

비즈니스 인텔리전스를 위한 데이터마이닝

제11장 연관성규칙 (Association Rule)

연관성 규칙 (Association Rule)의 개념

- 대규모 데이터 항목들 중에서 유용한 연관성과 상관관계를 찾는 기법
 - 상품 또는 서비스간의 관계를 살펴보고, 거래 데이터베이스로부터 유용한 규칙을 찾아내고자 할 때 이용
 - 동시에 구매될 가능성이 큰 상품들을 찾아냄으로써 시장 바구니 안의 구매물품 사이의 관계를 알아본다는 뜻에서 **장바구니 분석**(Market Basket Analysis) 또는 **친화성 분석** (Affinity Analysis)이라고도 함
 - “If ~ then ~” 형태의 간단하고 명확한 규칙들을 생성
 - 이해하기가 쉽고 명료하며 실질적으로 적용하기 쉬움
 - 데이터로부터 계산된 연관 규칙들은 확률에 근거하고 있음.
 - 상당히 높은 신뢰도를 갖는 규칙에 대해 특정 고객에게 그 규칙이 적용되지 않는다면 수상하게 판단할 수 있음
-

시장 바구니 분석 (연관성 규칙 발견) 예시

- 서로 다른 아이템의 구매 사이에 존재하는 의존성을 결정하기 위해 고객의 거래 데이터 베이스를 분석하는 것에서 비롯되었기 때문에 **장바구니 분석**으로 불림

Ex) Products in Shop Cart (One trip, Together)

