Time Series Data Mining Tool

Synopsis

Contents

1. Introduction
2. Key Challenges
3. Time Series Representation
   1. SAX (Symbolic Aggregate approximation)
4. Similarity Detection
   1. DTW (Dynamic time warping)
   2. Hierarchical Clustering with SAX
5. Time SeriesForecasting/Prediction
   1. Classic Algorithms - Moving average, geometric average.
   2. Wavelet Smoothing Algorithms - Fractal projection, active moving average
   3. Advanced Algorithms - Kernel smoothing, dynamic average, SAX based Algorithm.
6. Novel/Surprising/Anomalous Behavior Detection
   1. The Markov model-based technique using SAX

**1. Introduction**

A time series is a set of observations Xt , each one being recorded at a specific time t .discrete-time time series is one in which the set T of times at which observations are made is a discrete set. Continuous-time time series are obtained when observations are recorded continuously over some time interval, e.g., when T0 belongs [0,1]. Examples of time series are the daily closing value of the ECG readings and the annual flow volume of the Nile River at Aswan. Time series are very frequently plotted via line charts. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

Time series analysis can be applied to:

* real-valued, continuous data
* discrete numeric data
* discrete symbolic data (i.e. sequences of characters)

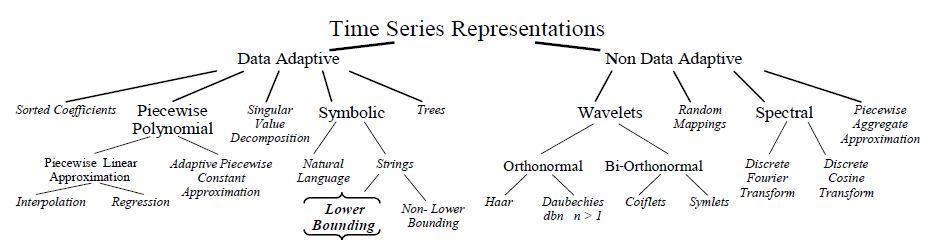
**2. Key Challenges**

The key challenges of time series analysis are listed below.

* Scalability Issues arising from huge data sets
* Calculating trends
* High Dimensionality of data sets

The last decade has seen the introduction of hundreds of algorithms to classify, cluster, segment and index time series. In addition, there has been much work on novel problems such as rule extraction, novelty discovery, and dependency detection. This body of work draws on the fields of statistics, machine learning, signal processing, information retrieval, and mathematics. It is interesting to note that with the exception of indexing, researches in the tasks enumerated above predate not only the decade old interest in data mining, but in computing itself. The key difference between the classic and data mining versions of these problems is simply one of size and scalability; time series data miners routinely encounter data sets that are gigabytes in size. A data mining approach to clustering time series, must explicitly consider the scalability of the algorithm In addition to the large volume of data, most classic machine learning and data mining algorithms do not work well on time series data due to their unique structure; it is often the case that each individual time series has a very high dimensionality, high feature correlation, and large amount of noise which present a difficult challenge in time series data mining tasks. Whereas classic algorithms assume relatively low dimensionality (for example, a few measurements such as “height, weight, blood sugar, etc.”), time series data mining algorithms must be able to deal with dimensionalities in the hundreds or thousands. The problems created by high dimensional data are more than mere computation time considerations; the very meanings of normally intuitive terms such as “similar to” and “cluster forming” become unclear in high dimensional space. The reason is that as dimensionality increases, all objects become essentially equidistant to each other, and thus classification and clustering lose their meaning. This surprising result is known as the “curse of dimensionality” and has been the subject of extensive research. The key insight that allows meaningful time series data mining is that although the actual dimensionality may be high, the intrinsic dimensionality is typically much lower. For this reason, virtually all time series data mining algorithms avoid operating on the original “raw” data; instead, they consider some higher-level representation or abstraction of the data.

Also, in practice the modeling of trends is a difficult task. It is also a task where failure has major implications in forecasting. The trend is often regarded as a dominant feature of the data and if the trend mechanism is poorly captured in an empirical model, we can expect forecasts from the model to carry forward the poor approximation. As the forecast horizon is extended and observations are subsequently collected and calibrated against the forecasts, the data drift steadily away from the given model. In short, one of the laws of modern time series is that ‘no one understands trends, but everyone sees them in the data’.

**3.Time Series Representation**

A hierarchy of all the various time series representations in the literature is shown in the above figure.. The leaf nodes refer to the actual representation, and the internal nodes refer to the classification of the approach.

#### SAX (Symbolic Aggregate approXimation)

Many high level representations of time series have been proposed for data mining, including

Fourier transforms, wavelets, eigen waves, piecewise polynomial models etc. Many researchers have also considered symbolic representations of time series, noting that such representations would potentially allow researchers to avail of the wealth of data structures and algorithms from the text processing and bioinformatics communities. While many symbolic representations of time series have been introduced over the past decades, they all suffer from two fatal flaws.

1. Firstly, the dimensionality of the symbolic representation is the same as the original data, and virtually all data mining algorithms scale poorly with dimensionality.
2. Secondly, although distance measures can be defined on the symbolic approaches, these distance measures have little correlation with distance measures defined on the original time series.

