

Разработка алгоритмов обработки временных рядов при работе со статистическими отчетными формами производственного сектора пенитенциарной системы

Д.С. Пономарев

*Научно-исследовательский институт Федеральной службы исполнения наказаний,
г. Москва*

*Ижевский государственный технический университет имени М.Т. Калашникова,
г. Ижевск*

Аннотация: На сегодняшний день в пенитенциарной системе Российской Федерации собраны достаточно обширные базы данных для производственного сектора. Собранные данные представляют собой временные ряды. Однако, при исследовании взаимных распределений параметров возникает ряд проблем, главная из которых состоит в том, что для разных параметров ведется различная система учета данных: в некоторых случаях учет данных ведется накопительным итогом в течении года, в других случаях – учитываются фактические значения (другими словами, одни временные ряды являются трендовыми, а другие – сезонными (циклическими)). Отличаются также и периоды учета данных: поквартально, поквартально или за год. Таким образом, на первый взгляд связанные между собой параметры, практически невозможно сопоставить. В работе предлагается ряд алгоритмов, которые позволили бы решить данную проблему. *Целью* работы явилась разработка новых алгоритмов, позволяющих сопоставить трендовые и сезонные временные ряды на примере производственного сектора пенитенциарной системы. В качестве *задач* исследования можно обозначить: классификацию параметров, которые учитываются как сезонные и как трендовые временные ряды; разработка алгоритмов для их сопоставления; исследование применимости полученных результатов.

Ключевые слова: алгоритм, обработка данных, python, временной ряд, пенитенциарная система, производственный сектор.

Введение

На сегодняшний день, в пенитенциарной системе ведутся активные разработки и внедрение научно-исследовательских, инженерных и расчетных работ [1-3]. Ведется активный сбор и учет информации, разработаны базы данных для учета из ведомственных статистических форм. Кроме того, в системе исполнения наказаний Российской Федерации существует обширная база данных, ориентированная на производственный сектор. Так как данные собираются с периодичностью в один месяц, квартал или год (в зависимости от форм учета), то собранные массивы данных представляют собой временные ряды. Собранная информация может быть использована для

дальнейшего прогноза в деятельности предприятий и производств пенитенциарной системы.

Однако, возникают определенные трудности при анализе взаимных распределений параметров. Основная проблема заключается в том, что для различных параметров используются разные системы учета данных. В некоторых случаях данные учитываются накопительным итогом за год, в других случаях учитываются фактические значения. Также различаются периоды учета данных: ежемесячно, ежеквартально или ежегодно. Следовательно, на первый взгляд связанные параметры практически невозможно сопоставить. Исследование собранной информации привело к следующим категориальным разделениям данных [4, 5].

1. Были выделены два вида временных рядов: трендовый; циклический или сезонный.

2. Выделены интервалы учета временных рядов: ежемесячно; поквартально; ежегодно.

Соответственно, в исследовании было поставлено два решения проблем: разработать алгоритмы, которые бы позволили сопоставить циклические и трендовые временные ряды; разработать алгоритмы, благодаря которым возможно конвертировать временные ряды с разным периодом учета данных.

Стоит при этом отметить, что учет данных за месяц может быть переведен в поквартальный или годовой, а поквартальный – только в годовой, без каких-либо дополнительных изменений в данных. Перевод можно сделать путем суммирования, либо расчета арифметического среднего (зависит от поставленной задачи исследования) [6].

Разработка алгоритмов обработки данных

Более предпочтительным будет перевод учета значений параметров в фактические значения. Для того, чтоб рассчитать фактические значения за

каждый месяц, здесь следует из интересующего периода вычесть предыдущий. При этом для января следует брать фактическое значение (так как он является первым месяцем в году). Поставленная задача схожа с одной из известных задач *computer science*, а именно – «*FizzBuzz*» [7, 8], но, в нашем случае, ключевым моментом в решении задачи является сравнение остатка от целочисленного деления на 12 с 0, таким образом, новая последовательность будет сформирована, исходя из (1-3):

$$x_n \bmod 12 \neq 0 \Rightarrow x_n^{\text{прирост}} = x_n - x_{n-1} \quad (1)$$

$$x_n \bmod 12 = 0 \Rightarrow x_n^{\text{прирост}} = x_n \quad (2)$$

$$f(x_n^{\text{прирост}}) = [x_1^{\text{прирост}}, x_2^{\text{прирост}}, \dots, x_n^{\text{прирост}}] \quad (3)$$

В результате был разработан следующий алгоритм (рисунок 1).

