202017999\_DM\_hw1

support와 confidence 변화에 따른 연관 규칙이 달라지는 정도에 관한 보고서 입니다. {}괄호 안의 내용은 전제(I집합)이며, -> 뒤의 내용은 결과(j원소) 입니다.

1. support = 0.01, confidence = 0.4 의 경우

{eggs, milk} -> mineral water with confidence 0.42 and interest 0.19

{eggs, chocolate} -> mineral water with confidence 0.41 and interest 0.17

{eggs, ground beef} -> mineral water with confidence 0.51 and interest 0.27

{spaghetti, milk} -> mineral water with confidence 0.44 and interest 0.21

{milk, ground beef} -> mineral water with confidence 0.50 and interest 0.26

{spaghetti, frozen vegetables} -> mineral water with confidence 0.43 and interest 0.19

{spaghetti, pancakes} -> mineral water with confidence 0.46 and interest 0.22

{ground beef, mineral water} -> spaghetti with confidence 0.42 and interest 0.24

{spaghetti, ground beef} -> mineral water with confidence 0.44 and interest 0.20

{chocolate, ground beef} -> mineral water with confidence 0.47 and interest 0.24

1. support = 0.015, confidence = 0.3 의 경우

{milk, mineral water} -> spaghetti with confidence 0.33 and interest 0.15

{milk, spaghetti} -> mineral water with confidence 0.44 and interest 0.21

{chocolate, mineral water} -> spaghetti with confidence 0.30 and interest 0.13

{ground beef, mineral water} -> spaghetti with confidence 0.42 and interest 0.24

{ground beef, spaghetti} -> mineral water with confidence 0.44 and interest 0.20

1. support = 0.017, confidence = 0.4 의 경우

{ground beef, mineral water} -> spaghetti with confidence 0.42 and interest 0.24

{spaghetti, ground beef} -> mineral water with confidence 0.44 and interest 0.20

1. support = 0.18, confidence = 0.4 의 경우

출력 값 없음.

-결론

support와 confidence에 대한 해석

support와 confidence의 값이 조금만 변해도 값이 크게 달라진다. support의 값을 증가시키면 값을 만족하는 아이템의 수가 줄어든다. confidence의 값도 마찬가지이다. 신뢰도를 의미하기 때문에 값을 높일수록 정확한 값을 얻을 수 있지만 규칙의 수는 감소하게 된다. 위 결과에서 support가 증가함에 따라 규칙에서 제외된 항목들은 전체 항목에서 차지하는 비율이 적음을 알 수 있고 confidence가 증가함에 따라 규칙에서 제외된 항목들은 신뢰도가 낮음을 알 수 있다.

interest의 해석

interest 값은 높을수록 규칙이 의미있다는 것을 뜻한다. 두 값을 비교해보자

A: {eggs, ground beef} -> mineral water with confidence 0.51 and interest 0.27(support 0.1)

B: {spaghetti, ground beef} -> mineral water with confidence 0.44 and interest 0.20(support 0.17)

A를 산 사람은 B를 산 사람보다 적지만 A가 신뢰도가 높고 흥미도도 높기 때문에 더 의미있는 규칙이라고 할 수 있다. 계란과 고기와 생수는 구매하는 경우가 많아 너무 보편적이어서 interest 값이 낮지만 원래 interest값의 의의는 생각하지 못했던 현상을 예측할 수 있다는 것에 있다.