

Выбор параметров при автоматическом выделении трендовых и периодических составляющих временного ряда в рамках подхода “Гусеница”-SSA

Александров Ф.И., theo@pdm.ras.ru
Голяндина Н.Э., nina@ng1174.spb.edu

Одним из мощных методов анализа и прогноза временных рядов является достаточно новый и быстро развивающийся метод “Гусеница”-SSA (метод разрабатывался как в России, так и независимо в Великобритании и США под разными именами, “Гусеница” и SSA - Singular Spectrum Analysis). Описание метода, его теории, примеры применения и ссылки на связанную с ним литературу можно найти в книгах [1-3]. Метод позволяет выделять интерпретируемые составляющие временного ряда, такие как тренд или периодические составляющие, сглаживать ряд, отделять сигнал от шума. Он не требует априорного задания параметрической модели ряда.

Этап сингулярного разложения предоставляет разбиение ряда на элементарные восстановленные ряды – аддитивные компоненты исходного ряда. Для того чтобы выделить искомую составляющую ряда, необходимо найти все элементарные восстановленные ряды, сумма которых её формирует. При поиске нужных компонент используются данные сингулярного разложения, назовем такой поиск процедурой идентификации.

Исходно, идентификация при применении метода проводилась интерактивно, в основном визуальным способом на основе теоретических сведений. Примеры такой идентификации приведены в [1]. С одной стороны, интерактивность является положительной стороной метода, так как дает возможность сознательного его применения со стороны пользователя и приводит к более качественному и глубокому анализу. С другой стороны, в ряде случаев (например, при необходимости анализа большого количества однотипных данных) возникает потребность в автоматизации идентификации составляющих ряда. На основе свойств разложения, используя теоретический аппарат метода “Гусеница”-SSA, были разработаны методы идентификации трендовых и гармонических (возможно, модулированных) компонент [4-5]. Методы позволяют автоматизировать процесс идентификации, а также предоставляют дополнительную информацию, которая может быть использована также и при интерактивной идентификации. Это удобно, например, при исследовании сложных рядов. Для демонстрации работы методов была написана программа [5], позволяющая выделять искомую составляющую автоматически.

Разработанные методы идентификации управляются заданием параметров и пороговых значений используемых ими критериев. Поэтому важной задачей является исследование методов с точки зрения выбора их параметров, проводимое с учётом как поставленной задачи, так и выбранной предметной области. При этом необходимо изучить результаты работы методов на модельных рядах, так как в случае реальных рядов результаты сложно интерпретировать. В данной работе были проведены исследования по выбору параметров и пороговых значений, в частности, с помощью средств статистического моделирования.

1. N.Golyandina, V.Nekrutkin, A.Zhigljavsky (2001) Analysis of Time Series Structure: SSA and Related Techniques, Boca Raton, Chapman&Hall/CRC.

2. J.B.Elsner, A.A.Tsonis (1996) Singular Spectrum Analysis. A New Tool in Time Series Analysis, Plenum Press, New York.
3. Д.Л.Данилов, А.А.Жиглявский (ред.) (1997) Главные компоненты временных рядов: метод “Гусеница”, Санкт-Петербург, изд-во СПбГУ.
4. R.Vautard, P.Yiou, and M.Chil (1992) Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, Physica D 58, с. 95-126.
5. Ф.Александров (2003) Выделение аддитивных компонент временного ряда на основе метода “Гусеница”, Санкт-Петербург, СПбГУ, дипломная работа.