

08

연관분석 개념

[연관분석 개념]

사건들의 발생이 있을 때, 조건부 확률을 통해 사건들의 관계,

규칙을 발견하고 분석하는 방법

00:09

8.2 연관분석

8.2.1 연관 분석(Association analysis)이란?

효율적인 상품 진열, 패키지 상품 개발, 교차판매 전략, 기획상품의 결정 등에 사용되고 있음

8.2.2 연관규칙분석

- 기업의 데이터베이스에서 상품 구매, 서비스 등 일련의 거래 또는 사건들 간의 규칙을 발견하기 위해 적용됨
- 서로 다른 두 아이템 집합이 얼마나 빈번하게 발생하는지를 알 수 있게 함
- 조건과 반응 형태(IF_THEN)로 이루어져 있음
- 장바구니 분석, 서열분석이 있음
 - 장바구니 분석 : 장바구니에 무엇이 같이 들어있는지에 대한 분석
 - 서열 분석 : A를 산 다음에 B를 산다

8.2.2.1 연관규칙 분석에 사용되는 척도

[연관분석에 사용되는 척도]

지지도, 신뢰도, 향상도에 대해 명확히 알아야 연관분석을

사용할 수 있습니다.

01:18

jupyter 8.2 연관분석개념 Last Checkpoint: 44분 전 (autosaved) Logout

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipykernel)

8.2.2.1 연관규칙 분석에 사용되는 척도

- 지지도 : 전체 거래 중 항목 A와 항목 B를 동시에 포함하는 거래의 비율로 정의

$$\text{support} = A \text{와 } B \text{가 동시에 포함된 거래수} / \text{전체 거래수}$$
- 신뢰도 : 항목 A를 포함한 거래 중에서 항목 A와 항목 B가 같이 포함될 확률로 연관성의 정도 파악 가능

$$\text{confident} = A, B \text{가 동시에 포함된 거래수} / A \text{를 포함하는 거래수}$$
- 향상도 : A가 구매되지 않았을 때 항목 B의 구매확률에 비해 A가 구매되었을 때 항목 B의 구매확률의 증가비

$$\text{Lift} = A \text{와 } B \text{가 동시에 포함된 거래수} / A \text{를 포함하는 거래수} * B \text{를 포함하는 거래수}$$

8.2.3 Apriori 알고리즘

가능한 모든 경우의 수를 탐색하여 지지도, 신뢰도, 향상도가 높은 규칙들을 찾아내는 방식이 가장 이상적이지만 아이템 수가 증가할수록 계산에 소요되는 시간이 기하급수적으로 증가하는 문제가 있음
아이템이 n개일 때 탐색해야 할 모든 경우의 수 : $n \cdot (n-1)$
이 때문에 빈발 집합(frequent item sets)만을 고려해서 연관 규칙을 생성하는 Apriori algorithm이 제안됨

- 최소지지도보다 큰 지지도 값을 갖는 품목의 집합을 빈발항목집단(frequent item set)이라고 함
- Apriori 알고리즘은 모든 품목집합에 대한 지지도를 전부 계산하는 것이 아니라, 최소 지지도 이상의 빈발항목집합을 찾은 후 그것들에 대해서만 연관규칙을 계산하는 것임

[Apriori 개념]

모든 상품을 탐색하면 비용이 많이 들어갑니다. 빈발항목집단을 만드는 것이 Apriori입니다.

02:51

jupyter 8.2 연관분석개념 Last Checkpoint: 한 시간 전 (autosaved) Logout

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (ipykernel)

$$\text{confident} = A, B \text{가 동시에 포함된 거래수} / A \text{를 포함하는 거래수}$$

- 향상도 : A가 구매되지 않았을 때 항목 B의 구매확률에 비해 A가 구매되었을 때 항목 B의 구매확률의 증가비

$$\text{Lift} = A \text{와 } B \text{가 동시에 포함된 거래수} / A \text{를 포함하는 거래수} * B \text{를 포함하는 거래수}$$

8.2.3 Apriori 알고리즘

가능한 모든 경우의 수를 탐색하여 지지도, 신뢰도, 향상도가 높은 규칙들을 찾아내는 방식이 가장 이상적이지만 아이템 수가 증가할수록 계산에 소요되는 시간이 기하급수적으로 증가하는 문제가 있음
아이템이 n개일 때 탐색해야 할 모든 경우의 수 : $n \cdot (n-1)$
이 때문에 빈발 집합(frequent item sets)만을 고려해서 연관 규칙을 생성하는 Apriori algorithm이 제안됨

- 최소지지도보다 큰 지지도 값을 갖는 품목의 집합을 빈발항목집단(frequent item set)이라고 함
- Apriori 알고리즘은 모든 품목집합에 대한 지지도를 전부 계산하는 것이 아니라, 최소 지지도 이상의 빈발항목집합을 찾은 후 그것들에 대해서만 연관규칙을 계산하는 것임
- 장점 : 1세대 알고리즘으로 구현과 이해가 쉬움
- 단점 : 지지도가 낮은 후보 집합을 생성 시 아이템의 개수가 많아지면 계산 복잡도가 증가함

(A,B)는 (A), (B)의 초집합(superset)이라고 함
마일 아이템 집합 {A, B}의 지지도가 사용자가 정한 최소 지지도 요건을 충족시키지 못했을 경우.