УДК 519.246.8

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ МЕТОДА SSA И АЛГОРИТМОВ НЕЙРОСЕТЕВОЙ КЛАССИФИКАЦИИ КОМПОНЕНТ

Д. В. БАДРЕТДИНОВ
Научный руководитель В. Г. ЗАМУРАЕВ, канд. физ.-мат. наук, доц.
Белорусско-Российский университет
Могилев, Беларусь

Задача разложения временного ряда на интерпретируемые аддитивные составляющие – тренд, периодическую (сезонную) компоненту и шум – является базовой в анализе сигналов и данных. Метод сингулярного спектрального анализа (SSA), в русскоязычной традиции известный как «Гусеница», предоставляет для этого непараметрический аппарат: базис разложения не задаётся заранее, а определяется структурой самого ряда через сингулярное разложение траекторной матрицы [1, 2]. Такой подход удобен там, где заранее неизвестна модель процесса, но требуется декомпозиция на интерпретируемые компоненты без жёстких априорных предположений. Обзорные сопоставления методологии SSA с родственными подходами к анализу рядов приведены, в частности, в [3].

На практике качество SSA существенно зависит от выбора длины окна вложения, от правила группировки элементарных компонент и от интерпретации сингулярных троек. При «ручном» сценарии возрастает роль субъективности. Для разработки и отладки программного средства полезно иметь контролируемую постановку: синтетические ряды с известными истинными значениями тренда, сезонной компоненты и шума позволяют объективно сравнивать восстановленные компоненты с эталоном и воспроизводить эксперимент.

В программе эта задача разделена на несколько частей: создание данных, выполнение SSA-разложения, группировка компонент, расчёт метрик и отображение результатов пользователю. Такое разделение упрощает разработку: вычисления не находятся внутри окон интерфейса, поэтому один и тот же алгоритм можно запускать как из графического приложения, так и из отдельных консольных сценариев.

Цель доклада – представить разработанное программное средство, которое позволяет сгенерировать синтетический ряд, выполнить SSA-декомпозицию, посмотреть промежуточные результаты метода, вручную сгруппировать компоненты и оценить качество восстановления тренда и сезонности.

Синтетический ряд в программе строится как сумма трёх частей: заданного тренда, сезонной составляющей в виде одной или нескольких гармоник и добавленного шума. Вместе с итоговым рядом сохраняются его истинные компоненты и параметры генерации. Пользователь может менять длину ряда, параметры тренда и гармоник, уровень шума, а на этапе SSA – длину окна и состав компонент, выбранных для восстановления.

Генератор поддерживает несколько видов тренда, набор гармоник с заданной амплитудой, частотой и фазой, а также разные варианты шума. Для

демонстраций используются готовые наборы параметров. Случайность при генерации фиксируется, поэтому один и тот же ряд можно повторно получить с теми же значениями. Это позволяет сравнивать результат SSA с заранее известными истинными компонентами.

Алгоритм SSA в программе выполняется по стандартным шагам: из исходного ряда строится траекторная матрица, затем проводится сингулярное разложение, выделяются элементарные компоненты, которые после группировки с помощью диагонального усреднения преобразуются обратно во временные ряды. Для проверки корректности реализации использованы контрольные примеры с известным аналитическим разложением. После этого пользователь может объединить компоненты в группы тренда, сезонности и шума. При этом доступны не только итоговые графики, но и промежуточные объекты метода: траекторная матрица, сингулярные числа и сингулярные векторы.

Работа алгоритма проверяется по нескольким признакам: матрицы должны иметь ожидаемые размеры, сумма всех восстановленных элементарных компонент должна совпадать с исходным рядом, а диагональное усреднение должно правильно восстанавливать ряд из матрицы. Группировка компонент выполняется вручную, но программа также может перебрать несколько вариантов, чтобы помочь найти более удачное разделение на тренд, сезонность и шум.

