

Data Analytics

Assignment -Association rule mining-

윤장혁 교수님

산업공학과

201811527

이영은

Week6

■ DVD 대여 데이터(data_week6.txt)를 활용하여 연관규칙 분석

1. **data처리과정**(txt파일 -> csv파일로 변환)
2. **Association Rule Mining**(연관규칙 분석) 의 과정으로 서술하였습니다.

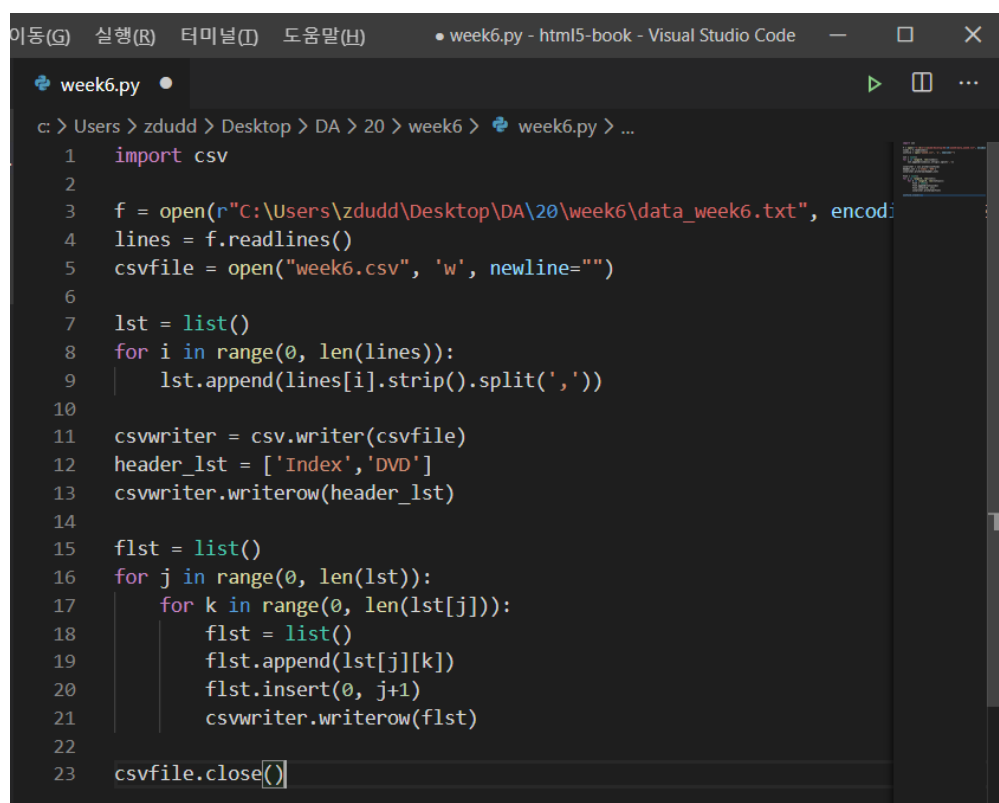
또한 data처리과정에서는 **python visual studio code**를 사용하였고 ARM에서는 **R studio arules package**를 사용하였으며 arules 시각화에서는 **arulesViz package**를 사용하였습니다.

(1) Data 처리 과정 (txt파일 -> csv파일로 변환)

우선 txt파일을 분석할 때 편리하게 하기 위하여 csv파일의 형태로 변환하였습니다.

이 과정은 python visual studio를 통해 실행하였습니다.

◆ Python 실행 코드



```
week6.py
c: > Users > zdudd > Desktop > DA > 20 > week6 > week6.py > ...
1  import csv
2
3  f = open(r"C:\Users\zdudd\Desktop\DA\20\week6\data_week6.txt", encoding='utf-8')
4  lines = f.readlines()
5  csvfile = open("week6.csv", 'w', newline='')
6
7  lst = list()
8  for i in range(0, len(lines)):
9      lst.append(lines[i].strip().split(','))
10
11  csvwriter = csv.writer(csvfile)
12  header_lst = ['Index', 'DVD']
13  csvwriter.writerow(header_lst)
14
15  flst = list()
16  for j in range(0, len(lst)):
17      for k in range(0, len(lst[j])):
18          flst = list()
19          flst.append(lst[j][k])
20          flst.insert(0, j+1)
21          csvwriter.writerow(flst)
22
23  csvfile.close()
```

