

사용자 기반의 협력필터링을 위한 퍼지 논리를 이용한 새로운 유사도 척도

이수정[†]

요 약

협력 필터링은 다수의 상업용 추천 시스템에서 구현되어 온라인 사용자들에게 성공적으로 서비스되고 있는 핵심적 기술이다. 이 기술은 현 사용자와 유사한 평가이력을 가진 다른 사용자들로부터 항목을 추천하기 때문에, 유사도 척도는 시스템 성능에 매우 큰 영향을 미친다. 본 연구에서는 기존 유사도 측정 방법의 문제점을 해결하고자 퍼지 논리에 입각하여 사용자 평가등급의 주관성 및 모호성과 사용자들의 평가 행태를 반영하는 새로운 유사도 척도를 제안한다. 성능 평가를 위한 다양한 실험을 실시하였고, 그 결과 제안 방법은 예측 정확도와 추천 정확도 면에서 우수한 성능 개선 효과를 보였다.

주제어 : 추천 시스템, 협력 필터링, 유사도 척도, 퍼지 논리, 사용자 기반 협력필터링

A New Similarity Measure using Fuzzy Logic for User-based Collaborative Filtering

Soojung Lee[†]

ABSTRACT

Collaborative filtering is a fundamental technique implemented in many commercial recommender systems and provides a successful service to online users. This technique recommends items by referring to other users who have similar rating records to the current user. Hence, similarity measures critically affect the system performance. This study addresses problems of previous similarity measures and suggests a new similarity measure. The proposed measure reflects the subjectivity or vagueness of user ratings and the users' rating behavior by using fuzzy logic. We conduct experimental studies for performance evaluation, whose results show that the proposed measure demonstrates outstanding performance improvements in terms of prediction accuracy and recommendation accuracy.

Keywords : Recommender System, Collaborative Filtering, Similarity Measure, Fuzzy Logic, User-based Collaborative Filtering

[†] 정 회 원: 경인교육대학교 컴퓨터교육과 교수
논문접수: 2018년 7월 30일, 심사완료: 2018년 9월 16일, 게재확정: 2018년 9월 19일

1. 서론

협력 필터링 시스템(collaborative filtering, CF)은 주로 온라인 상에서 사용자들에게 항목을 추천하기 위해 개발된 추천시스템(recommender system)의 주요 기술이다[1]. 항목이란 시스템에서 취급하는 소재에 따라 다양할 수 있는데, 서적, 영화, 음악 등이 그 예이다. 대표적인 상업용 추천 시스템으로서 아마존, Barnes&Noble 등이 있다.

협력 필터링 시스템은 크게 메모리 기반과 모델 기반으로 분류된다. 모델 기반 시스템은 사용자의 항목 평가 데이터로부터 모델을 구축하는데 이를 위해 다양한 기계학습 기술이 사용될 수 있다[1][2]. 본 연구에서는 구현이 보다 용이하다는 장점을 지닌 메모리 기반 시스템에 국한한다. 메모리 기반 CF의 기본 원리는 현 사용자가 미평가한 항목들에 대한 선호도를 예측하여 높은 예측치를 가진 항목들을 추천하는 것이다. 사용자의 항목에 대한 선호 여부는 시스템 상에서 직접 혹은 간접적으로 표현되는데, 직접적인 방법으로는 평가등급의 부여가 있고, 간접적인 방법으로는 항목 구매, 클릭 여부 등으로부터 선호도를 파악할 수 있다[1]. 결과적으로 메모리 기반 CF에서는 <표 1>과 같은 사용자-항목 매트릭스를 유지 관리하는데, 수치는 해당 항목에 대한 사용자의 선호 강도를 나타낸다.

메모리 기반 CF는 다시 사용자 기반과 항목 기반으로 분류할 수 있다. 사용자 기반 CF는 현 사용자와 유사한 선호 이력을 가진 다른 사용자들이 선호했던 항목들을 추천하는 것이다. 반면에 항목 기반 CF는 현 사용자가 선호했던 항목들과 유사한 항목들을 추천한다. 본 연구에서는 사용자 기반 CF 시스템에서의 유사도 측정 방법에 초점을 두고 있으나,, 항목 기반 CF에 대해서도 동일한 원리를 적용할 수 있다.