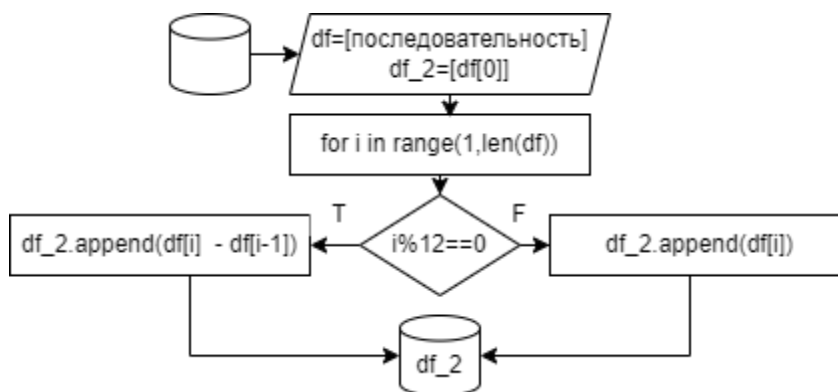


Рис. 1. – Алгоритм преобразования накопительного итога параметров в фактические значения (за каждый месяц, отсчет с начала года)

Однако, данное преобразование оказывается ошибочным, если учет показателей ведется не с начала года (не с первого месяца), поэтому будут более предпочтительными преобразования с прикреплением каждого значения к определенному месяцу. Для этого следует создать два списка: один из которых будет содержать в себе даты, а второй – значения. Безусловно, использование словарей [9, 10] в этом случае будет являться

отличным решением (т.е. использование связки «ключ-значение», где «ключами» будут являться даты, а «значениями» - непосредственно сами данные). Однако здесь, при практических расчетах, возникает ряд сложностей при работе с данными (например, когда нужно использовать как данные из поквартального учета, так и данные, которые учитываются за каждый месяц – и сопоставить их). Поэтому, были рассмотрены и другие подходы, помимо использования словарей. Одним из таких подходов является алгоритм, представленный на рисунке 2.

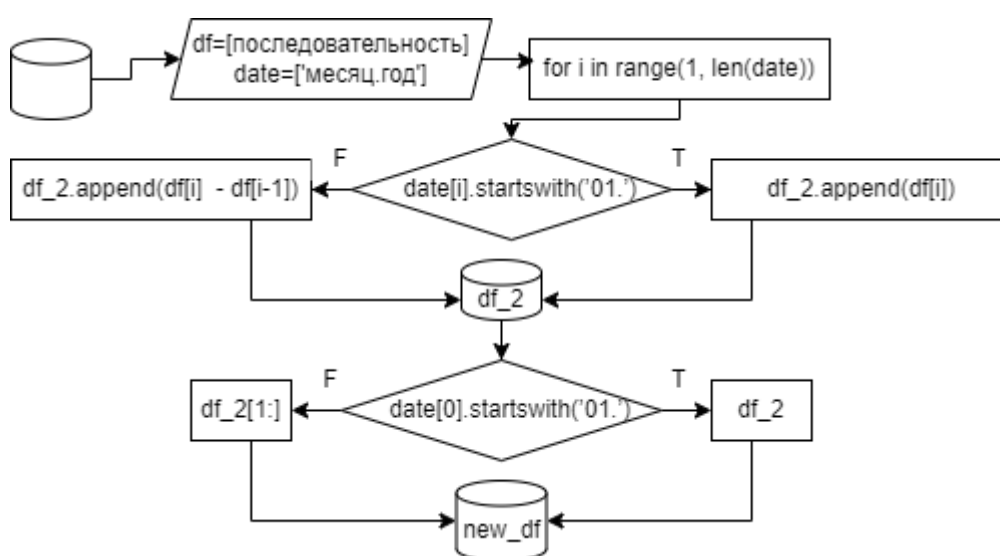


Рис. 2. – Алгоритм преобразования накопительного итога производственно-экономических показателей в фактические значения.

