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, the coefficients associated to the r level can be represented as a series $\{A_r, D_r, D_{r-1}, ..., D_2, D_1\}$. 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 and results in a reconstruction with less fidelity than the original signal $\overrightarrow{X}$.

4 Experimental evaluation

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.

4.1 Evaluation criteria

There is no explicit way to evaluate the quality of these time-series clustering systems, since it is of unsupervised learning. In supervised learning data set contains true classes of all data points, and they are used to calculate SSE(Sum of Square Errors) to evaluate the quality of the implementation. In contrast, evaluation of clustering implementation on unsupervised learning data set need another approach to internally calculate the quality of derived clusters. The most popular method is to get cohesion and separation. Cohesion means how close each node in a cluster $\sum_i \sum_j (x_{i,j} - m_j)^2$ where i is number of clusters, $x_{i,j}$ is a j_{th} node in the cluster i, and m_i is the center of the cluster i. On the other hand, separation means how far each cluster is from others, $\sum_k \sum_i (m_k - m_i)^2$. The higher those values are, the better nodes are clustered. Silhouette coefficient takes both concepts of cohesion and separation:

$$s = \{ 1 - a/b \ \text{ for } a \leq b, \ b/a - 1 \ \text{ otherwise} \tag{1}$$

where a is the average distance x_i, a random node in the cluster i, and other nodes in the cluster, b is the minimum value the average distances of x_i and nodes in another cluster k. The closer to one the value is, the better nodes are clustered. In this project we use Silhouette coefficient to measure the performance of our implementation. There are a few parameters to consider, such as clustering algorithm, level of complexity, in other words, feature extraction level, and a forced number of clusters. Silhouette score plots point which combination of parameters performs better on different type of time-series data sets.

4.2 Data description

4.2.1 Stock closing pricees

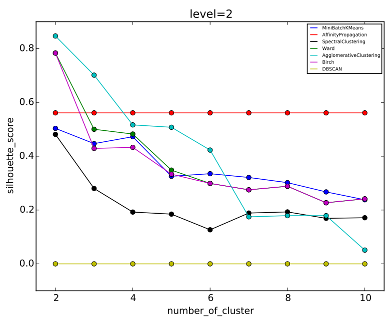
4.2.2 Historic exchange rates

4.2.3 Historic earthquake data

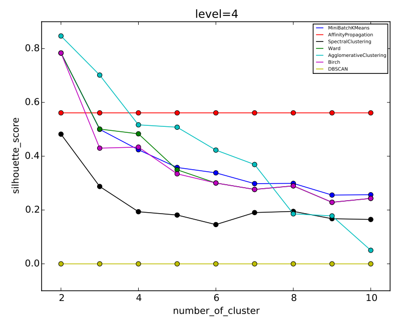
4.3 Performance evaluation

5 Conclusions

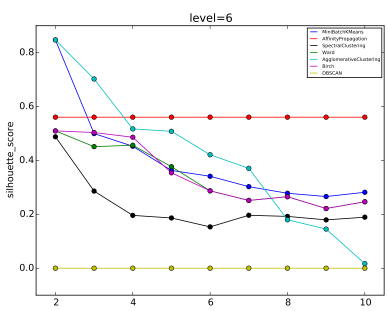
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum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



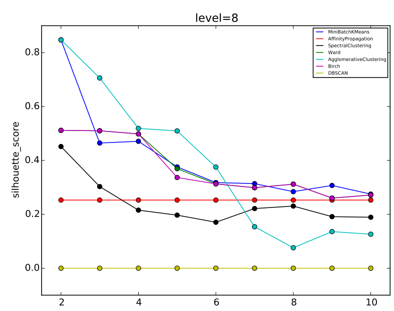
(a) Level 2



(b) Level 4

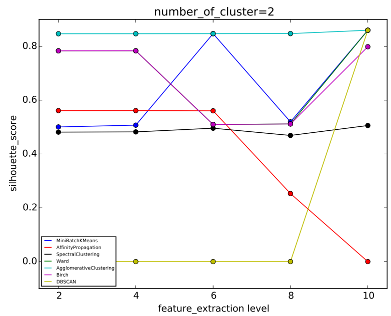


(c) Level 6

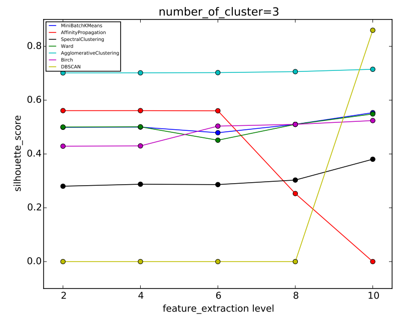


(d) Level 8

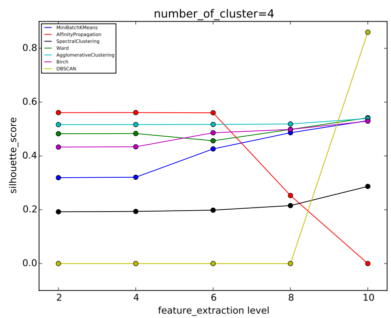
Figure 1: Silhouette score for various clustering levels.



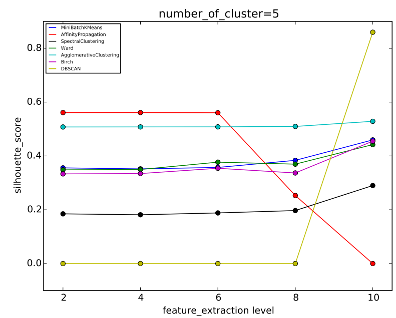
(a) Two clusters



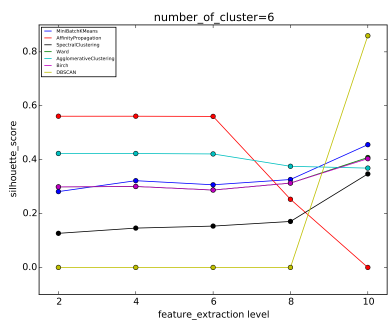
(b) Three clusters



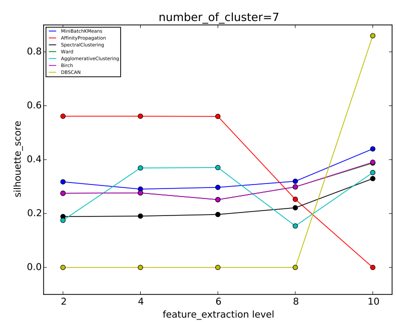
(c) Four clusters



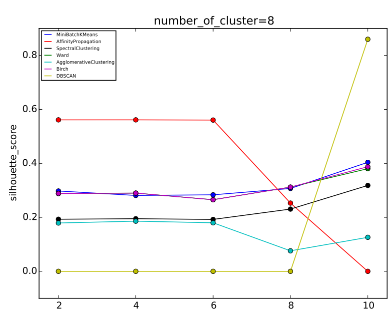
(d) Five clusters



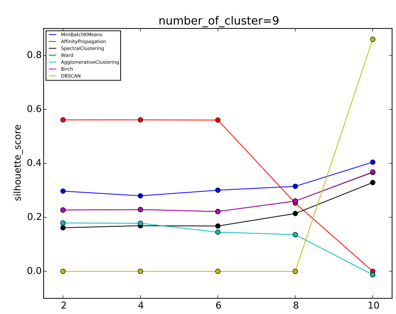
(e) Six clusters



(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.