Dynamic Time Warping

Steven Elsworth

School of Mathematics, University of Manchester

Friday 29th September, 2017

What is a time series?

A *time series* is a sequence of measurements of a quantity (e.g. temperature, pressure) over time.

What is a time series?

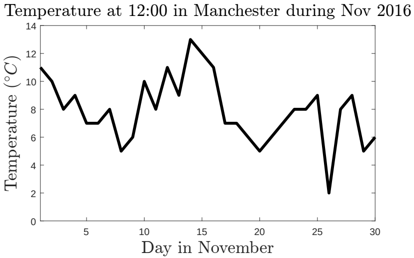
A *time series* is a sequence of measurements of a quantity (e.g. temperature, pressure) over time.

We do not need equally spaced time points.

What is a time series?

A *time series* is a sequence of measurements of a quantity (e.g. temperature, pressure) over time.

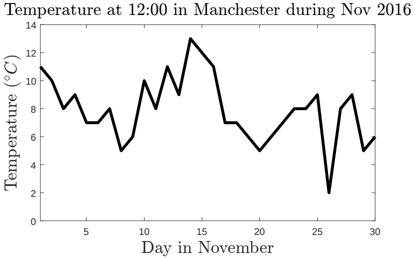
We do not need equally spaced time points.

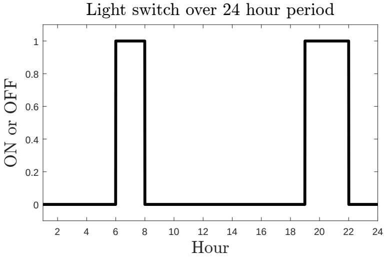


What is a time series?

A *time series* is a sequence of measurements of a quantity (e.g. temperature, pressure) over time.

We do not need equally spaced time points.





What is a distance measure?

Consider two time series $\mathbf{x} = \{x_i\}_{i=1}^{m}$ and $\mathbf{y} = \{y_i\}_{i=1}^{n}$. A distance/ similarity measure $d(\mathbf{x}, \mathbf{y})$ gives the 'distance/ similarity' between the time series $\mathbf{x}$ and $\mathbf{y}$. Some examples include:

What is a distance measure?

Consider two time series $\mathbf{x} = \{x_i\}_{i=1}^m$ and $\mathbf{y} = \{y_i\}_{i=1}^n$. A distance/ similarity measure $d(\mathbf{x}, \mathbf{y})$ gives the 'distance/ similarity' between the time series $\mathbf{x}$ and $\mathbf{y}$. Some examples include:

- Euclidean distance (on time series or transformed representation of time series)

What is a distance measure?

Consider two time series $\mathbf{x} = \{x_i\}_{i=1}^{m}$ and $\mathbf{y} = \{y_i\}_{i=1}^{n}$. A distance/ similarity measure $d(\mathbf{x}, \mathbf{y})$ gives the 'distance/ similarity' between the time series $\mathbf{x}$ and $\mathbf{y}$. Some examples include:

- Euclidean distance (on time series or transformed representation of time series)
- Dynamic Time Warping

What is a distance measure?

Consider two time series $\mathbf{x} = \{x_i\}_{i=1}^{m}$ and $\mathbf{y} = \{y_i\}_{i=1}^{n}$. A distance/ similarity measure $d(\mathbf{x}, \mathbf{y})$ gives the 'distance/ similarity' between the time series $\mathbf{x}$ and $\mathbf{y}$. Some examples include:

- Euclidean distance (on time series or transformed representation of time series)
- Dynamic Time Warping
- Longest Common Subsequence model

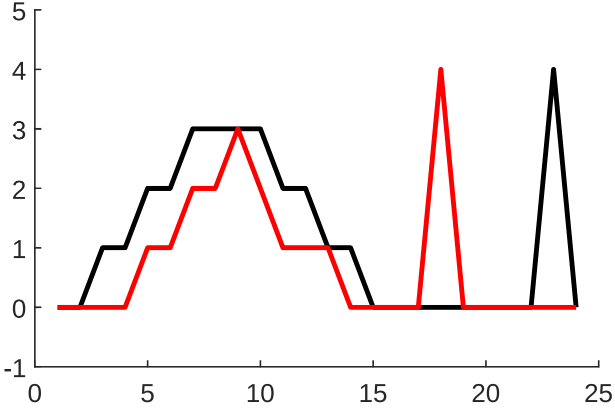
What is a distance measure?

