

6주차 과제

# Association Rule Mining



담당교수님: 윤장혁 교수님

과목명: Data Analytics

이름: 박민성

학번: 201611145

전공: 산업공학과

제출일: 2020.04.26

Objective: ARM분석에 DVD 대여 데이터를 활용해서 DVD의 추천에 활용하려고 한다.

계산 Tool은 파이썬을 활용했다.

```
201611145_박민성_week6.py ×
C: > Users > minisong > Desktop > 201611145_박민성_week6.py
1  import pandas as pd
2  from mlxtend.preprocessing import TransactionEncoder
3  from mlxtend.frequent_patterns import apriori
4  dataset=[]
5  import csv
6  f = open('data_week6.txt','r',encoding="utf-8-sig")
7  rdr = csv.reader(f)
8  for i in rdr:
9      dataset.append(i)
10
11  te = TransactionEncoder()
12  te_ary = te.fit(dataset).transform(dataset)
13  df = pd.DataFrame(te_ary, columns=te.columns_)
14  frequent_itemsets = apriori(df, min_support=0.1, use_colnames=True)
15
16  from mlxtend.frequent_patterns import association_rules
17  association_rules(frequent_itemsets, metric="confidence", min_threshold=0.3).to_csv("./결과데이터.csv", encoding='ms949')
```

Import Apriori를 이용해서 분석을 진행했다.

data\_wee6.txt를 list로 가져온 다음 분석을 진행했다.

support의 threshold는 0.1로 두고 confidence의 threshold는 0.3으로 두고 분석을 진행했다.그 결과값을 csv파일로 저장해서 엑셀로 데이터를 확인했다.

	A	B	C	D	E	F	G	H	I	J
1		antecedents(X)	consequents(Y)	antecedent support	consequent support	Support	confidence	lift	leverage	conviction
2	0	frozenset({'반지의제왕1'})	frozenset({'반지의제왕2'})	0.32	0.20	0.20	0.625	3.125	0.136	2.133
3	1	frozenset({'반지의제왕2'})	frozenset({'반지의제왕1'})	0.20	0.32	0.20	1.00	3.125	0.136	inf
4	2	frozenset({'반지의제왕1'})	frozenset({'해리포터1'})	0.32	0.36	0.16	0.50	1.389	0.045	1.28
5	3	frozenset({'해리포터1'})	frozenset({'반지의제왕1'})	0.36	0.32	0.16	0.444	1.389	0.045	1.224
6	4	frozenset({'해리포터2'})	frozenset({'반지의제왕1'})	0.20	0.32	0.12	0.60	1.875	0.056	1.70
7	5	frozenset({'반지의제왕1'})	frozenset({'해리포터2'})	0.32	0.20	0.12	0.375	1.875	0.056	1.28
8	6	frozenset({'쇼생크탈출'})	frozenset({'식스센스'})	0.28	0.28	0.12	0.429	1.531	0.042	1.26
9	7	frozenset({'쇼생크탈출'})	frozenset({'식스센스'})	0.28	0.28	0.12	0.429	1.531	0.042	1.26
10	8	frozenset({'식스센스'})	frozenset({'스타워즈'})	0.28	0.28	0.12	0.429	1.531	0.042	1.26
11	9	frozenset({'스타워즈'})	frozenset({'식스센스'})	0.28	0.28	0.12	0.429	1.531	0.042	1.26
12	10	frozenset({'어벤져스'})	frozenset({'스타워즈'})	0.48	0.28	0.28	0.583	2.083	0.146	1.728
13	11	frozenset({'스타워즈'})	frozenset({'어벤져스'})	0.28	0.48	0.28	1.00	2.083	0.146	inf
14	12	frozenset({'어벤져스'})	frozenset({'식스센스'})	0.48	0.28	0.16	0.333	1.19	0.026	1.08
15	13	frozenset({'식스센스'})	frozenset({'어벤져스'})	0.28	0.48	0.16	0.571	1.19	0.026	1.213
16	14	frozenset({'아바타'})	frozenset({'어벤져스'})	0.24	0.48	0.20	0.833	1.736	0.085	3.12
17	15	frozenset({'어벤져스'})	frozenset({'아바타'})	0.48	0.24	0.20	0.417	1.736	0.085	1.303
18	16	frozenset({'해리포터2'})	frozenset({'해리포터1'})	0.20	0.36	0.20	1.00	2.778	0.128	inf
19	17	frozenset({'해리포터1'})	frozenset({'해리포터2'})	0.36	0.20	0.20	0.556	2.778	0.128	1.80
20	18	frozenset({'해리포터2', '반지의제왕1'})	frozenset({'해리포터1'})	0.12	0.36	0.12	1.00	2.778	0.077	inf
21	19	frozenset({'해리포터2', '해리포터1'})	frozenset({'반지의제왕1'})	0.20	0.32	0.12	0.60	1.875	0.056	1.70
22	20	frozenset({'반지의제왕1', '해리포터1'})	frozenset({'해리포터2'})	0.16	0.20	0.12	0.75	3.75	0.088	3.20
23	21	frozenset({'해리포터2'})	frozenset({'반지의제왕1', '해리포터1'})	0.20	0.16	0.12	0.60	3.75	0.088	2.10
24	22	frozenset({'반지의제왕1'})	frozenset({'해리포터2', '해리포터1'})	0.32	0.20	0.12	0.375	1.875	0.056	1.28
25	23	frozenset({'해리포터1'})	frozenset({'해리포터2', '반지의제왕1'})	0.36	0.12	0.12	0.333	2.778	0.077	1.32
26	24	frozenset({'어벤져스', '식스센스'})	frozenset({'스타워즈'})	0.16	0.28	0.12	0.75	2.679	0.075	2.88
27	25	frozenset({'어벤져스', '스타워즈'})	frozenset({'식스센스'})	0.28	0.28	0.12	0.429	1.531	0.042	1.26
28	26	frozenset({'식스센스', '스타워즈'})	frozenset({'어벤져스'})	0.12	0.48	0.12	1.00	2.083	0.062	inf
29	27	frozenset({'식스센스'})	frozenset({'어벤져스', '스타워즈'})	0.28	0.28	0.12	0.429	1.531	0.042	1.26
30	28	frozenset({'스타워즈'})	frozenset({'어벤져스', '식스센스'})	0.28	0.16	0.12	0.429	2.679	0.075	1.47

