Unit Root Test를 기반으로 한 장기 시계열 데이터의 non-stationary 발생에 따른 추세 변화 검정 및 시각화 연구

Jae-seong Yoo

November 1, 2018

Authors

Jaeseong Yoo*
Ph.D Candidate
lv999@korea.ac.kr

Jaegul Choo
Assistant Professor
jchoo@korea.ac.kr

Dept. of Computer Science and Engineering
College of Informatics
Korea University

Goals

- 비정상(non-stationary) 장기 시계열 안에서도, 단기적으로 추세의 변화가 일시적인 것인지, 아니면 구조적으로 변한 것인지를 적시에 판단하는 것은 중요함.
- 이는 시계열 추세의 변화를 상시 감지하여, 변화에 맞는 적정한 수준의 대응을 할 필요가 있기 때문임.

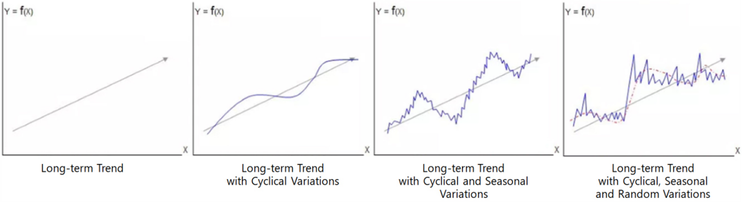
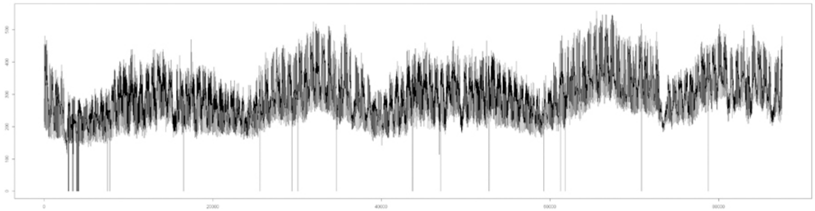
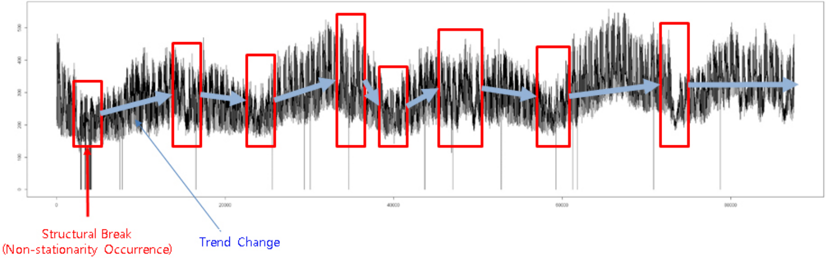


Figure: 시계열 데이터의 변동 요소에 따른 변화

Goals



Goals


Structural Break
(Non-stationarity Occurrence)
Trend Change

Non-stationary

우선, 시계열을 다음과 같은 확률과정으로 정의할 것이다.

$$\{y(s,t),\ s \in \mathfrak{S}, t \in \mathfrak{T}\}$$

여기서, $t \in \mathfrak{T}, y(s,t)$는 확률공간 $\mathfrak{S}$ 상의 확률변수
확률과정의 실현은 $t \in \mathfrak{T}$ 에 관한 각 $s \in \mathfrak{S}$에 대해 $y(s,\cdot)$로 주어진다.

따라서, 우리가 실제로 관찰하는 데이터는, 알려지지 않은 확률과정의 실현, 즉 데이터 생성 과정이라고 할 수 있다.

$$\{y\}_{t=1}^{T} = \{y_1, y_2, \cdots, y_t, \cdots, y_{T-1}, y_T\}, \quad (t = 1, \cdots, T \in \mathfrak{T})$$

Non-stationary

정상 프로세스(stationry process)의 정의

$$E[y_t] = \mu < \infty, \ \ \forall t \in \mathfrak{T}$$

$$E\left[(y_t - \mu)(y_{t-j} - \mu)\right] = \gamma_j, \ \ \forall t, j \in \mathfrak{T}$$

더욱 엄격한 정의

$$F\left\{y_1, y_2, \cdots, y_t, \cdots, y_T\right\} = F\left\{y_{1+j}, y_{2+j}, \cdots, y_{t+j}, \cdots, y_{T+j}\right\}$$

