

Improving Performance of Jaccard Coefficient for Collaborative Filtering

Sojung Lee*

Abstract

In recommender systems based on collaborative filtering, measuring similarity is very critical for determining the range of recommenders. Data sparsity problem is fundamental in collaborative filtering systems, which is partly solved by Jaccard coefficient combined with traditional similarity measures. This study proposes a new coefficient for improving performance of Jaccard coefficient by compensating for its drawbacks. We conducted experiments using datasets of various characteristics for performance analysis. As a result of comparison between the proposed and the similarity metric of Pearson correlation widely used up to date, it is found that the two metrics yielded competitive performance on a dense dataset while the proposed showed much better performance on a sparser dataset. Also, the result of comparing the proposed with Jaccard coefficient showed that the proposed yielded far better performance as the dataset is denser. Overall, the proposed coefficient demonstrated the best prediction and recommendation performance among the experimented metrics.

▶ Keyword : Collaborative Filtering, Recommender System, Similarity Measure, Jaccard Coefficient

1. Introduction

인터넷 상의 정보량이 극심하게 증대함에 따라 사용자들은 원하는 정보를 검색하기가 점차 어려워졌다. 이러한 문제점을 해결하기 위한 한 가지 방법으로서 추천 시스템(recommender system)이 대두되었다. 이 시스템은 주로 상업용으로써 고객이 선호할만한 상품을 제안하는 목적으로 활용되었다[1].

어떠한 정보를 이용하여 추천하는가에 따라 추천 시스템의 종류가 구분되는데, 크게 두 가지 범주로 나뉜다. 즉, 내용 기반 필터링(content-based filtering, CBF)과 협력 필터링(collaborative filtering, CF) 방식이 그것이다[2]. CBF 방식은 사용자가 이전에 선호하였던 항목들의 정보나 사용자 본인의 프로필 정보를 기초로 하여 유사한 특성을 가진 항목들을 추천한다. 이러한 원리로 인하여 CBF는 serendipity problem(우연한 발견의 문제)와 같은 주요한 단점을 갖는다[2][3]. 즉, 과거에 선호하였던 항목만을 지속적으로 추천받게 되고, 사용자의 선호와는 다른 새로운 특성이나 내용의 항목들은 접할 기회가 없다는 문제점이 있다.

이러한 CBF 방식과는 달리, CF는 현 사용자와 유사한 사용

자들의 선호 항목들을 종합하여 추천한다[2]. 즉, 유사한 흥미를 가진 다른 사용자들이 선호하였던 항목들 중에서 추천대상이 미구매한 상품에 대한 선호도를 예측한 후, 가장 높은 예측 선호도를 가진 항목들을 추천한다. 따라서, CBF의 serendipity problem은 근본적으로 해결될 수 있는데, 그 이유는 사용자가 이전에 선호하지 않았던 특성을 가졌다 할지라도, 그 특성 및 내용은 고려하지 않고, 다른 사용자들이 선호하였다는 사실 하나만으로 추천 항목에 포함될 수 있으므로, 새로운 특성의 항목들을 발견할 수 있기 때문이다. CF 시스템은 가장 성공적인 추천기법으로 알려져 있으며 현재까지 음악, 서적, 영화 등 다양한 분야에서 사용되고 있다.

이와 같이 CF 방식의 성공 요소로서 매우 중요한 것들 중의 하나는 유사한 특성을 지닌 다른 사용자들을 찾아내는 것과 정확한 유사도 값을 산출하는 일이다. 사용자 기반의 CF(user-based CF) 시스템에서는 두 사용자가 평가했던 항목들의 평가등급들을 토대로 유사도를 산출한다. 기존 연구에 따르면 가장 널리 사용되는 유사도 산출 방법으로서 피어슨 상관(Pearson correlation), 코사인 유사도(cosine similarity), 평

*First Author: Sojung Lee, Corresponding Author: Sojung Lee

*Sojung Lee(sjlee@gin.ac.kr), Dept. of Computer Education, Gyeongin National University of Education

*Received: 2016. 09. 20, Revised: 2016. 10. 10, Accepted: 2016. 11. 04.

균사승차이(mean squared differences) 등이 있다[2][3]. 그러나, 이들 방법에서 공통평가항목수가 매우 적을 경우, 즉, 데이터가 희소할 경우에 유사도 산출은 불가하거나 부정확할 수 있다[4][5]. 기존에 발표되었던 다양한 유사도 척도와 이들에 대한 장단점 분석은 여러 연구 결과를 통해 참조할 수 있다 [3][6].

