

일관성 기반의 신뢰도 정의에 의한 협업 필터링

Collaborative Filtering by Consistency Based Trust Definition

김형도(Hyoung-Do Kim)*

초 록

사용자간 유사도에 의한 협업 필터링에서 추천 품질이 안정적인 상태에 이르기 위해서는 많은 이웃들이 필요하다. 이것은 높은 사용자간 유사도가 제품에 대한 동일한 선호도를 일관되게 보장하지 못하기 때문이다. 유사하지 않은 사용자라 할지라도 제품 선택에서 사용자간에 일관성이 있다면, 선호도 예측에서 유용하게 사용될 수 있다. 이 논문에서는 일관성을 기준으로 신뢰도를 정의하고, 이를 기반으로 이웃을 선정하여 선호도를 예측하는 협업 필터링 방법을 제시한다. 이 방법에 의한 추천 품질이 안정적인 상태에 이르기 위해서 필요한 이웃들의 수가 유사도에 의한 방법보다 매우 적으며, 추천 품질 또한 우수하다.

ABSTRACT

Many neighbors are needed for making the recommendation quality better and stable in collaborative filtering. Furthermore, the quality is not so good mainly due to a reason that high similarity between two users does not guarantee the same preference to items considered for recommendation. Dissimilar users who have consistency in item selection can be useful for predicting preferences. This paper proposes a new collaborative filtering method, defining trust based on consistency for improving this phenomenon. Empirical studies show that such a method reduces the number of neighbors required to make the recommendation quality stable and the recommendation quality itself is also significantly improved.

키워드 : 추천, 협업 필터링, 일관성, 신뢰도

Recommendation, Collaborative Filtering, Consistency, Trust

* 한양사이버대학교 경영학부

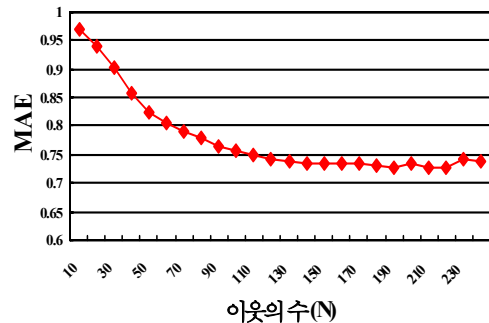
2008년 07월 15일 접수, 2009년 01월 30일 심사완료 후 2009년 02월 07일 게재확정.

1. 서론

추천 시스템은 사용자가 가장 적합한 정보를 발견할 수 있도록 도와 주는데 목적이 있다[1, 2, 20]. 이러한 추천 시스템 중에서 가장 성공적인 기술은 협업 필터링(Collaborative Filtering)[3, 5, 9, 16, 17]으로서, 이것을 크게 보면 메모리 기반의 협업 필터링과 모델 기반의 협업 필터링으로 구분할 수 있다. 전자는 예측을 하기 위해서 두 사용자 간의 또는 제품 간의 유사도를 측정하며, 가장 활발하게 사용되고 있는 추천 기법이다[13, 19, 21]. 이 기법에서의 유사도 측정 방법은 다양하며, 선정된 유사도 측정 방법은 예측의 정확성을 결정하게 된다[8]. 일반적으로 유사도가 미리 계산될 수 없기 때문에, 메모리 기반의 협업 필터링 시스템은 매 추천시마다 유사도를 계산해야 한다. 이러한 메모리 기반 협업 필터링의 단점은 문제의 크기가 확장되는데 대한 대응이 어렵고, 데이터 희소성에 대한 민감도가 낮다는 것이다. 모델 기반의 협업 필터링[10, 14]은 선형대수, 뉴럴 네트워크, 클러스터링 등을 기반으로 사전에 모델을 수립하여 신속하게 추천을 해줄 수 있다. 하지만 모델을 학습하거나 개선하는 단계에서는 많은 노력이 수반된다는 단점이 있다.

