

Anomaly Detection Library

Kapil Thakkar and Reshma Kumari

27th May, 2016

Contents

1	Window Based Correlation	2
1.1	Introduction	2
1.2	Related Functions	2
1.2.1	correlation(arr1, arr2, maxlag, pos, neg)	2
1.2.2	getMaxCorr(arar1,positive_correlation)	3
1.2.3	correlationAtLag(series1, series2, lag, window_size)	3
1.3	Description	3

Chapter 1

Window Based Correlation

1.1 Introduction

This technique is basically applied on two time-series. Let's say we have two time series as series1 and series2. So, in this method, we first find correlation at various lags between these two time series. User can specify minimum and maximum lag to consider. So, for all those values, we find correlation values.

After finding correlation values at all lags, we consider that lag at which correlation value is higher among all previously calculated correlation values at all lags. Let's say that lag be "x". So, depending upon that "x", we shift series1 or series2. If "x" is positive, we move series2 by "x" units and if it is negative than we shift series1 by $|x|$ units.

Now, we are ready to apply window correlation. Take window value, "w" as input. First window will be from 1st element to w'th element of both the time series after aligning by lag "x". Find correlation for this window between two time-series and save it in an array. Now, slide window by "w" elements and calculate correlation value again and so on. Now, we have correlation values at multiple windows.

Now, let's say both the series should have been positively correlated. So, what we do is, we choose threshold by MAD test if not provided to us, and find all correlation values which are below that threshold and report all those windows as anomaly.

1.2 Related Functions

1.2.1 correlation(arr1, arr2, maxlag, pos, neg)

This function calculates correlation between arr1 and arr2 at all possible lags between -maxlag to +maxlag, as specified by pos and neg parameters.

- Input Parameters

1. arr1 (*list*) : Input series 1 as a list of float values
2. arr2 (*list*) : Input series 2 as a list of float values
3. maxlag (*int*) : maximum (maxlag) and minimum (-maxlag) lag to consider while calculating correlation between arr1 and arr2

4. pos (*int*) : To consider positive lag or not, i.e. 1 to maxlag
5. neg (*int*) : To consider negative lag or not, i.e. -maxlag to -1

- Output (*list*) :
Returns list of tuples of the form

(lag, correlation value at this lag)

1.2.2 getMaxCorr(arr1, positive_correlation)

This function takes list of tuples of the form (lag, correlation value at this lag) as input. Returns lag value at which correlation value is maximum if positive_correlation is True, and returns lag at which correlation value is minimum if positive_correlation is False.

Basically, if both the series are positively correlated than we will be interested in maximum positive correlation or if both series are negatively correlated than we will be interested in minimum negative correlation, which is specified by positive_correlation parameter.

- Input Parameters

1. arr1 (*list*) : list of tuples of the form
(lag, correlation value at this lag)
i.e. correlation values at various lags
2. positive_correlation (*boolean*, “True” or “False”) :
 - True: If value of this parameter is True than it will return lag at which correlation value if maximum (positive)
 - False: If value of this parameter is False than it will return lag at which correlation value if minimum (negative)

- Output (*Tuple*) :
returns single tuple of the form (lag, correlation value at this lag), i.e. lag at which optimum correlation value is found along with correlation value.

1.2.3 correlationAtLag(series1, series2, lag, window_size)

This function first aligns two series by given lag. If lag is positive than it shifts start of series2 else start of series1. After aligning both the series according to lag, this function calculates correlation between both series at all windows.

window_size states size of the window. So, we will start with first window taking first window_size elements from each series and will calculate correlation. We will save this correlation value in list and will slide to next window. Next window will start after window_size elements. In such a way, we calculate, correlation at all windows and return the list of correlation values.

