Variance Change Point Detection

Santosh Srivastava LOPESANTOSH@GMAIL.COM
Ritwik Chaudhuri RITWICHA@IN.IBM.COM
Ankur Narang ANNARANG2@GMAIL.COM
Sahil Mathur SAHIMATH@IN.IBM.COM
Gyana Parija GYANA.PARIJA@IN.IBM.COM
IBM Research India

Abstract

We present a novel offline variance change point detection algorithm based on dynamic mode decomposition (DMD). The developed algorithm dynamic mode decomposition based variance change point detection (DVCPD) is completely data driven, doesn't require any knowledge of underlying governing equation or any probabilistic model assumption for time series. It uses a local adaptable window and sequential hypothesis testing to iteratively detect variance change points, where window's location and its size is automatically governed by acceptance and rejection of hypothesis. The DVCPD algorithm has been demonstrated to work robustly on different time series data sets and detects multiple variance change points accurately.

1. Introduction

Change point detection involves discovery of the points at which the stochastic behavior of a time series changes. Variance change point is an abrupt change in variance before and after the change point. When multiple variance change points arise, they could structurally break the time series. When a time series is structurally broken, one couldn't apply a reasonable modelling assumption with good accuracy. We consider the problem of iteratively detecting and handling variance change in sequential data. Given a fixed sample size of a sequential data, detecting all its variance change points accurately is a challenging problem and has important applications in many areas such as stock price (Tsay, 1988), oil & gas (Marti et al., 2015), economics, business analytics, and others. In sequential data from call centers, change points arise either due to known business calender like Thanksgiving, Christmas, and New Year or external events like movement in economy, merging and splitting of businesses. For example, estimating a change point and its impact due to an announcement of new iPhone in the market by Apple is a very difficult problem, and its inaccurate estimation generally costs hundreds of million of dollars. In these application areas, it is always desirable to search for the causes and sources of change points, so that such change point behavior can be properly analyzed and better understood; and if desired post detection prescriptive recommendation could be taken. Therefore, the task of finding change points has been the focus of considerable research in sequential data analysis.

We present a novel and efficient algorithm for variance change point detection called dynamic mode decomposition based variance change point detection (DVCPD) which uses a DMD based data-driven dynamical system (Schmid, 2011), local adaptable window, and sequential hypothesis testing to iteratively detect variance change point. (a) The location and size of the window is automatically governed by the acceptance and rejection of the hypothesis. Thus, the variance change is detected locally as well as globally unlike the well-known methods (Tsay, 1988) or (Inclán & Tiao, 1994) where the variance change is detected on the whole dataset.(b) DVCPD normalized data in a suitable way so that the presence of other variance changes, that might have been hidden due to masking, could be accurately detected. (c) Using real-world univariate datasets we demonstrate superior accuracy as compared to well-known prior-art.

2. Dynamic Mode Decomposition (DMD)

DMD is a data-driven dynamical system (Schmid, 2011) that works by extracting information from a sequence of data. It computes a linear map of a sequential data generated from a nonlinear process (in a least- squares sense), that describes the evolution of the dynamics over a small time interval. The eigenvalues and eigenvectors of this map capture the principal dynamics of the time snapshots. We briefly review DMD algorithm (Grosek & Kutz, 2014) in the follow-

Appearing in Proceedings of the 2^{nd} Indian Workshop on Machine Learning, IIT Kanpur, India, 2016.

ing Section.

2.1. DMD Algorithm

Let $x_t \in \mathbb{R}^J$ be data points collected at time $t = 1, 2, ..., T$. The data can be grouped into matrices as follows: $X = [x_1\ x_2\ ...\ x_{T-1}]$, $Y = [x_2\ x_3\ ...\ x_T]$. The Koopman operator A maps the data at time t to time $t+1$ such that $x_{t+1} = Ax_t$. The DMD algorithm estimates the Koopman operator A that best represents the data in X such that the columns of:

$$X = [x_1\ Ax_1\ A^2x_1\ ...\ A^{T-2}x_1],$$

form a Krylov space generated by x_1 and the matrix A. Thus, $AX = Y$. The last data point x_T is determined by $x_T = \sum_{t=1}^{T-1} k_t x_t + \epsilon$, where $k_t's$ are the coefficients of the Krylov space basis vectors and ϵ is the residual error. Notice that $Y = XS + \epsilon.e_{T-1}^T$, where e_{T-1} is the $(T-1)^{\text{th}}$ unit vector and S is a $(T-1) \times J$ matrix whose subdiagonal entries are 1 and the last column is $k_t, t \in [1..T-1]$.