Non-stationary

다음과 같은 시계열이 있다고 가정해보자.

$$y_t = \beta y_{t-1} + \epsilon_t$$

여기서 y_t는 시간 t에서의 값이고, ϵ_t는 오차항이다.

y_t를 계산하기 위해서는 y_{t-1}의 값이 필요하다. y_{t-1}을 구하려면 마찬가지로 다음과 같이 구하면 된다.

$$y_{t-1} = \beta y_{t-2} + \epsilon_{t-1}$$

모든 관측치에 대해서 이와 같이 구한다고 하면, y_t의 값은 결국 다음과 같이 구한다고 볼 수 있다.

$$y_t = \beta^n y_{t-n} + \sum \epsilon_{t-i} \beta^i$$

위의 방정식에서 β의 값이 1이면 (즉, 단위근이면) 예측 값은 $y_{t-n} + \sum \epsilon_{t-i}$와 같게 된다. 이는 분산이 시간에 따라 증가함을 의미한다. 정상 시계열에 대해 분산이 시간에 의존적이면 안된다. Unit root test는 $\beta = 1$인지 아닌지를 확인하여 계열의 단위근 유무를 검정하는 방법이다.

Unit Root Test

AR(1) 하에서의 단위근 검정(unit root test) 절차

$$AR(1) : y_t = \rho y_{t-1} + \nu_t, \quad \nu_t \sim N(0, \sigma_\nu^2)$$

$$\text{가설:} H_0 : \rho = 1, \quad H_1 : |\rho| < 1$$

이러한 검정을 위한 검정통계량은 다음과 같은 변형된 식으로부터 얻어진다.

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \nu_t$$

$$\Delta y_t = (\rho - 1) y_{t-1} + \nu_t = \gamma y_{t-1} + \nu_t \ \ \text{여기서, } \gamma \equiv \rho - 1$$

$$\Longrightarrow \begin{cases} H_0 : & \rho = 1 \\ H_1 : & \rho < 1 \end{cases} \text{or} \begin{cases} H_0 : & \gamma = 0 \\ H_1 : & \gamma < 1 \end{cases}$$

이 AR(1)에서 y_t를 차분한 $\Delta y_t = y_t - y_{t-1}$은 ϵ_t, 즉 white noise이므로 stationary가 된다.

Unit Root Test

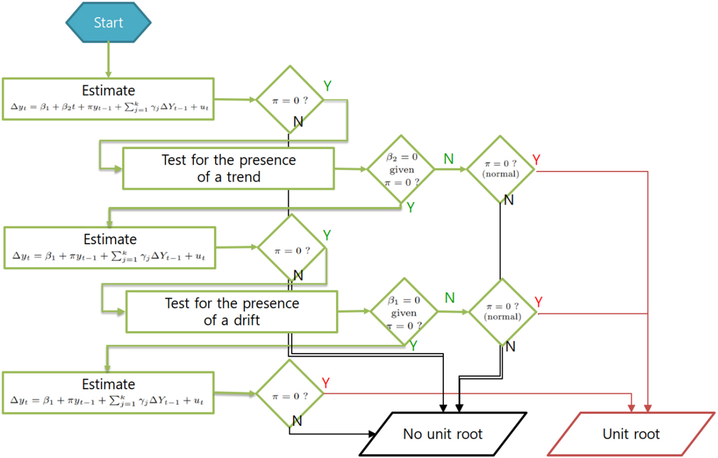


Figure: 단위근 검정 절차

Unit Root Test

다양한 종류의 stationarities와 테스트 결과를 해석하는 방법을 알아보자.

- Strict Stationary: Strict stationary는 stationary process의 수학적 정의를 충족시킨다. 엄격하게 정상성을 만족하는 경우, 평균, 분산 및 공분산은 시간에 의존적이지 않다. 비정상 시계열을 엄격하게 정상성을 만족하는 시계열로 변환할 수 있다면 예측이 아주 수월해질 것이다.
- Trend Stationary: 단위근이 없지만, 추세(trend)가 나타나는 시리즈를 trend stationary라고 한다. 추세가 제거되면 시리즈는 strict stationary가 된다. KPSS 테스트는 단위근이 없는 경우 시리즈를 strict stationary라고 본다.
- Difference Stationary: Differencing에 의해 strict stationary로 변환할 수 있는 시계열이 있을 수 있다. ADF test는 이러한 difference stationary에 있는 시계열을 검정하는 방법이라고 할 수 있다.

