

협력필터링의 데이터 희소성 해결을 위한 자카드 지수 반영의 유사도 성능 분석

이수정[†]

요 약

협력 필터링 시스템에서 데이터 희소성 문제의 해결을 위해 공통평가항목수를 반영하는 방법이 연구되었다. 이러한 방법으로 널리 알려진 자카드 지수는 기존의 유사도 척도와 결합되어 성능을 개선할 수 있었다. 그러나, 다양한 데이터 환경에서 여러 유사도 척도들과 각각 결합했을 때의 성능 개선 효과에 대한 분석 연구는 미미하므로, 본 연구는 이에 대한 분석을 목적으로 한다. 우선 자카드 지수 자체를 유사도 척도로 사용했을 때 희소한 데이터셋 상에서 전통적인 척도들보다 월등한 예측 성능을 보였고 추천 성능도 매우 우수하였다. 자카드 지수를 결합함으로써 기존 유사도 척도는 데이터 특성에 상관없이 성능이 대개 향상되었고, 특히 코사인 유사도는 희소한 데이터셋에서 가장 큰 향상을 이루었으나, 평균차이 제곱(Mean Squared Difference)의 유사도는 밀집된 데이터셋에서 오히려 저하된 예측 성능을 보였다. 따라서, 자카드 지수를 결합하여 사용하기 위해 데이터 환경 특성과 유사도 척도를 고려할 필요가 있다.

주제어 : 추천 시스템, 협력 필터링, 유사도 척도, 자카드 지수

Performance Analysis of Similarity Reflecting Jaccard Index for Solving Data Sparsity in Collaborative Filtering

Soojung Lee[†]

ABSTRACT

It has been studied to reflect the number of co-rated items for solving data sparsity problem in collaborative filtering systems. A well-known method of Jaccard index allowed performance improvement, when combined with previous similarity measures. However, the degree of performance improvement when combined with existing similarity measures in various data environments are seldom analyzed, which is the objective of this study. Jaccard index as a sole similarity measure yielded much higher prediction quality than traditional measures and very high recommendation quality in a sparse dataset. In general, previous similarity measures combined with Jaccard index improved performance regardless of dataset characteristics. Especially, cosine similarity achieved the highest improvement in sparse datasets, while similarity of Mean Squared Difference degraded prediction quality in denser sets. Therefore, one needs to consider characteristics of data environment and similarity measures before combining Jaccard index for similarity use.

Keywords: Recommender System, Collaborative Filtering, Similarity Measure, Jaccard index

[†] 정 회 원: 경인교육대학교 교수
논문접수: 2016년 6월 2일, 심사완료: 2016년 7월 14일, 게재확정: 2016년 7월 18일

1. 서론

추천 시스템은 인터넷 시대의 정보 과부하 문제를 해결하고 사용자의 정보 검색 시간을 절약해 줄 수 있는 유용한 방법으로 간주되어왔다. 상업적으로 또는 학계에서 다양하고 구체적인 관련 방안이 구현 및 연구되었고 많은 논문 등의 결과물이 발표되었다[1][2]. 예를 들어 아마존 시스템은 추천 시스템을 구현하여 상업 분야에서 성공적으로 활용한 대표적인 사례이다.

상업계의 추천 시스템은 대개 협력 필터링(Collaborative Filtering, CF)을 기본 원리로 채택하는데, 이는 시스템 사용자들이 항목 검색 및 선택에 있어서 협력적으로 작동하여 효율적으로 원하는 결과를 얻는데 용이한 방법이다. 현 사용자가 선호할만한 항목들을 추천해 주기 위한 CF 방식의 기본 아이디어는 현 사용자와 유사한 선호도를 지닌 것으로 파악되는 다른 사용자들의 항목 선호 이력을 참조하여, 가장 높은 선호도를 표했던 항목들을 추천하는 것이다. 엄밀하게 말하면 이러한 방식은 사용자 기반의 CF(user-based CF)라고 불린다[1][3]. 이와 대조되는 개념으로서 항목 기반 CF(item-based CF) 방식은 현 사용자가 과거에 선호하였던 항목들과 유사한 항목들을 추천한다.