그림 1 – txt -> csv 파일로 변환하는 과정 python code

◆ Python 실행 코드 설명

실행 코드	설명
import csv	■ csv파일을 다루기 위하여 csv를 import 하였습니다.
f = open(r"C:\Users\Wzdudd \Desktop\DA\20 \week6\data_week6.txt", encoding='utf8') lines = f.readlines() csvfile = open("week6.csv", 'w', newline="")	<ul style="list-style-type: none"> ■ data_week6.txt파일을 열어서 읽어서 f라는 변수를 만들어 할당하였습니다. ■ f를 한 줄 씩 읽어서 lines라는 리스트에 저장하였습니다. ■ txt파일을 csv로 변환하여 저장하기 위해 week6.csv라는 이름의 csv 형식 파일을 생성하여 csvfile에 할당하였습니다.
lst = list() for i in range(0, len(lines)): lst.append(lines[i].strip().split(','))	■ lst라는 list를 생성하였고 이lst에 lines의 i번째행을 ,를 기준으로 나누어 append하였습니다.
csvwriter = csv.writer(csvfile) header_lst = ['Index','DVD'] csvwriter.writerow(header_lst)	■ csvwriter를 이용하여 csvfile을 작성하기 위하여 우선 index와 DVD를 header로 갖는 리스트를 만들고 이를 csvwriter를 사용하여 csv파일에 저장하였습니다.
flst = list() for j in range(0, len(lst)): for k in range(0, len(lst[j])): flst = list() flst.append(lst[j][k]) flst.insert(0, j+1) csvwriter.writerow(flst) csvfile.close()	■ flst라는 리스트를 만들어서 lst에 있는 i번째에 있는 DVD 하나씩과 그 Index를 한 행으로 만들어 저장하였고, 이를 csvwriter를 사용하여 한 행씩 입력하였습니다.

R studio에서의 분석과정에서 transaction ID로 쓰일 수 있는 ID열을 만들기 위하여 Index열을 추가하였고, data_week6의 순서대로 1부터 25까지 Index를 부여하였습니다.

◆ Python 코드 실행 결과

실행결과는 “week6.csv”라는 이름으로 R studio에서 사용하기 위하여 제가 사용하고있는 폴더에 저장하였으며, 저장한 csv파일은 다음과 같습니다.



	A	B	C
1	Index	DVD	
2	1	식스센스	
3	1	반지의제왕1	
4	1	해리포터1	
5	1	해리포터2	
6	1	쇼생크탈출	
7	2	어벤저스	
8	2	스타워즈	
9	2	아바타	
10	3	반지의제왕1	
11	3	반지의제왕2	
12	4	어벤저스	
13	4	스타워즈	
14	4	식스센스	

그림 2 – week6.csv 파일 일부

(색 표시는 가독성을 높이기 위하여 보고서 작성 과정에서 추가하였습니다.)

python을 사용하여 data_week6.txt파일의 각 행 즉, transaction ID를 Index로하고, DVD에 하나씩 들어갈 수 있게 하였습니다.

즉 Index를 1로갖는 식스센스, 반지의제왕1, 해리포터1, 해리포터2, 쇼생크탈출 모두 같이 대여하였음을 의미합니다.


association rule mining을 하기 위해 R studio를 사용하였고, (2)에 자세하게 서술 하였습니다.

(2) Association Rule Mining

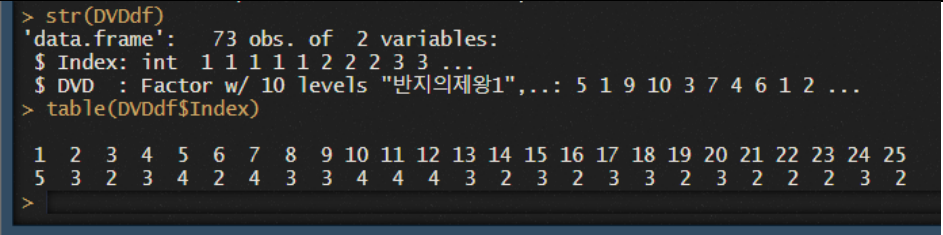
◆ R Studio 실행 코드

- ◆ R Studio에서는 분석 과정을 포함하고 있기 때문에, 한 부분 씩 나누어 결과를 첨부 하였습니다.

