

ИССЛЕДОВАНИЕ ВЛИЯНИЯ ШИРИНЫ ОКНА ДЕКОМПОЗИЦИИ В СИНГУЛЯРНОМ РАЗЛОЖЕНИИ

К.В. Новицкая, И.А. Ботыгин
Томский политехнический университет
E-mail: kvn8@tpu.ru

Введение

Результаты анализа данных методом сингулярного спектрального анализа (SSA) зависят от выбора параметров декомпозиции временного ряда. Один из основных таких параметров – это ширина окна декомпозиции L [1]. Величина этого параметра влияет на размер составляемой при декомпозиции матрицы, определяющей итоговые результаты анализа.

Экспериментальные исследования влияния ширины окна на разделимость составляющих компонентов проводились на данных, полученных из системы климатического мониторинга [2]. Анализировались метеорологические параметры за период с 01.10.2018 по 15.10.2019 годы. Полные данные содержат 6162772 значений по 17 параметрам.

Экспериментальные исследования 1

Для анализа структуры временных рядов бессмысленно выбирать ширину окна L больше, чем половина длины ряда ($L = N/2$, где N – это длина всего временного ряда) [1]. Чем ближе ширина окна к $N/2$, тем более детально разложение временного ряда. Маленькая ширина окна, как и слишком большая, может перемешать интерпретируемые компоненты.

В качестве инструментария использовался пакет Rssa языка программирования R [3]. Пакет Rssa предоставляет несколько методов сингулярного разложения (SVD). Методы «nutrlan» и «rproack» применяются для данных большого объема. Для данных небольшого размера в документации рекомендуется использовать методы «eigen» или «svd».

Влияние ширины окна на разложение временного ряда на пары собственных векторов представлено на рисунках 1 и 2. Графики пар собственных векторов помогают определить, какие векторы могли бы отвечать за сезонность.

Из рисунка 1 видно, что пары векторов 4-5 и 6-7 образуют окружности. Это значит, что составляющие исследуемого временного ряда – гармонические. Но так как эти окружности имеют большое количество спиралей внутри, форма сезонности сложная. Пара 5-6, вероятно, тоже указывает на гармоническую природу данных. Остальные векторы, предположительно, характеризуют шум.

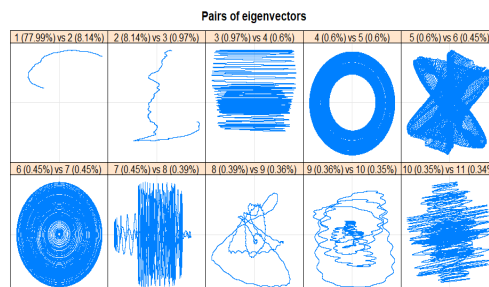


Рис.1. Пары собственных векторов, $L = N/2$

На рисунке 2 представлены результаты разложения с размером окна $L = N/3$. Пара векторов, отвечающая за сезонность теперь одна и точнее выделена (толщина линии окружности имеет меньший размер).

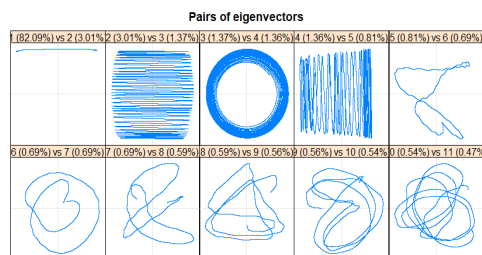


Рис. 2. Пары собственных векторов, $L = N/3$

Еще одним индикатором влияния ширины окна на разложение временного ряда являются W -корреляционные матрицы [1]. W -корреляционные матрицы позволяют определить наличие сильной или слабой связи между восстановленными компонентами. Из графика на рисунке 3 для ширины окна $L = N/2$ видно, что первые 7 компонентов (от $F1$ до $F7$), приблизительно, имеют сильную связь и соответствуют сглаженному варианту исходного ряда. Слабая связь (компоненты $F13$ - $F17$) указывает на наличие шума.

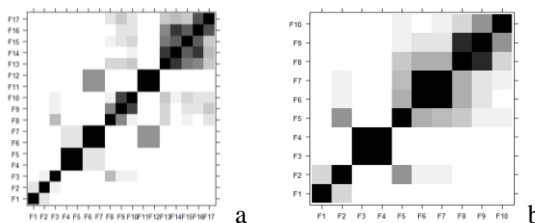


Рис. 3. Матрицы W -корреляций $L = N/2$ и $L = N/3$.

