

## ### 활용6-(1); 추천시스템: 엑셀작업 ###

	u1	u2	u3	u4
m1	1	4	2	2
m2	5	2	4	4
m3			3	
m4	2	5		5
m5	4	1		1
m6		2		

· u: 고객, m: 선호도

· 추천시스템; 빈칸의 선호도를 예측해서 가장 높은 선호도의 상품을 추천하는 시스템

· 우리가 사용할 알고리즘; 협력적 필터링 알고리즘(Collaborative filtering algorithm)

· 기존의 데이터를 하나 가려놓고 예측에 성공하는 지 평가

· 이웃 기반의 협력적 필터링 수식 (cf. U: 고객, J: 이웃한 고객, r: 상관계수)

$$\Rightarrow \hat{U}_x = \bar{U} + \frac{\sum_{j \in r_{aters}} (J_x - \bar{J}) r_{uj}}{\sum_{j \in r_{aters}} |r_{uj}|}, \quad \bar{J} = \frac{\sum_i J_i}{n}, \quad i \neq x$$

## #1; 모르는 값 예측

	A	B	C	D	E	F	G	H	I	J
1		Ken	Lee	Meg	Nam			Lee	Meg	Nam
2	1	1	4	2	2			-0.8	-1	0
3	2	5	2	4	4			0.8	1	0
4	3			3					0	0
5	4	2	5		5			-1.6		0
6	5	4	1		1			1.6		0
7	6	?	2	5				0.8	2	0
8	mean	3	3	3	3	sum		0.8	2	0
9	r		-0.8	1	0					
10	예측	4.555556								

## #2; 아는 값을 가리고 예측

	A	B	C	D	E	F	G	H	I
1		Ken	Lee	Meg	Nam			Lee	Nam
2	1	1	4	2	2			-0.83863	0.279543
3	2		2	4	4			0.838628	-0.55909
4	3			3					
5	4	2	5		5			-1.67726	-0.9784
6	5	4	1		1			1.677256	0.698857
7	6		2	5				0.838628	
8	mean	2.333333	3	3.333333	2.666667	sum		0.838628	-0.55909
9	r		-0.83863	#DIV/0!	-0.41931				
10	예측	2.555556							

· 상관계수; 최소한 2쌍 이상 존재해야 함

· #DIV/0!; 1쌍 밖에 없어서 상관계수를 못 구함 -> 이웃에서 제외

· 실제값=5 vs 예측값=2.56 -> 예측률 별로(∵ 데이터 개수가 너무 적음)

### #3; 실제 데이터에 적용

- ' 알고리즘; 협업적 필터링
- ' Data set; grouplens.org의 older dataset
- ' u.data; user id, item id, rating 등
- ' u.item; movie id, movie title, genre(0,1) 등
- ' u.user; user id, age, gender, occupation 등
- ' 평균, 분산 등 활용할 수 있게  $n, \sum x, \sum x^2$  자료 만들어 놓기
- ' For 추천시스템;  $u, m, \bar{U}, U_r, J, J_r, \bar{J}, cor$  등의 자료 필요

### ### 활용6-(2); 자료 변환 ###

#### #1; 자료 변환

- ' (1) 구분자 '|' -> ','(쉼표)
- ' (2) 성별 M/F -> 1/2

```
Sub kk()  
  Open "C:\Users\USER\For 통특\Wu.user" For Input As #1  
  Open "C:\Users\USER\For 통특\Wu.user2" For Output As #2  
  
  While Not EOF(1)  
    Line Input #1, a  
  
    For i = 1 To Len(a)  
      If Mid(a, i, 1) = "|" Then Mid(a, i, 1) = "," ' 구분자 | -> ,  
      If Mid(a, i, 1) = "M" Then Mid(a, i, 1) = "1" ' 남성 -> 1  
      If Mid(a, i, 1) = "F" Then Mid(a, i, 1) = "2" ' 여성 -> 2  
    Next i  
  
    Print #2, a  
  Wend  
  
End  
End Sub
```