메모리 기반의 협업 필터링의 일종인 사용자 기반의 협업 필터링은 각 상품에 대한 사용자의 선호도로부터 수집된 프로필을 생성함으로써 추천을 하게 된다. 선호도는 일반적으로 사용자에게 의해서 평가된 숫자 값으로 표현되는데, 어떤 사용자의 새로운 제품의 선호도를 예측하는 것은 이 제품에 대한 다른 사용자들의 평가치를 이용하여 계산된다. 따

라서, 제품에 대한 평가에서 이 사용자와 유사한 이웃들을 발견하는 것이 매우 중요하다 [11, 13]. 이웃을 발견하는 데 있어서 일반적으로 Pearson 상관계수가 사용된다. <그림 1>은 MovieLens 데이터집합[6]을 대상으로 Pearson 상관계수를 사용하여 가장 가까운 이웃 N명을 선정하여 추천을 한 경우의 평균절대예러(Mean Absolute Error : MAE로 약칭) 값을 정리한 것이다. 이 그림에서 보면, 안정적인 결과를 얻기 위해서는 100명 이상의 이웃이 필요함을 알 수 있다. 이렇게 많은 이웃이 사용되어야 한다면, 추천 시스템의 성능은 크게 악화될 수 있다. 이렇게 많은 이웃이 필요한 이유는 결국 Pearson 상관계수에 의한 이웃의 정의와 평가치의 예측이 부정확함을 의미하며, 많은 수의 사용자에게 의한 평탄화 효과가 필요함을 의미한다. 많은 수의 이웃이 필요하면, 결국 위에서 언급한 메모리 기반 협업 필터링의 단점이 확대되게 된다.



<그림 1> NEAREST-N의 평균절대예러(MAE)

현실 세계에서 살펴볼 때, 사용자간의 유사성보다는 두 사용자간의 일관성이 예측에는 더 유용할 수 있음을 알 수 있다. 예를 들어 어떤 사용자 A와는 정반대로 평가하는 사

용자 B가 있다고 한다면, B의 이러한 특성을 이용하여 사용자 A의 평가치를 예측할 수도 있을 것이다. 유사성을 기본으로 할 경우 B는 A의 이웃이 절대로 될 수 없다. 따라서 이웃 선정에서는 사용자간의 유사성보다는 일관성을 우선 고려해볼 필요가 있다. 이러한 일관성은 사용자간의 신뢰도로 확대하여 정의할 수 있다. 웹스터 사전에 의하면, 신뢰도는 성격, 능력, 힘, 사실에 대한 확실한 의존을 의미한다.

신뢰도 기반의 협업 필터링에 관한 연구는 최근에 관심을 받기 시작하고 있으나, 신뢰도의 정도를 결정하는 방법이 불명확하거나, 유사도를 신뢰도로 사용하거나, 이웃이 중간에 변경되지 못한다는 단점이 있다[15, 18, 21]. 이러한 문제점을 해결하기 위해서 이 논문에서는 엔트로피를 이용하여 일관성을 측정할 수 있는 신뢰도 정의를 제시하고, 이를 기반으로 이웃 선정 및 평가에 대한 예측을 처리할 수 있는 방법을 제시한다.

이 논문의 구성은 다음과 같다. 제 2장에서는 일관성을 기반으로 신뢰도를 정의하는 방법을 제시하고, 제 3장에서는 이러한 신뢰도를 기반으로 예측하는 여러 방법들을 설명한다. 제 4장에서는 사례 데이터 집합을 사용하여 실험 결과를 제시하고, 그 의미를 논한다. 마지막으로 제 5장에서는 결론을 맺고, 앞으로의 연구방향을 제시한다.

2. 일관성 기반의 신뢰도 정의

정보이론에서 엔트로피는 랜덤변수와 관련된 불확실성을 측정하기 위한 수단으로 사용

된다. 엔트로피는 데이터에 포함되어 있는 정보를 계량할 수 있어서, 데이터 압축(Data Compression)이나 의사결정나무(Decision Tree) 수립 등에 잘 적용되고 있다(x_1, \dots, x_M)의 값들을 취할 수 있는 이산 랜덤변수의 엔트로피는 식 (1)과 같이 표현할 수 있다. 여기서 만약 $M = 2$ 이고, $X = x_1$ 일 확률이 p 라고 한다면, 이 엔트로피는 식 (2)와 같이 표현될 수 있다. p 값에 따른 엔트로피의 변화는 <그림 2>와 같은 이차의 엔트로피 함수로 표현된다. 여기서 p 값이 0.5일 때 엔트로피 값이 최대이며, 이로부터 멀어질수록 엔트로피 값이 감소함을 알 수 있는데, 이것은 p 값이 0.5일 때 불확실성이 최대임을 나타낸다.