이와 같이 사용자 기반의 CF 시스템은 유사한 사용자들을 제대로 선정하는데에 그 성공 여부가 달려 있다고 해도 과언이 아니다. 따라서, 이제까지 다양한 유사도 측정 방식들이 심도 있게 개발되었다. 그러나, 이들 방식들은 대개 평가등급 수치 자체를 활용하기 때문에 평가등급에 내재된

사용자의 주관성이나 모호성을 고려하지 않는다는 문제점이 있다[3]. 또한, 각 항목에 대한 평가등급이 독립적으로 취급되어 전체 사용자들의 평가 행태를 반영하지 못한다[4][5]. 본 연구에서는 이러한 문제점들을 개선한 새로운 유사도 측정 방식을 제안한다. 퍼지 이론에 입각하여 평가등급의 퍼지값을 구하고 이를 이용한 유사도값을 산출한다. 또한, 각 항목을 평가한 전체 사용자들의 평가등급을 퍼지화하여 유사도를 산출하기 위한 가중치로 활용한다. 실험을 통하여 다양한 퍼지 함수에 대한 성능 측정을 실시하였고, 기존의 유사도 척도들과 비교하였으며, 그 결과, 제안 척도의 우수성을 입증하였다.

논문의 구성은 다음과 같다. 2절에서는 관련 연구를 기술하고, 3절에서 본 연구에서 제안하는 방법을 설명하며 4절에서 실험을 통한 성능을 입증하고 5절에서 논문의 결론을 맺는다.

<표 1> 사용자-항목 매트릭스의 예

item \ user	i1	i2	i3	i4	i5
u1	1	2	1	-	2
u2	-	2	3	-	-
u3	-	3	3	4	4

2. 관련 연구 분석

협력 필터링 시스템에서는 항목에 대한 평가기록이 유사한 사용자들은 향후 다른 항목에 대한 평가도 유사하게 부여할 것이라는 가정에 근거를 둔다. 따라서, 현 사용자가 미평가한 항목 x의 예상 평가치를 산출하기 위해서, 우선 모든 다른 사용자들을 대상으로 평가기록을 참조하여 현 사용자와의 유사도를 측정한다. 현 사용자 u의 인접 이웃 사용자들(Nearest Neighbors, NN)로부터 항목 x에 대한 평가치를 참조하여 x에 대한 예상 평가치를 구하는데, 인접이웃이란 미리 정해진 한계점을 넘는 유사도값을 가진 사용자들 또는 시스템 매개변수로서 미리 정해진 수의 사용자들을 일컫는다[6]. 평가치 예측값은 대개 다음의 계산식을 토대로 구한다[7].

$$\hat{r}_{u,x} = \bar{r}_u + \frac{\sum_{v \in NN_{u,x}} sim(u,v) \times (r_{v,x} - \bar{r}_v)}{\sum_{v \in NN_{u,x}} |sim(u,v)|}$$

$\hat{r}_{u,x}$: 사용자 u의 항목 x에 대한 예측 평가등급

\bar{r}_u : u의 모든 평가치들의 평균

$NN_{u,x}$: x에 대한 평가등급이 있는 u의 인접이웃 집합

$sim(u,v)$: 사용자 u와 v 간의 유사도값

$r_{v,x}$: 사용자 v의 x에 대한 평가등급

이상과 같이 현 사용자 u가 미평가한 모든 항목들에 대해 예측 평가치를 구한 후에, 가장 높은 예측치를 가진 항목들 순으로 추천이 이루어진다.

어떠한 사용자들이 인접이웃 집합에 속하는지에 따라 평가 예측치가 달라지기 때문에, 이에 따라 추천의 성능이 좌우된다. 따라서, 인접이웃을 결정하는 유사도 척도는 협력필터링 시스템에서 매우 중요한 요소로서, 다양한 유사도 척도들이 개발되었다[8][9]. 전통적인 유사도 척도로서 피어슨 상관도(Pearson correlation), 코사인 유사도(cosine similarity), 평균자승차이(mean squared differences) 등이 대표적이다. 또한, 상관도 기반의 변형 방식으로서, 제한 피어슨 상관도(Constrained Pearson Correlation), 스피어만 순위 상관도(Spearman Rank Correlation), 켄달의 타우(Kendall's τ) 상관도 등이 개발되었고, 사용자의 서로 다른 평가등급 스케일을 반영하기 위한 보정 코사인 유사도(Adjusted Cosine Similarity)가 개발되었다[1][9].