사용자 기반 CF 방법의 기본적인 절차는 구체적으로 다음과 같다[3][4]. 시스템은 항목들에 대한 사용자들의 과거 평가치 기록을 보유하고 있다.

1. 현 사용자 u 와 시스템의 다른 모든 사용자들 간의 유사도를 산출한다.
2. u 가 아직 구매하지 않은 항목들 각각에 대한 추천 여부를 결정하기 위하여, 임의의 미구매 항목 x 에 대한 u 의 평가예측값을 아래의 절차로써 구한다.
 - 2.1 x 를 평가했던 모든 사용자들 중 평가치를 참조할 u 의 이웃집합 N_u 를 선정한다. 이를 위해 u 와의 유사도를 기준으로 한다.
 - 2.2 x 에 대한 u 의 평가예측값 $r'_{u,x}$ 을 다음과 같이 구한다[5]. $sim(u, v)$ 는 u 와 v 간의 유사도 값이고, \bar{r}_u 는 u 와 v 간의 공통평가항목들에

대한 u 의 평가치 평균값이다.

$$r'_{u,x} = \bar{r}_u + \frac{\sum_{v \in N_u} sim(u,v) \times (r_{v,x} - \bar{r}_v)}{\sum_{v \in N_u} |sim(u,v)|} \dots\dots\dots(1)$$

3. 단계 2에서 산출된 평가예측치가 큰 항목들 순으로 추천한다.

항목 기반 CF에서의 추천 절차도 사용자 기반 CF 절차와 유사하다. 본 연구에서는 보다 기본적인 이해가 쉬운 사용자 기반 CF에 초점을 두었다. 위 절차에서 알 수 있듯이 유사도 산출은 이웃 집합과 평가예측치를 구하는데 있어서 매우 중요한 역할을 하므로, 사용자 간 정확한 유사도 값을 산출해야 한다. 가장 널리 사용되는 산출공식으로는 피어슨 상관과 코사인 벡터가 있다. 그러나, 이들 방법에서 공통평가항목수가 매우 적을 경우, 즉, 데이터가 희소할 경우에 유사도 산출은 불가하거나 부정확할 수 있다[4][6]. 이러한 데이터 희소성 문제(data sparsity problem)는 여러 상황에 기인하는데, 즉, cold start, 새로운 사용자 또는 항목의 등장 등이다.

위와 같은 문제를 해결하고자 다양한 방법들이 개발되었는데[7][8][9], 특히 공통평가항목수를 유사도 측정에 반영하는 간단한 방법들이 제안되었다[8][10][11]. 예로써, 자카드 지수(Jaccard index)라고 불리는 척도를 반영하여 보다 정확한 유사도 값을 산출하려는 연구가 실행되었다. 이 방법은 그 효율성과 확장 가능성 등의 측면에서 매우 유용하고 다수의 연구에서 언급되고 더욱 확장된 아이디어들이 제안될 정도로 주목되었으나[8], 근본적으로 어떠한 유사도 산출공식과 결합하는 것이 바람직한지, 또한 어떤 데이터 환경 하에서 좋은 결과를 낳는지 등에 대한 상세한 연구는 미흡하다. 따라서, 본 연구에서는 기존 유사도 공식들과 자카드 지수를 결합한 방법을 다양한 환경 조건 하에서 실험하여 기술한 연구 의문점들을 해결하고자 한다.

2. 배경

2.1 유사도 측정 방법

대표적인 방법으로 피어슨 상관(Pearson Correlation), 코사인 유사도(Cosine Similarity), 평균 제곱 차이(Mean Squared Difference, MSD)를 들 수 있다[1][12]. 두 사용자 u와 v 간의 피어슨 상관의 유사도는

$$cor(u, v) = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2 \sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \dots\dots\dots(2)$$

으로 계산하는데 이 때 I는 두 사용자의 공통평가 항목집합, $r_{u,i}$ 는 항목 i에 대한 u의 평가치이고, $\bar{r}_u = \frac{1}{I} \sum_{i \in I} r_{u,i}$ 이다.

