

ОБЗОР СРЕДСТВ ЯЗЫКА R ДЛЯ НЕПАРАМЕТРИЧЕСКОГО АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

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

Введение

Существует немало способов анализа временных рядов с использованием разных языков, библиотек и даже специальных программ. Выбор этих средств также зависит от исследуемых данных и цели, для которой они анализируются. Например, в простых задачах нет необходимости использовать специализированные языки и можно обойтись стандартным функционалом, в других случаях, в виду сложности анализа или большого объема данных, стоит использовать готовые среды и библиотеки.

Современные программные средства предоставляют широкий выбор инструментов для проведения статистического анализа, но даже в рамках одного языка присутствует множество различных инструментов. Язык R широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистического анализа.

Дополнительные пакеты для работы с языком R и их документация распространяются через CRAN (акроним Comprehensive R Archive Network) – сеть ftp и веб-серверов по всему миру, которые хранят идентичные, обновленные версии кода и документации для R.

В настоящей работе рассматриваются пакеты для использования сингулярного спектрального анализа (SSA), сингулярного разложения (SVD) и метода главных компонент (PCA) (рис. 1).

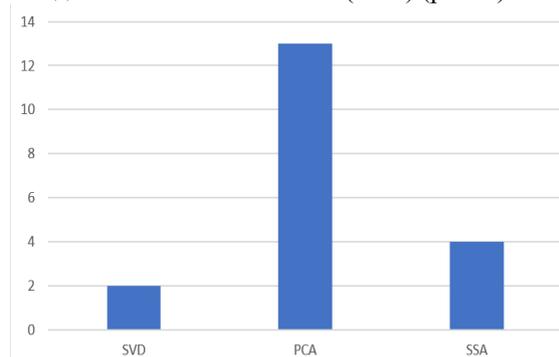


Рис. 1. Количество пакетов для каждого метода на CRAN

Пакеты для SSA

1. **ASSA** (Applied Singular Spectrum Analysis). Пакет содержит функции для моделирования и разложения временных рядов на главные компоненты с использованием сингулярного спектрального анализа.

В пакете всего 7 функций и 2 набора данных. Пакет ASSA реализует методы декомпозиции

временных рядов и моделирования, основанные на сингулярном спектральном анализе (SSA) и многомерном сингулярном спектральном анализе (MSSA). Текущая версия пакета включает инструменты, предназначенные для извлечения бизнес-циклов и вычисления трендовых линий.

2. **Rfssa** (Functional Singular Spectrum Analysis). В пакете представлены методы и инструменты для реализации функционального сингулярного спектрального анализа для временных рядов. В частности, реализованы функции для реконструкции рядов, построения графиков, а также функция для демонстрации приложений launchApp. Эта функция запускает приложение, которое может быть использовано, чтобы помочь исследователю лучше понять функциональный сингулярный спектральный анализ. Приложение позволяет пользователю выполнять одномерный или многомерный функциональный анализ сингулярного спектра (в зависимости от введенного типа параметра) на различных типах данных, включая имитированные и реальные данные, доступные через сервер. Для приложения требуется дополнительный пакет shiny.

Использование пакета начинается с декомпозиции временных рядов с использованием функционального сингулярного спектрального анализа. Затем требуется подходящая группировка главных компонент для реконструкции, которую можно выполнить, на основе визуального представления результатов декомпозиции. В качестве альтернативы можно изучить матрицу w -корреляций.

Заключительным этапом является реконструкция главных компонент в аддитивные объекты, сумма которых аппроксимирует исходный одномерный или многомерный функциональный временной ряд.

3. **Rssa**. В пакете собраны методы и инструменты сингулярного спектрального анализа, включая декомпозицию, прогнозирование и заполнение пробелов для одномерных и многомерных временных рядов.