Let the SVD of X be $U\Sigma V^T$, where $U \in \mathbb{R}^{J\times r}$, $V \in \mathbb{R}^{(T-1)\times r}$ have orthonormal columns, and $\Sigma \in \mathbb{R}^{r\times r}$ is diagonal. The parameter r is chosen to capture the fundamental structure and dynamics of the system represented by the data in X. Given ϵ is small one can estimate (Grosek & Kutz, 2014):

$$\begin{aligned} Y &\approx XS \approx U\Sigma V^T S \\ S &\approx V\Sigma^{-1}U^T Y. \end{aligned}$$

Using the similarity transform $V\Sigma^{-1}$, the matrix $\tilde{S} \approx U^T Y V \Sigma^{-1}$ can be derived, which is similar to the matrix S. Since, $AX = Y \approx XS$, some of the eigenvalues of the matrix S approximate the eigenvalues of the Koopman operator A. Also, $AU \approx U\tilde{S}$ and the eigenvectors of matrix $\tilde{S}$ approximate those of A. The complete algorithm is given below, in Algorithm 1.

Algorithm 1 DMD Algorithm

procedure DMD(data $= (x_1, x_2, ..., x_T)$)
  Arrange data into **matrices** X, Y
  $X = [x_1|x_2|...|x_{T-1}]$ and $Y = [x_2|x_3|...|x_T]$
  Compute **(reduced) SVD** of X, $X = U\Sigma V^T$
  Define matrix, $\tilde{A} \triangleq U^T Y V \Sigma^{-1}$
  Compute **eigenpairs** of $\tilde{A}$, writing $\tilde{A}w = \lambda w$.
  Each $\lambda \neq 0$ is a **DMD eigenvalue**
  DMD λ - mode is : $\varphi = \frac{1}{\lambda} Y V \Sigma^{-1} w$
end procedure

3. DVCPD Algorithm

We designed the variance change point detection algorithm so that it can accurately detect local and global change points in the given time series. For this, we consider varying window sizes. Varying window sizes are supposed to describe the local state of the underlying process. We start with smaller window sizes to detect true change point locally and then increase the window size up to a significant proportion of the whole time series so we can detect change points globally. We show that our algorithm accurately detects the variance change points as compared to typical algorithms (Tsay, 1988) (Inclán & Tiao, 1994).

3.1. Model Construction:

We fit DMD model on the given time series, $\{x_1, \ldots, x_T\}$, using Algorithm 1. (In general, we could also use other models as well such as: (a) the best possible $ARIMA(p, d, q)$ model and estimate the exact shocks $a = \{a_1, a_2, \ldots, a_T\}$ using maximum likelihood method (Box et al., 2008), or (b) State space model with a represented as estimated smoothed observation disturbance (Durbin & Koopman, 2012). We use the method of cumulative sum of squares ((Inclán and Tiao, 1994) (Inclán & Tiao, 1994)) on the shock $\{a_1, a_2, \ldots, a_T\}$, starting with $\alpha = \alpha_0$, to detect the potential variance change point k. After determining the value of k we split the time series observation into two subsections $\{x_1, x_2, \ldots, x_{k-1}\}$ and $\{x_k, x_2, \ldots, x_{[\alpha T]}\}$ and apply the Wald test statistics W^* to decided if the k^{th} point is a true variance change point or not. If the k^{th} is a variance change point, we move our window to the next section of data or otherwise increase the starting size of the window by $[\alpha T]$ and check for the potential change point and the true variance change point again. If two variance change points detected are within 10 data points of each other we choose the one which has smaller P-value. We iterate the algorithm for different values of α till $\alpha = \alpha_{max}$.

3.2. Iterations with Adaptable Window Sizes:

This section describes the procedure AdaptWin. We consider a subsection of the data by constructing a starting window of size $[\alpha T]$ in each iteration, where $\alpha \in \mathcal{P} = \{\alpha_0, \alpha_0 + \beta, \alpha_0 + 2\beta \ldots, \alpha_{max}\}$, and $[x]$ is the largest integer smaller than equal to x, and we search for the variance change point in this window.

