논문 19-44-05-20 The Journal of Korean Institute of Communications and Information Sciences '19-05 Vol.44 No.05
https://doi.org/10.7840/kics.2019.44.5.936

추천시스템의 성능향상을 위한 요인 분석 및 모형연구

구 지 현◆, 김 선 옥°

Factor Analysis and Model Studying for Improvement of Recommendation System

Jee Hyun Koo◆, Sun Ok Kim°

요 약

추천시스템은 고객의 취향에 적합한 상품을 선택하게 해주는 개인화서비스로 대량정보가 존재하는 인터넷시대에 그 역할은 점점 증가하고 있다. 추천시스템을 구축하는데 사용되는 알고리즘은 고객의 선호도를 예측해 주며, 이 알고리즘의 선호도 예측정확도인 MAE는 추천시스템의 신뢰도와 관련이 있기에 중요하다. 본 연구는 특정 고객정보가 MAE와 관련이 있는지를 조사하고 그 영향력을 분석하였다. 먼저, 추천시스템의 이웃기반 협력적 필터링 알고리즘으로 예측한 고객별 선호도에 대한 MAE를 분석하고, 이를 기본 자료로 하여 MAE가 낮은 그룹과 높은 그룹으로 구분하였다. 구분된 그룹이 특정 고객정보와 관련이 있는지를 알아보고, 그 영향력을 분석하기 위해 이분형 로지스틱 회귀분석을 실시하여 추천시스템의 예측정확도 향상을 위한 요인들로 연구모형을 제안하였다. 제안된 연구모형에서 고객의 응답수와 직업이 MAE와 관련이 있는 요인으로 분석되었다.

Key Words : Collaborative filtering, Recommender, Mean absolute error, Binary logistic regression, Prediction model

ABSTRACT

The recommendation system is a personalized service that allows you to select the product that suits your taste. Its role is increasing in the internet age where mass information exists. The algorithm used to build the recommendation system predicts customer preference and MAE, which is the prediction accuracy of the algorithm's preference, is important because it is related to the reliability of the recommendation system. This study analyzed whether specific customer information is related to MAE and analyzes its influence. The MAE of customer preference predicted by the neighborhood-based collaborative filtering algorithm of the recommendation system is analyzed, and classified as low and high MAE. To investigate whether the group is related to specific customer information and to analyze its influence, this binary logistic regression analysis was conducted to suggest a research model as a factor to improve the prediction accuracy of the recommendation system. In the proposed research model, the number of customer responses and occupation were analyzed as factors related to MAE.

◆ First Author : Sangji University Department of Computer Data Information, biostat9@naver.com, 정회원
° Corresponding Author : Halla University Department. of Information & Communication Software, sokim@halla.com, 종신회원
논문번호 : 201901-429-0-SE, Received January 31, 2019; Revised March 15, 2019; Accepted March 15, 2019

Ⅰ. 서 론

추천시스템은 고객에게 알맞은 상품을 추천하는 시스템으로 빅 데이터 처리를 위한 다양한 분야에서 사용되며, 추천시스템의 성능은 알고리즘에 의해 평가된다[1]. 각 상품에 대하여 고객의 선호도를 정확하게 예측하여, 선호도가 높은 상품을 추출하기에 추천시스템의 알고리즘은 성능 평가에 중요한 핵심이 된다[2,5,6]. 추천시스템의 알고리즘 중에서 협력적 필터링은 고객의 선호도를 예측하며 예측정확도가 높은 것으로 알려져 널리 사용되고 있다. 이 알고리즘은 고객이 구매하지 않은 상품들의 선호도를 예측하기 위해서는 그가 기존에 구매한 상품들에 평가한 선호도 정보들과 다른 고객들이 구매한 상품들에 평가한 선호도 정보들을 기초로 한다[7,8]. 따라서 고객정보량의 크기, 추천하고자 하는 고객과 동일한 상품을 구매한 다른 고객들의 정보량의 크기 등 알고리즘의 예측정확도에 영향을 미치는 요인을 분석하여 예측정확도를 향상 시키려는 다양한 연구가 지속적으로 진행되고 있다[3,4,10].