이러한 유사도 척도들은 사용자의 명확한 평가 등급을 기반으로 하기 때문에, 평가등급의 주관성과 모호성은 고려하지 않는다. 이러한 문제를 해결하기 위하여 퍼지 이론을 접목한 유사도 척도 개발 연구가 진행되었다[3][10][11][12]. Al-Shamri와 Al-Ashwal는 평가치와 편차를 퍼지값으로 변환하여 기존 유사도 척도와 통합한 유사도 척도를 제안하였고[3], Son은 사용자의 인구통계학적 특성을 활용한 퍼지 유사도 값을 피어슨 상관도와 통합하여 새로운 유사도 척도를 제안하였다[12]. 이들 연구 결과는 모두 기존 유사도 척도와 통합된 방식이나, 이와는 다르게 퍼지화된 평가값만을 이용하여 유사도를 계산한 결

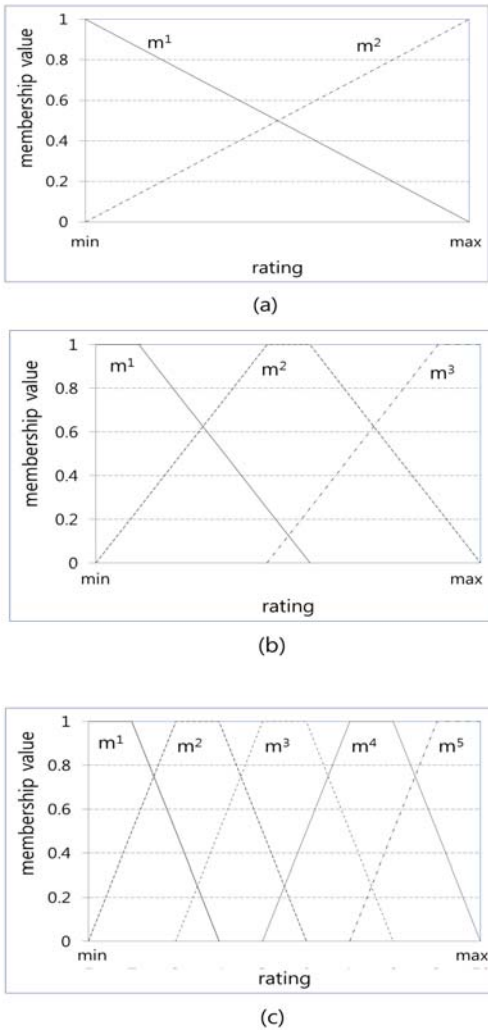
과 기존 척도에 비해 우수한 성능을 나타낸다는 연구결과도 발표되었다[13].

기존 유사도 척도들의 또다른 문제점은 각각의 평가등급을 독립적으로 취급하기 때문에, 사용자들의 전반적인 평가 행태를 고려하지 않는다는 점이다. 일부 연구 결과에서는 Shannon이 제안한 엔트로피(entropy)[14]를 활용하여 추가적인 정보를 개발하고 이를 기존의 유사도 척도에 병합함으로써 이러한 문제점을 개선하려 하였다. [5]의 방법은 두 사용자의 평가등급 엔트로피 차이값을 전통적 유사도 척도에 통합하여 유사도를 계산하였고, [15]은 사용자 평가등급 차이의 엔트로피를 피어슨 상관도와 접목하였다. 그러나, 이러한 연구들이 성능 개선에 도움이 되었을지라도, 각 사용자 중심의 방법이므로, 대개의 실제 시스템에서 처럼 사용자의 평가개수가 적을 때는 효과를 발휘하지 못한다. 사용자 중심이 아닌 각 항목에 대한 평가치들을 기준으로 하는 추가 정보를 정의하고 이를 평균자승차이 유사도 척도에 통합한 Bobadilla 외 2인[4]의 연구 결과가 있으나, 이 연구에서 제안한 특이성(singularity) 개념은 미리 정한 한계치에 의해 고평가등급과 저평가등급을 결정하기 때문에 이 한계치에 따라 시스템 성능이 좌우되며, 각 항목별 평가등급 데이터의 분포 특성을 반영하지 않는다는 문제점이 있다.