3.3. Change Points Selection:

Let m_i be the number of variance change points detected when Algorithm 2 was started by taking $\alpha = \alpha_0 + \beta(i-1)$, $i = 1, 2, \ldots, \gamma$, where, $\gamma = (\frac{\alpha_{max} - \alpha_0}{\beta} + 1)$. The number of variance change points present in the data is defined as $m =$ mode of $m_1, m_2, \ldots, m_\gamma$. Since

Algorithm 2 DVCPD: Variance change point

while no variance change is detected **do**
  DMD model construction (Algorithm 1)
  residual ← **DMD Model**
  procedure AdaptWin(data, residual)
    for $\alpha = \alpha_0$, $(\alpha_0 + 2 * \beta)$,..., α_{max} **do**
      while All data is read **do**
        Select **window**$(start, size)$
        Find **Impact point**(window) k
        Wald test(window, k)
        if rejected **then** variance CP
          move window to next section
        else Increase **window** size by $[\alpha T]$
        end if
      end while
    end for
    Determine **split points** and **base model**
  end procedure
  Time series ←**Normalization**
end while

we are applying our algorithm for each α from α_0 to α_{max}, the variance change points that were detected for some value $\alpha = \alpha_0$ should also be detected for other values of α which are close to α_0 if it has a significantly large impact factor. Then we determine those variance change points that are occurring in most of the iterations and take top m of those variance change points.

3.4. Base Model Selection & Normalization:

The main purpose of normalization is to remove the effect of already detected variance change point and rerun the above proposed algorithm on the updated data to capture new variance change points which could have been present in the data but remained hidden due to the effect of an already detected variance change point which has a very large impact factor.

Let us assume the process in the subsection i has the largest number of data points and $i \in \{2, 3, \ldots, m\}$. Let v_b be the variance of the process $\{x_{k_{i-1}}, x_{k_{i-1}+1}, \ldots, x_{k_i-1}\}$ in subsection i. Let v_l be the variance of the process in the subsection $i-1$ with data points $\{x_{k_{i-2}}, x_{k_{i-2}+1}, \ldots, x_{k_{i-1}-1}\}$ and let Let v_r be the variance of the process in the subsection $i+1$ with data points $\{x_{k_i}, x_{k_i+1}, \ldots, x_{k_{i+1}-1}\}$. Then the updated normalized new process is

$$x_t^* = \left[\bar{x} + \sqrt{\frac{v_b}{v_l}}(x_t - \bar{x})\right] 1_{(1 \le t < k_{i-1})} + x_t 1_{(k_{i-1} \le t < k_i)} + \left[\bar{x} + \sqrt{\frac{v_b}{v_r}}(x_t - \bar{x})\right] 1_{(k_i \le t \le T)}, \quad t = 1, 2, \ldots T$$

where, $\bar{x}$ is the average of all T data points in the original data set. Observe that due to this normalization variances in subsection $i-1$, i and $i+1$ of that updated data $\{x_1^*, x_2^*, \ldots, x_T^*\}$ are all same and in particular the variance is equal to v_b.

4. Experimental Results & Analysis

In this section, we present the experimental results on variance change point detection of univariate data on (a) **IBM stock price** (b) **Nile data**, (c) **Average handling time**.

4.1. IBM stock price

The data set considered here is the first difference of the IBM stock closing price from May 17, 1961 to November 2, 1962 as reported by (Box et al., 2008). They have identified an ARIMA(0, 1, 1) as the best model for this series, however they found that some evidence of possibility of inadequacy of the ARIMA(0, 1, 1) might be in part by change in variance. Therefore, this data set is extensively studied by various authors to illustrate their theory of variance change (Inclán and Tiao, 1994) (Inclán & Tiao, 1994), (Tsay, 1988) (Tsay, 1988).

Change Points Detected: The change points detected using DVCPD, Algorithm 2, are shown in Figure 1. Data points - 236, 279 are found to be the variance change points in the first iteration. Since the largest section on which variance remained stable is from data point 1 to data point 235, this section is taken as the base model. With respect to the base model the difference data is normalized to get the updated data. The plot of the normalized data is shown in each subplot of the Figure 1. The algorithm is then applied on the updated data, and 279^{th} point is found as the variance change point. And finally, on the renormalized data the algorithm detected 180^{th} as the variance change point.

