_t в предсказание значения x_{t+1} .

Рассмотрим алгоритм определения размера окна m с помощью методики box-counting.

1. Нормализуем значения временного ряда.
2. Построим таблицу T , содержащую обучающую выборку (X_1, X_2, \dots, X_l) .
3. Выберем размер ячейки ϵ .
4. Построим по таблице T таблицу M , имеющую структуру (имя, значение), где имя — индексы ячеек, значение — количество ячеек.
5. Рассчитаем таблицу N , просуммировав количество ячеек в таблице M по всем индексам, кроме последнего.
6. Рассчитаем таблицу L , просуммировав количество ячеек в таблице M по последнему индексу.
7. Рассчитаем кросс-энтропию по формуле (1) для $m = 2, 3, \dots$
8. Выберем то значение m , при котором кросс-энтропия стабилизируется.

Реализация данной методики позволила рассчитать размер окна намного эффективнее по вычислительным затратам, нежели его подбор во время обучения нейронной сети.

СПИСОК ЛИТЕРАТУРЫ

1. *Лукашин Ю. П.* Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Финансы и статистика, 2003, 416 с.
2. *Кричевский М. Л.* Интеллектуальный анализ данных в менеджменте. СПб.: СПб-ГУАП, 2005, 208 с.
3. *Ежов А. А., Шумский С. А.* Нейрокомпьютинг и его применение в экономике и бизнесе. М.: МИФИ, 1998, 224 с.