### ### 활용6-(3); 자료 추출 ###

- ' 개봉연도 추출 -> 직접 해보기

#### cf. 자료 확인하기

```
Sub kk()  
  Open "C:\Users\USER\For 통특\Wu.data" For Input As #1  
  While Not EOF(1)  
    Input #1, u, m, r, t  
    cnt = cnt + 1  
    Debug.Print u, m, r, t ' 실행결과, 뭔가 이상함  
    If cnt > 10 Then End
```

```
Wend
End
End Sub
```

```
Sub kk()
Open "C:\Users\USER\For 통특\Wu.data" For Input As #1
While Not EOF(1)
Line Input #1, a
cnt = cnt + 1
Wend
Debug.Print cnt ' 실행결과=1; 데이터가 한 줄로 되어있다고 인식
End
End Sub
```

· 원래 아스키코드 10번, 13번이 들어가 있어야 함; Enter + 커서 맨 앞으로 이동  
· 어떻게 해결? 코드 짜기 or 한글 파일로 저장 후 다시 txt로 변환

```
Sub kk()
Open "C:\Users\USER\For 통특\Wudata.txt" For Input As #1
While Not EOF(1)
Line Input #1, a
cnt = cnt + 1
Wend
Debug.Print cnt ' 실행결과=100000; 이제 정상적으로 인식
End
End Sub
```

```
Sub kk()
Open "C:\Users\USER\For 통특\Wudata.txt" For Input As #1
While Not EOF(1)
Input #1, u, m, r, t
cnt = cnt + 1
Debug.Print u, m, r, t
If cnt > 10 Then End
Wend
End
End Sub
```

· 따라서 정상적으로 변수가 불러와진 것을 확인가능

## 13week

### 활용6-(4); 영화별 고객 수 ###

#1; 분석 전 확인 작업

```
Sub kk()
Open "C:\Users\USER\For 통특\Wudata.txt" For Input As #1
Open "C:\Users\USER\For 통특\Wmov_u_cnt.txt" For Output As #2
```

```

While Not EOF(1)
    Line Input #1, a ' 줄바꿈 아스키코드가 적절히 들어가 있는지 확인하기 위해
    Input #1, u, m, r, t
    cnt = cnt + 1
    Debug.Print u, m, r, t ' Debug.? = Debug.Print
    If cnt > 10 Then End ' 11번째 줄까지 Print 변수를 잘 불러오니까 이제 제대로 분석
Wend

Debug.Print cnt

End
End Sub

```

### #1; 영화별 고객 수

```

Const movies = 1682 ' 영화의 개수 = 1682개
Dim mov(movies) ' 상수선언 습관화하기
Sub kk()
    Open "C:\Users\USER\For 통특\data.txt" For Input As #1
    Open "C:\Users\USER\For 통특\mov_u_cnt.txt" For Output As #2

    While Not EOF(1)
        Input #1, u, m, r, t
        mov(m) = mov(m) + 1 ' 영화마다(영화의 번호마다) 카운트
    Wend

    For i = 1 To movies
        Print #2, i, ","; mov(i) ' i번째 영화를 본 고객의 개수 저장
    Next i

End
End Sub

```

### ### 활용6-(5); 고객별 영화 수 ###

### #1; 고객별 영화 수

```

Const users = 943 ' 고객의 수 = 943명
Dim user(users)
Sub kk()
    Open "C:\Users\USER\For 통특\data.txt" For Input As #1
    Open "C:\Users\USER\For 통특\user_mov_cnt.txt" For Output As #2

    While Not EOF(1)
        Input #1, u, m, r, t
        user(u) = user(u) + 1 ' 고객마다(고객의 번호마다) 카운트
    Wend

```

```
For i = 1 To users
    Print #2, i; ","; user(i) ' i번째 고객이 본 영화의 개수 저장
Next i

End
End Sub
```

### 활용6-(6); 변수추가 ###

' 고객의 나이나 성별 등 uuser.txt의 정보를 udata.txt에 추가하기

#1; 자료 변환(복습)

```
Sub kk()
    Open "C:\Users\USER\For 통특\user.txt" For Input As #1
    Open "C:\Users\USER\For 통특\user2.txt" For Output As #2

    While Not EOF(1)
        Line Input #1, a
        For i = 1 To Len(a)
            If Mid(a, i, 1) = "|" Then Mid(a, i, 1) = ","
            If Mid(a, i, 1) = "M" Then Mid(a, i, 1) = "1"
            If Mid(a, i, 1) = "F" Then Mid(a, i, 1) = "2"
        Next i
        Print #2, a
    Wend

End
End Sub
```