Перейдем ко второй проблеме. В рассматриваемых данных присутствуют разные периоды учета данных (например, рассмотрим помесичный и поквартальный учет). В этом случае наиболее простым и рациональным будет перевод всех значений в поквартальный учет, т.е. для каждых трех месяцев следует вывести центральное значение данных (например, среднее арифметическое), записать его в новую последовательность и далее перейти к следующим трем месяцам.

Если сделать привязку к датам (например, к номеру месяца), как было представлено в предыдущем алгоритме, то будет возможно начинать отсчет с любого месяца в году. Реализация алгоритма может быть произведена следующим образом (рисунок 3).

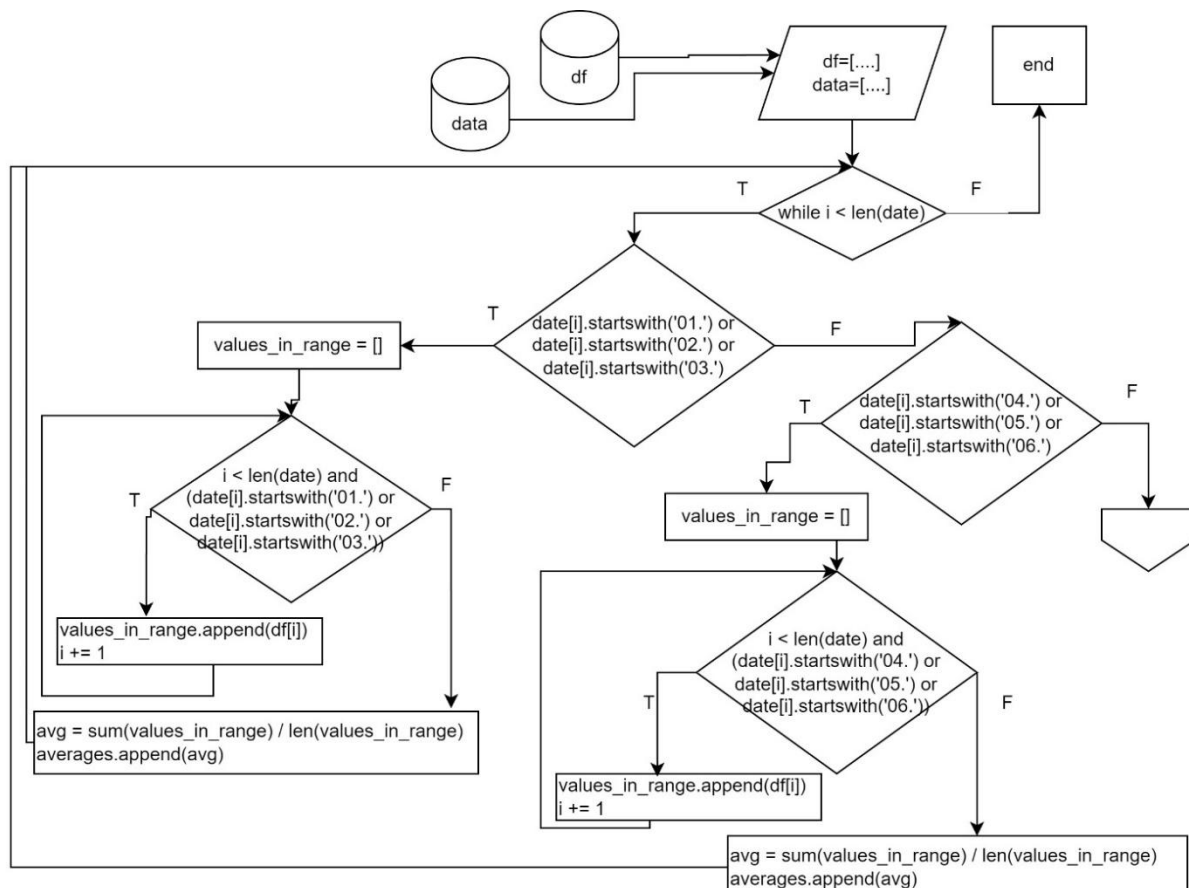


Рис. 3. – Алгоритм преобразования временных рядов для их сопоставления (на примере перевода учета данных за каждый месяц в поквартальный учет)

Аналогичным образом возможно сделать перевод ежемесячного учета данных в годовой или поквартальный в годовой. Таким образом, несмотря на разные периоды и методы учета параметров, создание единообразной базы данных для рассмотренных групп параметров представляется возможным.