Consider two time series $\mathbf{x} = \{x_i\}_{i=1}^{m}$ and $\mathbf{y} = \{y_i\}_{i=1}^{n}$. A distance/ similarity measure $d(\mathbf{x}, \mathbf{y})$ gives the 'distance/ similarity' between the time series $\mathbf{x}$ and $\mathbf{y}$. Some examples include:

- Euclidean distance (on time series or transformed representation of time series)
- Dynamic Time Warping
- Longest Common Subsequence model
- Threshold based distance measures

Example data

Consider the following time series:



Euclidean distance

Defined by $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=0}^{n} d(x_i, y_i)^2}$ where $d(x_i, y_i) = |x_i - y_i|$ or $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$.

Euclidean distance

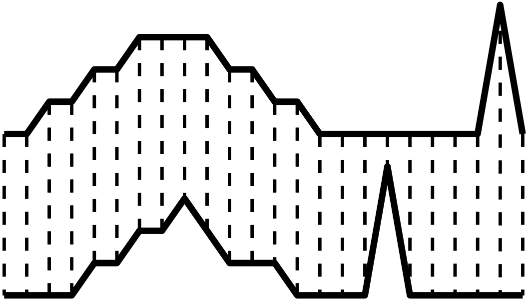
Defined by $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=0}^{n} d(x_i, y_i)^2}$ where $d(x_i, y_i) = |x_i - y_i|$ or $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$.

- Fast to compute
- Satisfies the triangle inequality
- Requires $n = m$
- Performs poorly when the time series have outliers.

Euclidean distance

Defined by $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=0}^{n} d(x_i, y_i)^2}$ where $d(x_i, y_i) = |x_i - y_i|$ or $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$.

- Fast to compute
- Satisfies the triangle inequality
- Requires $n = m$
- Performs poorly when the time series have outliers.



Dynamic Time Warping

First proposed by Berndt and Clifford (1994). Start by constructing $n \times m$ matrix D, where $D_{i,j} = d(x_i, y_j)$ where $d(x_i, y_i) = |x_i - y_i|$ or $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$.

Dynamic Time Warping

First proposed by Berndt and Clifford (1994). Start by constructing $n \times m$ matrix D, where $D_{i,j} = d(x_i, y_j)$ where $d(x_i, y_i) = |x_i - y_i|$ or $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$.
A *warping path* w is a contiguous set of matrix elements which defines a mapping between x and y that satisfies the following conditions:

Dynamic Time Warping

First proposed by Berndt and Clifford (1994). Start by constructing $n \times m$ matrix D, where $D_{i,j} = d(x_i, y_j)$ where $d(x_i, y_i) = |x_i - y_i|$ or $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$.

A *warping path* w is a contiguous set of matrix elements which defines a mapping between x and y that satisfies the following conditions:

1. Boundary conditions: $w_1 = (1, 1)$ and $w_k = (m, n)$ where k is the length of the warping path.

Dynamic Time Warping

First proposed by Berndt and Clifford (1994). Start by constructing $n \times m$ matrix D, where $D_{i,j} = d(x_i, y_j)$ where $d(x_i, y_i) = |x_i - y_i|$ or $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$.
A *warping path* w is a contiguous set of matrix elements which defines a mapping between x and y that satisfies the following conditions:

1. Boundary conditions: $w_1 = (1, 1)$ and $w_k = (m, n)$ where k is the length of the warping path.
2. Continuity: if $w_i = (a, b)$ then $w_{i-1} = (a', b')$ where $a - a' \leq 1$ and $b - b' \leq 1$.