CF 방식의 주요 단점인 데이터 희소성 문제(data sparsity problem)는 새로운 사용자 또는 항목의 등장, cold start 등의 상황으로 인해 발생한다[7]. 이는 근본적으로 사용자들의 평가 개수가 적으므로 신뢰할만한 유사도 값이 측정될 수 없기 때문이다. 이러한 문제를 해결하기 위한 다양한 방법들이 제안되어 왔으나, 기존의 유사도 척도에서 거의 활용되지 않았던 두 사용자의 공통평가항목수를 척도에 반영하려는 해결 방법이 주목을 받아 왔다. 자카드 계수(Jaccard coefficient)[8]는 공통평가항목수를 계상하는 유용한 척도로서 몇몇 연구에서 도입되어 기존 척도의 일부 단점을 보완하였다고 보고되었다 [6][7][9][10]. 본 연구에서는 이와 같이 기존 연구에서 발표된 새로운 유사도 척도의 주요 구성 요소였던 자카드 계수의 단점을 보완하여 그 성능을 개선한 새로운 계수를 제안한다. 또한 다양한 특성의 데이터셋을 활용한 실험을 통해 그 성능을 입증하고 기존의 계수 및 유사도 척도와 비교 분석하였다.

본문의 구성은 다음과 같다. 2절에서 기존의 자카드 계수 및 그와 관련된 계수에 대해 소개한다. 3절에서는 제안 방법을 설명한 후 4절에서 성능 측정 실험 결과를 제시하고, 5절에서 논문의 결론을 맺는다.

II. Preliminaries

1. Jaccard Coefficient

전통적인 유사도 척도는 두 사용자의 공통평가항목수가 적은 경우 신뢰할만한 유사도 값을 산출해내지 못하므로, 이를 보완하기 위한 척도로서 자카드 계수(Jaccard coefficient)가 개발되었다[8]. 이 척도는 두 사용자가 공통으로 평가한 항목수의 비율을 계산한다. I_u 를 사용자 u가 평가한 항목들의 집합이라고 할 때, 사용자 u와 v 간의 자카드 계수는 다음과 같이 산정한다.

$$Jaccard(u, v) = \frac{|I_u \cap I_v|}{|I_u| + |I_v| - |I_u \cap I_v|} \dots\dots\dots(1)$$

위 식에서 알 수 있듯이, 자카드 계수는 평가치를 전혀 반영하지 않고 단지 평가개수만을 고려하기 때문에, 유사도 산출을 위해 독립적으로 사용하는 데는 무리가 따른다. 따라서, Bobadilla 외 2인은 평균사승차이(mean squared differences)와 결합함으로써 개선된 유사도 척도를 제안하였고[9], Saranya 외 2인은 피어슨 상관도와 결합하여 성능 개선을 이

루었다고 보고하였다[10]. 이 밖에도 다양한 전통적 유사도 척도와 자카드 계수를 결합하여 비교 분석한 것 뿐만 아니라 독립적인 유사도 척도로서의 자카드 계수만의 성능을 조사 분석한 결과도 보고되었다[5].

2. Jaccard Uniform Operator Distance (JacUOD)

JacUOD는 [11]에서 제안한 지수로서 UOD에 자카드 계수를 곱의 형태로 결합한 것이다. $UOD(u, v)$ 는 두 사용자의 공통평가항목 각각에 대한 평가값 차이의 합을 사용자 u의 최대와 최소평가값 차이와 비교하여 그 비율을 산출한다. 전자가 작을수록, 즉, 두 사용자가 근접한 평가치를 부여할수록 UOD의 값은 커지도록 정의하였다. 구체적으로 다음과 같이 산출한다.

$$UOD(u, v) = \frac{\sqrt{m(r_{u, \max} - r_{u, \min})^2}}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u, i} - r_{v, i})^2}} \dots\dots\dots(2)$$

$$UOD(u, v) = \frac{\sqrt{m(r_{u, \max} - r_{u, \min})^2}}{0.9 + \sqrt{\sum_{i \in I_u \cap I_v} (r_{u, i} - r_{v, i})^2}} \dots\dots\dots(3)$$

위 식 (2)는 분모가 0이 아닐 경우, 즉, 적어도 하나의 공통평가항목에 대해서 서로 다른 평가치를 부여했을 경우이고 식 (3)은 그렇지 않은 경우에 사용한다. $r_{u, \max}$ 와 $r_{u, \min}$ 는 사용자 u가 부여한 최대평가값과 최소평가값이고, $r_{u, i}$ 는 항목 i에 부여한 평가값이다. 또한 m은 공통평가항목의 개수, 즉, $|I_u \cap I_v|$ 를 나타낸다.

