Dimension Reduction for Clustering Time Series Using Global Characteristics

Xiaozhe Wang[1], Kate A. Smith[1], and Rob J. Hyndman[2]

[1] Faculty of Information Technology, [2] Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia
{catherine.wang, kate.smith}@infotech.monash.edu.au
rob.hyndman@buseco.monash.edu.au

Abstract. Existing methods for time series clustering rely on the actual data values can become impractical since the methods do not easily handle dataset with high dimensionality, missing value, or different lengths. In this paper, a dimension reduction method is proposed that replaces the raw data with some global measures of time series characteristics. These measures are then clustered using a self-organizing map. The proposed approach has been tested using benchmark time series previously reported for time series clustering, and is shown to yield useful and robust clustering.

1 Introduction

Clustering time series has become an important topic in data mining [1,3], motivated by several research challenges including similarity searching of bioinformatics sequences. This paper focuses on "whole clustering" [2] using a variety of statistical measures to capture the time series global characteristics, which departs from the common methods of clustering time series based on distance measures applied to the actual values [1,2]. The proposed method seeks to provide a novel method for clustering time series with high dimensionality, varying lengths, and missing value. The dimension reduction is performed by a feature extraction process. These features are: trend, seasonality, serial correlation, non-linearity, skewness, kurtosis, self-similarity, and chaos. For additional dimension reduction and visualization, a self-organizing map (SOM) is used to cluster the features. A total of 15 measures are calculated (9 on the 'raw' data and 6 on the 'decomposed' data) and fed into the clustering process.

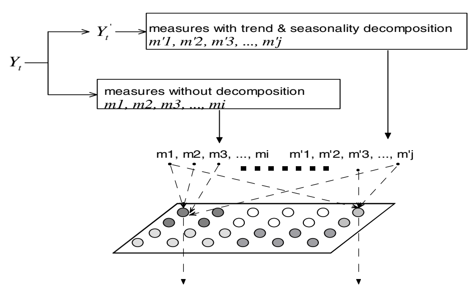

Fig. 1. Method framework with neural network SOM architecture for clustering

V.S. Sunderam et al. (Eds.): ICCS 2005, LNCS 3516, pp 792–795, 2005

2 Measuring Characteristics of Time Series

A time series is the simplest form of temporal data and is a sequence of real numbers collected regularly in time, denoted as $Y_1, \ldots, Y_n$. For each of the features, we attempted to find the most appropriate way to measure the presence of the feature, ultimately scaling the metric to (0,1) to indicate the degree of presence of the feature.

- **Trend and seasonality** are common features of time series, and it is natural to characterize a time series by its degree of trend and seasonality. Once the trend and seasonality of a time series have been measured, we can de-trend and de-seasonalize the time series to enable additional features such as noise or chaos to be more easily detected. We have used the basic decomposition model in [4]:

> - $Y^*_t=T_t+S_t+e_t$, where $Y^*_t=f_\lambda(Y_t)$, $f_\lambda(u)=(u^\lambda-1)/\lambda$ denotes a Box-Cox transformation, T_t denotes the trend at time t and S_t denotes the seasonal component at time t.;The parameter λ is chosen to make e_t as normal as possible. Where the minimum of $\{Y_t\}$ is non-negative, we choose $\lambda \in (-1,1)$ to minimize the Shapiro-Wilk statistic.
> - For seasonal data, the decomposition uses the STL procedure with fixed seasonality. If the data are nonseasonal, $S_t = 0$ and T_t is estimated using a penalized regression spline with smoothing parameter chosen using crossvalidation; The detrended data is $Y_t^{'} = f_\lambda^{-1}(Y_t^* - \hat{T}_t)$, the deseasonalised data is $Y_t^{''} = f_\lambda^{-1}(Y_t^* - \hat{S}_t)$ and the detrended and deseasonalised data is $Y_t^{'''} = f_\lambda^{-1}(Y_t^* - \hat{T}_t - \hat{S}_t)$ where $\hat{T}_t$ and $\hat{S}_t$ denote the trend and seasonal estimates.
> - The measures of trend and seasonality are: $1 - Var(Y_t^{'})/Var(Y_t)$ and $1 - Var(Y_t^{''})/Var(Y_t)$

- The **periodicity** of the seasonal pattern is used as an additional measure. We measure the period using the following algorithm:

> - Detrend time series using a regression spline with 3 knots.
> - Find $r_k = Corr(Y_t, Y_{t-k})$ (autocorrelation function) for all lags k up to 1/3 of series length. And look for peaks and troughs in autocorrelation function.
> - Frequency is first peak provided: a) there is also a trough before it, b) the difference between peak and trough is at least 0.1, c) the peak corresponds to positive correlation.
> - If no such peak is found, frequency is set to 1.

- To measure the degree of **serial correlation** of the data set, we use Q_h the Box-Pierce statistic [4] where $Q_h = n\sum_{k=1}^{h} r_k^2$.
- To measure **non-linear autoregressive structure**, nonlinear time series models have been used extensively in recent years to model complex dynamics not adequately represented using linear models [5]. We have used Teräsvirta's neural network statistic for nonlinearity [6].
- **Skewness** is a measure of lack of symmetry and **kurtosis** is a measure of a distribution's peakedness. For the standard normal distribution, the skewness and kurtosis are both zero. For univariate data Y_t, the skewness coefficient is

$S = n^{-1}s^{-3}\sum_{t=1}^{n}(Y_t - \bar{Y})^3$ and the kurtosis coefficient is $K = n^{-1}s^{-4}\sum_{t=1}^{n}(Y_t - \bar{Y})^4 - 3$, where $\bar{Y}$ is the sample mean and s is the sample standard deviation.

