

# 데이터가 보인다(6 - 7)

---

작성일 : 20.03.20.

작성자 : 문지선

## 6. 시뮬레이션

---

### 6.1. What-If 분석

- 미리 예상한 다양한 경우에 대해 결과를 계산해서 미래의 상황을 예측하기 위한 방법
  - What-If 분석을 이용하면 미래의 리스크나 수익을 정량적으로 파악할 수 있음
  - 주로 예측 가능한 상황에 대해 사전에 결과를 시뮬레이션할 때 이용
  - 변수를 변화시키면서 예상되는 모든 패턴의 결과를 계산
- 

### 6.2. 선형 계획법

- 복수의 제약 조건을 만족하는 최적의 값을 구하는 데이터 분석 방법
  - 제약 조건을 나타내는 식과 최적의 값을 계산하는 식 모두 1차식이 됨
  - 다양한 제약 조건에 대해 이익을 최대화 하는 방법을 수립할 때 이용
  - 제약 조건식 하에서 허용되는 변수 조합을 산출해내고, 이들 조합을 이용해 목적 함수에서 구할 수 있는 값이 최적이 되는 변수의 조합을 결정
    - 제약 조건식 : 변수가 취하는 값에 대한 식
    - 목적 함수 : 최적화하는 값을 계산하는 식
  - 변수의 수가 많을 경우에는 선형 계획법 중 하나의 방법인 심플렉스법을 이용
- 

### 6.3. 주성분 분석

- 데이터가 원래부터 갖고 있는 속성을 조합해서 데이터 전체의 분포를 가장 명확하게 설명할 수 있는 속성을 새롭게 작성하는 분석 방법
- 새롭게 작성된 속성은 원 속성의 가중 평균으로 표현되고 가중치는 데이터 분산이 최대가 되도록 하는 시뮬레이션에 의해 구해짐
- 데이터의 특징을 명확하게 비교하거나 그룹화할 때 자주 사용
- 주성분 분석에서는 모든 주성분 득점의 분산이 최대가 되도록 하는 주성분 부하량을 구함
  - 주성분 : 새롭게 작성된 데이터 전체의 특징에 대한 속성으로, 원 데이터의 속성의 가중 평균으로 나타냄
  - 주성분 부하량 : 가중 평균에서 사용되는 가중치
  - 주성분 득점 : 데이터를 주성분을 나타내는 가중 평균의 식에 대입해서 구해진 주성분의 값
- 주성분 득점 플롯 : 주성분 분석에서는 알아낸 주성분 득점을 분포도에 나타냄으로써 데이터를 분류하는데, 그 분포도를 주성분 득점 플롯이라고 함

## 7. 데이터 마이닝

---

### 7.1. 데이터 마이닝이란?

- 대량의 데이터로부터 눈에 보이지 않는 관련성이나 경향을 찾아내는 것을 목적으로 하는 애플리케이션
- 

### 7.2. 데이터마이닝을 수행하는 다양한 알고리즘

---

#### 7.2.1. 어소시에이션 룰

- 어떤 현상이 일어나면 동시에 다른 현상이 일어나는 것처럼 동시에 일어날 가능성이 높은 현상의 조합을 찾아내는 방법
  - 바스켓 분석 : 어소시에이션 룰에 기반한 상품 판매 데이터 분석
  - 소스 데이터 : 분석 대상이 되는 데이터
- 어소시에이션 룰에 의한 분석 결과는 일반적으로 발생 확률이 높은 것과 그 조합에 의해 표시
- 단순히 발생 횟수가 많은 조합을 확인할 경우에는 횟수가 많은 순으로 조합을 표시하는 경우도 있음

