

tsX: An R package for the exploratory analysis of a large collection of time-series

G. Subramaniam,¹ R. Varadhan,² S. Urbanek,¹ and S. Epstein¹

¹ AT&T Labs - Research, ² Division of Geriatric Medicine and Gerontology,
School of Medicine, Johns Hopkins University,

1 Abstract

We discuss **tsX**, an R package for exploring a large collection of time series. This was motivated by a time series data mining problem in telecommunication network data, where given a large number of time series the objective was to identify time series that have potentially “interesting behavior” (Subramaniam and Varadhan 2007 and 2008). Smooth representation of each time-series (which can be unequally spaced) is first obtained using automatic smoothing parameter selection procedures. Various features of the time-series are then derived based on the smoothed functions. Some of the useful features include mean value, scale, first and second derivatives, critical points, wiggleness, signal/noise ration, and potential outliers. A key feature of this package is that it provides a choice of different smoothing techniques and automatic smoothing parameter estimation procedures. A comprehensive simulation study of these smoothing techniques was performed to evaluate the performance of the smoothing techniques in terms of their ability to estimate the smooth underlying function and the first and second derivatives (Varadhan and Subramaniam 2009). This provided validation of the smoothing techniques available in **tsX** for exploratory use in time series data mining setting. **tsX** provides useful visualization techniques that provide the capability to easily identify and collect curves exhibiting interesting or anomalous behavior using interactive graphics.

References

- Subramaniam. G and Varadhan. R (2007). *Feature Extraction Using FDA*. JSM 2007, (Salt Lake City, UT,USA), Aug. 2007
- Subramaniam. G and Varadhan. R (2008). *Borrowing Strength in Time Series Data Mining*. JSM 2008, (Denver, CO,USA), Aug. 2008
- Varadhan. R and Subramaniam. G (2009). *Automatic Numerical Differentiation of Noisy, Time-Ordered Data in R*. UseR!2009, (Renne, France), Jul. 2009