추천해주고자 하는 사용자의 엔트로피는 과거 평가자료로부터 각 평가치의 횟수를 바탕으로 식 (3)과 같이 계산할 수 있다. 여기서 Z 은 평가치의 상태 수를 의미하며, N 은 총 평가횟수를 나타낸다. <표 1>과 같이 1점에서 5점까지 정수로 10회 평가를 하였다면, $Z = 5$ 이고, $N = 10$ 이 된다.

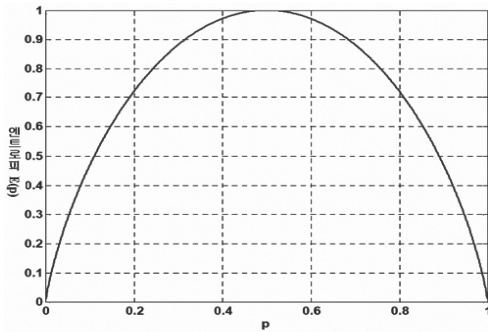
$$E(X) = - \sum_{i=1}^M \Pr(X = x_i) \log_2 \Pr(X = x_i) \quad (1)$$

$$\text{단, } \sum_{i=1}^M \Pr(X = x_i) = 1$$

$$E(p) = -p \log_2 p - (1-p) \log_2 (1-p) \quad (2)$$

$$\text{Entropy} = - \sum_{i=1}^Z \frac{n_i}{N} \log \frac{n_i}{N} \quad (3)$$

만약 추천해주고자 하는 사용자의 이웃이 주어진 상황이라면, 엔트로피는 좀 더 정교하게 계산될 수가 있다. 추천해주고자 하는



〈그림 2〉 2차 엔트로피 함수

〈표 1〉 Z과 N의 의미

평가치 i	횟수 n _i	비율 n _i /N
1	n ₁ = 2	n ₁ /N
2	n ₂ = 1	n ₂ /N
3	n ₃ = 3	n ₃ /N
4	n ₄ = 0	n ₄ /N
5 = Z	n ₅ = 4	n ₅ /N
합계	N = n ₁ + n ₂ + n ₃ + n ₄ + n ₅ = 10	1

사용자 u 와 그 이웃으로 선정된 사용자 v 의 평가치들을 비교하여 테이블 형태로 작성해 보면 〈그림 3〉과 같다. 이 상태에서의 엔트로피는 식 (4)와 같이 정리할 수 있다.

따라서 이웃이 주어질 경우 개선되는 엔트로피는 식 (3)에서 식 (4)를 차감한 값으로 계산할 수 있으며, 이러한 개선이 가장 큰 이웃을 선정하여 추천을 하게 되면 좋은 추천을 얻을 수 있을 것으로 판단된다. 왜냐하면, 엔트로피가 감소한다는 것은 불확실성이 감소한다는 의미이며, 불확실성이 감소한다는 것은 일관성이 증가한다는 것이기 때문이다.

하지만 식 (3)과 식 (4)는 동일한 차원이 아니기 때문에 식 (3)을 식 (4)와 같은 2차원

v에 의한 평가

u 에 의 한 평 가		1	2	3	4	5
	1	n ₁₁	n ₁₂	n ₁₃	n ₁₄	n ₁₅
	2	n ₂₁	n ₂₂	n ₂₃	n ₂₄	n ₂₅
	3	n ₃₁	n ₃₂	n ₃₃	n ₃₄	n ₃₅
	4	n ₄₁	n ₄₂	n ₄₃	n ₄₄	n ₄₅
	5	n ₅₁	n ₅₂	n ₅₃	n ₅₄	n ₅₅

〈그림 3〉 이웃이 주어진 경우의 횟수 테이블

Entropy($u|v$)

$$\begin{aligned}
 &= - \sum_{i=1}^Z \sum_{j=1}^Z \frac{n_{ij}}{N} \log \frac{n_{ij}}{N} \\
 &= - \sum_{i=1}^Z \sum_{j=1}^Z \frac{n_{ij}}{N} (\log n_{ij} - \log N) \\
 &= - \sum_{i=1}^Z \sum_{j=1}^Z \frac{n_{ij}}{N} \log n_{ij} + \sum_{i=1}^Z \sum_{j=1}^Z \frac{n_{ij}}{N} \log N \\
 &= \log N - \frac{1}{N} \sum_{i=1}^Z \sum_{j=1}^Z n_{ij} \log n_{ij}
 \end{aligned} \tag{4}$$

으로 재정리할 필요가 있다. 즉, 〈그림 4〉와 같이 가상의 이웃에 동일하게 분포되어 있는 것으로 가정할 수 있으며, 이 경우의 엔트로피는 식 (5)와 같이 계산할 수 있다.

