TrendCalculus

A data science for studying trends.

2014-12-16
Trends?!

Data science seems so focused on the micro scale:

deeper granularity

higher frequency...
Trends?!

My focus is broad patterns; big flock behaviours, and my objective is long range predictions.

Trends are a natural way to think, explain, and forecast.

Yet we lack tools to understand Trends, scientifically.

TrendCalculus is my unfinished research to that end.
What’s a trend?

“A Trend is defined by a shift in behaviour or mentality that influences a significant amount of people.” - Salomé Areias

“A Trend is the slow variation over a longer period of time, usually several years, generally associated with the structural causes affecting the phenomenon being measured.” - Eurostat
400+ years of trend discussion

What do you see?

Perhaps a shift in behaviour or mentality?
Maybe a drift in language use?
How do we quantify and study the trend?

-Wolfram Alpha
400+ years of trend discussion

What might cause Trend as a topic to be losing popularity?

- Wolfram Alpha
400+ years of trend discussion

What might cause Trend as a topic to be losing popularity?

Maybe traditional trend analysis is flawed and the collective knows it.
What if we could do better?

What would you do if you really understood trends and when they reversed?
Introduction

What is TrendCalculus?
TrendCalculus is our new, multi-scale trend reversal detection algorithm for streamed numeric data over all timeframes.

It’s pretty fast: $O(n)$
two timeframes:

- Yearly Trend Reversal
- Monthly Trend Reversal
What does output look like?

many multiple stacked timeframes build up into long term structures
What does output look like?

multiple stacked timeframes build up into long term structures...
Where does this fit?

Time Series Representations

Model Based
- Hidden Markov Models
- Statistical Models
  - Sorted Coefficients
  - Piecewise Polynomial Approximation
  - Piecewise Linear Approximation
    - Interpolation
    - Regression
  - Piecewise Piecewise Constant Approximation
  - Adaptive Piecewise Constant Approximation
  - Natural Language
  - Strings
  - Symbolic Aggregate Approximation
  - Non Lower Bounding
  - Value Based
  - Slope Based
- Piecewise Trend Partitioning

Data Adaptive
- Piecewise
- Singular Value Approximation
- Symbolic
- Trees
- Trees
- Wavelets
- Wavelets
- Orthonormal
- Orthonormal
- Bi-Orthonormal
- Bi-Orthonormal
- Daubechies
  - dnb
  - n > 1
- Coiflets
- Symlets
- Haar
- Chebyshev Polynomials
- Spectral
- Spectral
- Discrete Fourier Transform
- Discrete Cosine Transform
- Piecewise Aggregate Approximation
- Data Dictated
  - Grid
  - Clipped Data

TrendCalculus:
Is a multi-scale, bottom up, trend reversal detected, Piecewise Approximation that produces a hierarchical trend partitioning.

* Author:
Eamonn Keogh
Professor
Computer Science & Engineering Department
University of California - Riverside
Riverside, CA 92521
TrendCalculus:
It enables
“Multiscale Trend Analysis”
Multiscale Trend Analysis

What is MTA?
What does MTA offer?

If offers rich time series methods...

to better predict

to correlate timeseries

to index and compress

to do cross-scale retrieval of “motifs”

to build ‘episodic memory’ stores

to normalise signal extraction, reduce noise

to convert sub-symbolic data to rich symbolic data

the MTA paper is a good read:

Multiscale Trend Analysis

Ilya Zaliapin

Andrei Gabrielov

Vladimir Keilis-Borok

Revised: February 02, 2004
What are multi scale trends?

A time series is decomposed into local linear trends.

a) $X(t)$ is approximated by its global linear trend $L_0(t)$.

b) Detrended series $X_1(t) = X(t) - L_0(t)$ is approximated by the piecewise linear function $L_1(t)$, the whole analysis is then repeated at each of subintervals $[t_i, t_i+1]$.

c) Resulting hierarchy of trends. See Sect. 2 for details.
From the MTA paper....

a) MTD for Brownian walk

- Level 0: 1 interval
- Level 1: 3 intervals
- Level 2: 7 intervals
- ...  
- Level 10: 23 intervals

b) The corresponding hierarchy of trends

* these pictures are from the paper
The idea is to ignore noise

b) Reversals found on the scale of interest

* these pictures are from the paper
Build a tree of local trends

The trends are stacked in a hierarchy.

Like a b-tree, we index time series data into a shallow tree which is isn’t balanced per se, but partitions are interpretable and meaningful (not necessarily stationary)

* these pictures are from the paper
Trends are signed integers

Trending Together? = a correlation measure.
Multiply the trend signs at time \( t \).
If answer is +1 they are trending together...

* these pictures are from the paper
What does ByteSumo bring?

We created a Bottom Up algorithm, that detects Trend Reversals, aka “Knots”, at a Scale, based on a window, N.

Stacked, it creates multi-scale partitions over a stream of time series data.

