

03 시계열분석

(주)씨에스리 조경미 기술사
(kmicho@cslee.co.kr)

시계열 분석(Time Series Analysis)

Concept	국내총생산량, 강수량, 주가지수 등과 같이 연도별, 계절별, 월별 또는 일별로 시간의 흐름에 따라 순서대로 관측된 자료를 분석하는 기법
Keyword	정상성, AR, MA, ARMA, ARIMA, 정상시계열, 비정상시계열 시계열변동요인(추세, 순환, 계절, 불규칙), 차분, 평활화, 변수변화 등

빅데이터 분석 인프라, 분석 알고리즘, R 을 이용한 통계적 분석, Python 을 이용한 머신 러닝, 딥러닝까지 최근에 많은 관심을 가지고 있는데, 그 중 분석 알고리즘의 하나로 **시계열 분석은 시간의 흐름에 따른 데이터의 변화 추이 또는 패턴을 찾아 미래를 예측할 수 있는 분석 기법**입니다.

시계열 분석이 어디에 쓰이는지 궁금하다면 통계청에서 운영하는 국가통계포털(<http://kosis.kr>) 홈페이지를 통해서 시계열 분석이 적용되고 있는 다양한 분야의 데이터 및 통계를 확인할 수 있다. 해마다 발표되는 국내총생산(GDP), 소비자물가지수, 환율, 실업률과 같은 경제활동 관련된 분야, 강수량, 기온, 미세먼지 등 기상과 관련된 분야가 있고 기업에서는 해마다 매출, 영업이익 등을 예측하여 목표를 수립하고 한 해를 열심히 달려갑니다.

이렇게 시계열 분석이 많이 활용되고 있는데, 시계열 분석의 원리 및 개념(정상성, 변동요인, 변수변환, 지수평활화, 차분 등)을 알아보고, 시계열 분석의 대표적인 기법(AR, MA, ARMA, ARIMA 등)을 알아보고자 합니다.

1. 시간의 흐름에 따른 관측 자료의 분석, 시계열 분석의 개념

- 시간의 흐름에 따른 데이터의 변화 추이 또는 패턴을 찾아 미래를 예측하는 분석 기법

분야	활용 사례	특징
국가 경제	국내총생산(GDP), 소비자 물가지수, 환율, 실업률 등	예측기반 정책 수립
기상	강수량, 강우량, 기온, 습도, 오존 등	농업, 어업, 관광업 등과 연계
기업 경제	매출, 영업이익, 손익분기점 BEP 등	이익 극대화 및 외부환경 영향

2. 시계열 분석을 위해 알아야할 원리 및 개념

가. [원리 및 개념 1] 정상성(Stationary)

- 정상성은 시점에 상관없이 시계열의 특성이 일정하다는 것을 의미하며, 정상성은 아래 3 개의 조건을 말하여 정상성 조건을 하나라도 만족하지 못하는 경우의 시계열자료를 비정상 시계열이라고 함

- 가) 평균이 일정하다
- 나) 분산이 시점에 의존하지 않는다
- 다) 공분산은 단지 시차에만 의존하고 시점 자체에는 의존하지 않는다.

- 대부분의 시계열 자료는 비정상 시계열이며, 시간에 따라 전개되는 특성과 주기적으로 변동하는 특성을 볼 때, 추세변동요인, 순환변동요인, 계절변동요인 및 불규칙변동요인으로 구성됩니다.

나. [원리 및 개념 2] 추세변동요인, 순환변동요인, 계절변동요인 및 불규칙변동요인

번호	변동요인	설명
1	추세변동요인	-인구변화, 기술변화, 생산성 증대 등 장기적인 변동으로 통상 10년 이상의 변동 주기를 가지는 변동 요인
2	순환변동요인	-경기순환 등에 따라 반복되는 변동으로 2년~5년 주기로 변화하는 변동 요인
3	계절변동요인	-계절의 변화 및 각종 관습에 의해 생성되는 1년 주기로 반복되는 변동 요인
4	불규칙변동요인	-추세, 순환, 계절 변동요인이 아닌 돌발적이거나 원인 불명의 요인에 의거하여 발생하는 변동 요인

다. [원리 및 개념 3] 안정시계열과 불안정시계열

안정시계열(stationary)	불안정시계열(nonstationary)
-시계열의 움직임이 구간이 달라지더라도 매 구간별 특성이 동일한 시계열 -시계열의 평균과 분산에 시간에 따른 규칙적인 변화가 없고 주기적 변화도 없는 시계열	- 시계열의 평균 및 분산이 시간에 따라 변화하는 시계열 -추세변동요인과 계절변동요인이 뚜렷한 경제시계열은 대체로 불안정시계열임

- 시계열의 흐름을 보다 정확히 파악하기 위해 시계열에 대해 변수변환 등을 하기로 합니다.