이웃이 없는 경우의 평가

u 에 의 한 평 가		1	2	3	4	5	Total
	1	n ₁ /Z	n ₁ /Z	n ₁ /Z	n ₁ /Z	n ₁ /Z	n ₁
	2	n ₂ /Z	n ₂ /Z	n ₂ /Z	n ₂ /Z	n ₂ /Z	n ₂
	3	n ₃ /Z	n ₃ /Z	n ₃ /Z	n ₃ /Z	n ₃ /Z	n ₃
	4	n ₄ /Z	n ₄ /Z	n ₄ /Z	n ₄ /Z	n ₄ /Z	n ₄
	5 = Z	n ₅ /Z	n ₅ /Z	n ₅ /Z	n ₅ /Z	n ₅ /Z	n ₅

〈그림 4〉 이웃이 없는 경우의 횟수 테이블

$$\begin{aligned}
 & Entropy(u) \\
 &= -\sum_{i=1}^Z \sum_{j=1}^Z \frac{n_i}{ZN} \log \frac{n_i}{ZN} \\
 &= -\sum_{i=1}^Z \frac{Zn_i}{ZN} (\log n_i - \log ZN) \\
 &= -\sum_{i=1}^Z \frac{n_i}{N} \log n_i + \sum_{i=1}^Z \frac{n_i}{N} \log ZN \\
 &= \log ZN - \frac{1}{N} \sum_{i=1}^Z n_i \log n_i
 \end{aligned} \tag{5}$$

이상과 같이 두 사용자 간의 엔트로피 정의는 두 사용자 간에 일관성이 있는 평가가 많을수록 낮은 엔트로피 값을 가지게 됨을 의미한다. 즉, 두 사용자 간의 엔트로피 값이 낮을수록(불확실성이 낮을수록) 일관성이 높다고 말할 수 있다. 식 (4)가 두 사용자 간 일관성의 수준을 표현하는 반면에, 식 (5)는 한 사용자만의 일관성의 수준을 나타낸다. 바버(Baber)가 정의한 것처럼 신뢰도(Trust)를 “기술적으로 능력있는 역할의 수행에 대한 기대”이라고 본다면, 신뢰도를 식 (6)과 같이 엔트로피의 개선비율로 정의할 수 있다.

$$\begin{aligned}
 & Trust(u \rightarrow v) \\
 &= \frac{Entropy(u) - Entropy(u|v)}{Entropy(u)} \\
 &= \frac{(\log ZN - \frac{1}{N} \sum_{i=1}^Z n_i \log n_i) - (\log N - \frac{1}{N} \sum_{i=1}^Z \sum_{j=1}^Z n_{ij} \log n_{ij})}{\log ZN - \frac{1}{N} \sum_{i=1}^Z n_i \log n_i} \\
 &= \frac{(1 - \frac{1}{N \log ZN} \sum_{i=1}^Z n_i \log n_i) - (\frac{\log N}{\log ZN} - \frac{1}{N \log ZN} \sum_{i=1}^Z \sum_{j=1}^Z n_{ij} \log n_{ij})}{1 - \frac{1}{N \log ZN} \sum_{i=1}^Z n_i \log n_i} \tag{6} \\
 &= \frac{(1 - \frac{\log N}{\log ZN}) + \frac{1}{N \log ZN} (\sum_{i=1}^Z \sum_{j=1}^Z n_{ij} \log n_{ij} - \sum_{i=1}^Z n_i \log n_i)}{1 - \frac{1}{N \log ZN} \sum_{i=1}^Z n_i \log n_i}
 \end{aligned}$$

