Time Series Feature Detection

Evan Wojciechowski
Mentored by the MIT Haystack Astro- & Geoinformatics Group: Victor Pankratius, Justin Li, Cody Rude, Guillaume Rongier

August 10\textsuperscript{th}, 2017
Time Series Science Data

Time Series: any data coupled with time coordinates

Li, Justin D., et al. 2016 (Astro- & Geoinformatics group)

Data acquired from skdaccess (https://github.com/MITHaystack/scikit-dataaccess)
Time Series Science Data

Time Series: any data coupled with time coordinates

Li, Justin D., et al. 2016 (Astro- & Geoinformatics group)

Data acquired from skdaccess
(https://github.com/MITHaystack/scikit-dataaccess)
Fourier Transform

\[ f(\omega) = \int_{-\infty}^{\infty} e^{-2\pi it \omega} f(t) dt \quad f(t) = \frac{1}{T} \sum_{n=1}^{N} a_n e^{2\pi it \omega_n} f(\omega_n) \]

Akutan Volcano dE0607 [mm]
Fourier Transform

\[ f(\omega) = \int_{-\infty}^{\infty} e^{-2\pi it\omega} f(t) dt \quad f(t) = \frac{1}{T} \sum_{n=1}^{N} a_n e^{2\pi it\omega_n} f(\omega_n) \]

Akutan Volcano dE0607 [mm]
Fourier Transform works best with periodic events, but investigation into more general methods may provide better insights to other types of features present in data sets.
Wavelet Transform

\[ X(a, b) = \left( \frac{1}{\sqrt{a}} \right) \int_{-\infty}^{\infty} \Psi \left( \frac{t - b}{a} \right) x(t) dt \]

Wavelet functions act as eigenfunctions and form a basis for a data set

Haar

Biorthogonal (2.2, 3.1)

Daubechies (2, 7, 17)
Wavelet Transform

Wavelet functions act as eigenfunctions and form a basis for a data set.
Wavelet Transform

Wavelet functions act as eigenfunctions and form a basis for a data set.
Wavelets work best when the shape of the data is known or if one has time to find a good fit. Due to this intrinsic limitation wavelets is not the most general approach.
Empirical Mode Decomposition

EMD(data) = $\Sigma$(IMFs) + Residual

$$f(t) = \sum_{n=1}^{N} a_n(t) \exp \left[ i \int_{0}^{t} \omega_n(\tau) d\tau \right] + R_n(t)$$


IMFs: $\text{abs} (N_{\text{extrema}} - N_{\text{zero-point}}) \leq 1$, Mean Value (local extrema) = 0
Empirical Mode Decomposition

\[ f(t) = \sum_{n=1}^{N} a_n(t) \exp \left[ i \int_{0}^{t} \omega_n(\tau) d\tau \right] + R_n(t) \]


IMFs: \[ \text{abs}(N_{\text{extrema}} - N_{\text{zero-point}}) \leq 1 \], Mean Value (local extrema) = 0

\[ \text{EMD(data)} = \sum(\text{IMFs}) + \text{Residual} \]
Empirical Mode Decomposition

$\text{EMD(data)} = \sum (\text{IMFs}) + \text{Residual}$

$f(t) = \sum_{n=1}^{N} a_n(t) \exp \left[ i \int_{0}^{t} \omega_n(\tau) d\tau \right] + R_n(t)$


IMFs: $\text{abs}(N_{\text{extrema}} - N_{\text{zero-point}}) \leq 1$, Mean Value (local extrema) = 0

Linear decomposition allows for linear recomposition or summation
Empirical Mode Decomposition

Akutan Volcano dE0607 [mm]
Empirical Mode Decomposition

Individual IMFs can have extensive structure
IMF Analysis

Akutan Volcano dE0607 [mm]
High frequency components are often attributed to noise, leaving low frequency components as the representative for real data.
Periodicity

Kepler 009941662 Flux \([e^- / s]\)
Periodicity

Kepler 009941662 Flux [e^- / s]
Periodicity

Kepler 009941662

Time series

Fourier spectrum
Periodicity

More intuitive methods of finding periodicity could help understanding data thought to have inherent repeated behavior.
\[ r(t) = t - t_0, \quad \theta(t) = \frac{2\pi}{T} r(t), \quad z(t) = s(t) \]
Conclusions

- Time series features can fall under different categories, often needing separate methods of analysis
- Real data may contain noise that hampers understanding of features one is looking for, but can be mitigated in various ways
- Visualization is a good tool to help build intuition about features and present findings in a general manner
Thank You Haystack!

- Informatics group especially (bye Justin!)
- Questions/comments?