Dynamic Time Warping

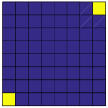
First proposed by Berndt and Clifford (1994). Start by constructing $n \times m$ matrix D, where $D_{i,j} = d(x_i, y_j)$ where $d(x_i, y_i) = |x_i - y_i|$ or $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$.

A *warping path* w is a contiguous set of matrix elements which defines a mapping between x and y that satisfies the following conditions:

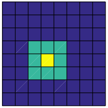
1. Boundary conditions: $w_1 = (1, 1)$ and $w_k = (m, n)$ where k is the length of the warping path.
2. Continuity: if $w_i = (a, b)$ then $w_{i-1} = (a', b')$ where $a - a' \leq 1$ and $b - b' \leq 1$.
3. Monotonicity: if $w_i = (a, b)$ then $w_{i-1} = (a', b')$ where $a - a' \geq 0$ and $b - b' \geq 0$.

Cont.

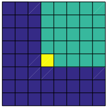
Boundary



Continuity

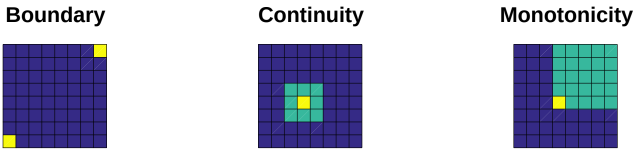


Monotonicity



We want $d_{DTW}(\mathbf{x}, \mathbf{y}) = \min \sum_{i=1}^{k} D(w_i)$.

Cont.



We want $d_{DTW}(\mathbf{x}, \mathbf{y}) = \min \sum_{i=1}^{k} D(w_i)$.
In practise we use dynamic programming to prevent construction of the whole matrix. Define $\gamma(i, j)$ to be the cumulative distance then

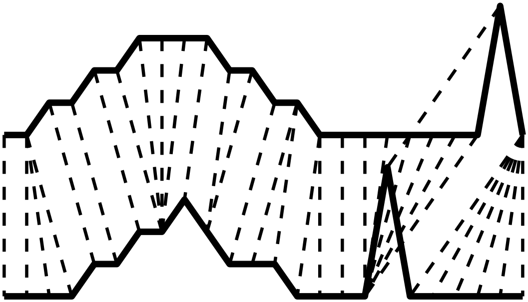
$$\gamma(i, j) = d(x_i, y_j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\}.$$

Cont.

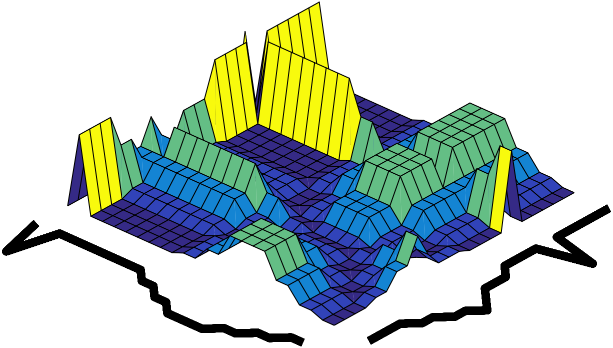
- Doesnt satisfy triangle inequality
- Slow to compute ($\mathcal{O}(nm)$ time and space complexity)

Cont.

- Doesnt satisfy triangle inequality
- Slow to compute ($\mathcal{O}(nm)$ time and space complexity)



3D graphical view of Dynamic Time Warping



What are distance measures used for?

What are distance measures used for?

- Indexing - Given time series $\mathbf{x}$, find time series $\mathbf{y}$ in database DB such that $d(\mathbf{x}, \mathbf{y})$ is minimal.

What are distance measures used for?

- Indexing - Given time series $\mathbf{x}$, find time series $\mathbf{y}$ in database DB such that $d(\mathbf{x}, \mathbf{y})$ is minimal.
- Clustering - Split database DB into groups using $d(\cdot, \cdot)$.

What are distance measures used for?

- Indexing - Given time series $\mathbf{x}$, find time series $\mathbf{y}$ in database DB such that $d(\mathbf{x}, \mathbf{y})$ is minimal.
- Clustering - Split database DB into groups using $d(\cdot, \cdot)$.
- Classification - Given clusters $DB_1, \cdots, DB_k$, which cluster does timeseries $\mathbf{x}$ belong to?