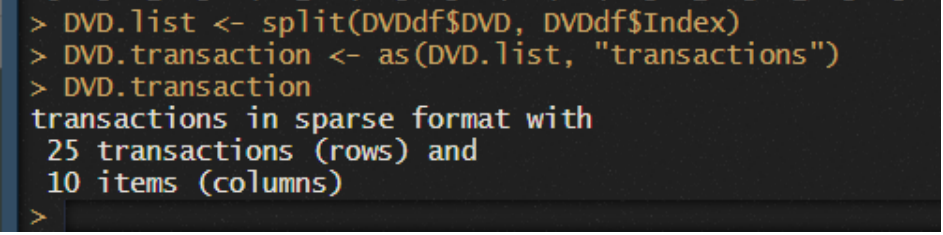
① csv 파일 읽기

실행 코드																																																																							
<pre>library(arules) DVDdf <- read.csv("C:/Users/zdudd/Desktop/DA/20/week6/week6.csv") DVDdf</pre>																																																																							
코드 설명																																																																							
<p>association rule mining을 하기 위하여 R의 arules package를 사용하였습니다. arules package를 로드하고, (1)에서 csv로 저장한 wee6.csv 파일을 로드 하였습니다.</p>																																																																							
실행 결과 (R studio 실행 결과 console 캡처)																																																																							
 <p>The screenshot shows the R Studio console with the following commands and output:</p> <pre>> library(arules) > DVDdf <- read.csv("C:/Users/zdudd/Desktop/DA/20/week6/week6.csv") > DVDdf</pre> <p>The output is a data frame with 34 rows and 2 columns: Index and DVD. The DVD column contains movie titles.</p> <table border="1"> <thead> <tr> <th>Index</th> <th>DVD</th> </tr> </thead> <tbody> <tr><td>1</td><td>식스센스</td></tr> <tr><td>2</td><td>반지의제왕1</td></tr> <tr><td>3</td><td>해리포터1</td></tr> <tr><td>4</td><td>해리포터2</td></tr> <tr><td>5</td><td>쇼생크탈출</td></tr> <tr><td>6</td><td>어벤져스</td></tr> <tr><td>7</td><td>스타워즈</td></tr> <tr><td>8</td><td>아바타</td></tr> <tr><td>9</td><td>반지의제왕1</td></tr> <tr><td>10</td><td>반지의제왕2</td></tr> <tr><td>11</td><td>어벤져스</td></tr> <tr><td>12</td><td>스타워즈</td></tr> <tr><td>13</td><td>식스센스</td></tr> <tr><td>14</td><td>어벤져스</td></tr> <tr><td>15</td><td>스타워즈</td></tr> <tr><td>16</td><td>식스센스</td></tr> <tr><td>17</td><td>해리포터1</td></tr> <tr><td>18</td><td>어벤져스</td></tr> <tr><td>19</td><td>아바타</td></tr> <tr><td>20</td><td>해리포터1</td></tr> <tr><td>21</td><td>해리포터2</td></tr> <tr><td>22</td><td>반지의제왕1</td></tr> <tr><td>23</td><td>반지의제왕2</td></tr> <tr><td>24</td><td>어벤져스</td></tr> <tr><td>25</td><td>스타워즈</td></tr> <tr><td>26</td><td>쇼생크탈출</td></tr> <tr><td>27</td><td>어벤져스</td></tr> <tr><td>28</td><td>스타워즈</td></tr> <tr><td>29</td><td>식스센스</td></tr> <tr><td>30</td><td>반지의제왕1</td></tr> <tr><td>31</td><td>식스센스</td></tr> <tr><td>32</td><td>어벤져스</td></tr> <tr><td>33</td><td>쇼생크탈출</td></tr> <tr><td>34</td><td>어벤져스</td></tr> </tbody> </table>		Index	DVD	1	식스센스	2	반지의제왕1	3	해리포터1	4	해리포터2	5	쇼생크탈출	6	어벤져스	7	스타워즈	8	아바타	9	반지의제왕1	10	반지의제왕2	11	어벤져스	12	스타워즈	13	식스센스	14	어벤져스	15	스타워즈	16	식스센스	17	해리포터1	18	어벤져스	19	아바타	20	해리포터1	21	해리포터2	22	반지의제왕1	23	반지의제왕2	24	어벤져스	25	스타워즈	26	쇼생크탈출	27	어벤져스	28	스타워즈	29	식스센스	30	반지의제왕1	31	식스센스	32	어벤져스	33	쇼생크탈출	34	어벤져스
Index	DVD																																																																						
1	식스센스																																																																						
2	반지의제왕1																																																																						
3	해리포터1																																																																						
4	해리포터2																																																																						
5	쇼생크탈출																																																																						
6	어벤져스																																																																						
7	스타워즈																																																																						
8	아바타																																																																						
9	반지의제왕1																																																																						
10	반지의제왕2																																																																						
11	어벤져스																																																																						
12	스타워즈																																																																						
13	식스센스																																																																						
14	어벤져스																																																																						
15	스타워즈																																																																						
16	식스센스																																																																						
17	해리포터1																																																																						
18	어벤져스																																																																						
19	아바타																																																																						
20	해리포터1																																																																						
21	해리포터2																																																																						
22	반지의제왕1																																																																						
23	반지의제왕2																																																																						
24	어벤져스																																																																						
25	스타워즈																																																																						
26	쇼생크탈출																																																																						
27	어벤져스																																																																						
28	스타워즈																																																																						
29	식스센스																																																																						
30	반지의제왕1																																																																						
31	식스센스																																																																						
32	어벤져스																																																																						
33	쇼생크탈출																																																																						
34	어벤져스																																																																						
<p>중간의 35-37 의 결과는 생략하였고, 총 25번, 73개의 DVD대여가 있음을 알 수 있습니다.</p>																																																																							

② csv 파일 자료의 구조 및 Index 정보 확인

실행 코드
<pre>str(DVDdf) table(DVDdf\$Index)</pre>
코드 설명
<p>str(DVDdf)를 사용하여 DVDdf의 정보를 찾았고, Index즉, 구매가 몇번 있었는지 찾기 위하여 table(DVDdf\$Index)를 사용했습니다.</p>
실행 결과 (R studio 실행 결과 console 캡처)
 <pre>> str(DVDdf) 'data.frame': 73 obs. of 2 variables: \$ Index: int 1 1 1 1 1 2 2 2 3 3 ... \$ DVD : Factor w/ 10 levels "반지의제왕1",...: 5 1 9 10 3 7 4 6 1 2 ... > table(DVDdf\$Index) 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 5 3 2 3 4 2 4 3 3 4 4 4 3 2 3 2 3 3 2 3 2 2 3 2 ></pre> <p>몇 번의 DVD 구매가 있었는지 확인하기 위하여 index를 확인한 결과 총25 번의 Index 정보가 있었고, str(DVDdf)결과에 따르면 Index와 DVD 두 column으로 이루어져 있는 dataframe형태임을 알 수 있습니다.</p>

