

**03. 연관분석(Association Analysis)**

#src: 9\_자연어처리/ch03\_연관분석 참고

; 데이터들 사이에서 '자주 발생하는 속성'을 찾고, 그 속성들 사이에 '연관성이 어느정도 있는지'를 분석하는 방법 #apriori, FP-growth, DHP알고리즘 등

**1) 연관분석 평가 척도**

(1) **지지도(Support)**; 연관 규칙이 얼마나 중요한지에 대한 평가; 조건 결과 항목 / 전체

(2) **신뢰도(Confidence)**; 연관 규칙을 얼마나 믿을 수 있는지 평가; 조건 결과 항목 / 조건 항목

(3) **향상도(Lift)**; 우연히 발생한 규칙 여부(상관 관계); 조건결과 신뢰도 / 결과 지지도

향상도<1: 음의 상관관계      향상도=1: 상관관계 없음      향상도>1: 양의 상관관계

**2) 연관분석**

(1) 연관 규칙 생성

apriori(); 연역적 알고리즘을 사용하여 연관규칙을 알아내는 함수

(2) 연관 규칙 형식: 아래와 같이 namedtuple 형식 #반복문 이용해서 조회해야 함

SupportRecord = namedtuple( 'SupportRecord', ('items', 'support'))

RelationRecord = namedtuple( 'RelationRecord', SupportRecord.\_fields + ('ordered\_statistics',))

OrderedStatistic = namedtuple( 'OrderedStatistic', ('items\_base', 'items\_add', 'confidence','lift',))

(3) 연관 규칙 조회

```
from apyori import apriori
rules = apriori(transaction, min_support=0.2, min_confidence=0.1)
rules = list(rules)
for row in rules:                                #규칙 조회를 위한 for문
    support = row[1]                              #지지도
    ordered_st = row[2]
    for item in ordered_st:
        lhs = ', '.join(x for x in item[0])       #조건(items_base)
        rhs = ', '.join(x for x in item[1])       #결과(items_add)
        confidence = item[2]                      #신뢰도
        lift = item[3]                            #향상도
        if lift != 1:
            print('{}=>{}, Wt{:5.3f}Wt{:5.3f} Wt{:5.3f}'.format(lhs, rhs, support, confidence, lift))
```

(4) 연관 규칙 평가: 지지도, 신뢰도, 향상도를 이용해 평가