

# [Data analysis] 기본적인 시계열 분석(Time Series Analysis)의 종류와 개념 설명 – 나무늘보의 개발 블로그

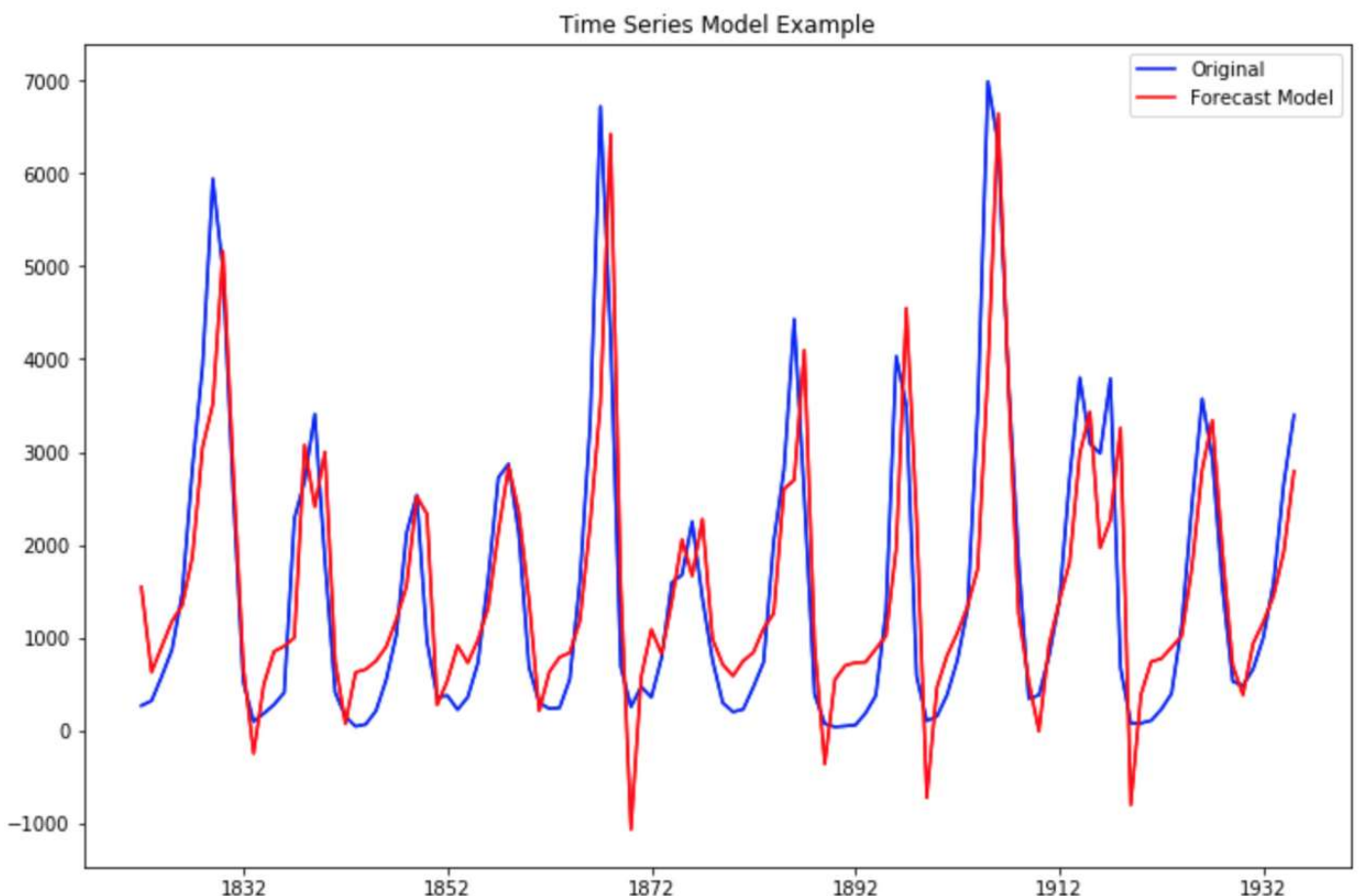
🌐 웹 클립



Data scientist/Data analysis

## [Data analysis] 기본적인 시계열 분석(Time Series Analysis)의 종류와 개념 설명

2021. 4. 25. 17:03 수정 삭제 공개



# 1.시계열 분석을 적용하는 대상

회귀분석과 시계열 분석의 차이는 시간을 고려한다는 차이점이다.

그러한 이유로 시계열 분석은 1종 데이터에 대해서도 자체적으로 분석이 가능하다.

## 2.대표적인 시계열 데이터의 특성과 모형 소개

시계열 정보 = 규칙성을 가지는 패턴 + 불규칙한 패턴의 결합

### 규칙성

- 자기상관성
- 이동 평균

자기상관성이란 ?

이전의 결과와 이후의 결과 사이에서 발생하는 자기상관성

이동평균이란?

이전에 생긴 불규칙한 사건이 이후 결과에 편향성을 초래하는 이동 평균

### 불규칙한 패턴

- White Noise

White Noise란?