본 연구는 고객에 대한 정보가 예측정확도에 미치는 영향력 정도를 분석하는데 초점을 두고 다음과 같은 분석을 실행하였다. 첫 번째 예측정확도와 고객정보와의 관련성 여부가 있는지를 분석하고, 두 번째 고객정보가 예측정확도에 미치는 영향력을 연구하고, 세 번째 영향력이 있는 고객정보를 이용하여 추천시스템의 예측 정확도 형상을 위한 연구모형으로 제안하였다.

Ⅱ. 연구방법

본 연구에서는 추천시스템의 알고리즘으로 이웃기반 협력적 필터링을 사용하였으며, 이 알고리즘의 예측 정확도를 알아보기 위한 성능 분석으로 MAE를 이용하였다. 예측정확도에 미치는 영향력과 요인에 대한 분석으로 이분형 로지스틱 회귀분석을 실시하였다.

2.1 이웃기반의 협력적 필터링

이웃기반의 협력적 필터링(neighborhood based collaborative filtering)[11]은 Resnick et al.이 제안한 선호도 예측방법으로 사용자기반의 협력적 필터링(user based collaborative filtering)이라고도 하며 고객 간의 선호도 유사성을 사용한 알고리즘이다. 상품을 추천하고자 하는 고객을 추천대상고객, 그 고객과 동일한 상품을 1개 이상 구매한 경험이 있는 고객을 이웃고객이라 하며, 알고리즘의 수식은 아래와 같고, 이를 간략히 NBCF라 정의한다[11].

$$\hat{U}_x = \overline{U} + \frac{\sum_{j \in raters} (J_x - \overline{J}) r_{uj}}{\sum_{j \in raters} |r_{uj}|} \quad (1)$$

where $\overline{J} = \frac{\sum_{i}^{n} J_i}{n}, \quad i \neq x$

$$r_{uj} = \frac{\sum (U - \overline{U})(J - \overline{J})}{\sqrt{\sum (U - \overline{U})^2 (J - \overline{J})^2}}$$

추천대상고객(u)가 상품 (x)에 예측할 선호도 $\hat{U}_x$를 알아보기 위해, 추천대상고객(u)와 이웃고객(j)들 간의 상품에 평가한 선호도 유사도를 사용한다. r_{uj}는 u와 j간의 상품들에 대한 선호도 유사도를 의미하며, 본 연구에서는 Pearson's 상관계수 (pearson correlation coefficient)를 사용하였다. 여기서, U는 u가 상품에 평가한 선호도이고, $\overline{U}$는 u가 구매한 상품에 평가한 선호도들의 평균이다. J는 j가 상품에 평가한 선호도이고, $\overline{J}$는 u에게 추천할 상품 x를 제외하고 계산된 j가 상품에 평가한 선호도들의 평균이다. n은 j가 선호도를 평가한 개수이며, u에게 추천하고자하는 상품 x의 선호도는 제외하고 평균을 계산하였다.

2.2 MAE(mean absolute error)

추천시스템의 알고리즘을 이용하여 예측값을 계산한 후에 그 예측값에 대한 정확도를 평가하는 방법으로 가장 일반적인 척도인 MAE를 사용하였으며 아래와 같다[9].

$$MAE = \frac{1}{n} \sum_{x}^{n} |R_{ux} - \hat{R}_{ux}| \quad (2)$$

MAE는 추천대상고객(u)가 상품(x)에 평가한 실제 선호도(R_{ux})와 알고리즘으로 예측된 선호도($\hat{R}_{ux}$)의 차이에 대한 평균값이다. 오차가 클수록 그 값은 커지고 예측정확도가 낮다고 평가하며, 오차가 작을수록 그 값은 작아지고 예측정확도가 높다고 평가한다. 여기서 n은 u가 상품에 선호도를 평가한 개수이다.