The SAX representation is unique in that it allows dimensionality/numerosity reduction, and it also allows distance measures to be defined on the symbolic approach that lower bound corresponding distance measures defined on the original series. This latter feature is particularly exciting because it allows one to run certain data mining algorithms on the efficiently manipulated symbolic representation, while producing identical results to the

algorithms that operate on the original data. In particular, this representation can be applied on various data mining tasks of clustering, classification, query by content, anomaly detection,

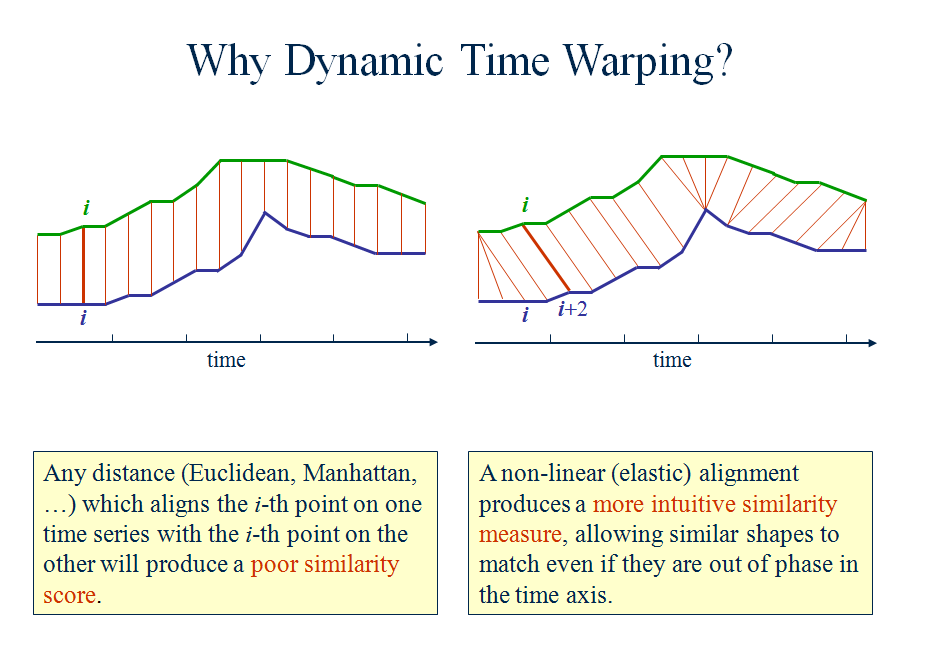
motif discovery, and visualization.

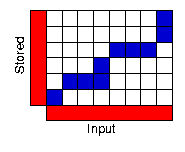
## 4. Similarity and Clustering

Given a query time series Q, a similarity finder will find the most similar time series in a given collection of time series. With Clustering, natural groupings of time series can be obtained.

#### Dynamic Time Warping - For Finding Similar Time Series

**Dynamic time warping** (DTW) is an algorithm for measuring similarity between two sequences which may vary in time or speed. For instance, similarities in walking patterns would be detected, even if in one video the person was walking slowly and if in another he or she were walking more quickly, or even if there were accelerations and decelerations during the course of one observation. DTW has been applied to video, audio, and graphics — *and we intend to apply DTW to environmental data*. Indeed, any data which can be turned into a linear representation can be analyzed with DTW.

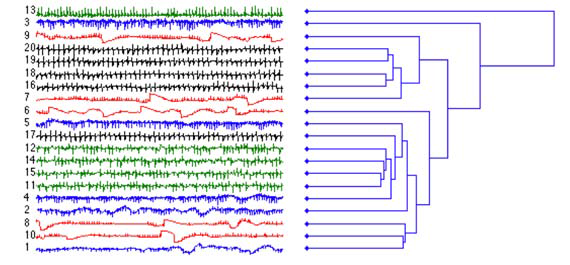




To find the best match between these two sequences we can find a path through the grid which minimises the total distance between them. The path shown gives an example. Here, the first and second elements of each sequence match together while the third element of the input also matches best against the second element of the stored pattern. This corresponds to a section of the stored pattern being stretched in the input. Similarly, the fourth element of the input matches both the second and third elements of the stored sequence: here a section of the stored sequence has been compressed in the input sequence. Once an overall best path has been found the total distance between the two sequences can be calculated for this stored template.

#### Hierarchical Clustering with SAX

Clustering is one of the most common data mining tasks, being useful in its own right as an exploratory tool, and also as a sub-routine in more complex algorithms



One of the most widely used clustering approaches is hierarchical clustering [11]. Hierarchical clustering computes pairwise distances of the objects (or groups of objects) and produces a nested hierarchy of the clusters. It has several advantages over other clustering methods. More specifically, it offers great visualization power withthe hierarchy of clusters, and it requires no input parameters.

## 4. Time Series and Forecasting/Prediction

A time series is a sequence of observations of a random variable. Hence, it is a stochastic process. Examples include the monthly demand for a product, the annual freshman enrollment in a department of a university, and the daily volume of flows in a river. Forecasting time series data is important component of operations research because these data often provide the foundation for decision models. An inventory model requires estimates of future demands, a course scheduling and staffing model for a university requires estimates of future student inflow, and a model for providing warnings to the population in a river basin requires estimates of river flows for the immediate future.