이중에서 우리가 관심있는 Support, confidence, lift만 남긴다.

Support가 0.1이 넘고 confidence가 0.3이 넘는 자료만 있는 것을 확인 할 수 있다.

	A	B	C	D	E	F
1		antecedents(X)	consequents(Y)	Support	confidence	lift
2	0	frozenset({'반지의제왕1'})	frozenset({'반지의제왕2'})	0.20	0.625	3.125
3	1	frozenset({'반지의제왕2'})	frozenset({'반지의제왕1'})	0.20	1.00	3.125
4	2	frozenset({'반지의제왕1'})	frozenset({'해리포터1'})	0.16	0.50	1.389
5	3	frozenset({'해리포터1'})	frozenset({'반지의제왕1'})	0.16	0.444	1.389
6	4	frozenset({'해리포터2'})	frozenset({'반지의제왕1'})	0.12	0.60	1.875
7	5	frozenset({'반지의제왕1'})	frozenset({'해리포터2'})	0.12	0.375	1.875
8	6	frozenset({'식스센스'})	frozenset({'쇼생크탈출'})	0.12	0.429	1.531
9	7	frozenset({'쇼생크탈출'})	frozenset({'식스센스'})	0.12	0.429	1.531
10	8	frozenset({'식스센스'})	frozenset({'스타워즈'})	0.12	0.429	1.531
11	9	frozenset({'스타워즈'})	frozenset({'식스센스'})	0.12	0.429	1.531
12	10	frozenset({'어벤져스'})	frozenset({'스타워즈'})	0.28	0.583	2.083
13	11	frozenset({'스타워즈'})	frozenset({'어벤져스'})	0.28	1.00	2.083
14	12	frozenset({'어벤져스'})	frozenset({'식스센스'})	0.16	0.333	1.19
15	13	frozenset({'식스센스'})	frozenset({'어벤져스'})	0.16	0.571	1.19
16	14	frozenset({'아바타'})	frozenset({'어벤져스'})	0.20	0.833	1.736
17	15	frozenset({'어벤져스'})	frozenset({'아바타'})	0.20	0.417	1.736
18	16	frozenset({'해리포터2'})	frozenset({'해리포터1'})	0.20	1.00	2.778
19	17	frozenset({'해리포터1'})	frozenset({'해리포터2'})	0.20	0.556	2.778
20	18	frozenset({'해리포터2', '반지의제왕1'})	frozenset({'해리포터1'})	0.12	1.00	2.778
21	19	frozenset({'해리포터2', '해리포터1'})	frozenset({'반지의제왕1'})	0.12	0.60	1.875
22	20	frozenset({'반지의제왕1', '해리포터1'})	frozenset({'해리포터2'})	0.12	0.75	3.75
23	21	frozenset({'해리포터2'})	frozenset({'반지의제왕1', '해리포터1'})	0.12	0.60	3.75
24	22	frozenset({'반지의제왕1'})	frozenset({'해리포터2', '해리포터1'})	0.12	0.375	1.875
25	23	frozenset({'해리포터1'})	frozenset({'해리포터2', '반지의제왕1'})	0.12	0.333	2.778
26	24	frozenset({'어벤져스', '식스센스'})	frozenset({'스타워즈'})	0.12	0.75	2.679
27	25	frozenset({'어벤져스', '스타워즈'})	frozenset({'식스센스'})	0.12	0.429	1.531
28	26	frozenset({'식스센스', '스타워즈'})	frozenset({'어벤져스'})	0.12	1.00	2.083
29	27	frozenset({'식스센스'})	frozenset({'어벤져스', '스타워즈'})	0.12	0.429	1.531
30	28	frozenset({'스타워즈'})	frozenset({'어벤져스', '식스센스'})	0.12	0.429	2.679