2.3 이분형 로지스틱 회귀분석

이분형 로지스틱 회귀분석(binary logistic regression analysis)은 질적 또는 양적 설명변수를 이용해 이분형인 반응변수를 설명하는 방법으로, 오즈(odds)에 로그를 취한 값을 종속변수로 사용하였으며, 아래와 같이 정의된다[9].

$$\log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_{2+} \cdots + \beta_n x_n \quad (3)$$

여기서 오즈는 종속변수가 발생할 확률(p)와 발생하지 않을 확률($1-p$)의 비율인 $p/(1-p)$을 의미한다. 종속변수의 범주가 1을 가진 확률로 전환한 식은 아래 식과 같으며, 이를 통해 종속변수의 특정 사건이 발생할 가능성 $P(y=1|x_1,x_2,\cdots,x_n)$을 예측할 수 있다.

$$P(y=1|x_{1,}x_{2,}\cdots,x_n) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_{2+}\cdots+\beta_n x_n)}} \quad (4)$$

본 연구는 이분형 로지스틱 회귀분석을 이용하여 영향력이 있는 요인들을 추출하여 예측정확도 향상을 위한 추출된 영향력 있는 요인으로 추천시스템의 연구모형을 제안하고자 한다.

Ⅲ. 연구자료 및 분포

3.1 연구자료

본 연구의 진행을 위해 Minnesota 대학에 있는 연구소인 GroupLens에서 공개한 MovieLens의 100k 자료를 사용하였다. 100k 자료는 943명의 고객이 1682편의 영화에 5점 척도로 선호도를 평가한 자료로 구성되어 있으며, 총 선호도를 평가한 응답수는 100,000개이다. 그리고 선호도 평가의 개수는 각 고객이 최소 20편의 영화에 평가해야 하는 자료로 구성되어 있다. 고객 943명의 영화 선호도에 응답한 자료를 사용하여 알고리즘 NBCF로 고객별 영화에 대한 선호도를 예측하고, 예측된 선호도의 정확도 평가척도인 MAE를 고객별로 분석한 후 연구 자료로 사용하였다. 연구 자료는 다시 고객별 MAE의 평균값을 기준으로 MAE가 높은 그룹(Group1)과 MAE가 낮은 그룹(Group2)으로 구분하였다. 즉, Group1은 예측정확도가 낮은 그룹이며, Group2는 예측정확도가 높은 그룹이다. 분류된 Group1과 Group2를 정리하면 다음과 같다.

Group1={u_{MAE} > MAE의 평균 | $u_{MAE}\in$ {영화에 선호도를 표시한 고객 u의 MAE} }

Group2={u_{MAE} ≦ MAE의 평균 | $u_{MAE}\in$ {영화에 선호도를 표시한 고객 u의 MAE} }

3.2 연구자료 분포

본 연구에서는 연구 자료의 고객 정보 중 먼저 성별로 분류하였으며, 성별 분포는 다음과 같으며, 남자에 비해 여자가 영화에 대한 응답수가 적음을 알 수 있다(표 1).

표 1. 성별 분포
Table 1. Distribution of gender

Classification	N	%
F	273	29.0%
M	670	71.0%
total	943	100.0%

각 고객별로 나이를 4분위로 나눈 기초 정보는 그림 1과 같다. 나누어진 나이별 그룹은 0-25세인 A1 그룹이 28.8%로 나머지 그룹보다 약간 높게 나타났으며, 26-31세인 A2그룹은 21.3%, 32-43세인 A3그룹은 25.7%, 44-73세인 A4그룹은 24.2%로 조사되었다.

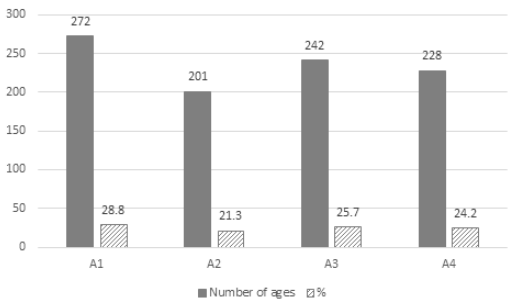

그림 1. 나이 분포
Fig. 1. Distribution of ages

다음은 고객의 정보 중에서 상품에 선호도를 평가한 응답수를 4분위수로 분류하였다. 그 결과는 표 2와 같다. 나누어진 응답수 그룹은 0-33개인 R1그룹, 34-65개인 R2그룹, 66-148개인 R3그룹, 149-737개인 R4그룹으로 분류되었다.

