Clustering of Time Series using Wavelet Transformations as a Feature Extraction Mechanism

Kathy Norman, Ssurey Moon, Felix Huang, Josué Kuri
UCSC Extension. 30164:(003) Machine Learning and Data Mining

May 27, 2015

1 Introduction

A time series is a sequence of data points indexed by time at regular intervals. This model is used to represent a wide range of metrics such as the daily closing price of stocks, temperature, precipitation, population, etc. In the context of machine learning and data mining, clustering of a large set of time series is an exploratory technique aimed at identifying and understanding underlying patterns.

Considering every point of a time series as a dimension results in a high dimensional space which clustering algorithms cannot handle easily. These algorithms depend on a distance measure as a basis to maximize cohesion and separation. In a high dimensional space, the contrast between the nearest and the farthest neighbor becomes smaller making it difficult for clustering algorithms to find meaninful groups [1].

Data dimensionality reduction is an approach to map a high dimensional space into a lower dimensional space such that the main characteristics of the data points in the original space are preserved and clustering on the lower dimensionality space results in meaningful groups. The two types of dimensionality reduction are feature selection and feature extraction. The former consists in selecting a subset of features from the original features. Feature extraction, on the other hand, generates a new set of features through a mapping function.

Feature extraction techniques commonly used include Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT). Of these techniques, SDV is the most effective at reconstructing time series with minimal error. However, its time complexity $O(mn^2)$, where m is the number of time series and n is the lenght of each time series, makes this a computally-intensive approach [2]. A Fast Fourier Transform (FFT) algorithm can compute DFT coefficients in $O(mn \log n)$ and DWT, using a spacial type of wavelet called *Haar wavelet* can achieve $O(mn)$ [2].

In this project we use a feature extraction approach based on DWT using the Haar wavelet as the basis for the transformation. We create a generic framework for clustering of time series using this approach and apply the framework to three types of time series: daily closing stock prices, daily values of exchange rates, and earthquake activity over time for various geographic regions. We use a silhouette coefficient as the criterion to evaluate the quality of the clusterings generated by the framework.

2 Wavelet transformation

Wavelet transformation is a time-frequency domain transformation technique for hierarchical decomposition of signals [3, 4]. The decomposition creates an approximation of the original signal that preseves the trend of the signal, as well as additional data sets that provide increasing levels of detail to reconstruct the original signal. This original signal can be reconstructed without loss of information by applying an inverse wavelet transform to the combination of the approximation signal and all the detail data sets.

Early work on wavelets originated with Morlet in the 1980s as a new tool for seismic signal analysis [5]. Further work by Morlet, Grossman, Meyer, Mallat and Daubechies [6, 7] broght the concept to the mainstream mathematics community with applications in signal processing, statistics and other areas. There is at present a vast body of literature about the foundations and applications of wavelets. The interested reader is referred to [8] and similar works for a comprehensive presentation of the field.

3 Wavelet-based feature extraction

Consider a time series $\overrightarrow{X} \in \mathbb{R}^n$ as an ordered sequence of $n \in 2^J, J \in \mathbb{N}$ numbers. After decomposing $\overrightarrow{X}$ at a resolution $r \in \{2, 4, 6, 8, 10, ...\}$, the coefficients associated to the r level can be represented as a sequence $\{A_r, D_r, D_{r-1}, ..., D_2, D_1\}$. The first element A_r is the approximation coefficients array, and the subsequent elements $D_r, D_{r-1}, ..., D_2, D_1$ are the details coefficients arrays. In this project we use the vector $\widehat{X} = \{A_r, D_r\}$ at specific levels $r \in \{2, 4, 6, 8, 10\}$ as feature vectors for the clustering of a set of time series. The cardinality $|\widehat{X}|$ decreases as r increases. It results, on one hand, in a reconstruction with less fidelity than the original signal $\overrightarrow{X}$ and, on the other hand, on clusterings of potentially better quality because of the smaller dimensionality of $\widehat{X}$.