3. 신뢰도 기반의 예측방법

추천해주고자 하는 사용자와 신뢰도가 높은 것으로 판정된 이웃 사용자(들)의 과거 평가자료를 이용하여 사용자의 평가치를 예측할 수 있다. 가장 단순한 방법은 신뢰도가 가장 높은 이웃을 선정하여, 그 이웃의 상품에 대한 과거 평가치를 이용하는 방법이다. 이것은 식 (7)과 같이 상품 x에 대한 사용자 u의 평가치 예측은 이웃 v의 x에 대한 과거 평가치 $r_{v,x}$ 을 그대로 이용하는 것으로, 이 방법을 약칭하여 NAIVE라고 칭한다.

$$\begin{aligned}
 & NAIVE : \\
 & p_{u,x} | r_{v,x} = r_{v,x}
 \end{aligned} \tag{7}$$

또 다른 방법으로는 식 (8)과 같이 상품 x에 대한 이웃 v의 평가치 $r_{v,x}$ 와 같이 이웃이 평가한 상품들에 있어서, 사용자 u의 과거 평가치들을 가중평균하여 예측치를 구하는 것으로, 이것을 앞으로 HISTORIC이라고 부른다.

$$\begin{aligned}
 & HISTORIC : \\
 & p_{u,x} | r_{v,x} = \frac{\sum_{k=1}^Z k \cdot n_{k, r_{v,x}}}{\sum_{k=1}^Z n_{k, r_{v,x}}} \quad \text{if } \sum_{k=1}^Z n_{k, r_{v,x}} \neq 0
 \end{aligned} \tag{8}$$

과거 예측치의 차이를 고려하여 예측하는 AVGDIF 방법도 가능하다. 식 (9)와 같이, 상품 x에 대한 이웃 v의 평가치 $r_{v,x}$ 를 기본으로 하되, 전체적으로 두 사용자의 평균 값의 차이를 고려하는 것도 좋은 평가치 예측 방법이 될 수 있다.

$$\begin{aligned}
 & \text{AVGDIFF} \\
 & p_{u,x} | r_{v,x} = \\
 & = r_{v,x} + \frac{\sum_{k=1}^Z \sum_{l=1}^Z (k-l) \cdot n_{k,l}}{\sum_{k=1}^Z \sum_{l=1}^Z n_{k,l}} \quad (9) \\
 & \cong r_{v,x} + \bar{r}_u - \bar{r}_v
 \end{aligned}$$

복합적으로 HISTORIC과 NAIVE를 결합하거나, HISTORIC을 AVGDIFF와 결합하는 것도 가능하다.

유사성을 기반으로 한 예측의 경우와 마찬가지로, 이상과 같이 각각의 이웃을 이용하여 산출된 값들을 결합하여 보다 좋은 결과를 얻을 수 있다. 식 (10)과 같이 평활법을 이용하면 된다.

$$\begin{aligned}
 & p_{u,x} = \frac{\sum_{v \in \text{Neighbor}(u)} \text{Trust}(u \rightarrow v) \cdot p_{v,x} | r_{v,x}}{\sum_{v \in \text{Neighbor}(u)} \text{Trust}(u \rightarrow v)} \quad (10) \\
 & v \in \text{Neighbor}(u), \\
 & \text{if } \text{Trust}(u \rightarrow v) > \theta
 \end{aligned}$$

신뢰도를 이용한 예측방법의 장점중의 하나는 신뢰도 전파(Propagation)를 이용할 수 있다는 것이다. 신뢰도 전파를 이용하면 직접적인 신뢰관계에 있지 않는 사용자들 간에도 신뢰도를 계산할 수 있게 되어, 희소성 문제를 어느 정도 극복할 수 있게 된다. 사용자 u 의 신뢰하는 이웃으로 선정된 사용자 v 가 해당 상품에 대한 평가를 한 자료가 없다면, 우리는 v 가 신뢰하는 이웃으로 생각하는 또 다른 사용자의 평가 자료를 이용할 수 있다는 점이다. 이 경우에도 여러 가지의 예측방

법이 가능한데, 이 논문에서는 3가지 방법을 고려한다. 먼저 p-NAIVE의 경우, 식 (11)과 같이 사용자 v 의 예측치 $p_{v,x}$ 를 그대로 이용하는 방법이다. p-AVGDIFF의 경우에는 식 (12)와 같이 $p_{v,x}$ 를 기본으로 하되, 두 사용자 간 평균치의 차이를 고려한다. 마지막으로 p-REESTIMATE는 $p_{v,x}$ 가 정수가 아닌 점을 고려하여, $p_{v,x}$ 의 위와 아래에 위치한 두 정수를 이용하여 예측치를 계산하고, 이를 가중 평균한 값이다.