- Input Parameters

1. series1 (*list*) : Input series 1 as a list of float values

2. `series2` (*list*) : Input series 2 as a list of float values
 3. `lag` (*int*) : lag at which
 4. `pos` (*int*) : To consider positive lag or not, i.e. 1 to maxlag
 5. `neg` (*int*) : To consider negative lag or not, i.e. -maxlag to -1
- Output (*list*) :
Returns list of tuples of the form

(lag, correlation value at this lag)

1.3 Description

Chapter 2

Window Based Correlation

2.1 Introduction

This technique is basically applied on two time-series. Let's say we have two time series as series1 and series2. So, in this method, we first find correlation at various lags between these two time series. User can specify minimum and maximum lag to consider. So, for all those values, we find correlation values.

After finding correlation values at all lags, we consider that lag at which correlation value is higher among all previously calculated correlation values at all lags. Let's say that lag be "x". So, depending upon that "x", we shift series1 or series2. If "x" is positive, we move series2 by "x" units and if it is negative than we shift series1 by $|x|$ units.

Now, we are ready to apply window correlation. Take window value, "w" as input. First window will be from 1st element to w'th element of both the time series after aligning by lag "x". Find correlation for this window between two time-series and save it in an array. Now, slide window by "w" elements and calculate correlation value again and so on. Now, we have correlation values at multiple windows.

Now, let's say both the series should have been positively correlated. So, what we do is, we choose threshold by MAD test if not provided to us, and find all correlation values which are below that threshold and report all those windows as anomaly.

2.2 Related Functions

2.2.1 correlation(arr1, arr2, maxlag, pos, neg)

This function calculates correlation between arr1 and arr2 at all possible lags between -maxlag to +maxlag, as specified by pos and neg parameters.

- Input Parameters

1. arr1 (*list*) : Input series 1 as a list of float values
2. arr2 (*list*) : Input series 2 as a list of float values
3. maxlag (*int*) : maximum (maxlag) and minimum (-maxlag) lag to consider while calculating correlation between arr1 and arr2

4. pos (*int*) : To consider positive lag or not, i.e. 1 to maxlag
5. neg (*int*) : To consider negative lag or not, i.e. -maxlag to -1

- Output (*list*) :
Returns list of tuples of the form

(lag, correlation value at this lag)

2.2.2 getMaxCorr(arar1,positive_correlation)

This function takes list of tuples of the form (lag, correlation value at this lag) as input. Returns lag value at which correlation value is maximum if positive_correlation is True, and returns lag at which correlation value is minimum if positive_correlation is False.

Basically, if both the series are positively correlated than we will be interested in maximum positive correlation or if both series are negatively correlated than we will be interested in minimum negative correlation, which is specified by positive_correlation parameter.

- Input Parameters

1. arr1 (*list*) : list of tuples of the form
(lag, correlation value at this lag)
i.e. correlation values at various lags
2. positive_correlation (*boolean*, “True” or “False”) :
 - True: If value of this parameter is True than it will return lag at which correlation value if maximum (positive)
 - False: If value of this parameter is False than it will return lag at which correlation value if minimum (negative)

- Output (*Tuple*) :
returns single tuple of the form (lag,correlation value at this lag), i.e. lag at which optimum correlation value is found along with correlation value.

2.2.3 correlationAtLag(series1, series2, lag, window_size)

This function first aligns two series by given lag. If lag is positive than it shifts start of series2 else start of series1. After aligning both the series according to lag, this function calculates correlation between both series at all windows.

window_size states size of the window. So, we will start with first window taking first window_size elements from each series and will calculate correlation. We will save this correlation value in list and will slide to next window. Next window will start after window_size elements. In such a way, we calculate, correlation at all windows and return the list of correlation values.

- Input Parameters

1. series1 (*list*) : Input series 1 as a list of float values

2. `series2` (*list*) : Input series 2 as a list of float values
 3. `lag` (*int*) : lag at which
 4. `pos` (*int*) : To consider positive lag or not, i.e. 1 to maxlag
 5. `neg` (*int*) : To consider negative lag or not, i.e. -maxlag to -1
- Output (*list*) :
Returns list of tuples of the form

(lag, correlation value at this lag)

2.3 Description