Качество восстановления оценивается отдельно для тренда и сезонности. Используются три метрики: MAE, RMSE и коэффициент корреляции Пирсона. MAE отражает среднюю абсолютную величину ошибки, RMSE более чувствителен к крупным отклонениям, а коэффициент корреляции характеризует линейную близость восстановленной компоненты к фактической. Для получения единого показателя качества значения RMSE для тренда и сезонности нормируются на размах соответствующей истинной компоненты, после чего вычисляется евклидова норма полученных величин. Нормировка устраняет необоснованное доминирование компоненты с большим абсолютным масштабом, а покомпонентная оценка не даёт малой суммарной ошибке скрыть значительные ошибки в отдельных компонентах.

Используются следующие метрики:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|;\ \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2};$$

$$r = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})^2}};$$

$$Q = \sqrt{\left(\frac{\mathrm{RMSE}_T}{\max y_T - \min y_T}\right)^2 + \left(\frac{\mathrm{RMSE}_S}{\max y_S - \min y_S}\right)^2},$$

где n – количество точек ряда; y_i – истинное значение компоненты в момент времени i; $\hat{y}_i$ – восстановленное значение этой же компоненты; $\overline{y}$ – среднее

значение истинной компоненты; $\overline{\hat{y}_i}$ – среднее значение восстановленной компоненты; RMSE_T, RMSE_S – корень из среднеквадратичной ошибки для тренда и сезонной компоненты соответственно; y_T , y_S – векторы истинных значений тренда и сезонной компоненты.

Программа написана на Python с использованием NumPy, Matplotlib, pandas и фреймворка графического интерфейса PySide6. Генератор данных, SSA-алгоритм, группировка компонент и расчёт метрик реализованы отдельно от интерфейса. Сначала создаётся синтетический ряд с известными компонентами, затем выполняются SSA-разложение и группировка, после чего строятся графики и отчёт с метриками. Помимо графического интерфейса, предусмотрены запуск из консоли, сохранение результатов в CSV и формирование сводных таблиц.

Практический результат работы – программа для повторяемых экспериментов с SSA. В ней можно сгенерировать ряд с известными компонентами, выполнить разложение, сгруппировать компоненты и оценить качество восстановления по метрикам. На текущем этапе основной акцент сделан на синтетических данных, ручной группировке и переборе вариантов группировки.

Интерфейс построен по шагам: задание параметров генерации, настройка SSA, просмотр сингулярных чисел и матриц, группировка компонент, сравнение с истинными компонентами и просмотр метрик. Такое устройство помогает показать, как длина окна и выбранные группы компонент влияют на качество восстановления.

Перспективы развития включают подключение реальных наборов данных, автоматизацию группировки компонент, перебор длины окна и использование методов машинного обучения. На этой основе в дальнейшем планируется разработать нейросетевой модуль для классификации элементарных компонент SSA на трендовые, сезонные и шумовые; в настоящий момент уже начат подбор подходящей архитектуры нейронной сети. Также планируется расширить набор контрольных сценариев, добавить экспорт отчётов и дополнительные способы диагностики SSA, например анализ взвешенных корреляций компонент.

Для проверки программы используются автоматические тесты. Они помогают контролировать, что SSA-разложение, восстановление компонент и расчёт метрик продолжают работать правильно после изменений в коде.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. **Голяндина, Н. Э.** Метод «Гусеница»-SSA: анализ временных рядов : учеб. пособие / Н. Э. Голяндина. – СПб. : СПбГУ, 2004. – 76 с.

2. **Голяндина, Н. Э.** Анализ временных рядов: структурный подход. Метод SSA / Н. Э. Голяндина, В. В. Некруткин, А. А. Жиглявский. – М. : ЛКИ, 2012. – 498 с.

3. **Hassani, H.** Singular Spectrum Analysis: Methodology and Comparison / H. Hassani // Journal of Data Science. – 2007. – Vol. 5. – P. 239–257.