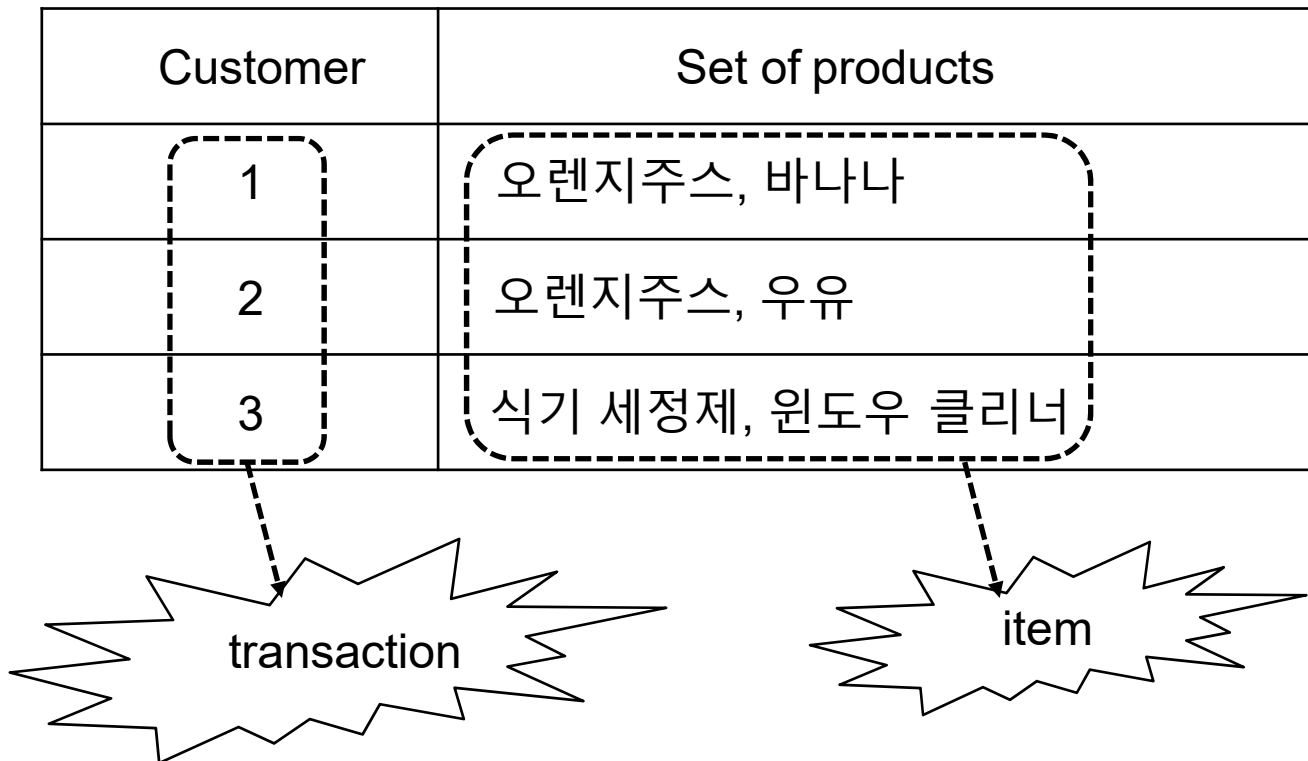


- 1) 구매자가 제품을 구매할 때 진열대에 배치된 이웃상품의 영향이 있었는가?
 - 2) 오렌지 주스와 식기 세정제 구입시 윈도우 클리너를 같이 구입하는가?
 - 3) 우유를 바나나 구입시 함께 구입하는가? 또한, 구입할 때 특정 브랜드를 선호 하는가?
 - 4) 식기 세정제를 어느 곳에 위치시켜야 판매고를 최대화할 수 있는가?
-

거래(transaction)와 항목(item)

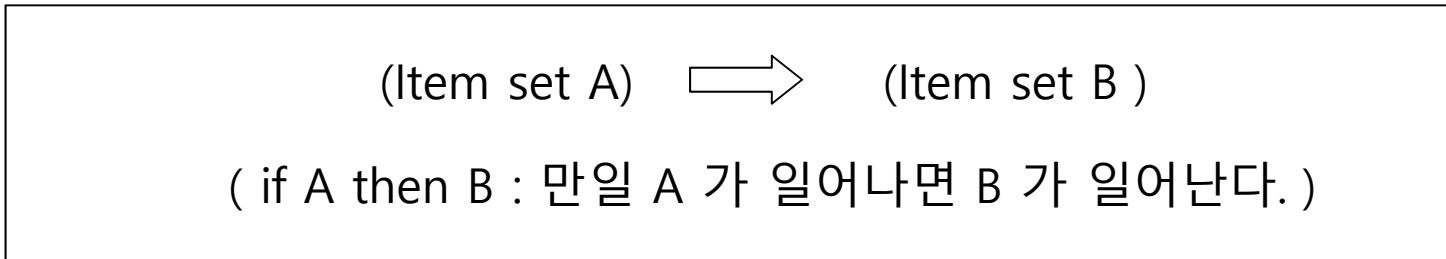
- Market Basket Analysis는 하나 또는 여러 개의 product나 service offering 의 거래와 이 거래에 대한 정보에서 시작.

Ex) 마크로의 Point-Of-Sale Transaction



연관성 규칙 (Association Rule)

- 어떤 Item 집합의 존재가 다른 Item 집합의 존재를 암시하는 것을 의미하며 다음과 같이 표시한다.



- 함께 구매하는 상품의 조합이나 서비스 패턴 발견하는데 이용
 - 특정 제품 또는 사건들이 동시에 발생 하는 패턴을 파악하는데 이용
Ex) 가정 용품 판매 기간 동안 같이 판매해야 하는 상품의 패턴 발견
 - 해석
 - 원인과 결과의 직접적인 인과관계가 아님
 - 둘 또는 그 이상 품목들 사이의 상호 관련성을 알려줌
-

연관성 규칙의 분석 원리

- 연관성 규칙 단계

- 후보규칙 생성: If ~ then 형식으로 항목들 사이의 모든 가능한 규칙들을 조사
 - If절은 "antecedent(조건부, 전건, 선행)", then절은 "consequent(결론부, 후건, 후행)"라고 함
 - 연관성 규칙에서 조건부와 결론부는 공통원소가 없는 항목들의 집합이어야 함
 - (예) 조건부: {축구공, 야구공} / 결론부: {야구공} (X)
 - p개의 서로 다른 항목을 이용하여 데이터베이스 내의 모든 가능한 항목들 사이의 조합을 만듦(단일 항목, 2개로 된 항목, 3개로 된 항목 등)
 - ※ 주의: 항목이 늘어날수록, 모든 조합들을 생성하기 위한 계산시간 증가
(p개의 아이템에 대하여 생성되는 규칙의 수는 $3^p - 2^{p+1} + 1$)
 - 후보규칙으로부터 조건부와 결론부 항목 집합 사이의 강한 의존성을 나타내는 규칙을 선택
-