$$\begin{aligned}
 & p\text{-NAIVE} : \\
 & p_{u,x} | p_{v,x} = p_{v,x} \quad (11)
 \end{aligned}$$

$$\begin{aligned}
 & p\text{-AVGDIFF} \\
 & p_{u,x} | p_{v,x} = p_{v,x} + \bar{r}_u - \bar{r}_v \quad (12)
 \end{aligned}$$

$$\begin{aligned}
 & p\text{-REESTIMATE} \\
 & p_{u,x} | p_{v,x} = \\
 & [p_{u,x} | \text{floor}(p_{v,x})] \times [p_{v,x} - \text{floor}(p_{v,x})] + \\
 & [p_{u,x} | \text{ceil}(p_{v,x})] \times [\text{ceil}(p_{v,x}) - p_{v,x}] \quad (13)
 \end{aligned}$$

4. 실험 결과

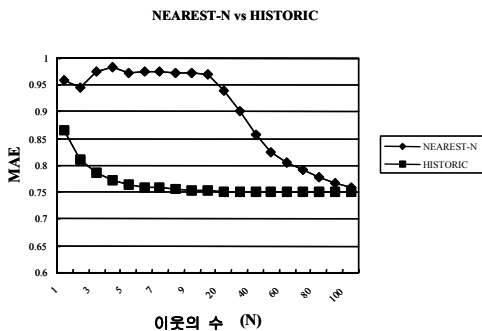
일관성 기반의 신뢰도 정의에 의한 협업 필터링 방법에 대하여 MovieLens 데이터집합[6]을 이용하여 모의 실험하였다. 이 데이터집합은 1682개 영화를 943명이 평가한 약 10만개의 기록을 포함한다. 이 방법에 의한 예측치의 품질을 평가하기 위하여, 데이터집합의 90%는 훈련용으로, 나머지 10%는 테스트용으로 사용되었다. 이 방법의 정확성을 측정하고 평가하기 위해서 평균절대예러(MAE)

를 사용한다. 이것은 식 (11)과 같이 계산된다. 이 식에서 N은 총 예측 횟수이고, e_i 는 항목 i에 대한 예측치와 실제 평가치의 차이를 나타낸다.

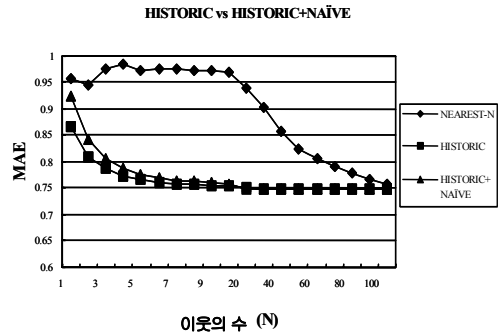
$$MAE = \frac{\sum_{i=0}^N |e_i|}{N} \quad (14)$$

먼저 HISTORIC의 결과를 NEAREST-N과 비교해보면, <그림 5>와 같이 MAE 값이 0.75수준으로 접근하는 이웃의 수 N이 매우 작음을 알 수 있다. 즉, NEAREST-N에서 100개 정도의 이웃을 사용해야 안정적인 MAE 값을 얻을 수 있는 반면에, HISTORIC에서는 10명 정도에서 안정적인 MAE 값을 얻을 수 있어서, 효율적인 추천이 가능함을 알 수 있다.

과거의 평가기록이 있으면 HISTORIC을 사용하고, 그렇지 못할 경우에는 NAÏVE를 사용하는 HISTORIC+NAÏVE를 HISTORIC과 비교해보면 <그림 6>과 같이 큰 차이가 없음을 알 수 있다.