코사인 유사도는 각 사용자의 평가등급을 벡터로 간주하고, 두 벡터 간의 코사인 값으로 유사도를 정의한다. 즉,

$$cos(u, v) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \times \|\vec{v}\|} = \frac{\sum_{i \in I} r_{u,i} r_{v,i}}{\sqrt{\sum_{i \in I} r_{u,i}^2} \sqrt{\sum_{i \in I} r_{v,i}^2}} \dots\dots\dots(3)$$

MSD는 공통평가항목 각각에 대해 두 평가치 차이의 제곱의 평균값이므로, 유사도는 1-MSD로 정의한다. 구체적으로 다음의 공식을 사용한다.

$$MSD(u, v) = \frac{1}{I} \sum_{i \in I} (r_{u,i} - r_{v,i})^2 \dots\dots\dots(4)$$

2.2 유사도와 데이터 희소성 문제

주요 유사도 측정 방법인 COR와 COS는 두 사용자의 공통평가항목들에 대한 평가치를 기반으로 유사도를 산출하기 때문에 공통평가항목수가 적으면 산출값에 대한 신뢰가 저하된다. 이와 같은 문제점은 상세히 분석 기술되었으므로 이 논문에선 생략한다[4][6].

유사도 산출에 있어서 데이터 희소성 문제를 해결하기 위하여 다수의 방법들이 연구되었다 [8][9][10][11]. 특히, 공통평가항목수를 반영한 척

도인 자카드 지수(Jaccard Index)[13]를 기존 유사도 척도와 결합하는 방식이 그 단순성과 효율성으로 인해 주목되어 왔다[11]. I_u 를 사용자 u가 평가한 항목 집합이라고 할 때, 사용자 u와 v간의 자카드 지수는 아래와 같이 산출한다[13].

$$Jaccard(u, v) = \frac{|I_u \cap I_v|}{|I_u| + |I_v| - |I_u \cap I_v|} \dots\dots\dots(5)$$

Bobadilla 외 2인의 연구에서는 기존의 여러 유사도 척도와 자카드 지수를 결합하여 다양한 환경에서 실험하였다[11]. 그 결과 MSD와 결합했을 때 가장 성능이 우수함을 밝혔다. 그러나, 단지 피어슨 상관의 결과와만 성능을 비교하여, 우수한 성능의 이유가 자카드 지수와 결합 때문인지 MSD 자체의 성능이 우수하기 때문인지에 대한 분석이 미흡하였다. 또한 평가범위가 가장 큰 데이터셋에서 가장 성능 결과가 좋지 않았는데 그 이유에 대해서도 분명한 원인 분석을 기술하지 않았다.

따라서 본 연구에서는 자카드 지수를 기존의 어떤 유사도 척도와 결합했을 때 가장 최적의 결과를 나타내며 데이터셋의 특성이 다를 때 성능은 어떤 변화를 보이는지 조사함으로써, 최적의 환경 조건과 유사도 척도를 알아내고자 한다.

