



19. 연관분석

chap19_AssociationAnalysis 수업내용

- 1) 연관분석 개요
- 2) 연관규칙 생성
- 3) Adult 내장 데이터를 이용한 연관규칙 생성
- 4) single format transaction 데이터 처리
- 5) basket format transaction 데이터 처리



1) 연관 분석 개요

● 연관분석

- 어떤 사건이 얼마나 자주 동시에 발생하는가를 표현하는 규칙 또는 조건
- 데이터베이스에서 사건의 연관규칙을 찾는 무방향성 데이터 마이닝 기법
- 마케팅에서 고객의 장바구니에 들어있는 품목 간의 관계 탐구
- y변수가 없는 비지도 학습에 의한 패턴 분석
- 사건과 사건 간 연관성(관계)를 찾는 방법(예:기저귀와 맥주)
예) 장바구니 분석 : 장바구니 정보를 트랜잭션이라고 하며,
트랜잭션 내의 연관성을 살펴보는 분석기법
- 분석절차 : 거래내역 -> 품목 관찰 -> 규칙(Rule) 발견



1) 연관 분석 개요

- **관련분야 : 대형 마트, 백화점, 쇼핑몰 판매자 -> 고객 대상 상품추천**

1. 고객들은 어떤 상품들을 동시에 구매하는가?
2. 라면을 구매한 고객은 주로 다른 어떤 상품을 구매하는가?

활용방안 : 위와 같은 질문에 대한 분석을 토대로 고객들에게

- 1) 상품정보 발송
- 2) 텔레마케팅를 통해서 패키지 상품 판매 기획,
- 3) 마트의 상품진열



1) 연관 분석 개요

What comes to your mind?

와인, 바나나, 사과 몇 알,
올리브, 오렌지주스, 꿀을
구입했군!

일반적으로 바나나와
와인을 같이 구매하나?
그렇다면, 와인의 브랜
드도 관련이 있나?

올리브와 오렌지 주스를 구입하면,
와인도 같이 사나?

고객의 신상 정보와
구매 내역이 서로
관계가 있을까?

보통은 같이 구매되는데
여기에는 없는 게 뭐가 있지?





1) 연관 분석 개요

● 연관규칙 평가척도

1. 지지도(support) : 전체자료에서 관련 품목의 거래 확률
 - $A \rightarrow B$ 지지도 식 = A 와 B 를 포함한 거래수 / 전체 거래수
 - A 를 구매한 후 B 를 구매하는 거래 비율
 - 순서 무관
2. 신뢰도(confidence) : A 가 구매될 때 B 가 구매될 확률(조건부 확률)
 - $A \rightarrow B$ 신뢰도 식 = A 와 B 를 포함한 거래수 / A 를 포함한 거래수
 - A 가 포함된 거래 중에서 B 를 포함한 거래의 비율
 - 순서 고려
3. 향상도(Lift) : 상품 간의 독립성과 상관성을 나타내는 척도
 - 향상도 식 = 신뢰도 / B 가 포함될 거래율
 - 향상도가 낮을 수록 두 상품의 상관성 낮음
 - 향상도가 클 수록 두 상품의 상관성 높음(빵과 버터)
 - 연관규칙에 의미가 있으려면 1보다 큰 값이어야 한다.
 - 값이 높을 수록 상품 간의 연관성이 높다.