앞에서 언급한대로, JacUOD는 자카드 계수와 UOD를 결합한 형태로 다음과 같이 정의한다.

$$JacUOD(u, v) = Jaccard(u, v) \times UOD(u, v) \dots\dots(4)$$

공식에서 알 수 있듯이, JacUOD는 공통평가항목 개수만을 고려한 자카드 계수의 단점을 개선하여, 공통항목에 대한 평가치 차이를 추가적으로 고려함으로써 성능을 보완하였으며, [10]에서 그 성능을 실험 조사하였다.

III. The Proposed Scheme

1. Motivation

Bobadilla 외 2인[9]의 연구에서는 MovieLens 1M(www.movielens.org)과 Netflix(www.netflixprize.com) 데이터셋 상에서 사용자들이 어떠한 평가치를 부여하는지 그 평균과 분포를 조사하였다. 이들 데이터셋은 모두 1부터 5까지의 정수값을 평가치로 허용한다. 실험 결과 사용자들은 대개 극단의 값을 피하고 허용 평가치 구간 내에서 상위의 값을 부여함을 알아냈다. 이같은 결과는 두 데이터셋에서 비슷하게 나타났는데, 4의 평가값을 부여한 사례수가 가장 많았고, 그 다음이

Table 1. Datasets for experiments

dataset	characteristics	number of ratings	matrix size (users×items)	rating range	sparsity level
MovieLens 1M		≥20 per user	1000 × 3952	1~5 (integer)	0.9607
BookCrossing		≥10 per user	1014 × 883	1~10 (integer)	0.9775
Jester		≥36 per user	998 × 100	-10~+10 (real)	0.2936

3, 5으로 나타났다. 또한 대부분의 항목에 부여된 평가치 차이도 평균적으로 1.2를 넘지 않는 것으로 드러났다.

이와 같은 연구 결과로부터 알 수 있는 것은 사용자가 부여한 평가값은 동일한 가치를 갖지 않는다는 사실이다. 즉, 값의 크기에 따라 적어도 희소성 측면에서 가치가 다르다. 따라서 두 사용자가 부여한 평가값이 동일하게 1인 경우가 동일하게 4를 부여한 경우보다 더욱 가치 있게 취급되어야 할 것으로 판단하여, 이러한 가정 하에 새로운 방법을 개발하였다.

2. Proposed Coefficient

앞 절에서 기술한 판단에 따라 기존 척도인 자카드 계수를 개선하는 방법을 고안하였다. 자카드 계수는 평가값의 범위와는 무관하게 두 사용자의 공통평가항목개수를 기준으로 산출하였으나, 본 연구의 방법에서는 두 사용자가 높은(혹은 낮은) 평가값을 얼마나 공통으로 많이 부여했는가를 고려한다. 즉, 본 제안 방법은 전체 평가값의 범위를 상, 중, 하의 셋으로 나누어 자카드 계수를 각각 적용한다. 즉, 평가값이 사전에 정한 특정값(Hbd) 이상인 구간, 또 다른 특정값(Lbd) 이하인 구간, 그리고 그 사이에 속한 구간에서 각각 공통평가항목수의 비율을 계산하여, 각 구간별 결과값의 평균을 산출한다. 구체적으로, 세 구간과 그와 연관된 자카드 계수를 아래와 같이 정의한다.

$$\begin{aligned}
 I_{L,u} &= \{i \in I_u \mid r_{u,i} \leq L_{bd}\} \\
 I_{M,u} &= \{i \in I_u \mid L_{bd} < r_{u,i} < H_{bd}\} \\
 I_{H,u} &= \{i \in I_u \mid r_{u,i} \geq H_{bd}\} \dots\dots\dots (5)
 \end{aligned}$$

$$\begin{aligned}
 Jaccard_L(u, v) &= \frac{|I_{L,u} \cap I_{L,v}|}{|I_{L,u} \cup I_{L,v}|} \\
 Jaccard_M(u, v) &= \frac{|I_{M,u} \cap I_{M,v}|}{|I_{M,u} \cup I_{M,v}|} \\
 Jaccard_H(u, v) &= \frac{|I_{H,u} \cap I_{H,v}|}{|I_{H,u} \cup I_{H,v}|} \dots\dots\dots (6)
 \end{aligned}$$