#### 7.2.2. 시퀀스

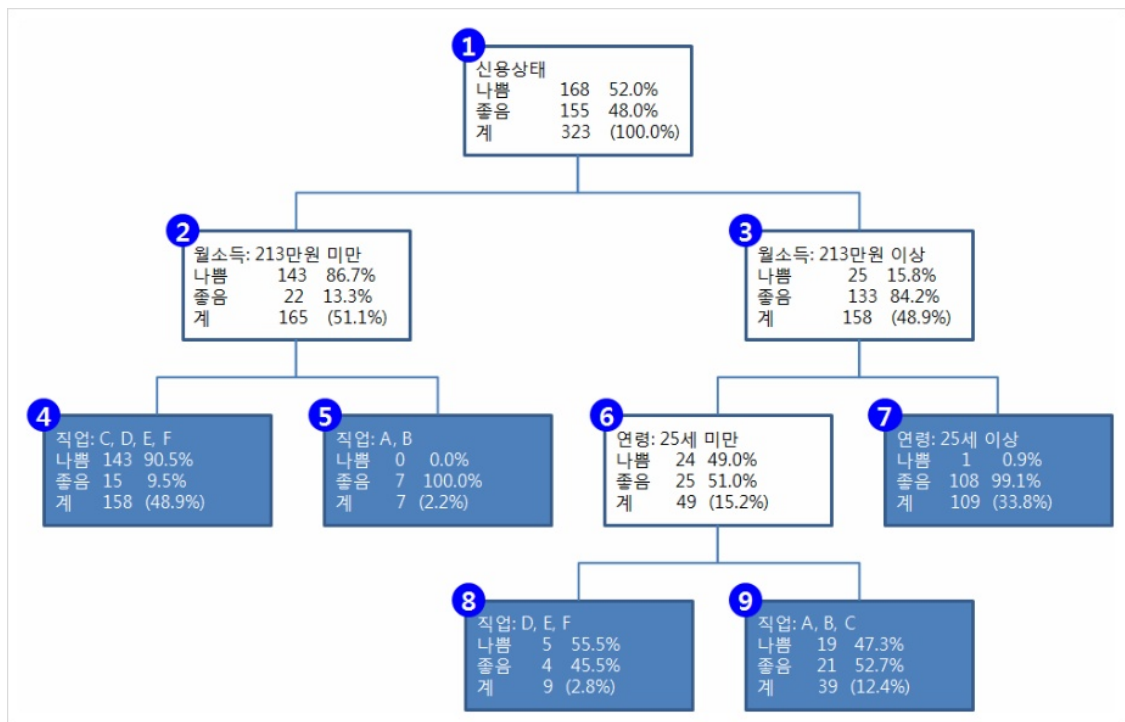
- 어떤 현상이 일어나면 다음에 다른 현상이 일어나는 것처럼 연속해서 일어날 가능성이 높은 현상의 조합과 일어날 순서를 찾아내기 위해 사용되는 것
- 시퀀스에 의한 분석 결과는 일반적으로 조합이 발생한 시계열순으로 표시
- 보다 고도화된 표시 방법으로 특정 현상 다음에 일어날 확률이 높은 현상을 관련도 형식으로 표현하는 방법도 있음

#### 7.2.3. 클러스터

- 복수의 속성을 갖는 집단의 구성요소를 클러스터로 나누기 위한 방법
  - 클러스터 : 비슷한 속성을 갖는 그룹
- 클러스터에 의한 분석 결과는 주로 분류된 클러스터 각각의 속성을 속성값 분포로 해서 표시
- 보다 고도화된 표시 방법으로 클러스터간 유사성을 관련도 형식으로 표시하는 방법도 있음

#### 7.2.4. 디시전 트리

- 특정 상황이 발생할지를 조사해서 발생 결과에 대해 영향을 미치게 하는 속성의 영향도와 상호 관련성을 시각화하기 위한 방법
- 디시전 트리에 의한 분석 결과는 일반적으로 영향도가 높은 속성부터 계층적으로 트리 형식으로 표시



- 특정 현상이 일어날지 여부에 대해 가장 영향도가 높은 속성을 찾아서, 이 속성에 의해 분류된 집단에 대해 다시 한 번 영향도가 높은 속성을 판정하기 위해 이용