- Processes with **long-rang dependence** have attracted a good deal of attention from probabilists and theoretical physicists. The definition of **self-similarity** most related to the properties of time series is the self-similarity parameter or Hurst exponent (H). H is estimated using $H=d+0.5$ by a class of fractional autoregressive integrated moving-average (FARIMA) processes of FARIMA(0,d,0) by maximum likelihood [7].
- Nonlinear dynamical systems often exhibit **chaos**, which is characterized by a Lyapunov Exponent (LE). For a one-dimensional discrete time series, we used the method by Hilborn [8] to calculate LE:

- We consider the rate of divergence of nearby points in the series by looking at the trajectories h periods ahead. Suppose Y_j and Y_i are two points such that $|Y_j - Y_i|$ is small. Then we define $\lambda(Y_i, Y_j) = h^{-1}\log|Y_{j+h} - Y_{i+h}| / |Y_j - Y_i|$.
- We estimate the Lyapunov exponent of the series by averaging these values over all i: $\lambda = n^{-1}\sum_{i=1}^{n}\lambda(Y_i, Y_j)$, where Y_j is the closest point to Y_i such that $i \neq j$ .Scaling transformations

In order to present the clustering algorithm with scaled data in the (0,1), we perform a statistical transformation of the data. For example, to map the correlation measure Q in the range $(0,\infty)$ to a scaled value q in the range (0,1) we use the transformation: $q = (e^{aQ}-1)\big/(b+e^{aQ})$, where a and b are constants chosen so that q satisfies the following conditions: q has 90th percentile of 0.10 when Y_t is standard normal white noise and q has value of 0.9 for a well-known benchmark data set.

3 Clustering and Experimental Evaluation

The central property of SOM is that it forms a nonlinear projection of a high-dimensional data manifold on a regular, low-dimensional (usually 2D) grid [9]. The clustered results can show the data clustering and metric-topological relations of the data items. It has a very powerful visualization output and is useful to understand the mutual dependencies between the variables and data set structure.

To determine the effectiveness of the measures in the proposed approach, we used the time series clustering benchmark datasets, “Reality check Dataset” (www.cs.ucr.edu/~eamonn/TSDMA). The data contains 14 time series of 1000 points in each which normalized into (0,1). Then 15 measures extracted from the data are used as inputs to the SOM. To compare with the benchmarking clusters generated using hierarchical clustering of the raw data points, we re-interpreted the clusters generated by the SOM into a hierarchical structure in (Fig. 2 right). Compared to the hierarchical clustering result (Fig. 2 left) [2], similar clusters have been obtained from our approach. But our clustering results are arguable better, or at least more intuitive. For example, series 1&4 and 9&10 have been grouped far from each other based on the hierarchical clustering using actual points (Fig. 2 left), but a visual inspection of these series shows that they are actually quite similar in character. Using the global measures as inputs, the clustering algorithm is aware of the "whole picture" and recognizes the similarity of these four time series (Fig. 2 right).

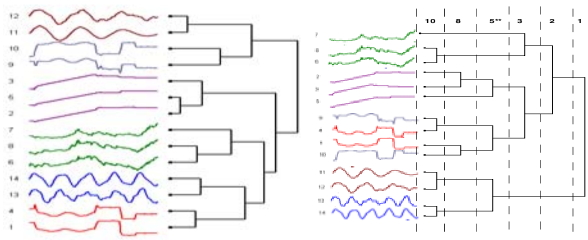

Fig. 2. Clustering results using hierarchical representation (left) and SOM (right)

4 Conclusions and Future Research

In this paper, we have proposed a new method for time series clustering and compared the results to a benchmarked time series clustering data. The empirical results demonstrate that our proposed clustering approach is able to cluster time series with high dimensionality, varying lengths, and missing value. Using only a finite set of global measures as input for clustering, we can still achieve useful clustering. In fact, the knowledge provided to the clustering algorithm by the global measures appears to benefit the quality of the clustering results. In future, a more comprehensive metrics to summarize the time series features will be explored and applied to more datasets.

References

1. Bradley, P. S., Fayyad, U. M.: *Refining Initial Points for K-means Clustering*. In: the 15th international conference on machine learning, Madison, (1998).
2. Keogh, E., Lin, J., Truppel, W.: *Clustering of Time Series Subsequences is Meaningless: Implications for Past and Future Research.* In: the 3rd IEEE International Conference on Data Mining, Melbourne.
3. Wang, C., & Wang, X. S.:*Supporting Content-based Searches on Time Series via Approximation.* In: the 12th international conference on scientific and statistical database management, Berlin, (2000).
4. Makridakis, S., Wheelwright, S. C., Hyndman, R. J.: Forecasting Methods and Applications (3rd Ed.): John Wiley & Sons, Inc., (1998).
5. Harvill, J. L., Ray, B. K., Harvill, J. L.: *Testing for Nonlinearity in a Vector Time Series.* Biometrika, 86, (1999), 728-734.
6. Blake, A. P., Kapetanios, G.: A Radial Basis Function Artificial Neural Network Test for Neglected Nonlinearity. The Econometrics Journal, 6(2), (2003), 357-373.
7. Haslett, J., & Raftery, A. E.: *Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource* (With Discussion). Applied Statistics, 38, (1989), 1-50.
8. Hilborn, R. C.: *Chaos and Nonlinear Dynamics: an Introduction for Scientists and Engineers.* New York: Oxford University Press, (1994).
9. Kohonen, T.: Self-Organizing Maps (Vol. 30): Springer Verlag, (1995).