③ transaction 타입으로의 데이터 변환

실행 코드
<pre>DVD.list <- split(DVDdf\$DVD, DVDdf\$Index) DVD.transaction <- as(DVD.list, "transactions") DVD.transaction</pre>
코드 설명
<p>연관성 분석을 하기 위해서는 데이터를 transaction 타입으로 변경 해 주어야 하기 때문에 data.frame 형식의 데이터 변환 작업을 하였고, data.frame을 Index별 list로 쪼갠 후에 transaction 타입으로 변환하였습니다.</p>
실행 결과 (R studio 실행 결과 console 캡처)
 <pre>> DVD.list <- split(DVDdf\$DVD, DVDdf\$Index) > DVD.transaction <- as(DVD.list, "transactions") > DVD.transaction transactions in sparse format with 25 transactions (rows) and 10 items (columns) ></pre> <p>25건의 대여 (Index)와 DVD 종류(10종류) 가 있었음을 알 수 있습니다.</p>

④ 연관성 규칙 생성

● 지지도 (Support)

- 전체 대여에서 특정 DVD A와 B가 동시에 대여되는 비중으로 해당 규칙이 얼마나 의미가 있는 규칙인지 나타냅니다.
- $Support = A와 B를 동시에 대여한 횟수 / 전체 대여 횟수$ 입니다.

● 신뢰도 (Confidence)

- A를 포함하는 대여 중 A와 B가 동시에 대여되는 비중으로 A 대여 라는 사건이 발생 했을 때 B 대여 라는 사건이 발생할 확률이 얼마나 높은지를 뜻합니다.
- $Conf(A \rightarrow B) = \frac{supp(A \cup B)}{supp(A)}$ 입니다.

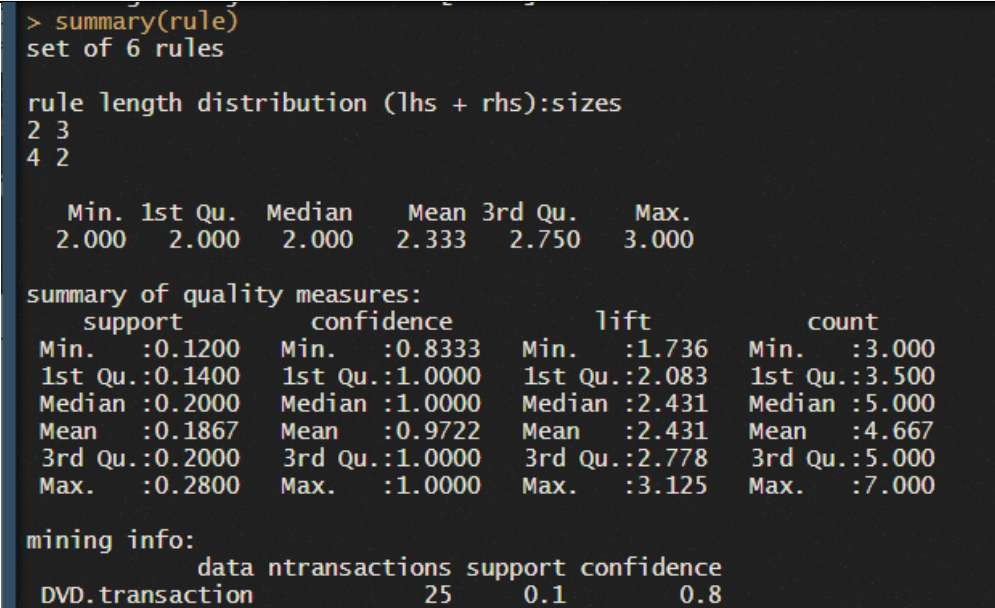
● 향상도 (Lift)

- A와 B가 동시에 대여된 비중을 A와 B가 독립된 사건일 때 동시에 거래된 비중으로 나눈 값입니다.

A와 B가 우연히 같이 대여 된 확률 보다 A와 B의 관계가 얼마나 더 높은지 보는 지표입니다.

- $Lift = \frac{conf(A \rightarrow B)}{supp(B)} = \frac{supp(A \cup B)}{supp(A) * supp(B)}$ 즉 $\frac{P(A \cap B)}{P(A) * P(B)} = P(B|A)/P(B)$ 입니다.
- DVD A와 B 사이에 아무런 상호 관계가 없다면 향상도는 1 입니다. 향상도가 1보다 클수록 이 관계는 우연히 일어나지 않았다는 지표입니다.

연관성 규칙을 생성한 결과는 다음과 같습니다.