최종적으로 제안 방법을 JacLMH로 명명하며 아래와 같이 산출한다.

$$\begin{aligned}
 JacLMH(u, v) &= (Jaccard_L(u, v) + Jaccard_M(u, v) \\
 &\quad + Jaccard_H(u, v))/3 \dots\dots\dots (7)
 \end{aligned}$$

IV. Performance Experiments

1. Background

제안 방법의 성능을 자카드 계수(Jaccard), JacUOD, 그리고 전통적 유사도 척도로서 대표적인 피어슨 상관도(Pearson Correlation Coefficient, PCC)의 성능과 비교 실험하였다. 실험에 사용된 데이터셋은 척도들의 형평성을 위해 서로 다른 다양한 특성을 가진 셋들로 정하였다. 표 1에 기술한 데이터셋들은 관련 연구에서 흔히 사용되는데[4][9][11][12][13], 평가 범위가 각기 다르며 희소성 수준 측면에서는 Jester가 가장 밀집된 평가데이터를 갖는다. MovieLens 1M은 원래 6,040명의 사용자가 평가한 데이터를 포함하나, 본 실험에서는 사용한 컴퓨터 시스템의 용량 문제로 인하여 1,000명의 사용자로 국한하여 실험하였다. 희소성 수준(sparsity level)은 1-총평가개수/행렬크기로 산출한다.

매 실험 결과의 신뢰도를 향상시키기 위하여 5회 크로스 확인(5-fold cross validation)을 실시한 후에 결과들의 평균값을 산출하였다. 각 실험은 전체 데이터를 훈련 데이터와 시험 데이터로 구별하여, 우선 훈련 데이터를 이용하여 각 사용자별로 큰 유사도를 가진 인접 이웃(topNN)들을 구한 후에, 시험 데이터를 이용하여 평가기준에 따른 결과를 산출하였다. 훈련 데이터와 시험 데이터의 비율은 80:20으로 구성하였다. 즉, 전체 평가 데이터 집합을 다섯 개의 부분집합으로 나눈 후, 매 실험에서 하나의 부분집합을 시험데이터로 정하고, 나머지 4/5의 데이터를 훈련 데이터로 정하여 진행하였다. 이와 같이 다섯 번의 실험을 실행한 후 얻어진 성능 결과들의 평균치를 최종 결과로서 본 논문에서 제시하였다.

특정한 한 사용자 u의 topNN을 구하는 상세 절차를 설명하자면, 우선 u의 훈련데이터 내의 평가치들과 다른 사용자 v의 훈련데이터 내 평가치들을 대상으로 유사도를 산출한다. 이 때 물론 유사도는 미리 선정된 유사도 척도에 따라 값을 산출하는 것이다. 이와 같이 모든 다른 사용자들 각각과 u와의 유사도를 산출한 후, 이들 유사도를 내림차순 정렬하면 가장 큰 유사도를 가진 topNN 개의 이웃 사용자들을 얻게 된다. 여기서 가장 적합한 크기의 topNN은 실험에 의하여 정할 수도 있지만, 대개 크기를 변화시켜 그에 따른 성능 변화를 측정한다.

관련 연구에서 성능 평가의 기준은 크게 예측 성능과 추천 성능의 두 측면을 고려한다. 예측 성능이란 사용자가 아직 미평가한 항목에 대해 시스템이 예측 평가치를 산출한 후 이를 그