Unit Root Test

ADF, KPSS 테스트 외에도 다른 검정 방법이 많지만 일단 이 두 검정만 진행하였다고 하면 다음과 같은 4가지 케이스가 나올 것이다.

- Case 1: 두 테스트 모두 시리즈가 non-stationary라는 결론을 내림
  → 시리즈가 non-stationary
- Case 2: 두 테스트 모두 시리즈가 stationary라는 결론을 내림
  → 시리즈가 stationary
- Case 3: KPSS는 stationary라는 결론을 도출했지만 ADF는 그렇지 않음
  → 트렌드를 제거하면 stationary
- Case 4: ADF는 stationary라는 결론을 도출했지만 KPSS는 그렇지 않음
  → 차분을 하면 stationary

Unit Root Test - ADF(Adjusted Dickey-Fuller) Test

DF 검정

DF 검정은 다음 각각의 모형을 설정하여 귀무가설*을 검정한다.
(귀무가설: $\gamma = 0$, 즉 해당 시계열이 비정상(non-stationary))

- 랜덤워크 과정:

$$\Delta y_t = \gamma y_{t-1} + \nu_t$$

- 랜덤워크 과정이 상수항을 가지는 경우:

$$\Delta y_t = a_0 + \gamma y_{t-1} + \nu_t$$

- 랜덤워크 과정이 상수항을 가지고 비확률 추세를 포함한 경우:

$$\Delta y_t = a_0 + \alpha_1 t + \gamma y_{t-1} + \nu_t$$

Unit Root Test - ADF(Adjusted Dickey-Fuller) Test

Adjusted DF 검정

DF 검정은 오차항에 자기상관이 있지 않다는 가정이 전제된다. 오차항에 자기상관이 있을 경우를 고려하여 모형을 설정하고 단위 검정을 하는 것을 adjusted DF 검정이라고 한다.

- 랜덤워크 과정:

$$\Delta y_t = \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + u_{1t}$$

- 랜덤워크 과정이 상수항을 가지는 경우:

$$\Delta y_t = a_0 + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + u_{1t}$$

- 랜덤워크 과정이 상수항을 가지고 비확률 추세를 포함한 경우:

$$\Delta y_t = a_0 + a_1 t + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + u_{1t}$$

Unit Root Test - ADF(Adjusted Dickey-Fuller) Test

- DF의 가정은 오차항이 독립적이며 동일한 분포를 한다는 것이다.
- 또한 ADF 검정은 설명변수에 시차를 갖는 차분 값을 포함함으로써 자기상관의 문제를 고려하고 있다.

Unit Root Test - PP(Phillips-Perron) Test

PP 검정은 시차를 갖는 차분 값의 포함 없이 자기상관을 고려하는 방법을 제시하였다.
PP 검정은 다음과 같은 모형을 설정하여 검정을 진행한다.

- Model 1:

$$y_t = \mu + \alpha y_{t-1} + \epsilon_t$$

- Model 2:

$$y_t = \mu + \beta \left[t - \frac{1}{2} T \right] + \alpha y_{t-1} + \epsilon_t$$

Unit Root Test - PP(Phillips-Perron) Test

- Model 1:

$$y_t = \mu + \alpha y_{t-1} + \epsilon_t$$

검정통계량

$$Z(\hat{a}) = T(\hat{a} - 1) - \hat{\lambda}\sqrt{m_{yy}}$$

$$Z(\tau_{\hat{a}}) = \left(\frac{\hat{s}}{\hat{\sigma}_{T\ell}}\right) t_{\hat{\sigma}} - \frac{\hat{\lambda}'_{T\ell}}{\bar{m}_{yy}^{1/2}}$$

$$Z(\tau_{\hat{\mu}}) = \left(\frac{\hat{s}}{\hat{\sigma}_{T\ell}}\right) t_{\hat{\mu}} - \frac{\hat{\lambda}'_{T\ell} m_y}{\bar{m}_{yy}^{1/2} m_{yy}^{1/2}}$$