고객의 정보 중에서 직업을 9개의 직업군으로 구분하여 빈도를 조사하였으며, writer(J8)의 직업이 4.8%로 가장 낮게 나타났으며 결과는 표 3과 같다.

표 2. 응답수 분포
Table 2. Distribution of responses

Classification	Number of responses	N	%
R1	0-33	246	26.1
R2	34-65	231	24.5
R3	66-148	232	24.6
R4	149-737	234	24.8
Total		943	100.0

표 3. 직업 분포
Table 3. Distribution of occupations

Classification	Occupation	N	%
J1	administrator	79	8.4
J2	educator	95	10.1
J3	engineer	67	7.1
J4	librarian	51	5.4
J5	other	105	11.1
J6	programmer	66	7.0
J7	student	196	20.8
J8	writer	45	4.8
J9	special occupation	239	25.3
Total		943	100.0

이중에서 special occupation은 고객의 수가 적은 직업들로 구성되었으며, 다음과 같은 직업들로 이루어진 집단이다. artist(28), entertainment(18), doctor(7), none(9), executive(32), healthcare(16), homemaker (7), lawyer(12), marketing(26), retired(14), salesman(12), scientist(31) 그리고 technician(27)로 이루어진 13개의 직업들이 포함되어 있으며 25.34%로 가장 높은 비율을 차지하고 있다. 또한 other의 직업은 Table 3에서 언급된 직업들을 제외한 직업들로 구성되어 있으며 11.13%를 차지하고 있고, student(J7)도 20.78%로 두 번째로 높은 직업군으로 보이고 있다.

Ⅳ. 자료 분석 결과

예측정확도의 평가척도인 MAE와 고객의 정보 중에서 성별과의 연관성을 살펴보기 위해 카이제곱 검정을 실시하였으며, 그 결과는 다음과 같다.

분석 결과 추천시스템의 예측정확도에 성별이 관련이 있음이 나타났으며, 통계적으로 유의한 영향이 있음이 조사되었다. 여자의 경우 예측정확도가 높은 그룹인 Group2는 51.6%로 예측정확도가 낮은 그룹인 Group1의 48.4% 보다 높으며, 남자도 예측정확도가 높은 그룹인 Group2는 58.8%로 Group1보다 높게 나타났다. 또한 성별에 대한 전체적인 자료도 추천시스템의 예측정확도가 높은 그룹인 Group2는 56.7%로 예측정확도가 낮은 그룹인 Group1의 43.3% 보다 높은 비율로 조사되었다.

표 4. 성별과 Groupi의 연관성 분석 (i=1,2)
Table 4. Association analysis for sex and Groupi (i=1,2)

Classification	Group1	Group2	Total	χ^2	p- value
F	48.4% (132)	51.6% (141)	100% (273)	4.048	*.044
M	41.2% (276)	58.8% (394)	100% (670)		
Total	43.3% (408)	56.7% (535)	100% (943)		

*p<0.05

표 5는 나이를 4분위수로 구분하여 추천시스템에서 예측정확도와의 관련성을 조사한 표이며, 예측정확도와 나이는 통계적으로 유의미한 수준으로 서로 관련성이 나타나지 않았다.

또한, 응답수를 4분위수로 그룹화한 Rj (j=1,2,3,4)와 MAE를 그룹화한 Group1과 Group2의 연관성 분석을 실시하였으며, 표 6과 같다.

분석한 결과, 검정통계량(χ^2)이 24.574이고, 유의확률(p-value)이 .000로 나타나, 고객이 평가한 응답수와 NBCF로 예측한 선호도에 대한 고객별 MAE는 서로 관련이 있는 것으로 분석되었다. 추천시스템의 예측정확도가 높은 그룹인 Group2에서 응답수가 R4(149-737개)인 그룹이 가장 높은 67.9%를 보였으

표 5. Ak와 Groupi의 연관성 분석 (k=1,2,3,4, i=1,2)
Table 5. Association analysis for Ak and Groupi (k=1,2, 3,4, i=1,2)