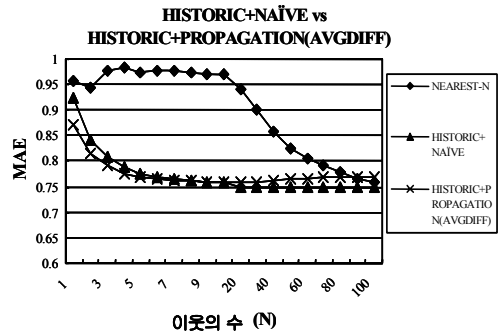


<그림 5> NEAREST-N과 HISTORIC의 MAE 비교



<그림 6> HISTORIC과 HISTORIC+NAÏVE의 MAE 비교

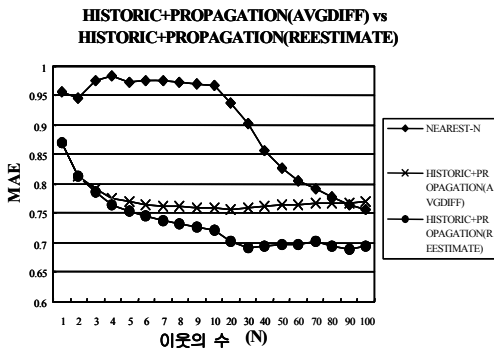
HISTORIC과 함께 과거의 평가기록이 없는 경우에만 AVGDIFF에 의한 PROPAGATION을 사용하게 되면, <그림 7>과 같이 이웃의 수가 10미만일 경우에는 HISTORIC보다 약간 좋은 결과인 반면, 10이상인 경우 오히려 나빠진다.



<그림 7> HISTORIC+NAÏVE와 HISTORIC+PROPAGATION(AVGDIFF)의 MAE 비교

HISTORIC+PROPAGATION(AVGDIFF)와 HISTORIC+PROPAGATION(REESTIMATE)에 의한 MAE 값은 <그림 8>과 같다. 전체적으로 PROPAGATION(REESTIMATE)가 더 우수함을 알 수 있다. 이웃의

수가 20명 이상일 경우에는 0.7이하의 MAE 값을 얻을 수 있다. MovieLens 데이터집합을 이용한 최근의 대표적인 협업 필터링 연구들[4, 7, 12]에서 MAE가 0.7이상임을 고려할 때, 이것은 매우 우수한 결과로 평가될 수 있다.



〈그림 8〉 HISTORIC+PROPAGATION(AVGDIFF)와 HISTORIC+PROPAGATION(REESTIMATE)의 MAE 비교

5. 결 론

이 논문은 일관성을 기반으로 하는 신뢰도를 정의하고 이를 이용한 이웃의 선정과 평가치 예측 방법을 제시하였다. 유사성 기반의 이웃 선정에 비하여 신뢰도 기반의 이웃 선정은 안정적인 결과를 얻기 위해 필요한 이웃의 수를 크게 줄일 수 있다. 추가적으로 신뢰도 전파는 협업 필터링의 희소성 문제(Sparsity Problem)를 감소시킴으로써 예측의 품질을 크게 개선할 수 있었다.

앞으로의 연구 방향으로 다차원의 신뢰도 전파와 분산 협업 필터링을 선정하여 추진하

고 있다. 현재는 1차원의 신뢰도 전파만을 대상으로 하고 있으나, 2차 이상의 다차원 신뢰도 전파를 통해서 보다 좋은 결과를 얻을 수 있을 것으로 예상하고 있다. 중앙집중적인 협업 필터링의 대안으로 등장하고 있는 분산 협업 필터링은 신뢰도가 높은 이웃들에 관한 소수의 빈도 테이블을 각각의 사용자가 자체적으로 유지함으로써, 부하가 한 곳에 집중되는 것을 방지하고 보안에 적극적으로 대비할 수 있는 장점을 가지고 있다.

참 고 문 헌

- [1] 박지수, 정석재, 이영훈, 김경섭, “퍼지 추론을 통한 규칙 기반의 보험상품 추천 및 설계 시스템 구현”, 한국전자거래학회지, 제12권, 제1호, pp. 99-122, 2007.
- [2] 이재식, 명훈식, “사례기반 추론을 이용한 인터넷 서점의 서적 추천시스템 개발”, 한국전자거래학회지, 제13권, 제4호, pp. 173-191, 2008.
- [3] Adomavicius, G. et al., “Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach,” ACM Transactions on Information Systems, Vol. 23, No. 1, pp. 103-145, Jan. 2005.
- [4] Ahn, H. J., “A New Similarity Measure for Collaborative Filtering to Alleviate the New User Cold-Starting Problem,” Information Sciences, Vol. 128, pp. 37-