What are distance measures used for?

- Indexing - Given time series $\mathbf{x}$, find time series $\mathbf{y}$ in database DB such that $d(\mathbf{x}, \mathbf{y})$ is minimal.
- Clustering - Split database DB into groups using $d(\cdot, \cdot)$.
- Classification - Given clusters $DB_1, \cdots, DB_k$, which cluster does timeseries $\mathbf{x}$ belong to?
- Query search - Given large time series $\mathbf{x}$ and small query time series $\mathbf{q}$ and distance d, find subsequences of $\mathbf{x}$ which are less than d away from $\mathbf{q}$.

How do we speed it up?

How do we speed it up?

1. Global Constraints

How do we speed it up?

1. Global Constraints
2. Lower bounds

How do we speed it up?

1. Global Constraints
2. Lower bounds
3. FASTDTW

Global Constraints

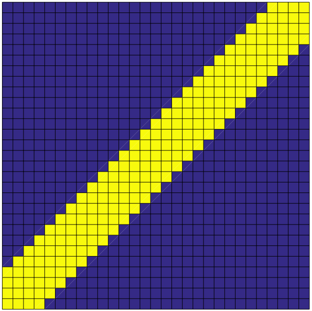


Figure: Sakoe-Chimba band

Global Constraints

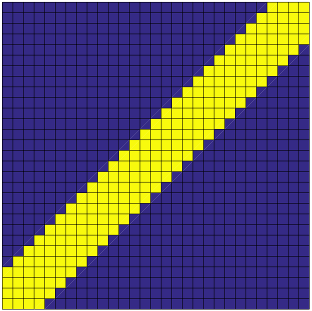


Figure: Sakoe-Chimba band

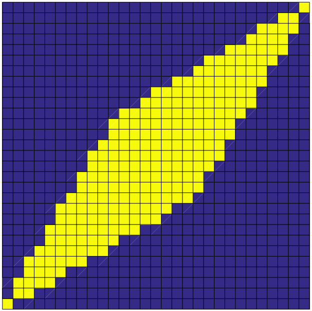


Figure: Itakura parallelogram

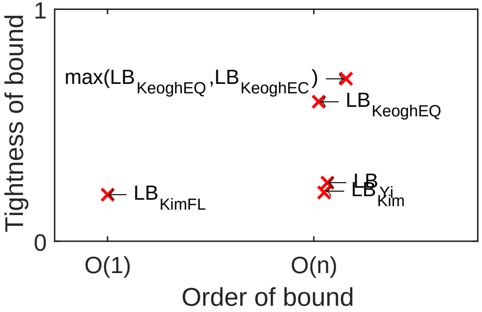
Lower Bounds

Many lower bounds have been developed. These are cheaper to compute but prone off unnecessary DTW distance measures. Two most popular are LB_{KIM} and LB_{KEOGH}.

LB_{kimFL} is the simplest and of $\mathcal{O}(1)$. It consists of finding the Eulidean distance of the first and last point in time series. LB_{kim} is of $\mathcal{O}(n)$ and consists of distance between maximum and minimum values of time series.

Lower Bounds

Many lower bounds have been developed. These are cheaper to compute but prone off unnecessary DTW distance measures. Two most popular are LB_{KIM} and LB_{KEOGH}.
LB_{kimFL} is the simplest and of $\mathcal{O}(1)$. It consists of finding the Eulidean distance of the first and last point in time series. LB_{kim} is of $\mathcal{O}(n)$ and consists of distance between maximum and minimum values of time series.



FASTDTW

FASTDTW was developed by Salvador and Chan. It achieves linear time and space complexity. Consists of three stages: Coarsening, projection, refinement.

FASTDTW

FASTDTW was developed by Salvador and Chan. It achieves linear time and space complexity. Consists of three stages: Coarsening, projection, refinement.