It’s fast, because we changed the definition of a Trend (?!)

Yes - We abandoned linear regressions...

Our definition is:

Rising = Higher Highs, Higher Lows

Falling = Lower Lows, Lower Highs
Let's see it in action!

Let's try the **FTSE 100**, extended back to 1935 via the FTSE 30 data.

**Time Series length:** 21499 records (daily closes)

This run uses window size of **n=200** (market days)

The process in Lua creates lots of intermediate calculations for each window size from n down to 1 ... so it should be slow....

**Total run time is ~13 seconds** on my mac.

Output is shown left: **51 major trend reversals** found that approximate the time series.

Alternatively, we could say we have “generalised the time series” into 51 important change points.

It’s true luajit can speed this up...

But is how else might we be able to to speed it up?
Let’s see it in action!

Let’s try another way. Stacking the calculations: i.e. Pipe output back through the algo again x3.

There is practically a magnitude improvement in performance when stacking.

With a setting of N=5, I just processed the stack of 4 runs in less than 2 seconds using straight lua on my mac for 21,499 input records.

that’s ~10k streamed records per second.

With luajit it will drop further! The partitions in the trend tree we calculated are:

level 4  = 0  trend reversals
level 3  = 28  trend reversals
level 2  = 249  trend reversals
level 1  = 2,079  trend reversals.
Let’s see it in action!

Here is the last 14 years of the stacked output. The 3 levels of partitions are seen nested:
Let's see it in action!

Here is the last 14 years of the stacked output. The 3 levels of partitions are seen nested:
Let’s see it in action!

Here is the last 14 years of the stacked output. The 3 levels of partitions are seen nested:
Let’s see it in action!

Zoom out. Here is from 1945 to Present. We see the 28 “level 3” partitions as red knots.
Let’s see it in action!

Zoom in. Here is 2013. Here we can see some of the 2,079 fine grain “level 1” reversals up close:
Let’s see it in action!

Zoom in. Here is 2013. Here we can see some of the 2,079 fine grain “level 1” reversals up close:
Let’s see it in action!

Zoom in. Here is 2013. Here we can see some of the 2,079 fine grain “level 1” reversals up close:

While we didn’t set out to generate piecewise linear regressions, (we abandoned regression remember) you can see the results are often not bad if we judge it on that basis.
Let’s see it in action!

Zoom in. Here is 2013. Here we can see some of the 2,079 fine grain “level 1” reversals up close:

All these partitions were created in that 2 second run, ~10k data points per second.
New Directions for Trend Calculus

Our research is uncovering whole new avenues of study.
A Rolling Trend score?

This involves moving away from fixed windows of N and to rolling arrays for all timeframes to N.

The information revealed is not trend reversals, but the underlying data used in their calculation.

I will output these internal arrays to feed deep learning algorithms as a form of “trend feature generator”.

For display, I turn values into symbols, and we can see rich patterns emerging from the trends across all scales.

Quants who reviewed this said: “ah, it shows the relationship of the price to the Pivot Points”
Here I present the trend score as a symbol for each timeframe to a max N to build a "multi scale trend map".
Rolling Trends - all timeframes

The columns are symbols representing the value of the rolling channels I calculate in my array for a value n. A timeframe becomes vertical stripes on the “trend map” from 1 .. n.
What are we seeing?

a small N is a short channel

bigger N is a long channel

When price above a channel, trend is up.
When below a channel, it’s down.

When the price is in a channel the trend is neutral.

here we see two timeframes.
The next steps are to use all these rich inputs to see if we can make long range predictions...

.. by for instance feeding deep learning algorithms with all these trends to predict future trend reversals.

It means I’ll use TrendCalculus to generate interesting trend features.

Lots of potential for further work.
Prediction

MTA was created by people predicting earthquakes
The 11 April 2012, M8.6 and M8.2 earthquakes off the west coast of Northern Sumatra did confirm an alarm TIP reported in January, in the regular 2010a Update of the M8-MSc predictions of the Global Test of M8 (Healy et al. 1992; password protected URL http://www.mitp.ru/en/restricted_global/2010a/2010am8.html; yellow outline in the attached figure). The earthquake epicenters missed the reduced area of alarm (red outline) diagnosed in the second approximation due to inapplicability of the MSc algorithm outside bulk distribution of seismic activity. Nevertheless, it appears remarkable that the reduced area is about the same as the area of the 11 April 2012 first-day aftershocks located at about the same latitudes.