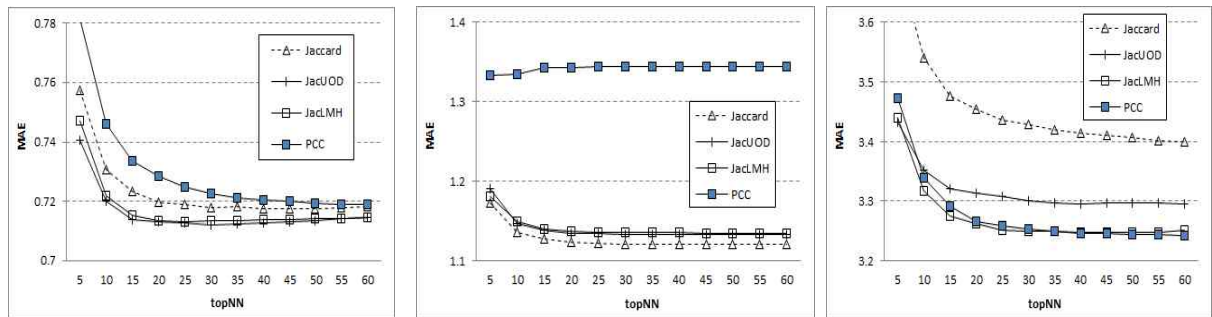


Fig. 1. MAE for varying topNNs: MovieLens (ML) (left), BookCrossing (BX) (middle), Jester (JT) (right)

사용자가 실제로 부여한 평가치와 비교해 보는 것으로서, 물론 실제치에 가장 근접한 예측치를 산출하게 되면 시스템의 성능이 우수한 것으로 판단한다. 기존 문헌에서 대개 MAE (Mean Absolute Error, 평균절대오차)를 예측 성능의 척도로 사용하는데, 평가 예측치와 실제치 차이의 절대값 평균으로 측정한다 [2]. 구체적으로 사용자 u 의 미평가 항목 x 에 대한 예측치는 대개 아래와 같이 산출하는데 이 때 $sim(u, v)$ 는 시스템에서 선정하여 사용하는 유사도 척도에 따른 u 와 v 간의 유사도값이고, \bar{r}_u 는 두 사용자의 공통평가항목들에 대한 u 의 평균 평가치이다 [12].

$$r'_{u,x} = \bar{r}_u + \frac{\sum_{v \in topNN_u} sim(u,v) \times (r_{v,x} - \bar{r}_v)}{\sum_{v \in topNN_u} |sim(u,v)|} \dots\dots\dots(8)$$

성능 평가 기준의 또다른 척도인 추천 성능을 평가하기 위해 대개 precision(P)과 recall(R)을 이용하는데, P는 시스템의 추천항목들 중 사용자가 실제로 선호하는 항목들의 비율을 나타내며, R은 사용자의 전체 선호항목들 중에서 시스템이 추천한 항목들의 비율이다 [13]. 시스템은 높은 평가예측치를 얻은 항목들을 우선적으로 추천한다. P와 R은 서로 상반된 결과치를 대개 보여주므로, 이들을 합한 척도인 F1을 종종 추천 성능의 지표로 활용하게 된다. F1은 P와 R의 조화평균값, 즉, $F1(2PR/(P+R))$ 으로 계산한다 [13].

마지막으로 3.2절에서 언급한 파라미터값의 설정을 위해 Lbd와 Hbd 값의 조합을 다양하게 변화시켜 실험한 후 MAE 결과를 측정하여 가장 최적의 성능을 나타내는 조합을 구하였다. 표 2는 실험에 사용한 값의 구간과 최종적으로 결정된 파라미터값을 제시한다. 이와 같이 결정된 값을 다음 절에 기술할 다양한 실험에 이용하였다.

Table 2. Lbd and Hbd values for each dataset

	실험값 구간		최종값	
	Lbd	Hbd	Lbd	Hbd
MovieLens 1M	1~3	3~5	2	5
BookCrossing	1~4	6~9	4	6
Jester	(Lbd, Hbd)를 (-7,7)~(-2.5,2.5)까지 각각 동일하게 0.5씩 변화하여 실험.		-3	3

2. Results

그림 1은 세 가지 서로 다른 특성의 데이터셋을 이용하여 실험한 MAE(평균절대오차)의 결과이다. 인접 이웃수를 변화시켜 각 척도의 성능 변화 추이를 알아보았다. 전반적으로 인접이웃수가 증가함에 따라 MAE는 초기에 급격히 감소하다가 안정화되는 경향을 보인다. 그림에서 PCC는 예상 밖의 결과를 보이는데, MovieLens(ML)와 BookCrossing(BX)에서 상대적으로 매우 저조한 성능을 보였고, 특히 후자의 데이터셋에서 더욱 그러하였는데, 이는 PCC가 매우 희소한 데이터셋에서 종종 불합리한 유사도 값을 산출하기 때문이다 [4][5]. BX 상에서 나타난 PCC 성능과 다른 척도들의 성능 차이가 매우 큰 것을 볼 때, 데이터 희소성이 PCC 유사도 척도에 얼마나 중요한 영향을 미치는지 확인할 수 있다.