Classification	A1	A2	A3	A4	Total
Group1	45.6% (124)	42.3% (85)	41.7% (101)	43.0% (98)	43.3% (408)
Group2	54.4% (148)	57.7% (116)	58.3% (141)	57.0% (130)	56.7% (535)
Total	100% (272)	100% (201)	100% (242)	100% (228)	100% (943)
χ^2			0.914		
p-value			*.822		

*p>0.1

표 6. Rj와 Groupi와의 연관성 분석 (j=1,2,3,4, i=1,2)
Table 6. Association analysis for Rj and Groupi (j=1,2, 3,4, i=1,2)

Classification	R1	R2	R3	R4	Total
Group1	52.8% (130)	48.1% (111)	39.7% (92)	32.1% (75)	43.3% (408)
Group2	47.2% (116)	51.9% (120)	60.3% (140)	67.9% (159)	56.7% (535)
Total	100% (246)	100% (231)	100% (232)	100% (234)	100% (943)
χ^2			24.574		
p-value			**.000		

**p<0.05

며, 영화에 대한 선호도 응답수를 전체 자료로 보면 추천시스템의 예측정확도가 높은 그룹인 Group2는 56.7%로 예측정확도가 낮은 그룹인 Group1의 43.3% 보다 높은 수치로 나타났다.

예측정확도의 평가척도인 MAE와 고객의 직업이 관련이 있는지 알아보자, 9개의 Jr (r=1,2,…,9)로 그룹화 한 고객의 직업과 MAE를 그룹화한 Groupi (i=1,2)와의 연관성 분석을 실시하였다.

표 7. Jr과 Groupi와의 연관성 분석 (r=1,2,…,9, i=1,2)
Table 7. Association analysis for Jr and Groupi (r=1,2, …,9, i=1,2)

Classification	Group1	Group2	Total	χ^2	p-value
J1	41.8% (33)	58.2% (46)	100% (79)	15.760	**.046
J2	40% (38)	60% (57)	100% (95)		
J3	32.8% (22)	67.2% (45)	100% (67)		
J4	33.3% (17)	66.7% (34)	100% (51)		
J5	46.7% (49)	53.3% (56)	100% (105)		
J6	39.4% (26)	60.6% (40)	100% (66)		
J7	40.8% (80)	59.2% (116)	100% (196)		
J8	62.2% (28)	37.8% (17)	100% (45)		
J9	48.1% (115)	51.9% (124)	100% (239)		
Total	43.3% (408)	56.7% (535)	100% (943)		

**p<0.05

그 결과, 표 7에서와 같이 검정통계량(χ^2)이 15.760이고, 유의확률(p-value)이 .046로 나타나, 고객의 직업과 NBCF로 예측한 선호도에 대한 고객별 MAE는 서로관련이 있는 것으로 분석되었다. 추천시스템의 예측정확도가 높은 그룹인 Group2에서 J3(engineer)이 가장 높은 67.2%를 보였으며, 직업에 대한 전체 자료로 보면 추천시스템의 예측정확도가 높은 그룹인 Group2는 56.7%로 예측정확도가 낮은 그룹인 Group1의 43.3% 보다 높은 수치로 나타났다.

추천시스템에서 고객들의 성별과 고객이 평가한 응답수와 고객의 직업은 추천시스템의 예측 성능에 관련이 있음이 분석되었다. 특히 성별에서는 남성이 여성보다 예측정확도가 높은 Group2에 많은 분포를 보였으며, 응답수가 많을수록 예측정확도가 높은 그룹에 속할 수 있다는 결과가 나왔다.

따라서 성별과 직업(Jr, r=1,2,…,9)과 응답수(Rj, j=1,2,3,4)가 MAE를 분류한 Groupi (i=1,2)에 얼마만큼 영향력이 있는지를 분석하고자 한다. 먼저 종속변수를 Groupi (i=1,2)로 하고 각각의 그룹을 0(=Group1), 1(=Group2)로 구분하였다. Group1은 예측정확도가 낮은 집단이고 Group2은 예측정확도가 높은 집단으로 분류되었고 범주형 자료인 성별, 직업(Jr, r=1,2,…,9)과 응답수(Rj, j=1,2,3,4)을 독립변수로 하여 영향력을 알아보기 위해 이분형 로지스틱 회귀분석을 실시하였다.