<div>실행 코드</div> <pre>rule = apriori(DVD.transaction) summary(rule) rule.list <- as.data.frame(inspect(rule)) rule.list <- rule.list[order(rule.list\$lift, decreasing = TRUE),] rule.list</pre>																																																	
<div>코드 설명</div> <p>규칙을 생성하여 6개의 규칙을 찾았습니다.</p> <p>paramter를 따로 설정하지 않았기 때문에 support 는 0.1, confidence는 0.8 이상인 DVD 사이의 관계가 나왔습니다.</p> <p>그 후에 상세히 보기 위하여 insect 함수를 사용하였고, 이를 lift가 높은 순서대로 정렬하였습니다.</p>																																																	
<div>실행 결과 (R studio 실행 결과 console 캡처)</div>  <pre>> summary(rule) set of 6 rules rule length distribution (lhs + rhs):sizes 2 3 4 2 Min. 1st Qu. Median Mean 3rd Qu. Max. 2.000 2.000 2.000 2.333 2.750 3.000 summary of quality measures: support confidence lift count Min. :0.1200 Min. :0.8333 Min. :1.736 Min. :3.000 1st Qu.:0.1400 1st Qu.:1.0000 1st Qu.:2.083 1st Qu.:3.500 Median :0.2000 Median :1.0000 Median :2.431 Median :5.000 Mean :0.1867 Mean :0.9722 Mean :2.431 Mean :4.667 3rd Qu.:0.2000 3rd Qu.:1.0000 3rd Qu.:2.778 3rd Qu.:5.000 Max. :0.2800 Max. :1.0000 Max. :3.125 Max. :7.000 mining info: data ntransactions support confidence DVD.transaction 25 0.1 0.8</pre> <p>lift가 높은 순으로 정렬한 결과는 다음과 같습니다.</p> <pre>> rule.list <- rule.list[order(rule.list\$lift, decreasing = TRUE),] > rule.list</pre> <table> <thead> <tr> <th></th><th>lhs</th><th>rhs</th><th>support</th><th>confidence</th><th>lift</th><th>count</th></tr> </thead> <tbody> <tr> <td>[2]</td><td>{반지의제왕2}</td><td>=> {반지의제왕1}</td><td>0.20</td><td>1.0000000</td><td>3.125000</td><td>5</td></tr> <tr> <td>[3]</td><td>{해리포터2}</td><td>=> {해리포터1}</td><td>0.20</td><td>1.0000000</td><td>2.777778</td><td>5</td></tr> <tr> <td>[5]</td><td>{반지의제왕1, 해리포터2}</td><td>=> {해리포터1}</td><td>0.12</td><td>1.0000000</td><td>2.777778</td><td>3</td></tr> <tr> <td>[4]</td><td>{스타워즈}</td><td>=> {어벤져스}</td><td>0.28</td><td>1.0000000</td><td>2.083333</td><td>7</td></tr> <tr> <td>[6]</td><td>{스타워즈, 식스센스}</td><td>=> {어벤져스}</td><td>0.12</td><td>1.0000000</td><td>2.083333</td><td>3</td></tr> <tr> <td>[1]</td><td>{아바타}</td><td>=> {어벤져스}</td><td>0.20</td><td>0.8333333</td><td>1.736111</td><td>5</td></tr> </tbody> </table>		lhs	rhs	support	confidence	lift	count	[2]	{반지의제왕2}	=> {반지의제왕1}	0.20	1.0000000	3.125000	5	[3]	{해리포터2}	=> {해리포터1}	0.20	1.0000000	2.777778	5	[5]	{반지의제왕1, 해리포터2}	=> {해리포터1}	0.12	1.0000000	2.777778	3	[4]	{스타워즈}	=> {어벤져스}	0.28	1.0000000	2.083333	7	[6]	{스타워즈, 식스센스}	=> {어벤져스}	0.12	1.0000000	2.083333	3	[1]	{아바타}	=> {어벤져스}	0.20	0.8333333	1.736111	5
	lhs	rhs	support	confidence	lift	count																																											
[2]	{반지의제왕2}	=> {반지의제왕1}	0.20	1.0000000	3.125000	5																																											
[3]	{해리포터2}	=> {해리포터1}	0.20	1.0000000	2.777778	5																																											
[5]	{반지의제왕1, 해리포터2}	=> {해리포터1}	0.12	1.0000000	2.777778	3																																											
[4]	{스타워즈}	=> {어벤져스}	0.28	1.0000000	2.083333	7																																											
[6]	{스타워즈, 식스센스}	=> {어벤져스}	0.12	1.0000000	2.083333	3																																											
[1]	{아바타}	=> {어벤져스}	0.20	0.8333333	1.736111	5																																											
<p>규칙은 총6개이고 lift(향상도)가 가장 높은 것은 반지의 제왕2=> 반지의 제왕1으로 반지의 제왕2를 대여하였을 때 반지의 제왕1을 대여할 때 우연적 확률이 아닐 확률이 가장 높음을 의미 합니다.</p>																																																	

다음으로 해리포터2 => 해리포터1
반지의 제왕1, 해리포터2 => 해리포터1
스타워즈 => 어벤져스
아바타 => 어벤져스가 있습니다.