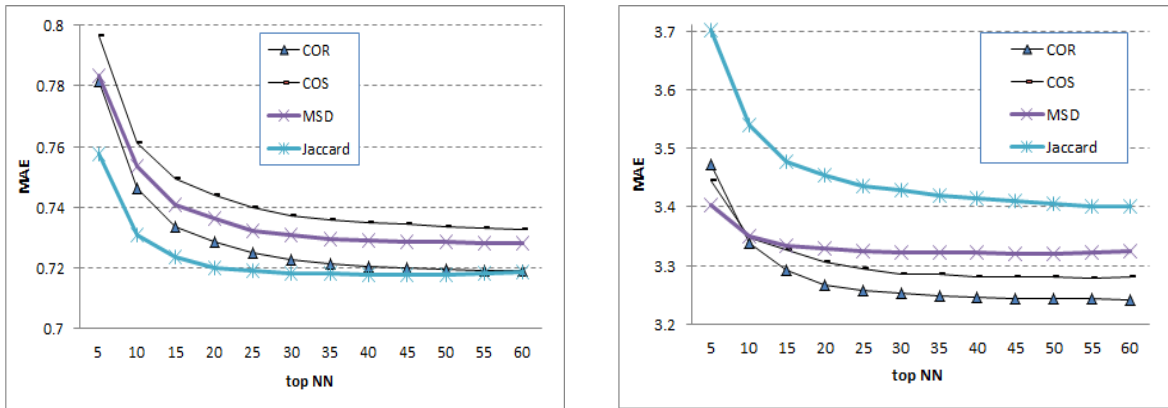
3. 실험 연구

3.1 실험 배경

면밀한 성능 분석을 위해 매우 상이한 특성을 지녔으며 관련 연구에서 널리 활용되는 두 데이터셋을 선택하였다. <표 1>에 이들의 특성을 상세 기술하였다. 희소성 수준(sparsity level)이란 사용자×항목 행렬 내 평가가 매겨지지 않은 요소의 비율을 의미한다. 연구 목적을 달성하기 위하여 의도적으로 매우 상반된 희소성 수준과 평가등급범위를 지닌 이들 데이터셋을 선정하여 실험

<표 1> 실험 데이터 집합

데이터셋	특성	평가개수	행렬크기 (사용자수×항목수)	평가범위	희소성수준
MovieLens		각 사용자 당 20개 이상	6040 × 3952	1~5의 정수	0.9581
Jester		각 사용자 당 36개 이상	24983 × 100	-10~+10의 실수	0.2936



[그림 1] 최인접이웃수의 변화에 따른 유사도 종류별 MAE 성능 변화: MovieLens(좌)와 Jester(우)

하였다.

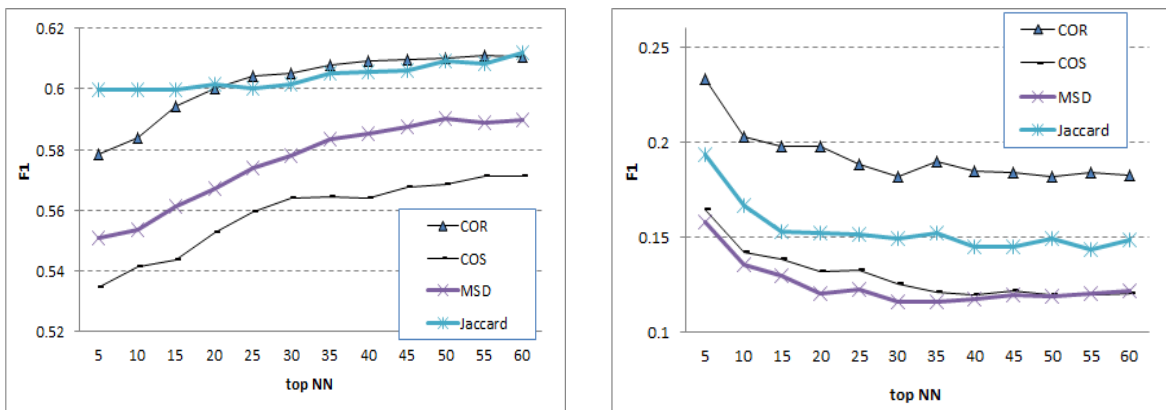
실험에 사용한 유사도 척도는 앞 절에서 언급한 COR, COS, MSD와 이들 각각에 자카드 지수를 곱한 유사도 척도인 COR*J, COS*J, MSD*J, 그리고 자카드 지수(Jaccard)로 하였다. 이들 중 MSD는 유사도값 산출시 실제로는 1-MSD로 계산한다. 자카드 지수와와 단순곱의 유사도 척도를 선택한 이유는, 2.2절에서 기술하였듯이 [11]의 연구를 기초로 하여 더욱 다양한 실험 조건과 측면에서 분석하기 위함이다. 즉, [11]에서 피어슨 상관도만을 성능 비교 대상으로 선택한 것과 달리, 본 연구에서는 각각의 유사도 척도가 자카드 지수를 결합함으로써 어떠한 환경에서 얼마나 성능 개선 효과가 있는지를 분석하였다.