연관성 규칙의 분석 원리

- [예제] 스마트폰 케이스 구매 거래

거래번호	구매한 폰케이스 색상			
1	빨간색	하얀색	초록색	
2	하얀색	오렌지색		
3	하얀색	파란색		
4	빨간색	하얀색	오렌지색	
5	빨간색	파란색		
6	하얀색	파란색		
7	하얀색	오렌지색		
8	빨간색	하얀색	파란색	초록색
9	빨간색	하얀색	파란색	
10	노란색			

- "If (antecedent) then (consequent)" 형태의 규칙을 생성
- "If 빨간색, 흰색, then 초록색"
 - 조건부(antecedent) = {빨간색, 흰색}, 결론부(consequent) = {초록색}
- 너무 많은 조합이 발생하기 때문에 빈번한(frequent) item set을 고려해야 함

연관성 규칙의 주요 지표

개념	산식	설명
지지도 (Support)	$\text{support}(A \Rightarrow B) = P(A \cap B)$	<ul style="list-style-type: none"> • 전체 거래 중에서 A와 B를 모두 포함하는 경우의 확률 • 전체적인 거래 규모에 대한 값 • 값이 클 수록 자주 발생하는 거래 • 규칙의 중요성에 대한 척도
신뢰도 (Confidence)	$\text{confidence}(A \Rightarrow B) = P(B A) = \frac{P(A \cap B)}{P(A)}$	<ul style="list-style-type: none"> • 항목 A가 일어난 상황에서 항목 B가 일어날 확률(조건부 확률) • A를 구매한 경우, 이 중에서 얼마나 항목 B구매로 이어지는지를 의미 • 값이 클 수록 A 구매시 B 구매율이 높음 • 규칙의 신뢰성에 대한 척도
향상도 (Lift)	$\begin{aligned} \text{lift}(A \Rightarrow B) &= \frac{\text{confidence}(A \Rightarrow B)}{P(B)} \\ &= \frac{P(A B)}{P(B)} = \frac{P(A \cap B)}{P(A)P(B)} \end{aligned}$	<ul style="list-style-type: none"> • 항목 A를 구매한 경우 그 거래가 항목 B를 포함하는 경우와 항목 B가 임의로 구매되는 경우의 비 • 항목 A와 B의 구매 패턴이 독립적인지, 서로 상관이 있는지를 의미 • 값이 1보다 크면 관련도 높음 값이 1이면 관련성 없음 값이 1보다 작으면 A선택시 B선택 안함

연관성 규칙의 장단점

장점	단점
<ul style="list-style-type: none">▪ 탐색적인 기법▪ 강력한 비목적성 분석기법▪ 결과가 명확하고 이해 용이▪ 자료구조와 계산과정이 간단▪ 수많은 상품 및 서비스 품목 간에 유의한 연관성을 갖는 구매패턴을 찾아줌▪ 이를 바탕으로 또 다른 연구가설을 탐지할 수 있는 가능성을 제공	<ul style="list-style-type: none">▪ 항목의 수를 결정하기 어려움▪ 드물게 발생하는 항목에 대해서 처리가 어려움▪ DBMS등과 같은 전산화 환경 미비시 동일한 거래를 추적하기 어려움▪ 항목의 수가 증가하면 계산시간이 급격히 증가▪ 비즈니스적으로 중요한 요소를 고려할 수 없어 현실적으로 중요한 연관규칙을 얻기에는 부족함

연관성 규칙 적용분야

- **교차 판매(Cross Selling)**
 - 상품간 연관성 분석을 통한 교차판매 규칙 제공
 - **상품 진열(Inventory Display)**
 - 구매가능성 분석을 통한 효율적 상품배치 및 고객동선 재설계
 - **부정탐지(fraud detection)**
 - 상당히높은신뢰도를갖는규칙에대해특정고객에게그규칙이적용되지않는다면수상하게판단할수있음
 - **Catalog Design**
 - 상품의 배치문제, 패키지 상품의 구성, 쿠폰 발행, 카탈로그의 구성, 신상품의 카테고리 선정
 - **제조분야 연구 논문**
 - 안진숙, & 손소영. (2007). 연관성 규칙을 활용한 제조기업 A/S 이용 고객 패턴 연구. *대한산업공학회 추계학술대회 논문집*, 1396-1401.
 - 이필립, 박영미, & 최양열. (2013). 조선 생산실행시스템 업무 분석을 위한 연관성 규칙 방법 적용. *한국 CDE 학회 학술발표회 논문집*, 117-122.
-