라. [원리 및 개념 4] 차분, 지수평활화, 변수변환

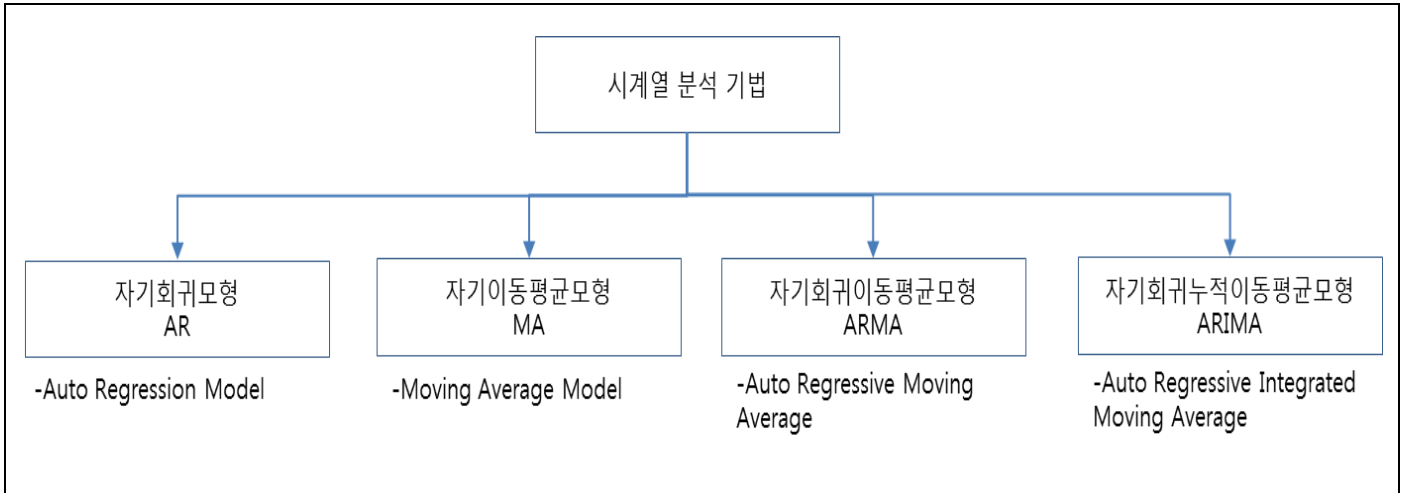
용어	설명
차분	-현재 시점 자료에서 과거 인접 시점의 자료를 차감하는 기법 -시계열의 차분으로 시계열이 전기에 비해 얼마나 증감했는지 파악
평활화	-주기가 짧은 변동 요인 즉 교란요인을 제거하여 시계열의 기초적 흐름 파악 - 중심화이동평균, 후방이동평균, 가중이동평균, 이중이동평균 등
변수변환	-로그(log)등 수학적 함수를 이용하여 큰 변동값은 작게, 작은 변동 값은 크게 만들어 주어 선형적 분석이 가능하게 하는 기법

- 정상성은 시점에 상관없이

3. 시계열 분석 기법 및 상세 모형

가. 시계열 분석 기법

- 시계열분석 기법으로는 자기회귀모형, 자기이동평균모형, 자기회귀이동평균모형, 자기회귀누적이동평균모형이 있습니다.



나. 시계열 분석의 상세 모형

상세모형	설명	수식
AR	- Auto Regression 자기회귀모형 -현 시점의 자료가 p 시점 전의 유한개의 과거 자료로 설명되는 모델	$Z_t = \theta Z_{t-1} + A_t$
MA	-같은 시점의 백색잡음과 바로 전 시점의 백색잡음의 결합으로 이루어진 모형	$Z_t = A_t - \theta_1 A_{t-1}$
ARMA	-Auto Regressive Moving Average 자기회귀이동평균모형 -AR, MA 모형을 동시에 포함하여 시계열모형의 모수를 줄여서 보다 효율적인 시계열 분석을 수행하는 모델 기법	
ARIMA	-Auto Regressive Integrated Moving Average 자기회귀누적이동평균모형 -시계열을 차분해서 ARMA 모형이 되는 모형	

- 이상 시계열 분석 모형은 선형 결합함수로 구성된 모형이고, 비선형 시계열 함수로 이선형 모델, TAR(Threshold Auto Regressive) 모형, GARCH(Generalized Auto Regressive Conditional Heteroscedasty) 모형도 있습니다.

스마트 팩토리(Smart Factory)의 빅데이터 분석 프로젝트를 진행할 때는 시계열 분석을 시간에 따른 흐름을 보는 정도로 진행해 봤었는데, 시계열분석을 알고 보니, 스마트 팩토리(Smart Factory)에서 주로 분석하는 온도, 진동, 전압, 압력 등의 데이터에서 온도 데이터는 계절의 영향을 받는 계절변동요인을 가지고 있고, 진동 데이터도 해수, 전력 등의 주변의 영향을 받아 불규칙변동요인의 특성을 가지기도 합니다. 이들 데이터는 특히 선형적이지 않고 비선형의 특성을 가지는 경우도 있으므로 다양한 시계열 분석 기법을 적용해 보면 좋을 것 같습니다.

[참고]

(도서) 예측방법론 / KNOU 이공희, 이한식, 장영재 공저

(도서) 데이터 분석 전문가 가이드 / 한국데이터진흥원

Contents connect communications!!

아이리포에 오시면 더 많은 지식을 가져가실 수 있습니다.

아이리포 IT지식창고 : <https://www.ilifo.co.kr/boards/knowledge>

아이리포 지덤시리즈 : <http://www.jidum.com>

아이리포 기술사/감리사 카페 : <http://cafe.naver.com/itlf>

서울시 마포구 상암동 1610번지, DDMC 3층 아이리포 교육센터

TEL: 02-303-9997 | MAIL: edu@ilifo.co.kr