성능 평가 지표로서 예측 정확도와 추천 정확도를 살펴보았는데, 전자의 대표적 지표로서 평가 예측치와 실제치 차이의 절대값인 평균절대오차 (Mean Absolute Error, MAE)[13]를 사용하였으

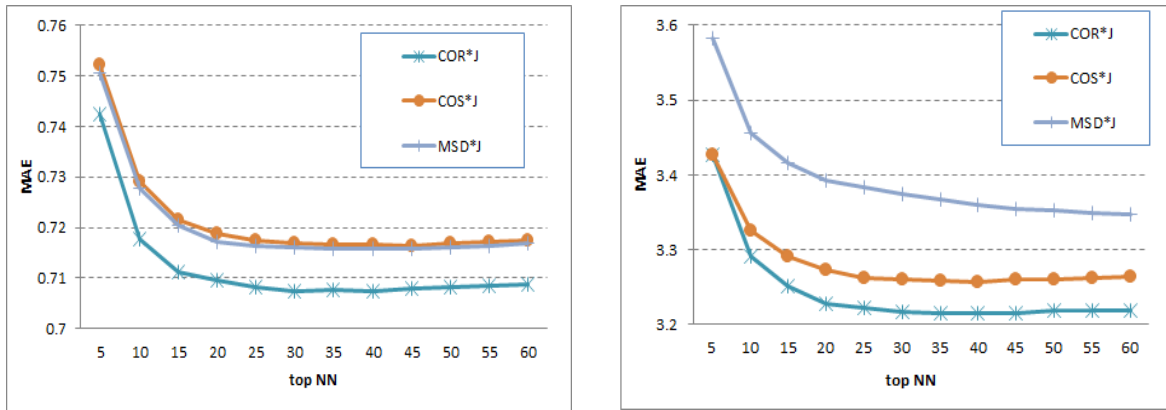
며, 후자의 지표로는 널리 사용되는 F1을 도입하였다. F1은 precision(P)과 recall(R)을 동일한 비중으로 결합한 조화평균($2PR/(P+R)$)으로 계산된다[14]. P는 시스템의 추천항목들 중 사용자 선호 항목들의 비율을 나타내며, R은 사용자의 전체 선호 항목들 중 시스템이 추천한 항목들의 비율이다 [14]. 본 실험에서는 평가치가 MovieLens에서는 4 점, Jester에서는 6점 이상이면 선호항목으로 간주하였다.

3.2 자카드 지수의 성능

본 연구의 목적은 공통평가항목수를 반영하는 지표인 자카드 지수가 기존 유사도 척도의 성능을 개선하는데 얼마나 기여할 수 있는지 또한 어떠한 환경에서 기여율이 가장 높은지를 파악하는 것이다. 그러므로, 우선 자카드 지수 자체를 유사도 척도로 하여 성능을 살펴보는 것이 우선 과제이다.



[그림 2] 최인접이웃수의 변화에 따른 유사도 종류별 F1 성능 변화: MovieLens(좌)와 Jester(우)



[그림 3] 최인접이웃수의 변화에 따른 자카드지수 결합 유사도 종류별 MAE 성능 변화: MovieLens(좌)와 Jester(우)

[그림 1]은 네 가지 유사도 척도를 사용하여 실험한 결과이다. 매우 뜻밖에도 자카드 지수는 그 자체만으로도 희소 데이터 환경인 MovieLens에서 가장 월등한 MAE 성능을 보였다. 반면 같은 환경에서 기존의 다른 두 척도보다도 COS는 가장 나쁜 성능을 보였는데, 이는 과거 연구된 바와 같이 데이터가 희소할 때 드러나는 COS의 약점 때문이다[4][6].