4 Experimental evaluation

The purpose of the experimental evaluation of the generic framework for clustering of time series is to assess the relative efficiency of the framework on a set of contrasting application domains ranging from financial markets to geological phenomena. Variations in the cardinality $|\widehat{X}|$ of feature vectors, the choosen clustering algorithm, as well as the number of clusters, are considered to assess what combination of algorithm and input parameters is the most appropriate for specific situations.

4.1 Evaluation criteria

The three main criteria used to evaluate clusters are [9]:

External. Measure the extent to which cluster labels match externally supplied labels.

Internal. Measure the goodness of a clustering without respect to external information.

Relative. Compare two different clusterings.

We evaluate our clustering framework using an internal criterion known as silhouette coefficient. The coefficient is defined as $S = 1 - a/b$, where a is a measure of cohesion -how closely related are objects in a cluster-, and b is a measure of separation -how distinct a cluster is from other clusters-. The value of S is typically in $[0, 1]$. A clustering with a silhouette coefficient close to 1 is deemed better than a clustering with a silhouette coefficient close to 0.

4.2 Data description

4.2.1 Stock closing prices

This data set consists of the time series of daily closing prices for one hundred stocks from diverse industries (Internet, telecommunications equipment and service providers, entertainment, media, airlines, etc.) between April 20th 2011 and May 16th 2015. The time series represent 1024 trading days, and are normalized using as basis the closing price of the first day of the interval (April 20th 2011).

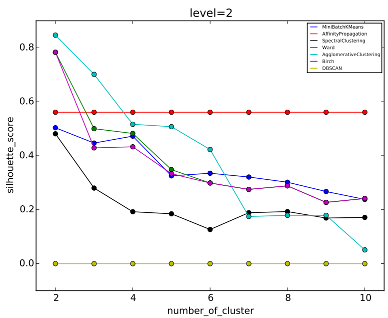
4.2.2 Historic exchange rates

4.2.3 Historic earthquake data

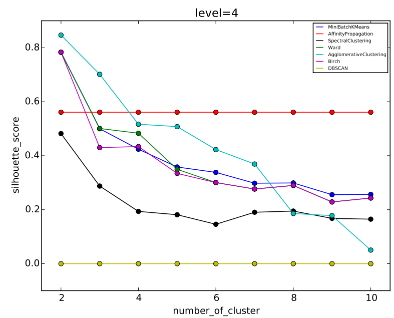
4.3 Performance evaluation

5 Conclusions

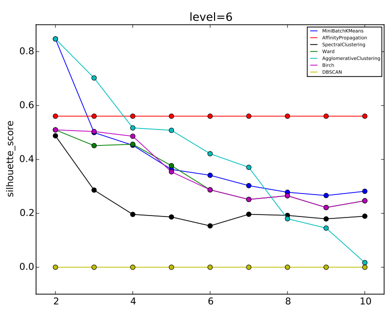
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.



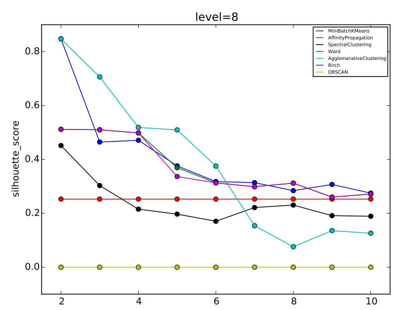
(a) Level 2



(b) Level 4

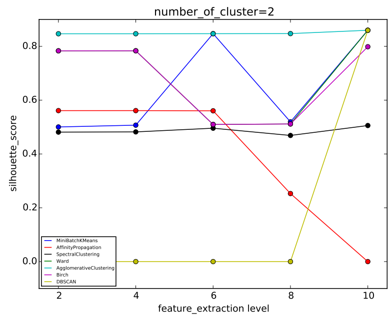


(c) Level 6

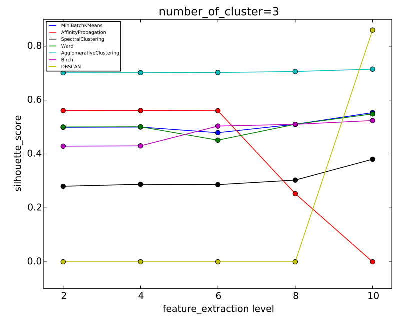


(d) Level 8

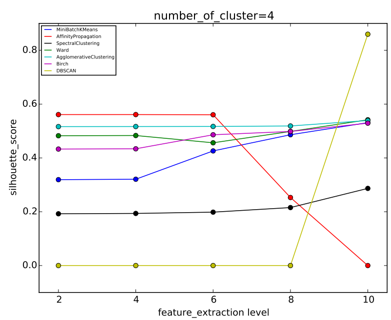
Figure 1: Silhouette score for various clustering levels.