표 8은 성별과 영화 응답수와 직업에 대한 MAE를 결정하는 모형에 대한 예측 분류표이다.

먼저 종속변수인 Group1과 Group2를 올바르게 분류할 확률인 모형 적중률은 전체적으로 60.1%로 조사되었다. 추천시스템의 예측정확도를 분류한 집단인 Group1과 Group2의 집단별 예측정확도는 각각 35.3%와 79.1%임을 알 수 있다. 추천시스템의 예측정확도가 높은 그룹인 Group2가 79.1%로 낮은 그룹인 Group1가 35.3%보다 높게 나타났다.

표 8. 분류표
Table 8. Classification table

observed	Predicted		Percentage correct
	Group1	Group2	
Group1	144	264	35.3
Group2	112	423	79.1
Overall percentage			60.1

또한, 표 9를 살펴보면 MAE를 결정하는 요인인 성별과 직업 그리고 영화에 평가한 응답수에 대한 모형의 설명력은 Nagelkerke 5.9%로 나타났으며, 모형의 적합도 검정인 Hosmer와 Lemeshow의 검정통계량(χ^2)은 3.045이고 유의확률(p-value)은 .931으로 추정된 모형은 적합하다고 판단된다.

표 9. 성별, 응답수, 그리고 직업의 -2Log 우도비
Table 9. -2Log Likeliwood ratio test of sex, responses and occupations

-2Log		1247.478
Nagelkerke R^2		.059
Hosmer & Lemeshow	χ^2	3.045
	p-value	.931

추천시스템의 예측정확도에 영향을 미치는 요인으로는 응답수와 직업이 분석 결과로 조사되었다(표 10). 성별에서 여자는 예측정확도와 음의 영향을 미치지만 통계적으로 충분하지 않다고 나타났다. 그리고 직업도 J4(librarian)만 예측정확도에 영향을 미치는 요인이며 나머지 직업군은 통계적으로 충분하지 않는 요인으로 분석되었다. MAE에 미치는 영향이 통계적으로 유의하다고 조사된 직업인 사서(J4)는 1.955배로 가장 많은 영향력이 미치는 것으로 나타났다. 0-33개의 응답수인 R1과 34-65개의 응답수인 R2가 유의확률(p-value)이 .000, 과 .001로 나타나 추천시스템의 예측정확도에 미치는 요인으로 분석되었으며, 응답수가 가장 적은 그룹인 R1은 0.422배, R2는 0.520배로 MAE에 통계적으로 유의한 영향을 미치는 것으로 나타났다.

표 10. 성별, 응답수, 그리고 직업에 의한 예측 모델
Table 10. Prediction model by number of sex, responses and occupations

Classification	B	p-value	Exp(B)
F	-.219	.156	.804
R1	-.863	*.000	.422
R2	-.653	*.001	.520
R3	-.332	.091	.718
J1	.340	.207	1.405
J2	.356	.154	1.428
J3	.513	.083	1.671
J4	.671	*.045	1.955
J5	.032	.895	1.032
J6	.244	.398	1.276
J7	.262	.186	1.300
J8	-.636	.061	.529
Constant	.632	*.001	1.881

*p<0.05

따라서 추천시스템의 MAE에 미치는 영향력의 정도는 다음과 같은 식으로 제안할 수 있다.

$$\ln\left(\frac{p(Group2)}{1-p(Group2)}\right)$$

$$= 0.422 \times R1 + 0.520 \times R2 + 1.955 \times J4 + 1.881$$

V. 결 론

추천시스템은 인터넷을 사용하는 시대에 많이 사용하는 시스템 중에 하나이다. 그중 추천시스템의 예측선호도에 대한 MAE는 추천시스템의 신뢰도와 관련이 있으므로 매우 중요하다. 본 연구는 943명의 고객들이 영화에 평가한 선호도 응답 자료를 바탕으로 영화에 대한 선호도를 예측하였다. 예측된 선호도의 정확도는 MAE로 측정하고, 예측 선호도의 정확도가 낮은 그룹인 Group1과 높은 그룹 Group2로 구분하였으며, Group2는 MAE가 적은 그룹으로 예측정확도가 높아 추천시스템의 성능향상에 영향력이 있는 그룹이다. 이 그룹의 특성을 분석하기 위해 고객정보 중에서 성별과 나이와 직업 그리고 영화에 대한 응답수를 분석하였다.