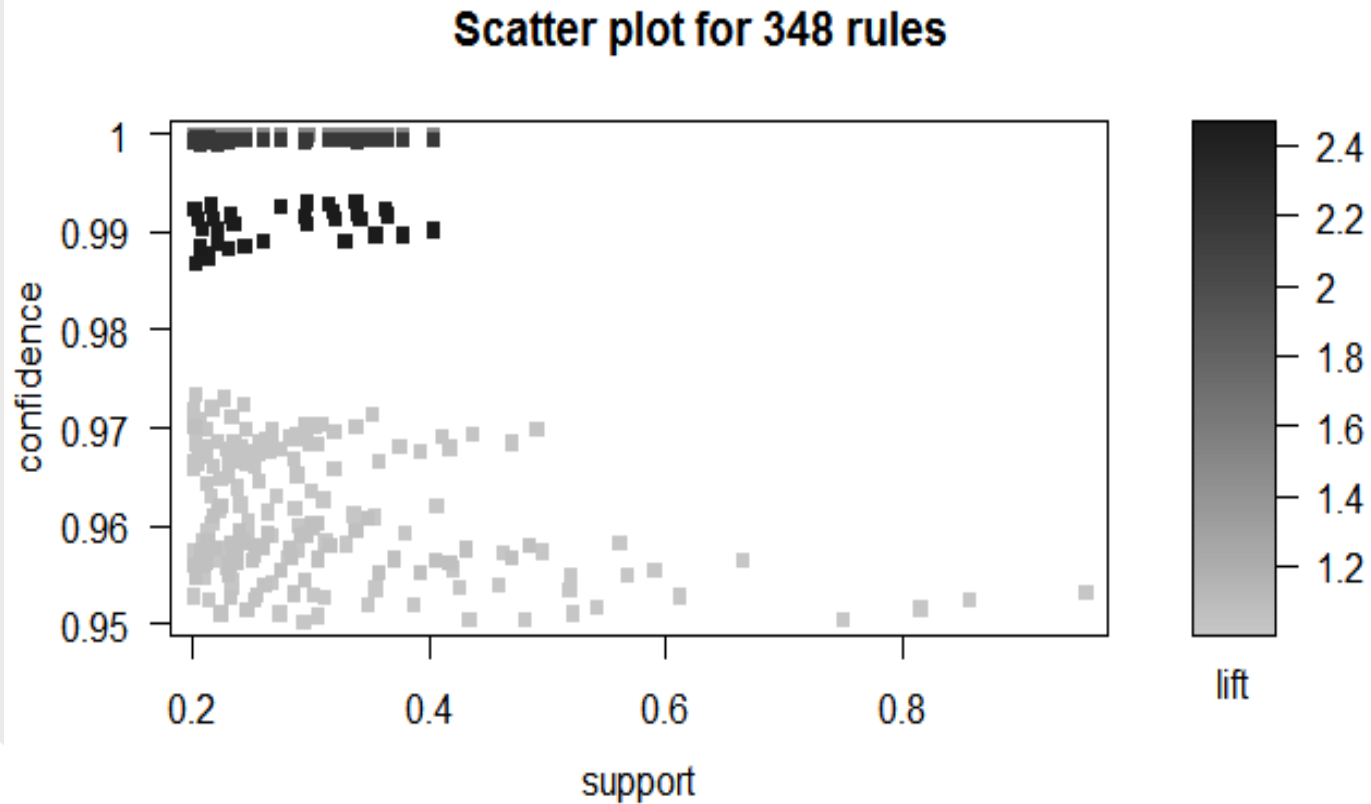


2) 연관 규칙 생성

lhs	rhs	support	confidence	lift
[1] {}	=> {과일}	0.3333333	0.3333333	1.000
[2] {}	=> {맥주}	0.3333333	0.3333333	1.000
[3] {}	=> {고기}	0.6666667	0.6666667	1.000
[4] {}	=> {라면}	0.6666667	0.6666667	1.000
[5] {}	=> {우유}	0.8333333	0.8333333	1.000
[6] {과일}	=> {고기}	0.1666667	0.5000000	0.750
[7] {고기}	=> {과일}	0.1666667	0.2500000	0.750
[8] {과일}	=> {라면}	0.1666667	0.5000000	0.750
[9] {라면}	=> {과일}	0.1666667	0.2500000	0.750
[10] {과일}	=> {우유}	0.1666667	0.5000000	0.600
[11] {우유}	=> {과일}	0.1666667	0.2000000	0.600
[12] {맥주}	=> {고기}	0.1666667	0.5000000	0.750
[13] {고기}	=> {맥주}	0.1666667	0.2500000	0.750
[14] {맥주}	=> {라면}	0.1666667	0.5000000	0.750
[15] {라면}	=> {맥주}	0.1666667	0.2500000	0.750