예상 밖으로 Jaccard의 성능이 BX에서 매우 뛰어난데, 평가치를 전혀 반영하지 않은 척도임에도 불구하고 공통평가항목 개수만으로 가장 월등한 MAE 성능을 보였다. 또한 BX 보다 희소성이 다소 향상된 ML에서도 다른 두 변종 척도(JacUOD와 JacLMH) 보다 성능이 다소 떨어지지만 PCC 보다는 우수한 성능 결과를 보였다. 반면에 Jester(JT)에서는 예상대로 Jaccard의 성능이 가장 크게 저조한데 이는 물론 데이터셋이 밀집되어 평가치를 전혀 고려하지 않은 Jaccard의 방법이 불리하기 때문이다.

한편 JacUOD와 JacLMH의 결과를 비교해보면 ML과 BX에서 거의 동일한 성능을 보인 반면, JT에서는 JacLMH가 훨씬 우수하였다. 따라서, 두 사용자의 평가치 차이의 크기를 고려한 JacUOD가 평가치는 전혀 고려하지 않은 채 구간별 공통평가항목수 비율만을 고려한 JacLMH에 비해, 희소한 데이터셋에서도 더 미세한 성능을 보이지 않고, 밀집된 데이터 환경에서는 오히려 저조한 성능을 보인 것이다. 특히, JacLMH는 JT에서 PCC와 거의 동일한 성능 결과를 보임으로써 평가치를 고려하지 않은 척도가 예상과는 달리 뛰어난 성능 결과를 나타낼 수 있다는 새로운 사실이 드러났다.

그림 2는 F1 결과를 나타내는데, 이를 위해 각 데이터셋이 허용한 평가치 범위의 80% 이상인 평가 예측치를 가진 항목들을 추천하였다. 즉, ML에선 평가예측치가 4, BX에선 8, JT에선 6 이상이면 해당 항목을 추천한다. MAE 결과와 유사한 결

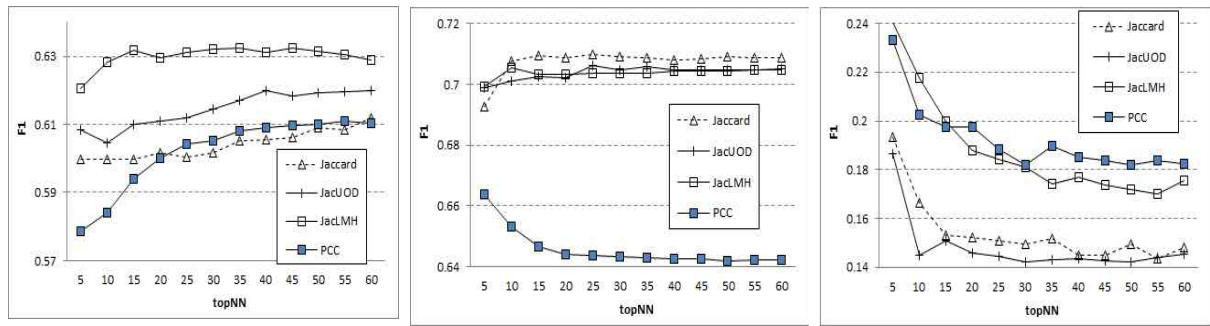


Fig. 2. F1 for varying topNNs: MovieLens(ML) (left), BookCrossing(BX) (middle), Jester(JT) (right)

과를 그림 2의 BX 결과에서 볼 수 있으며, ML에서는 JacLMH 결과가 다른 척도들에 비해 유독 우수한데 이는 MAE 결과와는 다소 상이하다. 또한 JT의 MAE 결과에서는 Jaccard 만이 상대적으로 매우 저조한 결과를 보였으나, F1 결과에서는 JacUOD가 Jaccard 보다도 다소 낮은 결과를 보여 모든 척도들 중에서 가장 성능이 낮았다. 결과적으로 JacLMH의 상대적 F1 성능은 MAE 결과보다 더욱 우수하였고, 반면에 JacUOD의 F1 성능은 MAE 결과에서의 비교 결과보다 더욱 좋지 못한 것을 알 수 있다.