- 불규칙한 패턴을 칭하며 평균이 0 이고 일정한 분산을 지닌 정규 분포에서 추출된 임의의 수치라고 규정

이러한 정규 분포의 가정은 모델의 해석을 편리하게 해준다.

대표적인 시계열 모델

- AR
- MA
- ARMA
- ARIMA

## 대표적인 시계열 모델 설명

### 1.자기상관(Autocorrelation) - AR 모형

어떤 Random Variable에 대해서 이전의 값이 이후의 값에 영향을 미치고 있는 상황  
ex) 이전에 값이 크면 이후에는 낮은 값이 나온다가어나하는 경향을 뜻한다.

자기상관성을 시계열 모형으로 구성한 것을 AR모형이라고 부르는데, 가장 간단한 형태가 바로 직전 데이터가 다음데이터 영향을 준다고 가정한 AR(1) 모형이다.

$$X(t) = \{a \cdot X(t-1) + c\} + u \cdot e(t)$$

이 식을 서술하면 시점 t에서 얻게 될 X(t)의 평균값은 시점 t-1에서 얻었던 X(t-1)의 값에 a를 곱하고 c를 더한 것 과 같다는 뜻이다.

$X(t)$  = 시점  $t$  에 얻게 될  $X(t)$ 의 평균값

$e(t)$  = White noise(평균이 0 이고 분산이 1인 정규분포에서 도출된 random 값)

$X(t)$  값은 평균이  $a \cdot X(t-1) + c$ 이며 분산이  $u$ 인 정규분포에서 도출되는 임의의 값이라는 뜻이다.

## 2.이동평균(Moving Average) - MA 모형

시간이 자날수록 어떠한 Random Varianabled의 평균값이 지속적으로 증가하거나 감소하는 경향이 생길수 있다.

ex) 여름에 가까워질수록 전기값이 올라가고 겨울이 될수록 낮아지는 경향

이러한 경우 평균이동이 있는 시계열 데이터가 될 가능성이 높다.

이동평균을 시계열 모형으로 구성한것을 MA 모형이라고 부른다. 이모델중 가장 간단한 형태가 MA(1) 모형이다.

$$X(t) = \{a \cdot e(t-1) + c\} + u \cdot e(t)$$

$X(t)$ 의 평균값은 시점  $t-1$ 에서 발생한 error  $e(t-1)$ 의 값에  $a$ 를 곱하고  $c$ 를 더한것과 같다는 뜻이다.

$X(t)$  값은 평균이  $a \cdot e(t-1)$ 이며 분산이  $u$ 인 정규분포에서 도출되는 임의의 값이라는 뜻이다.

AR과의 차이는 MA는 error값에 초점을 두고 AR은  $X(t-1)$ 값에 초점을 둔다.

## 2.3 ARMA(Autoregressive Moving Average) 모형

### 3.ARIMA(Autoregressive Moving Average) 모형

AR모형과 MA 모형을 합친것으로 연구기관에서 가장 선호되는 모델이다.

$$X(t) = \{a \cdot X(t-1) + c\} + \{b \cdot e(t-1) + c\} + u \cdot e(t)$$

원글자가 가장 좋아하는 모형은 ARMA(2,2)모형이다.

### 4.ARIMA(Autoregressive Integrated Moving Average) 모형

ARMA모형이 과거의 데이터들을 사용하는 것에 반해 ARIMA 모형은 과거의 데이터가 지니고 있는 추세데이터까지 반영하게 된다.

즉 Correlation 뿐 아니라 Cointegration까지 고려한 모델이다.

**Correlation** - 서로간의 선형관계

**Cointegration** - 추세관계

Cointegration은 시점이 고려되지 않으면 성립하지 않기 때문에 시계열 데이터에만 쓰이는 개념이다.

Correlation 0보다 크면 => x가 클때 y값도 큰값을 가진다.

Correlation 0보다 작으면 => x가 클때 y값은 작은값을 가진다.

Cointegration 0 보다 크면 => x의 값이 이전 값보다 증가하면 y 값도 증가한다.

Cointegration 0 보다 작으면 => x의 값이 이전값보다 증가하면 y 값은 감소한다.

ARIMA 모델은 자기 자신의 추세만 고려하기 때문에 White Noise의 추세는 고려하지는 않는다.

ARMA모델과 비슷하다고 생각해서 많은 사람들은 ARIMA 모델을 선호하지 않는다.

## 'Data scientist > Data analysis' 카테고리의 다른 글

[Data analysis] 기본적인 시계열 분석(Time Series Analysis)의 종류와 개념 설명

[Data analysis] 시간의 흐름에 따라 패널 데이터 분석 (데이터 분석의 힘 chapter.5)

[Data analysis] 계단식 변화가 있는 곳엔 집군분석 (데이터 분석의 힘 chapter.4)

[Data analysis] 급격한 변화의 경계선을 찾는 RD 디자인 (데이터 분석의 힘 chapter.3)

[Data analysis] 최선의 데이터 분석법, RCT (데이터 분석의 힘 chapter.2)

[Data analysis] 인과 관계와 상관 관계 (데이터 분석의 힘 chapter.1)

AR

ARIMA

Arma

Ma

시계열

시계열 모델

시계열 모델의 종류



나아무늘보

혼자 끄적끄적하는 블로그 입니다.