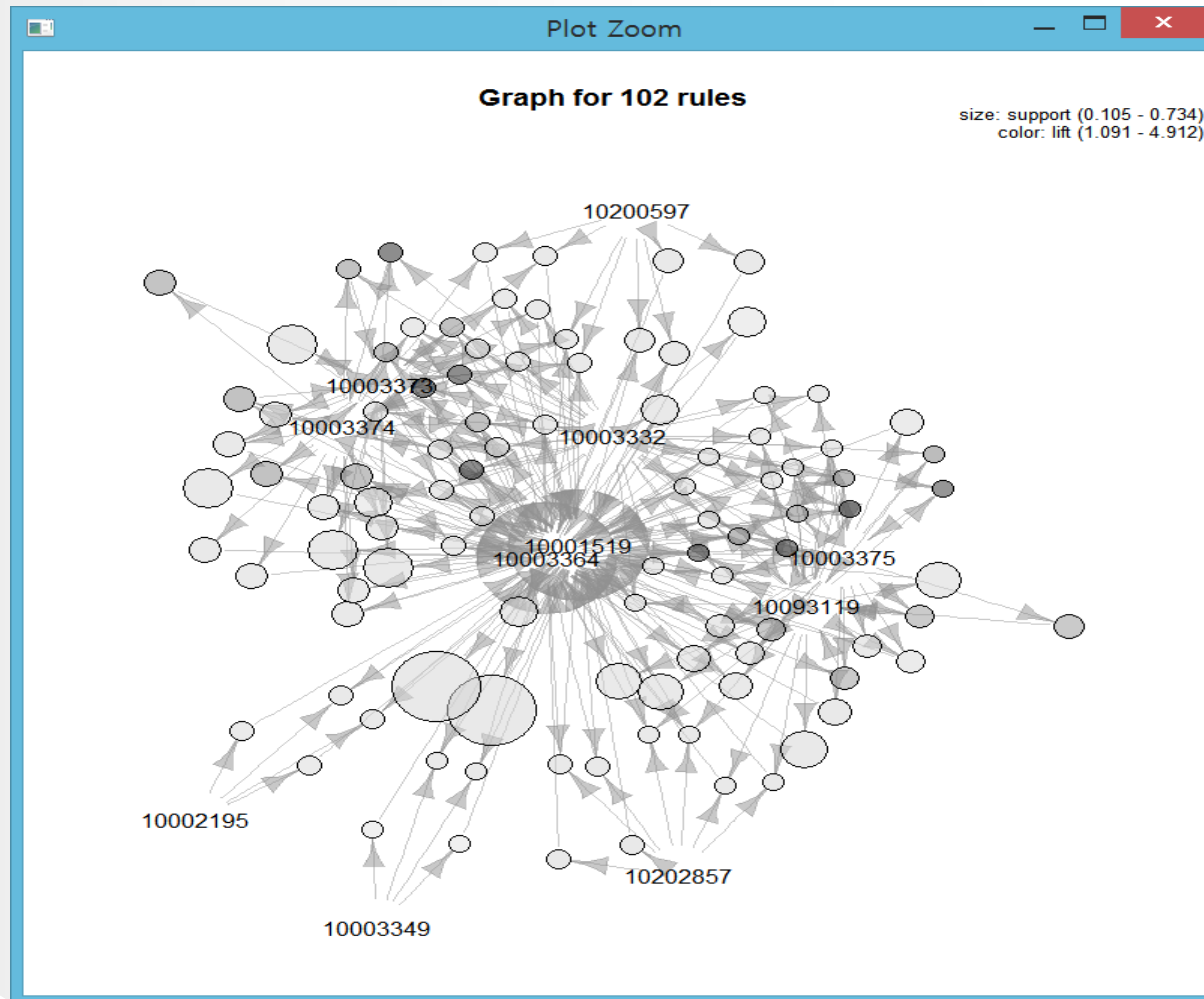


3) Adult 내장 데이터를 이용한 연관규칙 생성



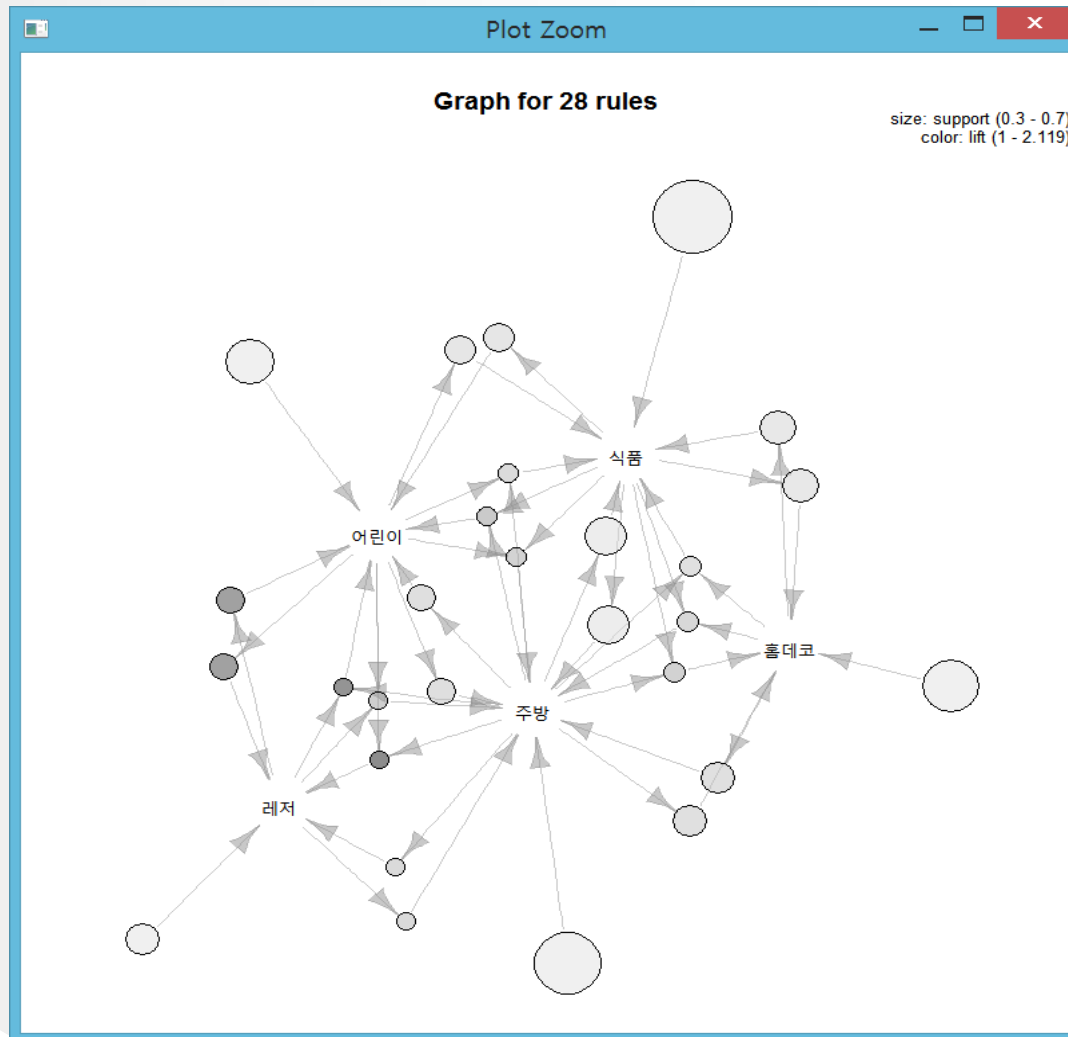


