0.1. *Чирихин К.С.* Теоретико-информационный метод интеграции различных алгоритмов прогнозирования временных рядов

Задача прогнозирования временных рядов имеет множество приложений в разных областях человеческой деятельности и обладает большой практической значимостью [1]. Известно, что для её решения могут успешно использоваться методы сжатия данных [2]. К настоящему времени было предложено большое количество разнообразных алгоритмов сжатия, основанных на разных подходах, и в общем случае заранее неизвестно, какой из них наиболее эффективен для прогнозирования определённого временного ряда. Однако, можно объединить различные алгоритмы сжатия в один метод прогнозирования таким образом, что на результат работы объединённого метода наибольшее влияние оказывает алгоритм, лучше остальных сжимающий ряд (и, следовательно, способный лучше других находить имеющиеся в нём закономерности). Описание такого способа объединения можно найти, например, в [3]. Это позволяет «выбрать» наиболее подходящий алгоритм для прогнозирования данных «автоматически».

В то же самое время, существует огромное число других методов прогнозирования временных рядов — экспоненциальное сглаживание, нейронные сети, модели авторегрессии-скользящего среднего и др. Как и в случае с разными алгоритмами сжатия, точность каждого из этих методов зависит от прогнозируемых данных. В настоящей работе мы описываем модификацию, способную преобразовать любой из подобных методов в метод сжатия. Она позволяет применять вышеупомянутый способ объединения алгоритмов с целью выбора наиболее подходящего из них для прогнозирования ряда. Также мы рассматриваем способ сокращения накладных расходов по времени вычислений, необходимых для выбора наилучшего алгоритма.

Опишем основную идею предлагаемой модификации. Сначала вещественный временной ряд $X = x_1, x_2, \ldots, x_t$ преобразуется к ряду с конечным алфавитом $X^{[n]} = x_1^{[n]}, x_2^{[n]}, \ldots, x_t^{[n]}$ с помощью процедуры квантования — отрезок возможных значений ряда разбивается на конечное число n пронумерованных интервалов, и каждое значение ряда заменяется номером интервала, в которое оно попадает. Затем, используя произвольный метод прогнозирования γ, мы можем получить оценку вероятности

$$p^*_\gamma(X^{[n]}) = \prod_{i=1}^{t} p^*_\gamma(x_i^{[n]} | x_1^{[n]}, \ldots, x_{i-1}^{[n]}).$$

Если $\gamma(x_1^{[n]}, \ldots, x_{i-1}^{[n]}) \in x_i^{[n]}$, то выберем $p^*_\gamma(x_i^{[n]} | x_1^{[n]}, \ldots, x_{i-1}^{[n]})$ близкой к 1, иначе близкой к 0. Затем мы можем получить нужную нам длину кодового слова для $X^{[n]}$ как $-\log_2 p^*_\gamma(X^{[n]})$. Мы выполнили программную реализацию описываемого подхода. На данный момент поддерживается работа с 7 «настоящими» алгоритмами сжатия и моделью Хольта-Уинтерса (её описание может быть найдено, например, в [4]). В работе рассматриваются примеры прогнозирования реальных и искусственных данных, приводятся случаи, когда подобный подход приводит к повышению точности прогноза.

Работа выполнена при финансовой поддержке РФФИ (гранты № 19-37-90009, № 19-47-540001).
Научный руководитель — д.т.н. Рябко Б. Я.

Список литературы

[1] Box G. E., Jenkins G. M., Reinsel G. C., Ljung G. M. Time series analysis: forecasting and control / John Wiley & Sons, 2015. 712 p.
[2] Ryabko B., Astola J., Malyutov M. Compression-based methods of statistical analysis and prediction of time series / Switzerland: Springer International Publishing, 2016. 144 p.
[3] Chirikhin K. S., Ryabko B. Y. Application of artificial intelligence and data compression methods to time series forecasting // Proc. Intern. Workshop «Applied Methods of Statistical Analysis. Statistical Computation and Simulation — AMSA'2019». Novosibirsk: NSTU publisher, 2019. P. 553–560.
[4] Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice. OTexts, 2018. 382 p.