Выводы

В статье представлены первые шаги по работе с временными рядами, которые можно сформировать из форм статистической отчетности

производственного сектора пенитенциарной системы. Представлено решение задач по преобразованию временных рядов для последующего исследования взаимных распределений. Предложен ряд алгоритмов, благодаря которым представляется возможность сопоставить данные, учет которых ведется разными методами. Исследование временных рядов – является достаточно обширной темой для изучения. В качестве продолжения данной работы могут быть рассмотрены исследования, которые направлены на прогноз временных рядов исходя из их взаимного распределения, цикличности и трендов.

Литература

1. Царькова Е.Г. Математическая модель управления надежностью системы интеллектуального видеонаблюдения на охраняемых объектах уголовно-исполнительной системы // Инженерный вестник Дона, 2022, № 8. URL: ivdon.ru/ru/magazine/archive/n8y2022/7854.
2. Царькова Е.Г. Применение нейросетевого моделирования в управлении процессами ликвидации чрезвычайных ситуаций на территории учреждений уголовно-исполнительной системы // Инженерный вестник Дона, 2022, № 6 URL: ivdon.ru/ru/magazine/archive/n6y2022/7706.
3. Царькова Е.Г. Методы поддержки принятия решений в деятельности органа исполнительной власти по обеспечению надежности ведомственной информационной среды // Инженерный вестник Дона, 2022, № 7. URL: ivdon.ru/ru/magazine/archive/n7y2022/7814.
4. Пономарев Д.С., Горохов М.М., Пономарев С.Б. «Медико-социальная адаптация как фактор повышения качества производственно-трудовой деятельности осужденных». Отчет о НИР / ФКУ НИИ ФСИН России, 2023. С. 5-40.
5. Nielsen A. Practical Time Series Analysis // O'Reilly Media, 2021. 544 p.

6. Пономарёв Д.С., Горохов М.М., Пономарёв С.Б. Обработка форм статистической отчетности Федеральной службы исполнения наказаний на основе методов разведочного анализа данных библиотек языков "Python" и "R" // Вестник Воронежского института ФСИН России. 2023. № 2. С. 106-112.
7. Cory Althoff. The Self-Taught Computer Scientist: The Beginner's Guide to Data Structures & Algorithms // Wiley, 1st edition. 2021. 224 p.
8. George Heineman. Learning Algorithms: A Programmer's Guide to Writing Better Code // O'Reilly Media; 1st edition, 2021, 278 p.
9. Lutz M. Learning Python. 5th Edition Vol.1 O'Reilly, 2019. 832 p.
10. Aditya Bhargava. Grokking Algorithms: An illustrated guide for programmers and other curious people // Manning; 1st edition, 2016. 258 p.

References

1. Czar`kova E.G. Inzhenernyj vestnik Dona, 2022, № 8. URL: ivdon.ru/ru/magazine/archive/n8y2022/7854.
2. Czar`kova E.G. Inzhenernyj vestnik Dona, 2022, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2022/7706.
3. Czar`kova E.G. Inzhenernyj vestnik Dona, 2022, № 7. URL: ivdon.ru/ru/magazine/archive/n7y2022/7814.
4. Ponomarev D.S., Goroxov M.M., Ponomarev S.B. FKU NII FSIN Rossii, 2023. pp. 5-40.
5. Nielsen A. Practical Time Series Analysis // O'Reilly Media, 2021. 544 p.
6. Ponomaryov D.S., Goroxov M.M., Ponomaryov S.B. Vestnik Voronezhskogo instituta FSIN Rossii. 2023. № 2. pp. 106-112.
7. Althoff Cory Wiley, 1st edition. 2021. 224 p.
8. George Heineman. Learning Algorithms: A Programmer's Guide to Writing O'Reilly Media; 1st edition, 2021, 278 p.
9. Lutz M. Learning Python. 5th Edition Vol.1 O'Reilly, 2019. 832 p.



10. Aditya Bhargava. Grokking Algorithms: An illustrated guide for programmers and other curious people. Manning; 1st edition, 2016. 258 p.

Дата поступления: 9.01.2024

Дата публикации: 16.02.2024