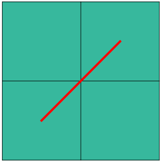


Figure: 1/4

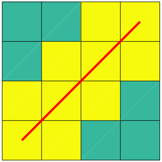


Figure: 1/2

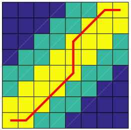


Figure: 1/1

FASTDTW

FASTDTW was developed by Salvador and Chan. It achieves linear time and space complexity. Consists of three stages: Coarsening, projection, refinement.

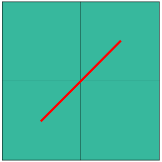


Figure: 1/4

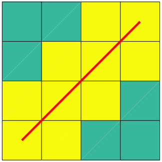


Figure: 1/2

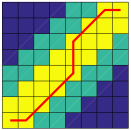
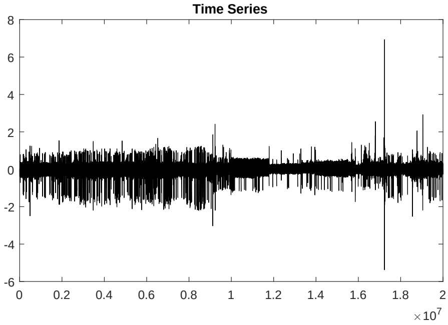


Figure: 1/1

Doesnt always find optimal solution.

Query search example

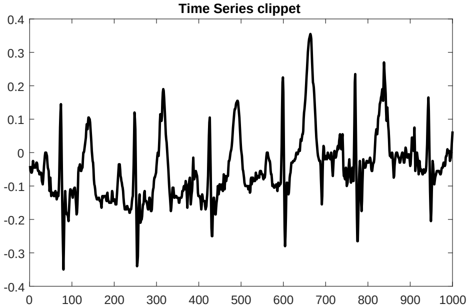
Consider the following time series.



This is an electrocardiogram of an 84-year old female. Recording is 22 hours and 23 minutes long containing 110,087 beats. End cropped for visual purposes.

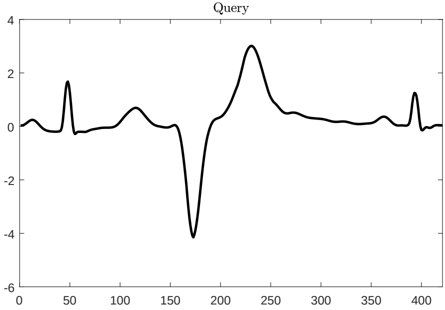
Time series: $\mathbf{x}$

250 samples per second giving a total length of 20,145,000. Here is a small 4s clippet (1000 data points).



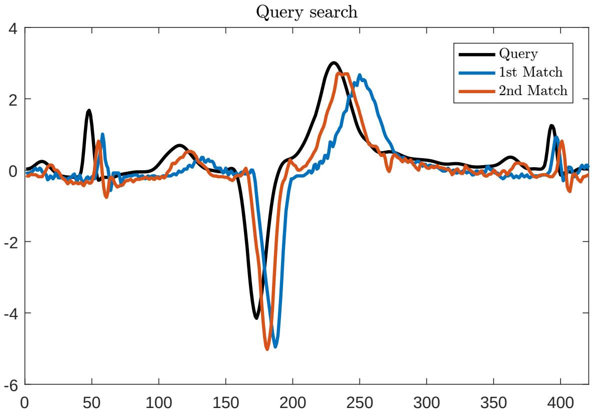
Query: $\mathbf{q}$

Premature venticular contraction. These are premature heartbeats originating from the ventricles of the heart, possibly due to high blood pressure or heart disease. Query containing 421 data points, about 1.7 seconds.



Results

Find the following two closest matches.



References

- DTW : Berndt and Clifford (1994)
- Sakoe-Chimba band: Sakoe and Chimba (1978)
- Itakura parallelogram: Itakura (1975)
- Bounds and UCR suite: Rakthanmanon et al. (2012)
- FastDTW: Salvador and Chan (2007)
- Last worked example:
  `http://www.cs.ucr.edu/$\sim$eamonn/UCRsuite.html`