3. 제안 유사도 척도

본 연구에서는 기존 유사도 척도들의 두 가지 중요한 문제점들을 동시에 다루고자 한다. 첫째는 사용자 평가등급의 주관성 및 모호성 문제이며, 둘째는 각 항목을 평가한 전체 사용자들의 평가 행태를 반영하는 문제이다. 본 연구의 제안 척도에서는 이들 문제의 해결을 위해 퍼지 이론을 이용한다.

우선 첫번째 문제에 대한 해결 방안으로서, 각 사용자 평가등급을 퍼지값으로 변환하여 유사도를 산출한다. [그림 1]은 다음 절의 실험에서 사용할 세 가지 퍼지함수를 제시한다. min과 max는 시스템에서 허용한 평가등급의 최소값과 최대값이다. 평가등급을 r이라 할 때, [그림 1](a)에 따르면, r에 대응하는 퍼지값은 $\langle m^1(r), m^2(r) \rangle$ 로



[그림 1] 제안 유사도 척도에서 사용한 퍼지 함수들

표기하고, 이는 각각 r의 낮은 정도와 높은 정도를 나타내며, $m^1(r)=(max-r)/(max-min)$, $m^2(r)=(r-min)/(max-min)$ 이다. [그림 1]의 나머지 퍼지 함수들을 사용할 때도 마찬가지로 변환한다.

제안 방법은 평균자승차이(mean squared differences, MSD)를 기본으로 하여 퍼지화된 평가등급들 간의 유사도를 측정한다. 우선 MSD에 대해 설명하자면, $r_{u,i}$ 를 사용자 u의 항목 i에 대한 평가등급이라고 하고, $I_{u,v}$ 를 두 사용자 u와 v의 공통평가항목집합이라고 할 때, 두 사용자 간 MSD는 다음과 같이 계산된다.

$$MSD(u, v) = 1 - \frac{1}{|I_{u,v}|} \sum_{i \in I_{u,v}} (r'_{u,i} - r'_{v,i})^2 \dots\dots(1)$$

이 때 $r'_{u,i}$ 는 $r_{u,i}$ 를 [0, 1] 사이의 값으로 변환

한 값으로서 $r'_{u,i} = (r_{u,i} - min)/(max - min)$ 으로 계산한다.

제안 방법은 원 평가등급 대신에 퍼지값을 이용하는 것이므로, 다음과 같이 산출 공식을 제안한다.

$$MSD_F(u, v) = 1 - \frac{1}{k|I_{u,v}|} \sum_{i \in I_{u,v}} \sum_{j=1}^k (m^j(r_{u,i}) - m^j(r_{v,i}))^2 \dots\dots\dots(2)$$

위 식에서 $r_{u,i}$ 의 퍼지소속값들은 $\langle m^1(r_{u,i}), \dots, m^k(r_{u,i}) \rangle$ 이며, k는 퍼지집합개수이고,

$\frac{1}{k} \sum_{j=1}^k (m^j(r_{u,i}) - m^j(r_{v,i}))^2$ 는 각 퍼지집합 소속값의 차이 자승의 평균을 의미한다.

이제 두 번째 문제에 대해 언급하자면, 특정 공통평가항목 x와 y에 대한 두 사용자의 평가등급 차이가 동일하다고 할지라도, 각 항목을 평가한 모든 사용자들의 평가등급 범위가 다르다면 차이가 동일하지 않은 것으로 취급하고자 한다. 예를 들어, 사용자 u의 두 항목에 대한 평가등급이 모두 3이고, 사용자 v는 모두 5라고 하자. 항목 x에 대한 모든 평가등급의 범위는 [1, 5]라고 하고, 항목 y에 대해서는 [1, 10]이라고 가정한다면, 두 사용자 u와 v의 평가등급 간의 차이는 두 항목에 대해서 동일한 값으로서 2이지만, 항목 x에 대해서 두 사용자의 평가등급 차이가 상대적으로 더욱 크다고 볼 수 있다.

