데이터 마이닝

15강연관성분석11

통계·데이터과학과 장영재 교수



01 연관성분석 관련 R 함수 02 연관성분석의 시각화 관련 R 함수 03 R 사용 예제



1. 연관성분석 관련 R 함수



- 1 str 함수
 - I 함수의 구조str(x)
 - Ⅰ 기능R 객체(object)의 내부 구조를 출력



- 2 read.transactions 함수
 - 함수의 구조read.transactions(file, format = c("basket", "single"), sep = NULL, cols = NULL)
 - Ⅰ 기능자료를 읽어 transactions 객체(object)형태로 저장

▮옵션

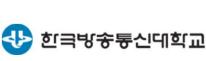
- file: 거래자료(transaction data)가 저장된 파일 이름
- format : 자료가 저장된 형태를 지정. "basket"으로 지정해 주면 한 줄에 하나의 거래가 기록되어 있는 데이터임을 의미하며 "single"을 사용하면 각 줄이 거래와 항목으로 이루진 데이터임을 의미
- sep: 데이터의 관측값 구분자
- cols: format이 "single"인 경우 거래와 항목의 열을 나타내는 벡터. format이 "basket"인 경우 거래의 열을 나타내는 숫자



- 3 apriori 함수
 - 함수의 구조 apriori(data, parameter = NULL, appearance = NULL, control = NULL)
 - 기능 apriori 알고리즘을 이용하여 연관규칙을 도출

▮옵션

- data: 거래자료가 저장된 transactions 객체(object) 이름
- parameter : apriori를 실행하기 위한 모수를 가진 APparameter
 객체(object) 또는 name list





- 4 inspect 함수
 - 함수의 구조 inspect(x, ...)
 - Ⅰ 기능 거래자료와연관규칙을 출력
 - 옵션
 - x: 연관규칙, 거래자료 또는 항목행렬



2. 연관성분석의 시각화 관련 R함수



02 연관성분석의 시각화관련 R 함수

- 1 plot 함수 (R의 arulesViz 패키지 활용)
 - ▮ 함수의 구조 plot(x, method = c("grouped", "graph", "paracoord"))
 - ▮ 기능 거래자료와연관규칙을시각화

▮ 옵션

- x: 연관규칙, 거래자료 또는 항목행렬
- method
- "grouped"로 지정하면 연관규칙의 조건(LHS)과 결과(RHS)을 기준으로 그래프 를 출력(원의 크기는 각 규칙의 지지도를 색상의 진하기는 향상도(lift)를 나타냄)
- "graph"로 지정하면 품목들 간의 연관성을 그래프로 출력 (화살표의 두께는 지지도를 화살표의 색상의 진하기는 향상도(lift)를 나타냄)
- "paracoord"로 지정하면 각 물품 간의 연관관계를 병렬적으로 확인할 수 있음

3. R 사용 예제