#2; 나이와 성별 변수 추가

```
Const users = 943
Dim user_age(users)
Dim user_sex(users)

Sub kk()
    Open "C:\Users\USER\For 통특\user2.txt" For Input As #1
    While Not EOF(1) ' 주의; 혹시 1줄로 이루어지진 않았는지 확인
        Input #1, u, age, sex, jop, zip ' zip; 주식
        user_age(u) = age ' user의 나이 저장
        user_sex(u) = sex ' user의 성별 저장
    Wend

    Open "C:\Users\USER\For 통특\udata.txt" For Input As #2
    Open "C:\Users\USER\For 통특\u_age_sex.txt" For Output As #3
    While Not EOF(2)
        Input #2, u, m, r, t
        Print #3, u; ","; m; ","; r; ","; user_age(u); ","; user_sex(u)
```

```
Wend
End
End Sub
```

### #3; 영화의 개봉일 변수 추가 -> 직접 해보기

### 활용6-(7); 분류 ###

#### #1; 성별에 따라 자료 분류

```
Sub kk()
Open "C:\Users\WUSER\For 통특\Wu_age_sex.txt" For Input As #1
Open "C:\Users\WUSER\For 통특\Wu_age_sex_F.txt" For Output As #2
Open "C:\Users\WUSER\For 통특\Wu_age_sex_M.txt" For Output As #3
```

```
While Not EOF(1)
Input #1, u, m, r, age, sex
If sex = 1 Then ' 남성일 때 2번 파일에 저장
cntM = cntM + 1
Print #2, u, m, r, age, sex
End If
If sex = 2 Then ' 여성일 때 2번 파일에 저장
cntF = cntF + 1
Print #3, u, m, r, age, sex
End If
Wend
```

```
Debug.Print cntM, cntF ' 각 성별의 빈도수
```

```
End
End Sub
```

#### #2; 고객별로 자료 분류

```
Sub kk()
Open "C:\Users\WUSER\For 통특\Wu_age_sex.txt" For Input As #1

While Not EOF(1)
Input #1, u, m, r, age, sex

Open "C:\Users\WUSER\For 통특\Wu" + Trim(Str(u)) + ".txt" For Append As #2
Print #2, u, m, r, age, sex
Close #2 ' 닫는 거 잊지 말기!
Wend
```

```
End
End Sub
```

#### #3; 나이대별로 자료 분류

```

Sub kk()
    Open "C:\Users\WUSER\For 통특\Wu_age_sex.txt" For Input As #1
    Open "C:\Users\WUSER\For 통특\Wu_age_10.txt" For Output As #2
    Open "C:\Users\WUSER\For 통특\Wu_age_20.txt" For Output As #3
    Open "C:\Users\WUSER\For 통특\Wu_age_30.txt" For Output As #4
    Open "C:\Users\WUSER\For 통특\Wu_age_40.txt" For Output As #5

    While Not EOF(1)
        Input #1, u, m, r, age, sex
        If age < 20 Then Print #2, u, m, r, age, sex ' 10대
        If (20 <= age) And (age < 30) Then Print #3, u, m, r, age, sex ' 20대
        If (30 <= age) And (age < 40) Then Print #4, u, m, r, age, sex ' 30대
        If age >= 40 Then Print #5, u, m, r, age, sex ' 40대
    Wend

End
End Sub

```

### 활용6-(8); 전체 평가점수의 평균 ###

#1; 전체 rating의 평균

```

Sub kk()
    Open "C:\Users\WUSER\For 통특\Wudata.txt" For Input As #1

    While Not EOF(1)
        Input #1, u, m, r, t
        cnt = cnt + 1
        u_rating_sum = u_rating_sum + r
    Wend

    u_rating_mean = u_rating_sum / cnt ' 평균
    Debug.Print cnt, u_rating_sum, u_rating_mean ' 총 개수, 합계, 평균

End
End Sub

```

### 활용6-(9); 고객별 평가점수의 평균 ###

' 협력적 필터링에서 평균 -> 해당 상품(예측하려는 상품)은 제외한 평균  
' (전체 합계 - 해당 평점) / (전체 개수 - 1)

#1; 고객별 rating의 평균

```

Const users = 943
Dim u_rating_cnt(users)
Dim u_rating_sum(users)