#### Classical Algorithms

#### **Simple Moving Average**

The Simple Moving Average smooth past data by arithmetically averaging over a specified period and projecting forward in time. This is normally considered a smoothing algorithm and has poor forecasting results in most cases.

#### **Geometric Moving Average**

The Geometric Moving Average smooth past data by geometrically averaging over a specified period and projecting forward in time. This is normally considered a smoothing algorithm and has poor forecasting results in most cases.

#### **Triangular Moving Average**

The Triangular Moving Average is a weighted moving average with weights that form a triangular shape. The projection technique is the same of the Simple Moving Average. This is normally considered a smoothing algorithm and has poor forecasting results in most cases.

#### **Parabolic Moving Average**

The Parabolic Moving Average is a weighted moving average with weights that form a parabolic shape. The projection technique is the same of the Simple Moving Average. This is normally considered a smoothing algorithm and has poor forecasting results in most cases.

#### **Double Moving Average**

The Double Moving Average applies in sequence for two times the Simple Moving Average algorithm. This is normally considered a smoothing algorithm and has poor forecasting results in most cases.

#### **Exponential Moving Average**

The Exponential Moving Average is summarized by the equation:

X't = αXt + (1-α)X't-1

it is a weighted moving average with weights that decrease exponentially going backwards in time. This is normally considered a smoothing algorithm and has poor forecasting results in most cases.

#### **Wavelet Smoothing and Forecasting**

#### **Frequency Identification** One of the biggest issues when dealing with algorithms that require seasonality indexes is to compute the seasonalities. All competitors require you to compute these magic numbers and even the most advanced packages aren't able to tell this simple and intuitive figure: the seasonality of a time series. The Frequency Identification algorithm computes the frequencies that are inside the input time series.

#### **Haar Denoising**

Removing noise from a time series is always difficult and current algorithms (averages, exponential averages, etc...) always introduce a lag in data or change the statistical properties of the underlying time series. Haar Denoising is able to remove the noise that is in the time series using the Haar Wavelet transform and a proprietary algorithm.

#### **Daubechies Linear Denoising**

Like Haar Denoising the Daubechies Linear Denoising applies a Daubechies Wavelet transform and a proprietary algorithm to linearly denoise the time series.

#### **Wavelet Forecasting**

This algorithm uses the Daubechies Wavelet transform to produce a forecast with a proprietary algorithm.

**Advanced Algorithms**

#### **Fractal Projection**

The Fractal Projection algorithm represents a new approach to forecasting. This algorithm is able to project to the future a pattern stretching it in times or values as appropriate.

#### **Active Moving Average**

The Active Moving Average belongs to the class of Exponential Moving Averages but its calculation is quadratic in nature and is much quicker than standard Moving Averages in "following" the original signal.

# **Kernel Smoothing**

Kernel smoothing weights every single data point in a time-series with weights coming from a generating function.

#### **Gaussian Kernel Smoothing**

The Gaussian Kernel Smoothing is a classical Kernel Smoothing algorithm.

The kernel function is the following:

K(t) = e-λ \* t2

## 6. Time Series and Anomaly Detection

#### The Markov model-based technique using SAX to detect Novel/Surprising/Anomalous Behavior

A simple idea for detecting anomalous behavior in time series is to examine previously observed normal data and build a model of it. Data obtained in the future can be compared to this model and any lack of conformity can signal an anomaly. In order to achieve this, a statistically sound scheme can be combined with an efficient combinatorial approach. The statistically scheme is based on Markov chains and normalization. Markov chains are used to model the “normal” behavior, which is inferred from the previously observed data. The time- and space-efficiency of the algorithm comes from the use of suffix tree as the main data structure. Each node of the suffix tree represents a pattern. The tree is annotated with a score

obtained comparing the support of a pattern observed in the new data with the support recorded in the Markov model. This apparently simple strategy turns out to be very effective in discovering surprising patterns. We plan to use a simple symbolic approach using SAX.

**References**

## [1] The Markov model-based technique using SAX to detect Novel/Surprising/Anomalous Behavior - <http://cs.gmu.edu/~jessica/SAX_DAMI_preprint.pdf>

[2] Dynamic Time Wrapping Technique <http://web.science.mq.edu.au/~cassidy/comp449/html/ch11s02.html>

[3]<http://www.cs.gmu.edu/~jessica/Lin_BOP_SSDBM.pdf>

[4] Microsoft Time Series Forecasting Algorithm - <http://msdn.microsoft.com/en-us/library/ms174923.aspx>

[5]Time Series Forecasting Algorithms - <http://www.ipredict.it/ForecastingMethods.aspx>

[6]Time Series and Forecasting Techniques i<http://www.me.utexas.edu/~jensen/ORMM/supplements/units/time_series/time_series.pdf>

[7]SAX: a Novel Symbolic Representation of Time Series <https://www.cs.gmu.edu/~jessica/SAX_DAMI_preprint.pdf>