여기서

$$m_y = T^{-3/2} \sum y_t$$

$$\bar{m}_{yy} = T^{-2} \sum (y_t - \bar{y})^2$$

$$m_{yy} = T^{-2} \sum y_t^2$$

$$\hat{\lambda} = 0.5(\hat{\sigma}_{T\ell}^2 - \hat{s}^2)$$

Unit Root Test - PP(Phillips-Perron) Test

- Model 2:

$$y_t = \mu + \beta \left[t - \frac{1}{2} T \right] + \alpha y_{t-1} + \epsilon_t$$

검정통계량

$$Z(\hat{a}) = T(\hat{a} - 1) - \frac{\hat{\lambda}}{M}$$

$$Z(\tau_{\tilde{a}}) = \left(\frac{\tilde{s}}{\tilde{\sigma}_{T\ell}} \right) t_{\hat{\sigma}} - \frac{\tilde{\lambda}'_{T\ell}}{M^{1/2}}$$

$$Z(\tau_{\tilde{\mu}}) = \left(\frac{\tilde{s}}{\tilde{\sigma}_{T\ell}} \right) t_{\tilde{\mu}} - \frac{\tilde{\lambda}'_{T\ell} m_y}{\left[M^{1/2} (M + m_y^2) \right]^{1/2}}$$

$$Z(\tau_{\tilde{\beta}}) = \left(\frac{\tilde{s}}{\tilde{\sigma}_{T\ell}} \right) t_{\tilde{\beta}} - \frac{\tilde{\lambda}'_{T\ell} \tilde{\sigma}_{T\ell} \left(\frac{1}{2} m_y - m_{ty} \right)}{\left[\left(\frac{M}{12} \right)^{1/2} \bar{m}_{yy}^{1/2} \right]^{1/2}}$$

여기서

$$M = (1 - T^{-2}) m_{yy} - 12 m_{ty}^2 + 12(1 + T^{-1}) m_{ty} m_y - (4 + 6T^{-1} + 2T^{-2}) m_y^2$$

$$m_{ty} = T^{-5/2} \sum t y_t$$

Unit Root Test - ERS(Elliott-Rothenberg-Stock) Test

앞선 두 가지 단위근 검정의 단점은 실제 데이터 생성 과정이 계수가 1에 가까운 AR(1) 과정인 경우 검정력이 낮아진다는 것이다.

ERS 검정은 Dickey-Fuller 단위근 검정을 변형시켜 검정력을 향상시키기 위한 방법이다.

이 방법의 기각역은 Elliott, Rothenberg and Stock이 표로 제시하였다.

Unit Root Test - SP(Schmidt-Phillips) Test

DF 검정의 도 다른 단점은 불필요한 파라메터가 확정되지 않았다는 것이다.
(또는, 확정되었다 해도, 대립가설 하에서 다른 해석을 하게 된다.)

SP 검정은 귀무가설과 대립가설 하에서 동일한 일련의 불필요한 파라메터를 정의하는 Lagrange multiplier 검정 방법을 제안하였다.
또한 이 검정 방법에서는 선형 추세보다 높은 다항식을 고려한다.

$$y_t = \alpha + Z_t\delta + x_t, \quad Z_t = (t, t^2, \cdots, t^p)$$

$$x_t = \pi x_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$

Unit Root Test - SP(Schmidt-Phillips) Test

$$y_t = \alpha + Z_t\delta + x_t, \quad Z_t = (t, t^2, \cdots, t^p)$$
$$x_t = \pi x_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$

검정통계량은 위 회귀 식을 실행함으로써 구성된다.

$$\Delta y_t = \Delta Z_t\delta + u_t$$

1. $\tilde{\psi}_x = y_1 - Z_1\tilde{\delta}$를 계산한다.
2. $\tilde{S}_t = y_t - \tilde{\psi}_x - Z_t\tilde{\delta}$를 정의한다.
3. 검정을 위한 회귀 식을 다음과 같이 고려한다.

$$\Delta y_t = \Delta Z_t\gamma + \phi\tilde{S}_{t-1} + \nu_t, \quad \nu_t\text{는 오차항}$$

4. SP 검정의 검정통계량을 다음과 같이 계산한 후, 계상된 검정통계량을 바탕으로 검정을 한다.

$$Z(\rho) = \frac{\tilde{\rho}}{\hat{\omega}^2} = \frac{T\tilde{\phi}}{\hat{\omega}^2}$$