```

```

Sub kk()
    Open "C:\Users\USER\For 통특\udata.txt" For Input As #1
    Open "C:\Users\USER\For 통특\u_rating_mean.txt" For Output As #2

    While Not EOF(1)
        Input #1, u, m, r, t
        u_rating_cnt(u) = u_rating_cnt(u) + 1 ' 각 고객별 빈도수
        u_rating_sum(u) = u_rating_sum(u) + r ' 각 고객별 평점 합
    Wend

    For i = 1 To users
        Print #2, i, u_rating_sum(i), u_rating_cnt(i), u_rating_sum(i) / u_rating_cnt(i)
    Next i

End
End Sub

```

## #2; 고객별 rating의 평균 for 협력적 필터링

' 협력적 필터링에서의 평균 -> 해당 상품(예측하려는 상품)은 제외한 평균  
' (전체 합계 - 해당 평점) / (전체 개수 - 1)

```

Const m = 1682, u = 943
Dim rating(m, u)

```

```

Sub kk()
    Open "C:\Users\USER\For 통특\wu_mean.txt" For Output As #3
    Open "C:\Users\USER\For 통특\udata.txt" For Input As #1
    While Not EOF(1)
        Input #1, tr_u, tr_m, tr_r, tr_t
        rating(tr_m, tr_u) = tr_r ' 2차원 배열; 해당 고객의 해당 영화에 대한 평점
    Wend
    Close #1

    Open "C:\Users\USER\For 통특\udata.txt" For Input As #2
    While Not EOF(2) ' 특정 고객에 대해서 영화를 다 돌리면 다시 다른 고객으로
        Input #2, te_u, te_m, te_r, te_t

        For i = 1 To m ' 특정 고객에 대해서 영화를 가지고 돌림
            If rating(i, te_u) > 0 Then ' 평가 점수가 0보다 크다 = 평가를 했다
                r_sum = r_sum + rating(i, te_u)
                r_cnt = r_cnt + 1
            End If
        Next i

        r_mean = (r_sum - rating(te_m, te_u)) / (r_cnt - 1) ' (전체 평점-해당 영화 평점)/(전체 개수 - 1)
        Print #3, te_u, te_m, rating(te_m, te_u), r_mean
    Wend

```



End  
End Sub

### 활용6-(10); 이웃수 ###

· 파일 다운이 안 됨!!!

- For 협력적 필터링
- $u$ (고객) /  $m$ (영화) /  $J$ (이웃)
- $J_r$ (이웃의 평가점수)
- $u_r$ (고객의 평가점수)
- $\bar{K}$ (고객의 예측하려는 영화를 제외한 평균)
- $\bar{J}$ (이웃의 예측하려는 영화를 제외한 평균)
- $J_r$ (이웃의 평가점수)
- $u_r$ (고객의 평가점수)
- Cor(고객과 이웃의 상관계수)

#1; 먼저 파일이 정상적으로 열리는지 확인

```
Sub kk()  
    Open "C:\Users\USER\For 통특\final_text.txt" For Input As #1  
    While Not EOF(1)  
        Input #1, u, m, j, bar_K, jr, ur, cor, bar_j  
        s = s + 1  
        Debug.Print u, m, j, bar_K, jr, ur, cor, bar_j  
        If s > 10 Then End  
    Wend  
End  
End Sub
```

#2; 이웃수

- 해당 영화를 본 이웃을 알고 싶음
- 이웃 = 각 고객이 본 영화에 대해서 영화마다 같이 본 고객
- 이웃수 = 각 고객이 본 영화에 대해서 영화마다 같이 본 고객의 수

```
Const u = 943  
Dim cnt(u, u), c(u)
```

```
Sub kk()  
    Open "C:\Users\USER\For 통특\final_test.txt" For Input As #1  
    Open "C:\Users\USER\For 통특\final_test_neighbor.txt" For Output As #2  
  
    While Not EOF(1)  
        Input #1, u, m, j, bar_K, jr, ur, cor, bar_j  
        cnt(u, j) = cnt(u, j) + 1 ' for 이웃수  
    Wend
```

```
For i = 1 To u
  For j = 1 To u
    If cnt(i, j) >= 1 Then c(i) = c(i) + 1 ' cnt(i, j)가 1이상 = 같이 본 영화 -> 이웃으로 카운트
  Next j
Next i
```

```
For i = 1 To m
  Print #2, i, c(i) ' c(i); i번째 사람에 대한 이웃수
Next i
```

```
End
End Sub
```

### 활용6-(11); 두 고객 간의 상관계수 ###

· 직접 해보기

· 항상 데이터를 만들어놓고 필요한 것만 뽑아서 쓰는 것을 추천