The 11 April 2012 great earthquakes have ruptured the conjugate faults, about 300 and 500 km each in the oceanic lithosphere of Indo-Australian plate. Both are strike-slip intra-oceanic-plate events with epicenters in an area of sparse seismicity, some 100 km and 200 km to the southwest of the major seismic belt of the subduction zone next to the complex junction of India, Australia, India, and Burma plates. These events continue a series that can be attributed to the 28 December 2004, M9.1 Sumatra-Andaman mega-thrust, followed by the 26 March 2005, M8.6 great Nias earthquake. In course of the Global Test of M8 a segment of the subduction zone from Burma to Southern Sumatra was recognized as capable of producing magnitude M8.0+ event starting from July 2005-January 2006, which prediction was already confirmed with a pair of the great 12 September 2005, M8.9 Southern Sumatra and M8.1 KEPULAUAN MENTAWAI REGION, INDONESIA earthquakes (http://www.mitp.ru/en/restricted_global/2005a/m8confirmed.html).

(Note: The M8 algorithm provides prediction in the first approximation, and the algorithm MSc, if the data permit, narrows down the area covered by alarm. Both apply to the null approximation delivered by identifying earthquake-prone zones, e.g. “active fault zones”, “D-intersections or knots”, etc.)
Predicting Earthquakes?!

The 11 April 2012, M8.6 and M8.2 earthquakes off the West Coast of Northern Sumatra did confirm an alarm TIP reported in January, in the regular 2012a Update of the M8-MSc predictions of the Global Test of M8 (Healy et al. 1992; password protected URL http://www.mitp.ru/en/restricted_global/2012a/2012am8.html; yellow outline in the attached figure). The earthquake epicenters missed the reduced area of alarm (red outline) diagnosed in the second approximation due to inapplicability of the MSc algorithm outside the bulk distribution of seismic activity. Nevertheless, it appears remarkable that the reduced area is about the same as the area of the 11 April 2012 first-day aftershocks located at about the same latitudes.

The 11 April 2012 great earthquakes have ruptured the conjugate faults, about 300 and 500 km each, in the oceanic lithosphere of Indo-Australian plate. Both are strike-slip intra-oceanic plate events with epicenters in an area of sparse seismicity, some 100 km and 200 km to the southwest of the major seismic belt of the subduction zone next to the complex junction of India, Australia, Sofia, and Burma plates. These events continue a series that can be attributed to the 26 December 2004, M9.1 Sumatra-Andaman mega-thrust, followed by the 28 March 2005, M8.6 great Nias earthquake. In course of the Global Test of M8 a segment of the subduction zone from Burma to Southern Sumatra was recognized as capable of producing magnitude M8.0+ event starting from July 2005-January 2006, which prediction was already confirmed with a pair of the great 12 September 2007, M8.5 Southern Sumatra and M8.0 Kepulauan Mentawai Region, Indonesia earthquakes (http://www.mitp.ru/en/restricted_global/2007b/m8t5confirmed.html).

(Note: The M8 algorithm provides prediction in the first approximation, and the algorithm MSc, if the data permit, narrows down the area covered by alarm. Both apply to the null approximation delivered by identifying earthquake-prone zones, e.g. “active fault zones”, “D-intersections or knots”, etc.)

I think they did it by finding unusual divergences “(Un)Correlations” between trends in different geophysical measures and these were found to be precursors to major earthquakes...

see here for detail:

Temporal (Un)correlations Between Coda Q and Seismicity: Multiscale Trend Analysis

http://link.springer.com/article/10.1007%2Fs00024-004-2643-x
THANK YOU
Andrew Morgan is a practicing Senior Enterprise Data Architect and Data Scientist and currently is designing a data science practice and platform for a top 4 audit firm client. He is also the CEO of ByteSumo, a data science Consultancy.

He is a specialist in data processing languages, data platform design, emerging data technologies, exotic data structures, data science methods, technical architecture, and data security systems.

He founded ByteSumo to build a data science led consultancy that has the experts and tools needed to transform and disrupt traditional enterprises.

<table>
<thead>
<tr>
<th>(curr. client role)</th>
<th>2014 - 2015</th>
<th>Interim Head of Data Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>ByteSumo</td>
<td>2013 - present</td>
<td>CEO</td>
</tr>
<tr>
<td>Capgemini</td>
<td>2010 - 2013</td>
<td>Senior Enterprise Architect, BIM</td>
</tr>
<tr>
<td>Thomson Reuters</td>
<td>2006 - 2010</td>
<td>Architect, Senior Technologist</td>
</tr>
<tr>
<td>Aprimo (now Teradata)</td>
<td>2005 - 2006</td>
<td>Senior Consultant</td>
</tr>
<tr>
<td>dunnhumby</td>
<td>1999 - 2000</td>
<td>Database Consultant</td>
</tr>
<tr>
<td>Elf Gas &amp; Power UK</td>
<td>1995 - 1999</td>
<td>Operational Dev. Executive</td>
</tr>
</tbody>
</table>

Bachelor of Arts, Geography. University of Toronto. 1994
Attribution

Salomé Areias:  http://salomeareias.com/what-is-a-trend/


MultiScale Trend Analysis:

Eamonn Keogh:
Time Series Representations - a slide found in the tutorials found here:  http://www.cs.ucr.edu/~eamonn/tutorials.html