여기서,

$$\hat{\omega}^2 = \frac{T^{-1}\sum_{i=1}^{T}\hat{\epsilon}_t^2}{T^{-1}\sum_{i=1}^{T}\hat{\epsilon}_t^2 + 2T^{-1}\sum_{s=1}^{\ell}\sum_{t=s+1}^{T}\hat{\epsilon}_t\hat{\epsilon}_{t-s}}$$

Unit Root Test - KPSS(Kwiatkowski-Phillips-Schmidt-Shin) Test

앞서 소개한 검정 방법들은 귀무가설이 단위근 과정이었지만, KPSS 검정은 귀무가설이 정상 과정 (stationary process)이다.
KPSS 검정에서는 다음과 같은 모델을 고려한다.

$$y_t = \xi t + r_t + \epsilon_t, \;\; \epsilon_t \sim N(0, \sigma_u^2)$$

$$r_t = r_{t-1} + u_t$$

여기서,

- r_t는 랜덤 워크이다.
- $\xi = 0$이면, 이 모델은 결정론적 회귀변수만 남아 상수로 간주한다.

Unit Root Test - KPSS(Kwiatkowski-Phillips-Schmidt-Shin) Test

1. 레벨이나 추세를 테스트할지 여부에 따라 상수 또는 추세에 대해 y_t의 회귀식을 세운다.
2. 이 식으로부터 잔차의 부분 합을 다음과 같이 계산한다.

$$S_t = \sum_{i=1}^{t} \hat{\epsilon}_i, \quad t = 1, 2, \cdots, T$$

3. 그러면, 검정통계량은 다음과 같이 구할 수 있다.

$$LM = \sum_{t=1}^{T} \frac{S_t^2}{\hat{\sigma}_\epsilon^2}$$

여기서, $\hat{\sigma}_\epsilon^2$은 위 단계에서 구한 오차 분산의 추정치

Trend testing and visualization based on Unit root test

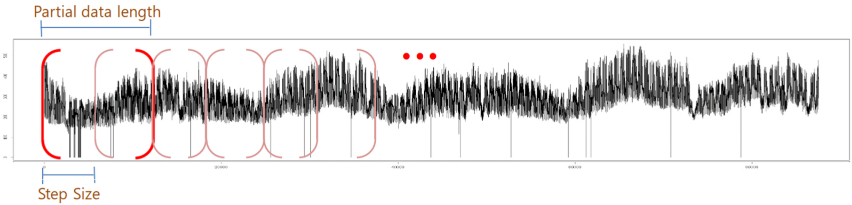


Figure: 시계열의 partial data length와 step size를 이용한 분할

Trend testing and visualization based on Unit root test

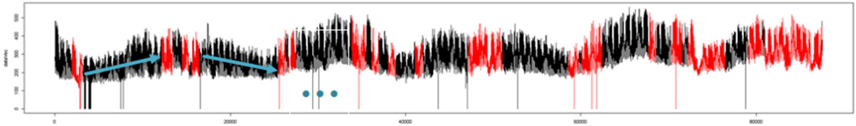
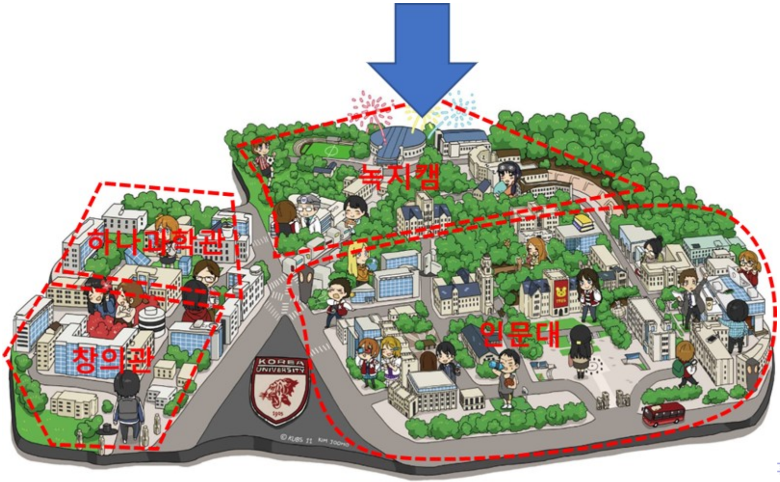