4.2. Nile data

We consider the classic example of change point data set, the minimum water levels of the Nile river during the AD 622-1284, measured at the island of Roda, near Cairo, Egypt. Several authors have reported evidence supporting a change point in this data around the year AD 722 (Ray and Tsay, 2002).

Change Points Detected: Table 1 presents the results from multiple iterations of DVCPD and the corresponding change points detected on Nile data. DVCPD detects 720 and 805 points as the variance change points. A change point, which is a variance change point around AD 722, agrees with the previous result which uses sequential Bayesian one step ahead

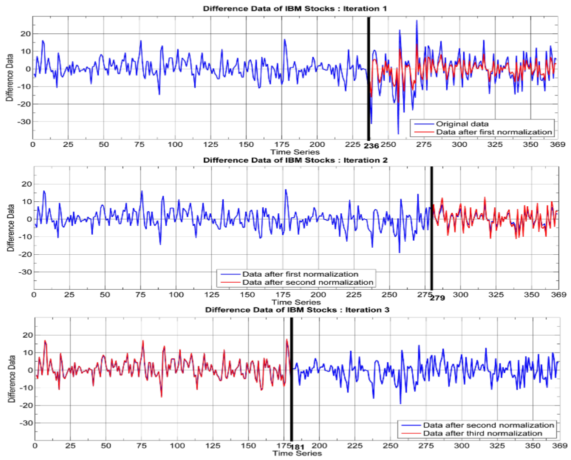

Figure 1. Difference data of IBM stock price: plots show variance change points as thick vertical lines and normalized series in each of the three iteration.

Table 1. Variance change points from Nile and AHT data

Dataset	Variance change points
Nile	720, 805
Average handling time	56, 276, 355, 786, 983

prediction in the presence of change points (Saatci et al., 2010).

4.3. Average handling time

As a final example, we consider the time series of average handling time (AHT) of an agent from 01-08-2009 to 09-30-2012 of a call center. AHT is the average time taken by an agent to answer a call from a customer. It is important to determine the time instances when the variance changes occur in AHT. If the variation in AHT increases consistently from the previously seen pattern, that means agents are taking more time to answer calls, and also within the same time period some other calls are answered in relatively small time. For some calls AHT is so high, it may lead to large number of calls in a queue and more often than not important calls may get dropped without being answered. Hence when the variation within AHT is large, it is important to have enough number of agents deployed to answer all the calls. Only after the variance change points are detected in a time series, one can schedule enough agents to answer calls within the specified time.

Change Points Detected: Results of multiple variance change point detection, using DVCPD, are shown in the Table 1. In the first iteration, the variance change points obtained are: 56, 276 and 786. In the next iteration the new change points obtained are: 355 and 983.

5. Conclusions

We proposed the DVCPD algorithm which accurately captures the variance change locally and globally and uses normalization to capture hidden change points. By using DMD, it provides model-free approach that makes it flexible while still being accurate. On multiple univariate data sets, empirical results show the efficacy of the DVCPD algorithm. It detects variance change points that are similar or better than prior approaches (Tsay, 1988) (Inclán & Tiao, 1994) and owing to DMD provides superior performance over ARIMA and State Space.

6. Acknowledgments

The authors would like to thank Abha Srivastva for helpful discussion and proof reading.

References

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. *Time series analysis: forecasting and control.* Wiley Series in Probability and Statistics, 4^{th} Edition, 2008.

Durbin, J. and Koopman, S. J. *Time series analysis by state state space methods.* Oxford University Press, 2^{nd} Edition, 2012.

Grosek, J. and Kutz, J.N. Dynamic mode decomposition for real time background/foreground separation in video. *arXiv*, pp. 1404–7592, 2014.

Inclán, C. and Tiao, G. C. Use of cumulative sums of squares for retrospective detection of change of variance. *Journal of the American Statistical Association*, Vol 89(427):913–923, 1994.

Marti, L., Sanchez-Pi, N., Molina, J. M., and Garcia, A. C. B. Anomaly detection based on sensor data in petroleum industry applications. *Sensors*, 15:2774 – 2797, January 2015.

Saatci, Y., Turner, R., and Rasmussen, C. E. Gaussian process change point models. *Proceedings of the 27^{th} International Conference on Machine Learning, Haifa, Israel*, 2010.

Schmid, P. Application of the dynamic mode decomposition to experimental data. *Experimental Fluids*, 50:1123–1130, 2011.

Tsay, R.S. Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7:1–20, 1988.