**Анализ и прогнозирование больших временных рядов в реальном времени**

***Коковин П.П.***

*студент*

*Московский государственный университет имени М.В. Ломоносова,*

*физический факультет, Москва, Россия*

*E-mail:* *kokovin.pp20@physics.msu.ru*

В современном мире растет значимость временных рядов, поскольку они широко используются в экономике, финансах, метеорологии, медицине и других областях. Способность анализировать изменения и тенденции во времени становится важной задачей, целью которой является принятие более обоснованных решений. Эффективное предсказание требует умения прогнозировать тренды, выявлять аномалии и различные выбросы в данных. Также объем собранной информации постоянно растет [1], и для ее обработки требуется все больше ресурсов и времени. Помимо того, так как новые измерения постоянно добавляются к уже существующим, необходимо применять методы, которые позволяют анализировать обновляемые данные в реальном времени. Именно поэтому важно обращать внимание на качество получаемых данных ввиду того, что иногда возникают сбои в работе измерительных устройств, что приводит к потере или искажению части данных и усложнению составления прогнозов.

Для прогнозирования стационарного временного ряда, рассмотрим авторегрессионную модель [2][3]:

$$y\_{n}=α\_{1} ∙ y\_{n-1}+α\_{2}∙ y\_{n-2}+\cdots +α\_{k}∙ y\_{n-k}+ε\_{n}.$$

Здесь $y\_{n+1},y\_{n} $,$ y\_{n-1}$- измерения на шагах n+1, n и n-1 соответственно, $ε\_{n+1}$- нормально распределенный шум с нулевым средним.

Соотношения позволяющие оценить авторегрессионные коэффициенты можно записать в матричном виде:

$$\left(\begin{matrix}y\_{n}\\\vdots \\y\_{k}\end{matrix}\right)=\left(\begin{matrix}y\_{n-1}&       y\_{n-2}         y\_{n-k}      \\\vdots &\vdots       \cdots       \vdots \\y\_{k-1}&y\_{k-2}              y\_{0}\end{matrix} \right)∙\left(\begin{array}{c}\begin{matrix}α\_{1}\\α\_{2}\end{matrix}\\\vdots \\α\_{k}\end{array}\right)+\left(\begin{matrix}ε\_{n}\\ε\_{n-1}\\ \vdots \\ε\_{k}\end{matrix}\right)$$

Или:

$$Y=B ∙a+ε.$$

Таким образом, задача прогнозирования сводится к поиску коэффициентов регрессии (столбца а) путем минимизации функционала:

$$|\left|Y-B∙a\right||^{2}\~min\_{a}.$$

Оказывается, для минимизации этого функционала нет необходимости накапливать и хранить весь ряд. Достаточно лишь выделить из него нужную информацию фиксированного объема [4]. На основании этой специальной формы представления информации можно найти оценки коэффициентов АР, минимизирующие функционал, оценить дисперсию $ε\_{i} $и построить прогноз на требуемый временной горизонт. Более того, по мере поступления новых данных, накопленная информация может быть обновлена, а оценки коэффициентов и прогноз пересчитаны в реальном времени. Для этого на всех временных отрезках ведется сбор информации в удобном и компактном виде, а для получения полной информации, отдельные фрагменты информации суммируются.

Для стабильной работы многие современные алгоритмы требуют отсутствия пропусков в данных. В свою очередь, в данной работе был создан алгоритм, использующий методы обработки больших данных, нечувствительный к наличию пропусков. Таким образом, была не только решена проблема влияния потери или искажения части данных на достоверность прогнозов, но и упрощена процедура обработки данных.

Рисунок Пример предсказания временного ряда по всей истории (красная заливка - коридор погрешности)

**Литература**

1. Chernykh Elena, Golubtsov Peter, Shapkina Natalia: Prediction of meteorological quantities using a hybrid time series processing method// [IX International Conference on Information Technology and Nanotechnology (ITNT-2023)](https://istina.msu.ru/conferences/567681490/)
2. Лукашин Ю.В: Адаптивные методы краткосрочного прогнозирования временных рядов // Учеб. Пособие. – М.: Финансы и статистика, 2003.
3. Бокс Д., Дженкинс Г., Левшин А. Л.: Анализ временных рядов: Прогноз и управление. Вып. 2. – Мир, 1974.
4. Golubtsov P. V.: The concept of information in big data processing //Automatic Documentation and Mathematical Linguistics. – 2018. – Т.52. – No. 1. – С. 38-43.