이러한 상대적인 차이를 계상하기 위하여, 우선 각 사용자의 평가등급을 해당 항목의 모든 평가등급에 대한 상대적인 퍼지값으로 변환하고, 유사도 산출 대상의 두 사용자의 모든 공통평가항목들에 대해서 이와 같이 구한 퍼지값들의 평균차이를 가중치로 정의한다. 마지막으로, 식 (2)에 이러한 가중치를 부과하여 최종적인 유사도값을 산출한다.

구체적으로, 항목 i에 대한 $r_{u,i}$ 의 퍼지소속값들을 $\langle m^1(r_{u,i}), \dots, m^k(r_{u,i}) \rangle$ 로 표기하고, $r_{i,max}$ 과 $r_{i,min}$ 을 항목 i에 대한 각각 최대와 최소 평가등급이라고 하자. k가 2일 때, [그림 1](a)과 같은 두 개의 퍼지집합을 사용할 때, min과 max 대신에 $r_{i,min}$ 와 $r_{i,max}$ 로 각각 대체한다. 즉, $m^1(r_{u,i}) = (r_{i,max} - r_{u,i}) / (r_{i,max} - r_{i,min})$ 이고

$m^2(r_{u,i}) = (r_{u,i} - r_{i,\min}) / (r_{i,\max} - r_{i,\min})$ 이다. 이 와 같이 변환된 두 사용자 u 와 v 의 퍼지 평가등 급 간에 차이가 클수록 큰 가중치를 갖도록 아래 와 같이 가중치를 정의한다.

$$w_i(u, v) = \frac{1}{k} \sum_{j=1}^k |m^j(r_{u,i}) - m^j(r_{v,i})| \dots \dots \dots (3)$$

큰 가중치는 작은 유사도값을 산출하게 되도록 MSD_F 공식에 가중치를 반영한다. 최종적으로 가 중치를 반영한 사용자 u 와 v 를 위한 유사도값 산 출 공식을 아래와 같이 제안한다.

$$MSD_{FF}(u, v) = 1 - \frac{1}{k|I_{u,v}|} \sum_{i \in I_{u,v}} w_i(u, v) \sum_{j=1}^k (m^j(r_{u,i}) - m^j(r_{v,i}))^2 \dots \dots \dots (4)$$

4. 성능 측정 실험

4.1 실험 배경

제안 방법의 성능 측정을 위해 관련 연구에서 널리 사용되는 공개 데이터셋인 Jester를 활용한 다. 이 데이터셋의 허용 평가범위는 -10 부터 +10 사이의 실수값이며, 사용자수가 998명, 항목 수가 100, 각 사용자당 평가개수는 36~100개인 특 성을 지녔다. 대개의 다른 데이터셋은 1부터 5까 지의 정수 평가치를 허용하나, Jester는 매우 넓은 범위를 사용하므로, 사용자의 평가 행태의 차이가 쉽게 드러날 수 있다.

데이터셋의 전체 데이터 중에서 80%를 활용하 여 각 사용자의 인접이웃집합을 구하였고, 나머지 20%는 다양한 성능 척도의 결과를 산출하는데 사 용하였다. 대개 협력 필터링 시스템의 성능은 크 게 두 가지 측면, 즉, 예측 정확도와 추천 정확도 로서 평가된다[16]. 본 연구에서는 예측 정확도의 대표적인 평가 척도로서 관련 연구에서 주로 사 용하는 MAE (Mean Absolute Error)[7]를 도입하 였고 이 척도는 다음과 같이 정의된다.

$$MAE = \frac{\sum_x |r_{u,x} - \hat{r}_{u,x}|}{N}$$

$r_{u,x}$: 항목 x 에 대해 사용자 u 가 부여한 등급

$\hat{r}_{u,x}$: 항목 x 에 대한 사용자 u 의 등급 예측값

N : 성능 평가 대상 항목의 총개수

예측 정확도 측정을 위한 또다른 척도로서 RMSE(Root Mean Squared Error)를 도입하였다. 이 척도는 넷플릭스 대회[6]를 계기로 많이 활용 되어 왔는데, MAE보다 예러 크기를 더욱 증폭 반영하며 다음과 같이 정의한다.

$$RMSE = \sqrt{\frac{1}{N} \sum_x (r_{u,x} - \hat{r}_{u,x})^2}$$

추천 정확도의 측정을 위해서는 문헌에서 주로 사용되는 정밀도(precision)와 재현율(recall)을 도 입하였다. 정밀도는 모든 추천 항목들 중에서 사 용자가 선호하는 항목들 개수의 비율이며, 재현율 은 모든 사용자의 선호 항목들 중 실제로 추천된 항목들 개수의 비율이다. 사용자의 선호 여부를 결정하는 한계값은 4로 선정하였다. 즉, 항목 평 가값이 4 이상이면 사용자가 선호하는 것으로 가 정하였다.