분석결과 성별과 직업과 응답수가 MAE와 관련이 있으며, 이중에서 직업과 응답수는 추천시스템의 예측정확도에 영향을 미치는 요인으로 분석되었으며 성별은 MAE에 영향을 미치는 요인으로 충분하지 못하다는 결과를 도출했다. 특히, 추천을 실행할 때 사서(J4)는 1.955배로 가장 높은 영향력을 가지고 있는 것으로 나타났으며, 영화에 대한 응답수가 적을수록 그 영향력이 적음이 조사되었다. 하지만 응답수가 66개 이상인 경우에는 MAE에 영향을 미치는 요인으로 통계적으로 충분하지 못하다는 결과를 도출했다. 향후 연구는 응답수와 직업을 미리 알았을 경우 제안한 연구 모형을 추천시스템에 적용하여 자료의 희소성문제에 응답수와 직업군을 이용한 성능 향상을 연구과제로 진행할 예정이다.

References

[1] J. W. Kim and G. H. Park, "Personalized group recommendation using collaborative filtering and frequent pattern," *J. KICS*, vol. 41, no. 7, pp. 768-774, Jul. 2016.

[2] S. O. Kim and J. H. Koo, "A study on the model based on the number of responses and rank fitting accuracy in collaborative filtering," *J. Korean Data Anal. Soc.*, vol. 19, no. 4(B), pp. 1907-1915, Aug. 2017.

[3] Y. A. Kim and G. W. Park, "An efficient extended query suggestion system using the analysis of users' query patterns," *J. KICS*, vol. 37C, no. 7, pp. 619-626, Jul. 2012.

[4] Y. J. Kim, D. J. Shin, W. Y. Shin, and C. H. Hang, "Rating information-aided denoising autoEncoder for effective collaborative filtering," *J. KICS*, vol. 43, no. 8, pp. 1357-1367, Aug. 2018.

[5] J. H. Koo, "A study on collaborative filtering in responding to fluid responses for recommend product ranking," *J. Korean Data Anal. Soc.*, vol. 20, no. 4, pp. 1873-1882, Aug. 2018.

[6] J. H. Koo, A. R. Choi, S. O. Kim, and H. C. Lee, "An alternative method of missing data for improving rank fitting in collaborative filtering," *J. Korean Data Anal. Soc.*, vol. 18, no. 1(B), pp. 207-216, Feb. 2016.

[7] J. H. Koo, J. W. Park, and H. S. Choi, "The effect of customers' experience of diverse goods and selection of popular commodity on recommendation system," *J. Korean Data Anal. Soc.*, vol. 17, no. 6(B), pp. 3097-3106, Dec. 2015.

[8] H. C. Lee, "Improved algorithm for user based recommender system," *J. Korean Data & Inf. Sci.*, vol. 17, no. 3, pp. 717-726, Sep. 2006.

[9] H. Y. Lee, *Research methodology*, Cheongram, pp. 694-714, 2012.

[10] J. H. Moon, Y. H. Jang, Y. C. Choi, J. G. Kim, and J. C. Park, "Case study of big data-Based agri-food recommendation system according to types of customers," *J. KICS*, vol. 40, no. 5, pp. 903-913, May 2015.

[11] P. Resnick, N. J. Iacovou, M. Silvestret, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, pp. 175-186, Chapel Hill, North Carolina, USA, Oct. 1994.

구 지 현 (Jee Hyun Koo)

1995년 2월 : 상지대학교 통계학부

1997년 8월 : 숙명여자대학교 경영 석사

2014년 8월 : 한국외국어대학교 경영과학 박사

<관심분야> 추천시스템, 빅데이터

[ORCID:0000-0001-9856-2498]

김 선 옥 (Sun Ok Kim)

1988년 2월 : 덕성여자대학교 이학 학사

1991년 2월 : 서강대학교 이학 석사

1997년 2월 : 서강대학교 이학 박사

<관심분야>추천시스템, 암호학

[ORCID:0000-0002-9665-4214]