보다 밀집된 데이터 환경을 갖는 Jester에서는 자카드 지수가 가장 나쁜 성능을 보였는데, 그 이유는 다른 세 척도는 신뢰할만한 유사도값을 산출하기에 충분한 공통평가항목수를 갖기 때문으로 판단된다. 종합적으로, MAE 성능에 있어서 자카드 지수를 제외하고 COS가 데이터셋 희소성의 영향을 가장 많이 받는 것으로 나타났고, 그 다음이 COR이며, MSD는 가장 안정적이었으나 밀집된 데이터 환경에서는 오히려 상대적으로 더욱 저하된 성능을 보였다.

F1 성능에 있어서 [그림 2]에서 보듯이 자카드 지수의 성능이 데이터셋에 관계없이 매우 우수하며, COS나 MSD는 낮은 성능을 나타냈다. MAE에서와 달리 자카드 지수의 성능이 Jester 셋에서

도 COR만을 제외하고 다른 척도들보다 더욱 우수한데 이는 F1값 산출 특성상 특정 평가치 기준(Jester의 경우 6점) 이상이면 선호 항목으로 포함시키므로, MAE에서처럼 정확한 값이 아니어도 좋은 성능으로 간주될 수 있기 때문으로 판단된다. MAE에서와 마찬가지로 COR는 대체로 가장 우수한 F1 성능을 보임을 알 수 있다.

3.3 자카드 지수와 기존 유사도의 결합 성능

앞 절에서 자카드 지수의 성능은 특히 MovieLens에서 가장 뛰어났음이 확인되었다. 따라서 이를 나머지 세 유사도 척도와 각각 결합했을 때 기존 척도의 성능이 얼마나 향상될지 알아보았다. [그림 3]에서와 같이 MAE 성능의 순위는 자카드 지수를 결합하지 않았을 때와 거의 유사하다. 즉, 두 데이터셋에서 COR*J의 성능이 가장 뛰어나고 MSD*J의 성능이 가장 뒤떨어졌다.

<표 2>는 자카드 지수를 결합했을 때의 성능 향상 정도를 원 성능에 대비하여 상대적으로 계산하여 제시하였다. MAE 성능이므로 낮은 수치가 더욱 좋은 향상도를 나타낸다. MovieLens에서

<표 2> 최인접이웃수 변화에 따른 각 유사도 종류별 Jaccard 결합 방식의 MAE 향상률 비교

데이터셋	종류 \ topNN	5	10	15	20	25	30	35	40	45	50	55	60
MovieLens	COR*J/COR	0.950	0.962	0.969	0.974	0.977	0.979	0.981	0.982	0.983	0.984	0.985	0.986
	COS*J/COS	0.944	0.958	0.962	0.966	0.970	0.972	0.974	0.975	0.976	0.977	0.978	0.979
	MSD*J/MSD	0.958	0.966	0.972	0.974	0.978	0.980	0.981	0.982	0.982	0.983	0.984	0.984
Jester	COR*J/COR	0.986	0.986	0.988	0.989	0.989	0.989	0.990	0.991	0.991	0.992	0.992	0.993
	COS*J/COS	0.995	0.993	0.989	0.990	0.991	0.992	0.992	0.992	0.994	0.993	0.995	0.995
	MSD*J/MSD	1.053	1.032	1.024	1.019	1.017	1.016	1.013	1.012	1.010	1.009	1.008	1.007

<표 3> 최인접이웃수 변화에 따른 각 유사도 종류별 Jaccard 결합 방식의 F1 향상을 비교

데이터셋	종류	topNN												
		5	10	15	20	25	30	35	40	45	50	55	60	
MovieLens	COR*J/COR	1.060	1.068	1.055	1.049	1.042	1.039	1.036	1.033	1.033	1.029	1.026	1.020	
	COS*J/COS	1.130	1.106	1.102	1.092	1.078	1.074	1.071	1.072	1.071	1.068	1.070	1.071	
	MSD*J/MSD	1.095	1.087	1.076	1.067	1.049	1.046	1.038	1.037	1.038	1.038	1.042	1.038	
Jester	COR*J/COR	1.085	1.138	1.058	1.102	1.108	1.136	1.068	1.078	1.093	1.059	1.064	1.074	
	COS*J/COS	1.043	0.964	0.912	0.976	0.920	1.033	1.002	1.032	1.026	1.013	1.019	1.043	
	MSD*J/MSD	1.173	1.156	1.171	1.212	1.142	1.234	1.206	1.195	1.192	1.170	1.146	1.151	