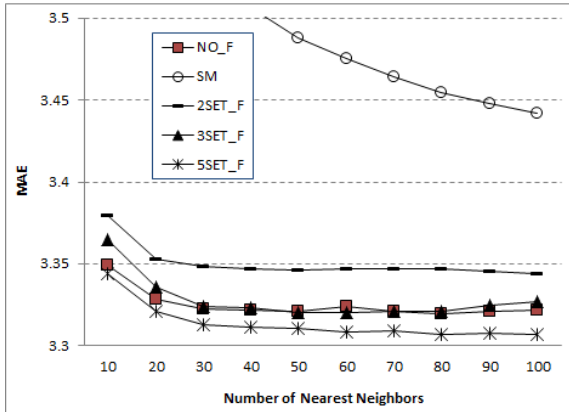
실험에 사용된 유사도 척도는 모두 다섯 가지 로서, 다음과 같다.

1. 'NO_F': 식 (1)의 MSD
2. '2SET_F': 식 (4)에 [그림 1](a)의 퍼지소속 값을 적용 (식 (3)에도 [그림 1](a) 적용).
3. '3SET_F': 식 (4)에 [그림 1](b)의 퍼지소속 값을 적용 (식 (3)에도 [그림 1](b) 적용).
4. '5SET_F': 식 (4)에 [그림 1](c)의 퍼지소속 값을 적용 (식 (3)에도 [그림 1](c) 적용).
5. 'SM': [4]에서 제안한 척도.

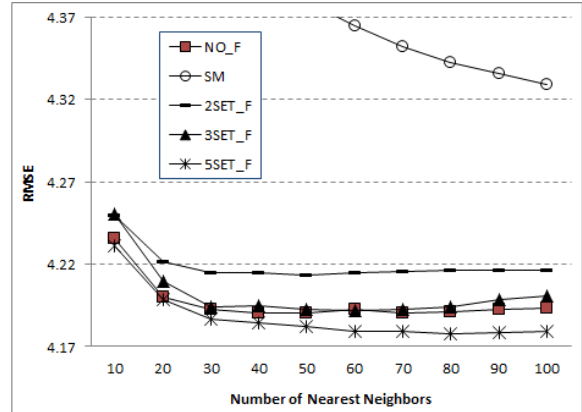
SM 척도는 각 항목에 대한 사용자들의 평가 행태를 반영한 혼치 않은 연구 결과이므로, 본 연 구의 제안 방법과 성능 비교를 위하여 포함하였 다.

4.2 실험 결과

[그림 2]는 인접이웃수의 변화에 따른 예측 정 확도를 보여준다. 미평가 항목에 대한 평가치를 참조할 인접이웃수가 증가함에 따라 예러 크기는 감소하며, 대략 30명 이상에 대해서 MAE와 RMSE 모두 안정적인 결과를 보임을 알 수 있다. 매우 특이한 점은 SM의 결과인데, SM은 MSD를 기반으로 하고 여기에 특이성(singularity) 값을



(a)



(b)

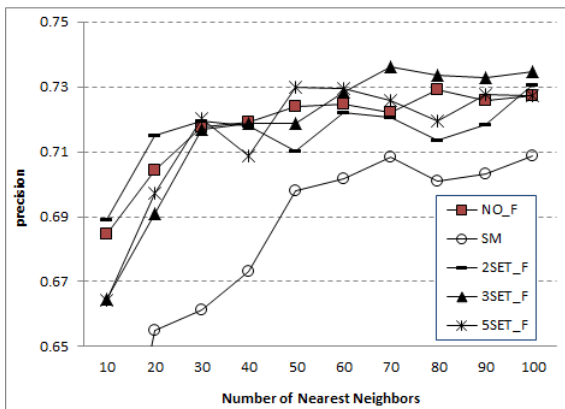
[그림 2] 유사도 척도들의 예측 정확도 비교: (a) Mean absolute error (b) Root mean squared error