(3) Association Rule Mining 결과 분석

◆ (2)의 결과에 따르면

LHS	RHS	support	confidence	lift
반지의 제왕2	반지의 제왕1	0.20	1	3.125
해리포터2	해리포터1	0.20	1	2.778
반지의 제왕1, 해리포터2	해리포터1	0.12	1	2.778
스타워즈	어벤져스	0.28	1	2.083
스타워즈, 식스센스	해리포터1	0.12	1	2.083
아바타	어벤져스	0.20	0.83	1.736

의 관계가 있음을 알 수 있습니다. (**support > 0.1, confidence > 0.8 일 때**)

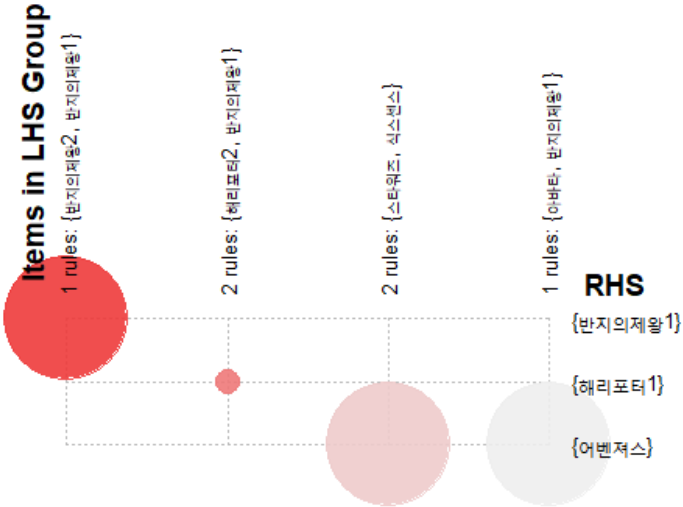
반지의 제왕과 해리포터의 경우 시리즈 물 이기 때문에 분석을 하기 전, 해당 DVD 는 같이 대여 될 확률이 높을 것으로 예상 했고, 이 예상과 맞게 각 시리즈 물이 LHS와 RHS 관계를 갖고 있음을 알 수 있습니다.

총 25번의 DVD 대여 데이터로 분석을 하였기 때문에 분석 결과가 많이 나오지 않았지만 ((5)에서 그래프를 그리기 위해 파라미터를 support = 0.1, confidence = 0.1로 설정하여 그래프를 그려보았습니다.) 더 많은 데이터를 다룬다면 더 유의미한 결과가 나올 것이라고 생각되며, 많은 데이터로 분석을 진행 할 경우에 그 분석 결과를 가지고 DVD진열에 사용하거나, 같이 대여가 많이 되는 영화를 토대로 추천을 해주거나, 프로모션 행사를 할 수 있을 것이라고 생각 됩니다.

다음 과정에서 위의 과정에서 구한 6가지 규칙을 시각화 하였습니다.

(4) Association Rule Mining 결과 시각화

① Grouped matrix for 6 rules

실행 코드
<pre>library(arulesViz) plot(rule) plot(sort(rule, by = "support"), method = "grouped")</pre>
실행 결과
<p style="text-align: center;">Grouped Matrix for 6 Rules</p> <div style="text-align: right; margin-right: 50px;"> Size: support Color: lift </div>  <p>LHS가 x축에, RHS가 y축에 설정되었고, Grouped matrix-based visualization은 clustering을 사용하여 grouped되었습니다. 원의 크기가 지지도 이며, 색은 향상도 입니다.</p> <p>* (7) 참고 자료의 Package 'arulesViz' CRAN의 4p를 참고하였습니다.</p>

② parameter(support = 0.1, confidence = 0.1로) 변경 후 Graph

실행 코드

```
rule1 <- apriori(DVD.transaction, parameter =
                  list(support = 0.1, confidence = 0.1, minlen = 2))
summary(rule1)
rule1.list <- as.data.frame(inspect(rule1))
plot(rule1, method = "graph", interactive = T)
```