세 척도의 향상도가 모두 비슷하지만 COS에서 조금 더 좋은 향상도를 나타냈는데, 이는 COS의 단점을 자카드 지수가 가장 효율적으로 보완하기 때문으로 보인다. 한편 Jester에서는 데이터가 이미 밀집되어 있으므로 향상도가 MovieLens에 비해 떨어지고 자카드 지수와 결합의 효과가 전반적으로 미미한데, 주목할 것은 MSD는 자카드 지수와 결합하는 것이 MAE 성능 측면에서 오히려 해가 되는 것으로 나타났다. 그러므로, 데이터셋의 특성과 원 유사도 척도의 종류에 따라 자카드 지수와 결합 여부를 결정하여 사용하는 것이 필요하다.

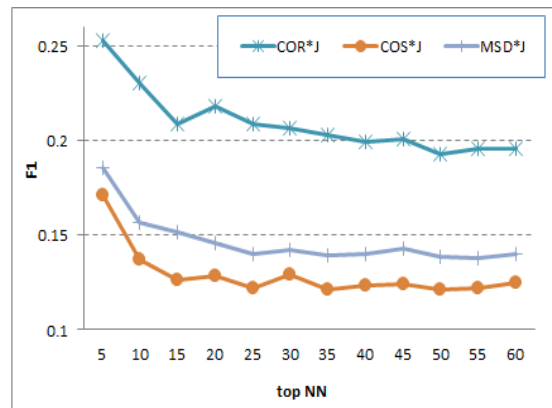
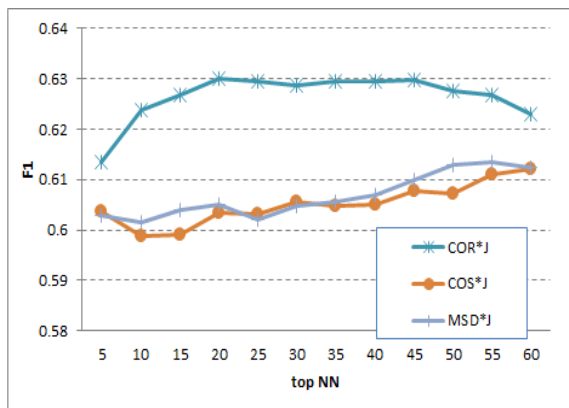
[그림 4]에 F1의 성능을 비교하였다. MAE에서와 별반 다르지 않게 COR*J의 성능이 월등히 높았고 나머지 두 척도는 거의 유사한 성능을 보였으나 Jester 셋에서 COS*J의 성능이 다소 현격하게 떨어짐을 볼 수 있다. <표 3>의 향상도는 원 유사도 척도 대비 자카드 지수 결합 척도의 F1 성능 개선 효과를 나타낸다. 숫자가 클수록 향상도가 클수록 의미한다. 모든 척도에서 F1 성능이 향상되었으나, MAE 성능에서와 마찬가지로 MovieLens에서는 COS에 조금 더 좋은 효과를

보인 반면에, Jester에서는 MSD에 가장 이로운 결과를 보였다. 이는 [그림 2]에서 보듯이 MSD가 가장 저하된 F1 성능을 보이는데, 이보다 훨씬 좋은 성능을 보이는 자카드 지수가 MSD의 저하된 성능을 상승시킬 수 있었기 때문으로 판단된다.

4. 결론

본 연구에서는 협력 필터링 기반의 추천 시스템에서 데이터 희소성을 해결하기 위한 주요 방안이었던 자카드 지수를 이용하는데 있어서 어떠한 환경과 기존의 어떤 유사도 척도와 결합하는 것이 성능 개선에 유익할지 분석하였다.