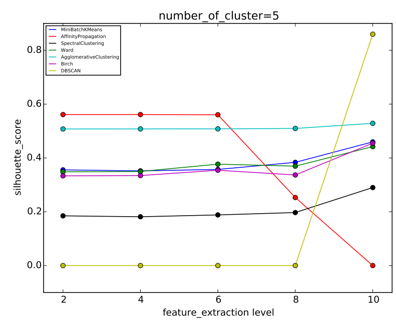
(a) Two clusters



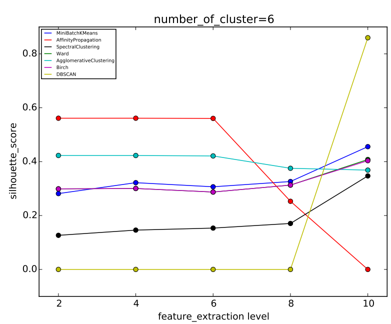
(b) Three clusters



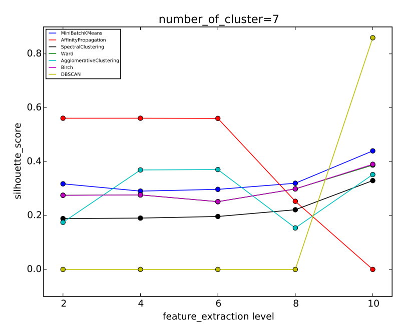
(c) Four clusters



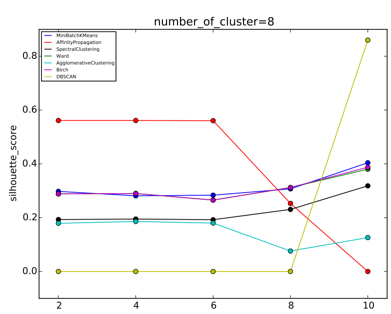
(d) Five clusters



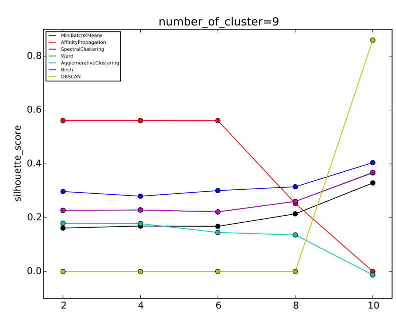
(e) Six clusters



(f) Seven clusters



(g) Eight clusters



(h) Nine clusters

Figure 2: Silhouette score for various clustering levels.

References

[1] K. Beyen, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is nearest neighbor meaningful? In *Proceedings of the 7th International Conference on Database Theory*, pp. 217-235, 1999.

[2] H. Zhang, T. B. Ho, Y. Zhang, M.-S. Lin. Unsupervised Feature Extraction for Time Series Clustering Using Orthogonal Wavelet Transform. In *Informatica*, Volume 30, pp. 305-319, 2006.

[3] C. K. Chui. *An Introduction to Wavelets.* Academic Press, San Diego, 1992.

[4] I. Daubechies. Ten Lectures on Wavelets. *SIAM*, Philadelphia, PA, 1992.

[5] J. Morlet, G. Arens, E. Fourgeau and D. Giard. Wave propagation and sampling theory, Part 1: Complex signal land scattering in multilayer media. *Journal of Geophysics,* 47:203-221, 1982.

[6] J. M. Combes, A. Grossman and P. Tchamitchian, editors. *Wavelets, Time-Frequency Methods and Phase Space.* Springer-Verlag, Berlin 1989.

[7] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics,* 41:909-996, 1988.

[8] C. Sidney Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer,* Prentice Hall, NJ 1998.

[9] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining,* Pearson, First Edition, 2005.