실행 결과

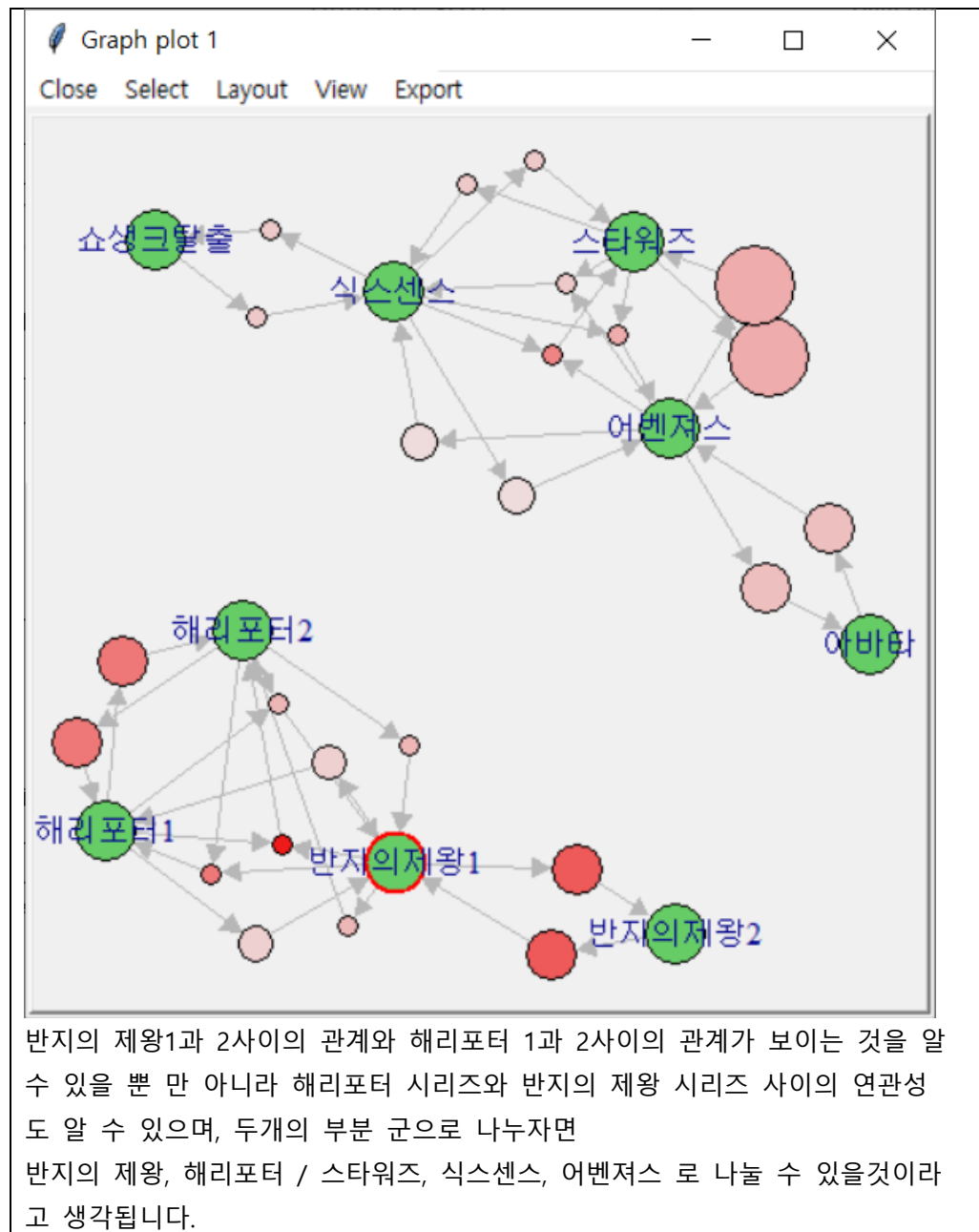
```
rule length distribution (lhs + rhs):sizes
 2  3
18  6

    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    2.00   2.00   2.00   2.25   2.25   3.00

summary of quality measures:
      support      confidence      lift      count
Min.   :0.12   Min.   :0.3333   Min.   :1.190   Min.   :3.0
1st Qu.:0.12   1st Qu.:0.4286   1st Qu.:1.531   1st Qu.:3.0
Median :0.14   Median :0.5774   Median :1.875   Median :3.5
Mean   :0.16   Mean   :0.6284   Mean   :2.049   Mean   :4.0
3rd Qu.:0.20   3rd Qu.:0.7708   3rd Qu.:2.703   3rd Qu.:5.0
Max.   :0.28   Max.   :1.0000   Max.   :3.750   Max.   :7.0

mining info:
      data ntransactions support confidence
DVD.transaction      25      0.1      0.1
> rule1.list <- as.data.frame(inspect(rule1))
      lhs      rhs      support confidence lift      count
[1] {아바타}      => {어벤져스}    0.20    0.8333333 1.736111 5
[2] {어벤져스}      => {아바타}    0.20    0.4166667 1.736111 5
[3] {반지의제왕2}    => {반지의제왕1} 0.20    1.0000000 3.125000 5
[4] {반지의제왕1}    => {반지의제왕2} 0.20    0.6250000 3.125000 5
[5] {해리포터2}      => {반지의제왕1} 0.12    0.6000000 1.875000 3
[6] {반지의제왕1}    => {해리포터2}  0.12    0.3750000 1.875000 3
[7] {해리포터2}      => {해리포터1}  0.20    1.0000000 2.777778 5
[8] {해리포터1}      => {해리포터2}  0.20    0.5555556 2.777778 5
[9] {쇼생크탈출}      => {식스센스}    0.12    0.4285714 1.530612 3
[10] {식스센스}        => {쇼생크탈출}  0.12    0.4285714 1.530612 3
[11] {스타워즈}        => {식스센스}    0.12    0.4285714 1.530612 3
[12] {식스센스}        => {스타워즈}    0.12    0.4285714 1.530612 3
[13] {스타워즈}        => {어벤져스}    0.28    1.0000000 2.083333 7
[14] {어벤져스}        => {스타워즈}    0.28    0.5833333 2.083333 7
[15] {식스센스}        => {어벤져스}    0.16    0.5714286 1.190476 4
[16] {어벤져스}        => {식스센스}    0.16    0.3333333 1.190476 4
[17] {반지의제왕1}    => {해리포터1}  0.16    0.5000000 1.388889 4
[18] {해리포터1}      => {반지의제왕1} 0.16    0.4444444 1.388889 4
[19] {반지의제왕1, 해리포터2} => {해리포터1} 0.12    1.0000000 2.777778 3
[20] {해리포터1, 해리포터2} => {반지의제왕1} 0.12    0.6000000 1.875000 3
[21] {반지의제왕1, 해리포터1} => {해리포터2}  0.12    0.7500000 3.750000 3
[22] {스타워즈, 식스센스} => {어벤져스}    0.12    1.0000000 2.083333 3
[23] {스타워즈, 어벤져스} => {식스센스}    0.12    0.4285714 1.530612 3
[24] {식스센스, 어벤져스} => {스타워즈}    0.12    0.7500000 2.678571 3
```