가중치로 통합하여 유사도를 산출한다. 즉, MSD의 부족한 점을 보완하고자 개발된 것임에도 불구하고, [그림 2]에서 보이듯이 MSD 자체(NO_F)보다도 현저히 저조한 성능 결과를 나타냈다. 그 이유는 Jester 데이터셋은 매우 밀집되어 있어, 각 항목에 대한 평가개수가 322~998로 상당히 많으므로, 평가치의 특이성 개념이 유효하게 작용하지 않기 때문인 것으로 판단된다. 또한 MAE 보다 RMSE에서 다른 유사도 척도들보다 상대적으로 더욱 저조한 성능을 보이므로, SM을 통해 산출되는 개별적인 에러 크기가 더욱 크다는 것을 알 수 있다.

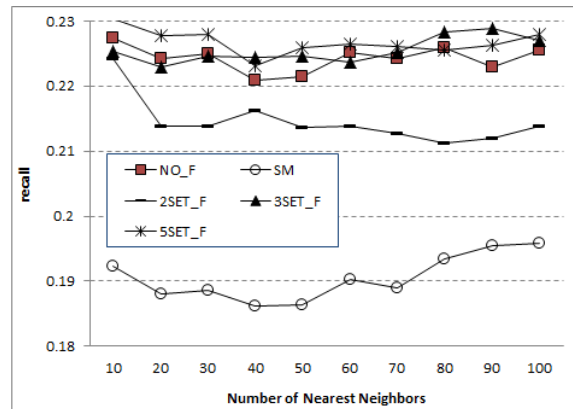
[그림 2]에서 제안 방법의 성능은 두 개의 퍼지 집합을 사용했을 때 원래의 방법(NO_F)보다 낮았

으나, 퍼지 집합 개수를 증가시키에 따라 성능이 좋아짐을 확인할 수 있다. 즉, 3SET_F의 성능은 NO_F와 거의 동일하였고, 5SET_F의 성능은 가장 우수함을 보였다. 이는 Jester 데이터셋의 평가 범위가 20으로서 매우 크므로, [그림 1](c)와 같이 많은 수의 퍼지집합을 사용하여 평가등급을 변환하는 것이 두 평가등급 간의 차이를 더욱 정밀하게 표현할 수 있기 때문인 것으로 분석된다.

[그림 3]은 각 유사도 척도를 사용한 협력 필터링 시스템의 추천 성능을 표시한다. 정밀도와 재현율은 정의에 따라 서로 역의 관계를 가지므로 [그림 3]에서 정밀도가 증가함에 따라 재현율은 감소하는 경향을 보인다. 예측 정확도의 결과와 마찬가지로 SM의 결과가 가장 저조하나, 다른 유



(a)



(b)

[그림 3] 유사도 척도들의 추천 정확도 비교: (a) 정밀도 (b) 재현율

사도 척도들에 비해 인접이웃 수의 증가에 따라 성능 개선의 속도가 다소 가파르게 진행되는 경향을 볼 수 있다. 이는 정밀도와 재현율 모두에서 확인된다.

다른 유사도 척도들의 추천 성능을 살펴보면 예측 성능 결과와 다소 다른 양상을 보이는데, 인접이웃수가 80을 넘을 때 3SET_F의 성능이 대체로 미미하지만 가장 우수하고, 5SET_F와 NO_F는 전반적으로 뚜렷한 성능 차이를 나타내지 않는다. 이는 추천 성능의 측정을 위해서는 한계값인 4보다 크기만 하면 사용자가 선호하는 것으로 간주하므로, 예측 정확도에서와 같은 정밀한 측정을 요구하지 않기 때문에 퍼지 집합 개수의 차이가 큰 영향을 주지 않는 것으로 판단된다. 그러나, 퍼지 집합 개수가 매우 적은 수인 둘인 경우에는 2SET_F의 결과에서 볼 수 있듯이 3SET_F나 5SET_F보다 현저히 저하된 성능을 나타내는 것으로 확인되었다.

