

**ГЕНЕРАЦИЯ СИНТЕТИЧЕСКИХ ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДОВ ПУТЕМ
КЛАСТЕРИЗАЦИИ РЕЖИМОВ**

Захаров К.А. (Университет ИТМО), **Ставинова Е.А.** (Университет ИТМО)

Научный руководитель – м.н.с. Ставинова Е.А.
(Университет ИТМО)

Введение. Задача прогнозирования временных рядов является одной из самых актуальных в последние десятилетия. Исследователи заинтересованы в прогнозах значений как финансовых индикаторов, так и цен на активы и процентных ставок. Однако, существенной сложностью данной области является доступ к данным финансовых временных рядов. Как правило, данные такого типа являются закрытыми, доступ к ним предоставляется на платной основе; из-за этих причин возникает проблема недостатка (отсутствия) исторических данных. Поскольку большинство подходов к прогнозированию финансовых временных рядов основано на методах машинного обучения, которые требуют большого объема исторических данных для получения модели приемлемого качества, то для решения данной проблемы зачастую используются синтетические временные ряды. В последнее время данная задача становится популярной, однако лишь небольшое количество методов ее решения охватывает необходимые свойства финансовых временных рядов.

Основная часть. На первом этапе исследования был проведен аналитический обзор научных источников по теме генерации синтетических временных рядов. Были выделены основные недостатки существующих подходов и, в особенности, проблемы существующих методов генерации финансовых временных рядов. Такие ряды отражают динамику перераспределения ресурсов в экономике. Они обладают характерными свойствами, которые мешают их генерации: тяжелые хвосты распределения, убывающая автокорреляция, нестационарность, отсутствие периодичности, эволюционное развитие со скачками, природа которых не ясна или ее невозможно учесть при построении моделей [2].

Вторая часть исследования направлена на разработку метода генерации синтетических временных рядов. Предложенный метод основан на кластеризации режимов временного ряда и последующее независимое обучение генеративных моделей на кластерах. В качестве генеративных моделей были использованы Fourier Flows [1] (модель, основанная на потоках Фурье) и QuantGAN [3] (генеративная состязательная сеть). Также предложена модификация архитектуры сети QuantGAN с использованием дополнительных контролирующего обучение элементов: супервизора, задача которого приблизить полученные синтетические данные к исходным, и второго дискриминатора для классификации рядов, получаемых после прохождения супервизора. Таким образом, генерируемые синтетические временные ряды имеют меньше взрывных значений и являются более близкими к исходному временному ряду, чем в случае генерации с помощью QuantGAN.

Выводы. Работа предложенного метода продемонстрирована с помощью экспериментального сравнения с методами Fourier Flows и QuantGAN на финансовых временных рядах, находящихся в открытом доступе. Результаты показали, что разработанный метод демонстрирует более высокое качество генерации финансовых временных рядов по ряду критериев. Первым критерием является сходство распределений синтетического и исходного ряда, которые были оценены с помощью таких характеристик как дивергенция Йенсена-Шеннона, статистика двухвыборочного теста Колмогорова-Смирнова, коэффициенты асимметрии и эксцесса, среднее квадратов модулей значений спектральной плотности. Вторым критерием является совпадение точек экстремума в синтетическом и исходном

временном ряде – данный критерий был проанализирован путем построения графиков Q–Q. Третьим критерием является наличие затухающей автокорреляционной функции синтетического ряда, поведение которой схоже с аналогичной функцией исходного ряда. Далее были проведены эксперименты по сравнению результатов прогнозирования исходного ряда при обучении прогнозной модели на синтетических временных рядах. Для сравнения результатов использовался ящик с усами, показывающий разброс ошибок MSE на кросс-валидационных выборках из исходного ряда. Разброс ошибок показал, что предложенный метод не ухудшает значения прогнозов, а в некоторых случаях даже улучшает. Результатом работы является открытый код с методом генерации. Разработанный метод можно использовать в случае недостатка данных для решения задач прогнозирования финансовых временных рядов методами машинного обучения.

Список использованных источников:

1. Alaa, A., Chan, A.J., van der Schaar, M.: Generative time-series modeling with fourier flows. In: International Conference on Learning Representations (2020).
2. Chakraborti, A., Toke, I.M., Patriarca, M., Abergel, F.: Econophysics review: I. empirical facts. *Quantitative Finance* 11(7), 991–1012 (2011).
3. Wiese, M., Knobloch, R., Korn, R., Kretschmer, P.: Quant gans: deep generation of financial time series. *Quantitative Finance* 20(9), 1419–1440 (2020).

Захаров К.А. (автор)

Подпись

Ставинова Е.А. (научный руководитель)

Подпись