좀 더 많은 관계의 결과를 보기 위하여 support 와 confidence를 모두 0.1로 낮추었고, 그결과 24가지의 rules가 나왔습니다. 따라서 이 자료들로 graph를 그려보았고, 결과는 다음과 같습니다.



support와 confidence를 낮게 잡아서 분석 한 결과 24개의 rule이 나왔지만, 이는 높은 support와 confidence를 갖지는 않으므로, 더 많은 데이터가 있었으면 하는 아쉬움이 있었습니다.

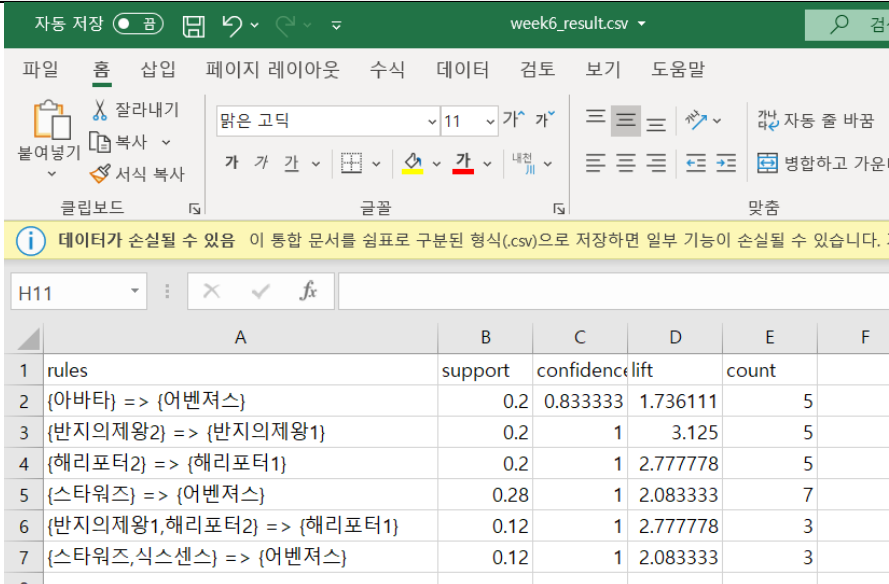
(5) Association Rule Mining 결과 csv 파일 저장

실행 코드

```
write(rule, file = "C:/Users/zdudd/Desktop/DA/20/week6/week6_result.csv",
      sep = ",",
      quote = TRUE,
      row.names = FALSE)

rule_df <- as(rule, "data.frame")
```

실행 결과



rule의 결과를 week6_result.csv파일을 만들어 저장하였습니다.

(6) R Studio 실행 코드 전체

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R.R* x Untitled1* x Untitled2* x
Source on Save Run Source
2 DVDdf <- read.csv("C:/Users/zdudd/Desktop/DA/20/week6/week6.csv")
3 DVDdf
4 str(DVDdf)
5 table(DVDdf$Index)
6
7 DVD.list <- split(DVDdf$DVD, DVDdf$Index)
8 DVD.transaction <- as(DVD.list, "transactions")
9 DVD.transaction
10
11 rule <- apriori(DVD.transaction)
12
13 summary(rule)
```

```

15 rule.list <- as.data.frame(inspect(rule))
16 rule.list <- rule.list[order(rule.list$lift, decreasing = TRUE),]
17 rule.list
18
19 install.packages("arulesViz")
20
21 library(arulesViz)
22 plot(rule)
23
24 plot(sort(rule, by = "support"), method = "grouped")
25
26 #Graph for association rules
27 plot(rule, method = "graph", control = list(type="items"))
28 plot(rule, method = "graph", control = list(type="items"),
29       vertex.label.cex = 0.5,
30       edge.arrow.size = 0.3,
31       edge.arrow.width = 2,
32       )
33
34 write(rule,
35       file = "C:/Users/zdudd/Desktop/DA/20/week6/week6_result.csv",
36       sep = ",",
37       quote = TRUE,
38       row.names = FALSE)
39
40 rule_df <- as(rule, "data.frame")
41
42
43 rule1 <- apriori(DVD.transaction, parameter =
44                 list(support = 0.1, confidence = 0.1, minlen = 2))
45 summary(rule1)
46 rule1.list <- as.data.frame(inspect(rule1))
47 plot(rule1, method = "graph", interactive = T)
48

```

(7) 참고 자료



cran.r-project.org>arulesViz의 Package 'arulesViz' -CRAN을 참고하였습니다.