- 51, 2008.
- [5] Chen, A., "Context-Aware Collaborative Filtering System : Predicting the User's Preferences in Ubiquitous Computing," Proceedings of CHI 2005, Portland, USA, Apr. pp. 2-7, 2005.
- [6] GroupLens Research, "MovieLens Datasets," <http://www.grouplens.org/taxonomy/term/14>.
- [7] Herlocker, J. L. et al., "An Algorithmic Framework for Performing Collaborative Filtering," Proceedings of the 22nd Annual International ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 230-237, Berkeley, USA, 1999.
- [8] Herlocker, J. L. et al., "Evaluating Collaborative Filtering Recommender Systems," ACM Transactions on Information Systems, Vol. 22, No. 1, pp. 5-53, 2004.
- [9] Kim, H.-N. et al., "Error-Based Collaboration Filtering Algorithm for Top-N Recommendation," Proceedings of APWeb 2007 and WAIM 2007(LNCS 4505), pp. 594-605, Huang Shan, China, June, pp. 16-18, 2007.
- [10] Kim, T.-H., and Yang, S.-B., "An Improved Neighbor Selection Algorithm in Collaborative Filtering," IEICE Transactions on Information and Systems, Vol. E88-D, No. 5, pp. 1072-1076, 2005.
- [11] Konstan, J. et al., "GroupLens : Applying Collaborative Filtering to Usenet News," Communications of the ACM, Vol. 40, No. 3, pp. 77-87, 1997.
- [12] Lee, H. C., Lee, S. J., and Chung Y. J., "A Study on the Improved Collaborative Filtering Algorithm for Recommender System," Proceedings of the 5th International Conference on Software Engineering Research, Management and Applications (SERA2007), pp. 297-304, 2007.
- [13] Lemire, D., and Maclachlan, A., "Slope One Predictors for Online Rating-Based Collaborative Filtering," Proceedings of the SDM'05, Newport Beach, USA, Apr, pp. 21-23, 2005.
- [14] Massa, P., and Avesani, P., "Trust-Aware Collaborative Filtering for Recommender Systems," Proceedings of the International Conf. on Cooperative Information Systems 2004.
- [15] Mohan, B. K., Keller, B. J., and Ramakrishnan, N., "Scouts, Promoters, and Connectors : The Roles of Ratings in Nearest Neighbor Collaborative Filtering," Proceedings of EC'06, pp. 250-259, Ann Arbor, Michigan, USA, Jun, pp. 11-15, 2006.
- [16] Nakamoto, R. et al., "Tag-Based Contextual Collaborative Filtering," Proceedings of DEWS2007, pp. 25-30, Hiroshima, Japan, 2007.
- [17] Papagelis, M, Plexousakis, D., and Kutsuras, T., "Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inference," Proceedings of iTrust 2005, pp. 224-239, 2005.

- [18] Sarwar B. et al., "Analysis of Recommendation Algorithms for e-Commerce," Proceedings of the 2nd ACM Conf. on Electronic Commerce, pp. 158-167, Minneapolis, USA, 2000.
- [19] Sarwar, B., Karypis, G., Konstan, J., and Reidl, J., "Item-Based Collaborative Filtering Recommendation Algorithms," Proceedings of the 10th Int'l WWW Conf., pp. 285-295, Hong Kong, May 1-5, 2001.
- [20] Shardanand, U., and Maes, P., "Social Information Filtering : Algorithms for Automating 'Word of Mouth'," Proceedings of the ACM CHI Conf. on Human Factors in Computing Systems, pp. 210-217, Denver, USA, May 1995.
- [21] Weng, J. et al., "Trust-Based Agent Community for Collaborative Recommendation," Proceedings of AAMAS'06, pp. 1260-1262, Hakodate, Hokkaido, Japan, May 8-12, 2006.

저 자 소 개



김형도

1985

1987

1992

1993~1999

2000~2002

2002~현재

2003~현재

2004~2006

2007~현재

관심분야

(E-mail : hdkim@hycu.ac.kr)

서울대학교 산업공학과 (학사)

한국과학기술원 경영과학과 (석사)

한국과학기술원 경영과학과 (박사)

(주)데이콤 EC인터넷 연구/기술 팀장

아주대학교 정보통신전문대학원 교수

전자상거래표준화통합포럼 전자문서기술위원회 부위원장

한양사이버대학교 경영학부 교수

ebXML 전문위원회 위원장

한양사이버대학교 경영학부장

전자상거래, 비즈니스 프로세스, 디지털 워터마킹, e-러닝,
데이터마이닝 등