V. Conclusions

자카드 계수는 기존의 전통적인 유사도 척도와 결합되어 협력 필터링의 근원적인 문제점인 데이터 희소성 문제를 다소 해결하기 위해 활용되어 왔다. 본 연구에서는 이러한 자카드 계수의 문제점을 분석하고 이를 보완하여 개선된 성능의 새로운 계수를 제안하였다. 제안된 계수를 단일 유사도 척도로 협력 필터링에 적용하여 다양한 특성의 데이터셋을 이용한 실험 결과, 전통적인 피어슨 상관도의 유사도 척도보다 월등한 예측 및 추천 성능을 나타냈다. 또한 매우 희소한 데이터셋 상에서는 자카드 계수보다 매우 근소하게 저조한 성능을 보였으나, 데이터셋이 밀집될수록 제안 계수의 성능이 더욱 우수하였다.

REFERENCES

- [1] D. Jannach, Z. Karakaya, and F. Gedikli, "Accuracy Improvements for Multi-criteria Recommender Systems," Proc. of the ACM Conf. Electronic Commerce, pp. 674-689, 2012.
- [2] X. Su and T.M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," Advances in Artificial Intelligence, 2009.
- [3] A. Bellogin and A. P. de Vries, "Understanding Similarity Metrics in Neighbour-based Recommender Systems," Proceedings of the 2013 Conference on the Theory of Information Retrieval, 2013.
- [4] H. Ahn, "A New Similarity Measure for Collaborative Filtering to Alleviate the New User Cold-starting Problem," Information Sciences, Vol. 178, No. 1, pp. 37-51, Jan. 2008.
- [5] S. Lee, "Performance Analysis of Similarity Reflecting Jaccard Index for Solving Data Sparsity in Collaborative Filtering," The Journal of Korean Association of Computer Education, Vol. 19, No. 4, pp. 59-66, July 2016.
- [6] H. Liu, Z. Hu, A. Mian, H. Tian, and X. Zhu, "A New User Similarity Model to Improve the Accuracy of Collaborative Filtering," Knowledge-Based Systems, Vol. 56, pp. 156-166, Jan. 2014.
- [7] J. Bobadilla, F. Ortega, A. Hernando, and J. Bernal, "A Collaborative Filtering Approach to Mitigate the New User Cold Start Problem," Knowledge-Based Systems, Vol. 26, pp. 225-238, Feb. 2012.
- [8] G. Koutrica, B. Bercovitz, and H. Garcia, "FlexRecs: Expressing and Combining Flexible Recommendations," Proc. of the ACM SIGMOD Int'l Conf. on Management of Data, pp. 745-758, 2009.
- [9] J. Bobadilla, F. Serradilla, and J. Bernal. "A New Collaborative Filtering Metric That Improves the Behavior of Recommender Systems," Knowledge-Based Systems, Vol. 23, pp. 520-528, Aug. 2010.
- [10] K. G. Saranya, G. S. Sadasivam, and M. Chandralekha, "Performance Comparison of Different Similarity Measures for Collaborative Filtering Technique," Indian Journal of Science and Technology, Vol. 9, No. 29, Aug. 2016.

- [11] H-F. Sun, et al., "JacUOD: A New Similarity Measurement for Collaborative Filtering," *Journal of Computer Science and Technology*, Vol. 27, No. 6, pp. 1252-1260, Nov. 2012.
- [12] B. Jeong, J. Lee, and H. Cho, "Improving Memory-based Collaborative Filtering via Similarity Updating and Prediction Modulation," *Information Sciences*, Vol. 180, No. 5, pp. 602-612, Mar. 2010.
- [13] M. Gao, Z. Wu, and F. Jiang, "UserRank for Item-based Collaborative Filtering Recommendation," *Information Processing Letters*, Vol. 111, pp. 440-446, Apr. 2011.

Authors



Soojung Lee received the B.S. degree in Mathematics Education from Ewha University, Korea in 1985. She received M.S. and Ph.D. degrees in Computer Science from Texas A&M University,

U.S.A, in 1990 and 1994, respectively.

Dr. Lee joined the faculty of the Department of Computer Education at Gyeongin National University, Gyunggi-do, Korea, in 1998, as a professor. She is interested in recommender systems, information filtering, data mining techniques, and computer education.