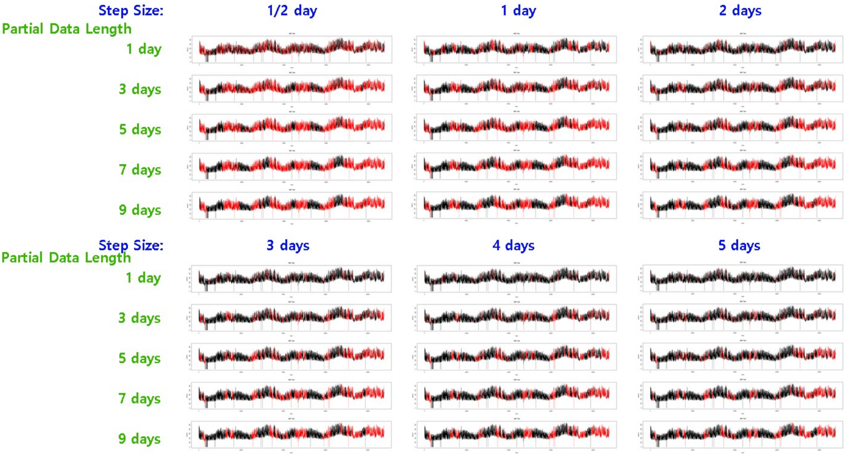
Figure: 단위근 검정 결과를 바탕으로 일부 부분 데이터를 강조

Experiments

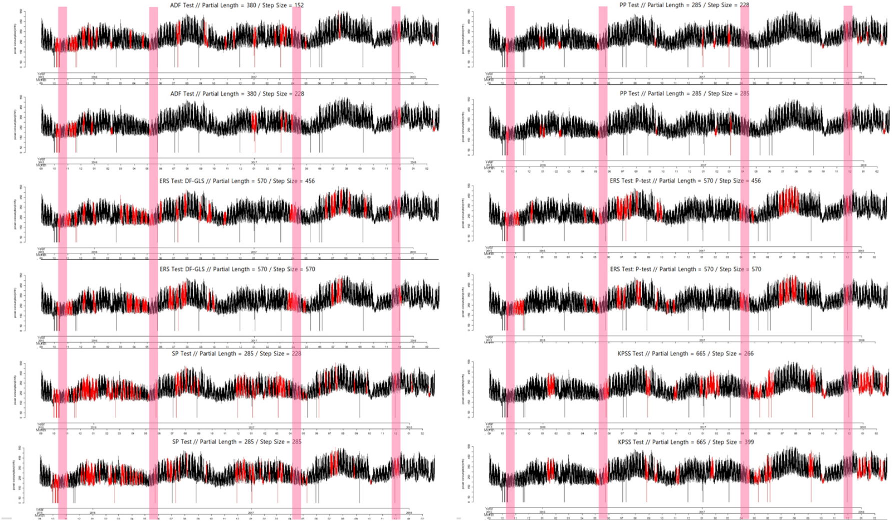
- 고려대학교 녹지캠퍼스의
  2015년 9월 1일부터 2018년 2월 28일까지
  15분 단위로 측정한 전력 사용량(kWh) 데이터를 사용하였다.



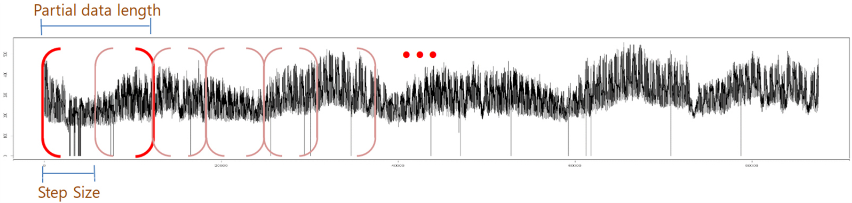
Experiments



Experiments



Remaining Works



Partial Length, Step Size 조절, Trend reduction, Differencing 등을 사용자가 interactive하게 조절할 수 있는 장치 개발중

* 주재걸, 유재성, "전력 사용량의 실시간 structural break 발생 여부 탐지 장치", (대한민국 특허출원 예정)