5. 결론

본 연구에서는 협력 필터링 시스템을 위한 새로운 유사도 척도를 제안하였다. 제안 방법은 기존 방법의 두 가지 주요한 문제점을 모두 다루었다는 점에서 차별화된다. 첫째는 사용자 평가등급의 주관성 및 모호성이며, 둘째는 전체 사용자의 평가 행태를 반영한 것이다. 퍼지 이론을 활용하여 새로운 유사도 척도를 개발하였고, 다양한 퍼지 함수에 대해 실험한 결과 예측 정확도와 추천 정확도 측면 모두에서 우수한 성능을 나타냈다.

실험에 사용한 퍼지 함수는 모두 세 종류로서 최적의 성능 도출을 목표로 하여 정의된 것은 아니므로, 향후 또다른 형태의 퍼지 함수를 활용하여 성능 조사를 실시하는 연구가 필요할 수 있다. 또한 서로 다른 퍼지 함수를 복합적으로 적용한 유사도 척도에 대한 실험도 필요하다고 본다. 마지막으로, 본문에서 사용한 데이터셋과는 다른 특성을 가진 데이터셋들을 사용한 실험도 향후 수행할 계획이다.

참 고 문 헌

- [1] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence, 2009*.
- [2] 이수정 (2016). 사용자 기반의 협력필터링 시스템을 위한 유사도 측정의 최적화. **컴퓨터 교육학회논문지**, 19(1), 111-118.
- [3] Al-Shamri, M.Y.H. & Al-Ashwal, N.H. (2014). Fuzzy-weighted similarity measures for memory-based collaborative recommender systems. *Journal of Intelligent Learning Systems and Applications, 6*, 1-10.
- [4] Bobadilla, J., Ortega, F., & Hernando, A. (2012). A collaborative filtering similarity measure based on singularities. *Information Processing and Management, 48*(2), 204 - 217.
- [5] Kwon, H.-J., Lee, T.-H., & Hong, K.-S. (2009). Improving prediction accuracy using entropy weighting in collaborative filtering. *Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing*, (pp. 40 - 45).
- [6] CACHEDA, F., CARNEIRO, V., FERNANDEZ, D., & FORMOSO, V. (2011). Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on Web, 5*(1), 1-33.
- [7] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 175-186). ACM
- [8] Bellogin, A. & de Vries, A.P. (2013). Understanding similarity metrics in neighbour-based recommender systems. *Proceedings of the Conference on the Theory of Information Retrieval*.

- [9] Saranya, K.G., Sadasivam, G.S., & Chandralekha, M. (2016). Performance comparison of different similarity measures for collaborative filtering technique. *Indian Journal of Science and Technology*, 9(29), 1-8.
- [10] Boulkrinat, S., Hadjali, A., & Mokhtari, A. (2013). Towards recommender systems based on a fuzzy preference aggregation. *Proceeding of the Eighth Conference of the European Society for Fuzzy Logic and Technology* (pp. 146-153).
- [11] Herrera-Viedma, E.S.-G., Olivas, J.A., Cerezo, A., & Romero, F.P. (2011). A Google wave-based fuzzy recommender system to disseminate information in university digital libraries 2.0. *Information Sciences*, 181(9), 1503-1516.
- [12] Son, L.H. (2014). HU-FCF: A hybrid user-based fuzzy collaborative filtering method in recommender systems. *Expert Systems with Applications*, 41, 6861-6870.
- [13] S. Lee. (2017). Similarity measures using fuzzified ratings for collaborative filtering. *Frontiers in Artificial Intelligence and Applications*, 299, 269-274.
- [14] Shannon, C.E. (1951). Prediction and entropy of printed English. *The Bell System Technical Journal*, 30, 50 - 64.
- [15] Wang, W., Zhang, G., & Lu, J. (2015). Collaborative filtering with entropy-driven user similarity in recommender systems. *International Journal of Intelligent Systems*, 30(8), 854 - 870.
- [16] Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5-53.



이수정

1985 이화여자대학교
수학교육과 (이학사)
1990 미국 Texas A&M 대학교
컴퓨터공학과 (석사)

1994 미국 Texas A&M 대학교 컴퓨터공학과
(박사)

1994~1998 삼성전자 통신개발실 선임연구원
1998~현재 경인교육대학교 컴퓨터교육과 교수
관심분야: 컴퓨터교육, 추천시스템, 웹마이닝

E-Mail: sjlee@ginue.ac.kr