Lift를 내림차순으로 정리해보면 <결과 표>를 얻을 수 있다

A	B	C	D
	antecedents(X)	consequents(Y)	lift
20	frozenset({'반지의제왕1', '해리포터1'})	frozenset({'해리포터2'})	3.75
21	frozenset({'해리포터2'})	frozenset({'반지의제왕1', '해리포터1'})	3.75
0	frozenset({'반지의제왕1'})	frozenset({'반지의제왕2'})	3.125
1	frozenset({'반지의제왕2'})	frozenset({'반지의제왕1'})	3.125
16	frozenset({'해리포터2'})	frozenset({'해리포터1'})	2.778
17	frozenset({'해리포터1'})	frozenset({'해리포터2'})	2.778
18	frozenset({'해리포터2', '반지의제왕1'})	frozenset({'해리포터1'})	2.778
23	frozenset({'해리포터1'})	frozenset({'해리포터2', '반지의제왕1'})	2.778
24	frozenset({'어벤져스', '식스센스'})	frozenset({'스타워즈'})	2.679
28	frozenset({'스타워즈'})	frozenset({'어벤져스', '식스센스'})	2.679
10	frozenset({'어벤져스'})	frozenset({'스타워즈'})	2.083
11	frozenset({'스타워즈'})	frozenset({'어벤져스'})	2.083
26	frozenset({'식스센스', '스타워즈'})	frozenset({'어벤져스'})	2.083
4	frozenset({'해리포터2'})	frozenset({'반지의제왕1'})	1.875
5	frozenset({'반지의제왕1'})	frozenset({'해리포터2'})	1.875
19	frozenset({'해리포터2', '해리포터1'})	frozenset({'반지의제왕1'})	1.875
22	frozenset({'반지의제왕1'})	frozenset({'해리포터2', '해리포터1'})	1.875
14	frozenset({'아바타'})	frozenset({'어벤져스'})	1.736
15	frozenset({'어벤져스'})	frozenset({'아바타'})	1.736
6	frozenset({'식스센스'})	frozenset({'쇼생크탈출'})	1.531
7	frozenset({'쇼생크탈출'})	frozenset({'식스센스'})	1.531
8	frozenset({'식스센스'})	frozenset({'스타워즈'})	1.531
9	frozenset({'스타워즈'})	frozenset({'식스센스'})	1.531
25	frozenset({'어벤져스', '스타워즈'})	frozenset({'식스센스'})	1.531
27	frozenset({'식스센스'})	frozenset({'어벤져스', '스타워즈'})	1.531
2	frozenset({'반지의제왕1'})	frozenset({'해리포터1'})	1.389
3	frozenset({'해리포터1'})	frozenset({'반지의제왕1'})	1.389
12	frozenset({'어벤져스'})	frozenset({'식스센스'})	1.19
13	frozenset({'식스센스'})	frozenset({'어벤져스'})	1.19

### <결과 표>

29개의 데이터 모두 lift 값이 1이 넘으므로

X의 itemset과 Y의 itemset이 positively correlated(양의 상관관계)를 갖는다.

<결과 표>에 있는 모든 itemset X는 support 값이 0.1을 넘고 confidence는0.3을 넘고 lift값은 1이 넘는다.

따라서, 모든 <결과 표>에 있는 itemset X를 대여하는 고객에게 itemset Y를 추천해 줄 수 있다.

Ex) 반지의제왕1, 해리포터1 대여하는 고객에게 해리포터2를 추천해준다.