Это наиболее полный и понятный в использовании пакет для SSA анализа временных рядов. Обычно использование пакета начинается с декомпозиции временного ряда с помощью функции ssa. После этого требуется соответствующая группировка элементарных временных рядов. Это можно сделать эвристически на основе визуального анализа. Также можно рассмотреть так называемую W -корреляционную матрицу. Следующий шаг включает в себя

реконструкцию временных рядов с использованием выбранной группировки. Шаг заканчивается оценкой частоты (parestimate) и прогнозированием временного ряда с использованием включенных функций forecast, rforecast, vforecast.

4. **Spectral.methods**. Пакет содержит некоторые реализации сингулярного спектрального анализа для заполнения промежутков и спектрального разложения временных рядов. Здесь реализован итерационный метод заполнения пробелов SSA Кондрашова и Гиля [1]. Все вычисления SSA выполняются с помощью усеченного и быстрого алгоритма SSA Коробейникова, представленного в пакете Rssa.

В пакете, в основном, находятся функции для заполнения пробелов во временных рядах и фильтрации (или декомпозиции) временных рядов. Кроме дополнительных функций для SSA анализа, в данном пакете намного больше функций для построения различных графиков результатов работы.

Пакеты для SVD

В основе SSA лежит сингулярное разложение (Singular Value Decomposition, SVD) исследуемого временного ряда. Пакетов для SVD всего два: RobRSVD и svdvisual.

Пакет RobRSVD содержит функции для вычисления SVD, regularized SVD, robust SVD и robust regularized SVD методов. Можно применять для расширения функций другого пакета для SSA анализа.

Пакет Svdvisual включает некоторые инструменты визуализации.

Пакеты для PCA

Пакетов для метода главных компонент в языке R. намного больше. Ниже представлены их краткие описания:

- **bootSVD**: реализует быстрый и точный bootstrap анализ главных компонент и сингулярное разложение для многомерных данных.

- **hdpca**: анализ главных компонент в многомерных (High-Dimensional Data) данных. Содержит расширенные функции для анализа такого вида данных.

- **juvcoords**: предоставляет функции для стандартизации и применение вида трансформации матриц с помощью белого шума, а также для метода основных компонент (PCA). Основное преимущество этого пакета перед альтернативами типа prcomp заключается в том, что juvcoords позволяет легко конвертировать данные между исходными и преобразованными координатами.

- **crpca**: методы для выполнения общего анализа главных компонент (Common PCA).

- **onlinePCA**: онлайн PCA для многомерных и функциональных данных с использованием

методов возмущений, инкрементных методов низкого ранга и методов стохастической оптимизации.

- **SPCAvRP**: реализует алгоритм SPCAvRP, разработанный и проанализированный в работе [1].

- **SuperPCA** (Supervised Principal Component Analysis): используется для уменьшения размерности сложных данных под контролем вспомогательной информации.

Заключение

В данной работе были рассмотрены большая часть пакетов для применения методов SSA, PCA и SVD в языке R. Некоторые найденные пакеты (например, для метода PCA) являются узко направленными – разработаны под некоторые собственные методы исследователей.

Пакеты для SVD стоит использовать при необходимости более углубленного или специального анализа. Среди всех пакетов для SSA более полный и эффективный функционал имеет пакет Rssa. Однако, остальные пакеты также можно использовать при детальных исследованиях.

Список использованных источников

1. Kondrashov, D. & Ghil, Michael. (2006). Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Processes in Geophysics*.
2. Gataric, Milana & Wang, Tengyao & Samworth, Richard. (2017). Sparse principal component analysis via random projections. [Электронный ресурс] / ResearchGate – URL: https://www.researchgate.net/publication/321873491_Sparse_principal_component_analysis_via_random_projections
3. Документация пакетов в CRAN. [Электронный ресурс] / CRAN. – URL: <https://cran.r-project.org/web/packages/>
4. Unsupervised learning for time series data: Singular spectrum versus principal components analysis. [Электронный ресурс] – URL: <http://karthur.org/2017/learning-for-time-series-ssa-vs-pca.html>
5. Analysis of time series structure: SSA and related techniques / N. Golyandina, V. Nekrutkin, and A. Zhigljavsky.
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
7. 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))