4) single format transaction 데이터 처리





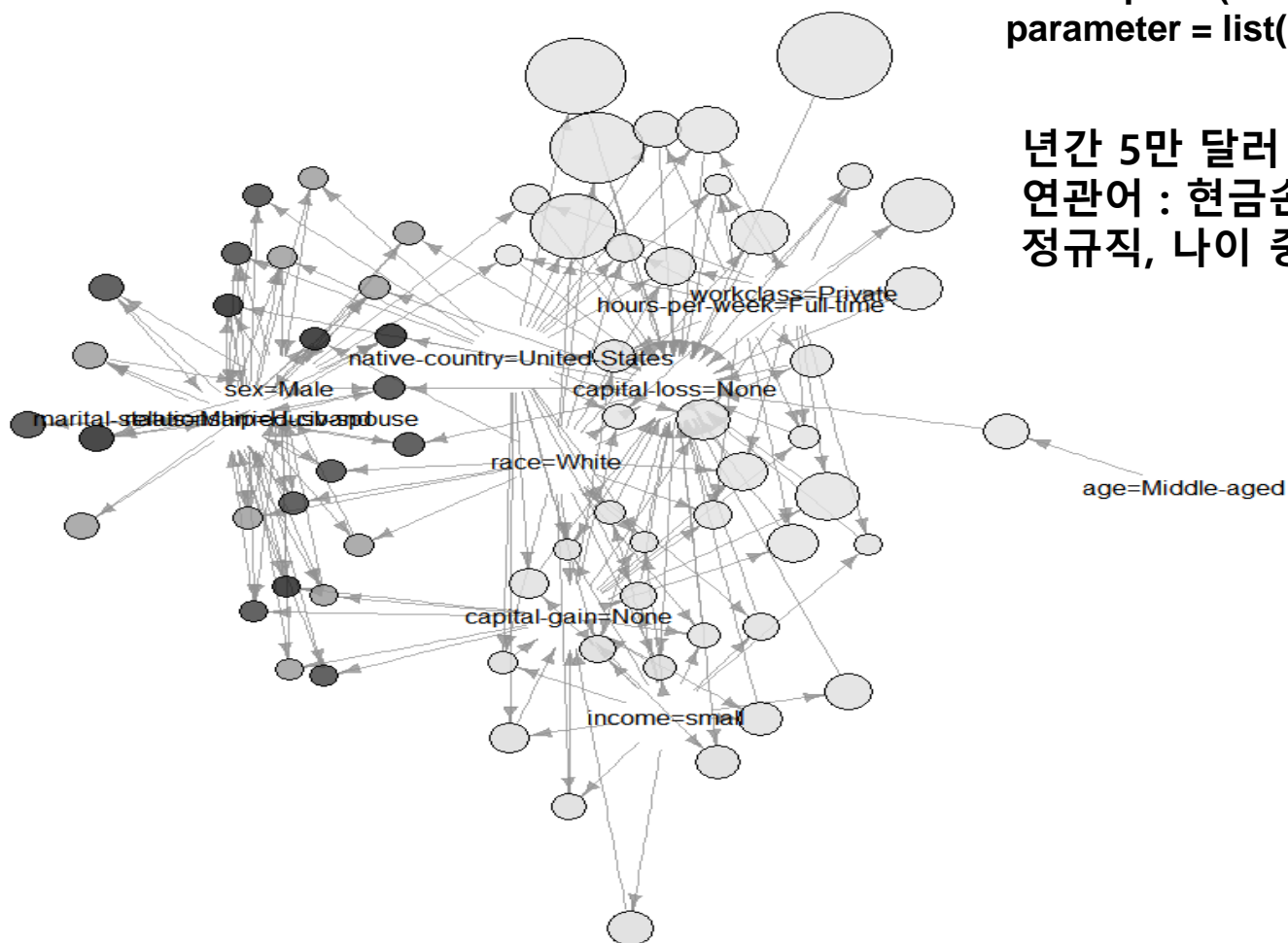
5) basket format transaction 데이터 처리





연관분석

Graph for 67 rules



```
ar1<- apriori(Adult,  
parameter = list(supp=0.35, conf=0.95))
```

년간 5만 달러 이상의 연봉 수령자
연관어 : 현금손실 무, 백인, 미국,
정규직, 나이 중년