Можно заметить, что пары $F4$ - $F5$, $F6$ - $F7$ на матрице W -корреляций для окна $L = N/2$ имеют

сильную взаимосвязь. И на графиках пар собственных векторов для окна $L = N/2$ те же пары векторов отвечают за сезонность. При этом пара векторов F5-F6 (рис.1), предположительно, указывающих на гармоническую природу данных имеют слабую взаимосвязь на W-корреляционной матрице (рис. 3). Таким образом, можно исключить группы векторов, не несущие никакой полезной информации.

Экспериментальные исследования 2

В этих экспериментах продемонстрированы динамика разделимости групп векторов в зависимости от следующих значений ширины окна: $L = 20, L = 50, L = 70, L = 100, L = 120, L = 150$ (рис.4 – рис.6), Полная длина ряда $N = 200$ значениям.

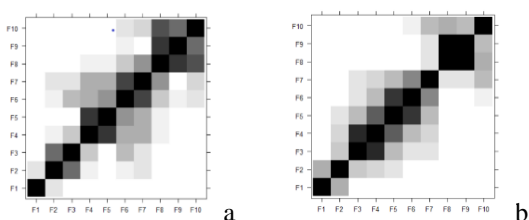


Рис. 4. Матрицы w-корреляций, $L = 20$ и 50

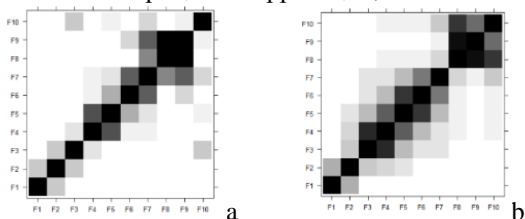


Рис. 5. Матрицы w-корреляций, $L = 120$ и 150

Слишком большая или слишком маленькая ширина окна L ведет к менее детальному SVD разложению, вследствие симметричности разложения при минимальных и максимальных L (левые сингулярные векторы тождественны правым сингулярным векторам). Соответственно, реконструкция и прогнозирование при таких результатах разложения тоже дают неточные результаты.

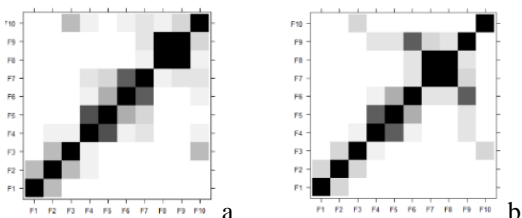


Рис. 5. Матрицы w-корреляций, $L = 70$ и 100

Видно, что при $L = 100$ W-корреляционная матрица (рис. 5, b) представляет собой наиболее четкую линию диагонали. Это значит, что можно выделить ведущие гармоники сезонности; остальные компоненты имеют относительно

большие значения w-корреляции. Также это значит, что больше компонентов можно выделить, по сравнению с более «зашумленными» матрицами.

Заключение

В данной работе было проведено исследование результатов SSA в зависимости от ширины окна декомпозиции. Оптимальная ширина окна должна быть, примерно, равна половине длины исследуемого ряда. Слишком малые значения ширины окна декомпозиции ухудшают разделимость данных. W-корреляционные матрицы помогают интерпретировать графики пар собственных векторов и исключить ненужные группы векторов. Ширину окна больше половины длины ряда или слишком маленькие значения нет смысла использовать, т.к. это сильно искажает результаты разложения.

Список использованных источников

1. Analysis of time series structure: SSA and related techniques / N. Golyandina, V. Nekrutkin, and A. Zhigljavsky.
2. Система климатического мониторинга // URL: <http://mon.imces.ru> (дата обращения: 15.01.2020).
3. Документация пакета Rssa. [Электронный ресурс] / CRAN. – URL: <https://cran.r-project.org/web/packages/Rssa/Rssa.pdf>
4. SSA (метод). [Электронный ресурс] / Википедия. – URL: [https://ru.wikipedia.org/wiki/SSA_\(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4\)](https://ru.wikipedia.org/wiki/SSA_(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4))
5. SSA (метод). [Электронный ресурс] / Википедия. – URL: [https://ru.wikipedia.org/wiki/SSA_\(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4\)](https://ru.wikipedia.org/wiki/SSA_(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4))
6. Golyandina, Nina & Korobeynikov, Anton. (2014). Basic Singular Spectrum Analysis and Forecasting with R. Computational Statistics & Data Analysis. [Электронный ресурс] / ResearchGate – URL: https://www.researchgate.net/publication/228092069_Basic_Singular_Spectrum_Analysis_and_Forecasting_with_R