실험 결과 자카드 지수 자체를 유사도로 사용하는 것은 뜻밖에도 매우 우수한 성능을 보였는데, MovieLens와 같은 희소한 데이터셋에서 다른 유사도 방법들에 비해 가장 월등한 MAE 성능을 보였고 F1 성능도 매우 우수하였다. 따라서, 자카드 지수를 기존 유사도 척도와 결합한 결과도 대체로 모든 지표의 성능을 향상시켰다. 그러나, MSD는 Jester 데이터셋에서 자카드 지수와 결합이 오히려 MAE 성능을 저하시켰다. 결론적으



[그림 4] 최인접이웃수의 변화에 따른 자카드 지수 결합 유사도 종류별 F1 성능 변화: MovieLens(좌)와 Jester(우)

로 데이터 희소성의 영향을 가장 많이 받는 척도는 COS로 드러났으므로 희소한 데이터셋에서 자카드 지수와의 결합의 잇점을 MAE와 F1 성능 측면에서 모두 가장 크게 누리는 척도임을 알 수 있었다. 이와 같이 성능 개선 여부가 데이터 환경과 유사도 척도에 따라 다르므로, 자카드 지수를 유사도 척도와 결합시에 면밀한 사전 검토 및 주의가 필요하다.

참 고 문 헌

- [1] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence, 2009*, 4.
- [2] Aamir, M. & Bhusry, M. (2015). Recommendation system: state of the art approach. *International Journal of Computer Applications, 120*(12), 25-32.
- [3] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering, 17*(6), 734-749.
- [4] Lee, S. (2015). A strategy for neighborhood selection in collaborative filtering-based recommender systems. *Journal of KIISE, 42*(11), 1380-138.
- [5] Resnick, P. et al. (1994). GroupLens: an open architecture for collaborative filtering of Netnews. *Proc. of the ACM Conf. Computer Supported Cooperative Work*, 175-186.
- [6] Ahn, H. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences, 178*(1), 37 - 51.
- [7] Chen, C. C., Wan, Y.-H., Chung, M.-C., & Sun, Y.-C. (2013). An effective recommendation method for cold start new users using trust and distrust networks. *Information Sciences, 224*, 19-36.
- [8] Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. *Knowledge-Based Systems, 26*, 225-238.
- [9] Liu, H., Hu, Z., Mian, A., Tian, H., & Zhu, X. (2014). A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems, 56*, 156-166.
- [10] Jamali, M., & Ester, M. (2009). TrustWalker: a random walk model for combining trust-based and item-based recommendation. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 397-406.
- [11] Bobadilla, J., Serradilla, F., & Bernal, J. (2010). A new collaborative filtering metric that improves the behavior of recommender systems. *Knowledge-Based Systems 23*, 520-528.
- [12] Sanchez, J. L., Serradilla, F., Martinez, E., & Bobadilla, J. (2008). Choice of metrics used in collaborative filtering and their impact on recommender systems. *Proceedings of the IEEE International Conference on Digital Ecosystems and Technologies*, 432-436.
- [13] Koutrica, G., Bercovitz, B., & Garcia, H. (2009). FlexRecs: expressing and combining flexible recommendations. *Proc. of the ACM SIGMOD Int'l Conf. on Management of data*, 745-758.
- [14] Gao, M., Wu, Z., & Jiang, F. (2011). UserRank for item-based collaborative filtering recommendation. *Information Processing Letters, 111*(9), 440-446.



이수정

1985 이화여자대학교
수학교육과(이학사)

1990 Texas A&M 대학교
컴퓨터과학과(석사)

1994 Texas A&M 대학교 컴퓨터과학과(박사)

1994~1998 삼성전자 통신개발실 선임연구원

1998~현재 경인교육대학교 컴퓨터교육과 교수

관심분야: 컴퓨터교육, 추천시스템, 정보필터링

E-Mail: sjlee@gin.ac.kr