

# 소셜 트러스트 클러스터 효과를 이용한 견고한 추천 시스템 설계 및 분석\*

노기섭,<sup>1\*</sup> 오하영,<sup>2\*</sup> 이재훈<sup>3</sup>  
<sup>1</sup>청주대학교, <sup>2</sup>아주대학교, <sup>3</sup>서울대학교

## Design and Analysis a Robust Recommender System Exploiting the Effect of Social Trust Clusters\*

Giseop Noh,<sup>1\*</sup> Hayoung Oh,<sup>2\*</sup> Jaehoon Lee<sup>3</sup>  
<sup>1</sup>Cheongju University, <sup>2</sup>Ajou University, <sup>3</sup>Seoul National University

### 요약

추천시스템(Recommender System, RS)는 정보 과잉 공급 상태에서 사용자들에게 최적화된 정보를 제공하는 시스템이다. RS의 핵심은 사용자의 행동 결과를 정확하게 예측하는 것이다. 이러한 예측을 위해 Matrix Factorization (MF) 방식이 초기에 사용되었으며, 최근 SNS의 발달에 따라 Social Information을 추가적으로 활용하여 예측 정확도를 높이고 있다. 본 논문에서는 기존 연구에서 간과되었던 RS 내부 trust cluster를 이용하여 추가적으로 성능을 향상시키고, trust cluster의 특성에 대하여 분석한다. 기존 방법론 3가지와 비교한 결과 본 논문에서 제안하는 방식이 가장 높은 정확도를 보임을 확인하였다.

### ABSTRACT

A Recommender System (RS) is a system that provides optimized information to users in an over-supply situation. The key to RS is to accurately predict the behavior of the user. The Matrix Factorization (MF) method was used for this prediction in the early stage, and according to the recent SNS development, social information is additionally utilized to improve prediction accuracy. In this paper, we use RS internal trust cluster, which was overlooked in previous studies, to further improve performance and analyze the characteristics of trust clusters.

**Keywords:** Trust Cluster, Social Relation, Robust Recommender System, Prediction Accuracy

## 1. 서론

스마트 기기의 확산과 정보통신 기술의 급속한 발

달은 사회의 여러 분야에 많은 영향을 미치고 있다. 개별 사용자들은 끊임없이 정보를 생성하고 있으며, 이러한 정보들은 정보통신 기술을 활용하여 가상공간에 저장되거나 또 다른 사용자 또는 스마트 기기와 공유된다. 스마트 기기의 광범위한 보급은 개별 사용자들이 온라인으로 연결되는 결과를 가져 왔으며 거대한 online social network (OSN) 형성을 야기했다. OSN을 효과적으로 사용하고 관리하기 위하여 사용되는 서비스 또는 기술을 social network service (SNS)라고 한다. 이러한 문화적 변화는 사람들의 삶의 방식을 오프라인에서 온라인으로 변화시켰다.

Received(11. 14. 2017), Modified(12. 20. 2017), Accepted(12. 28. 2017)

\* This work was supported by the Ajou University research fund and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2017R1D1A1B03035557).

† 주저자, kafa46@gmail.com

‡ 교신저자, hyoh79@gmail.com (Corresponding author)

이런 거대한 정보의 생성으로 인터넷 사용자들은 정보의 홍수 시대를 맞이하게 되었고 특정 개별 사용자는 과도하게 제공되는 정보로부터 자신에게 필요한 정보를 취사선택하는 것에 어려움을 느끼기도 한다.

이러한 정보 과 제공으로 인한 문제를 information overload problem (IOP) 라고 하며[1], 이를 극복하여 사용자에게 최적의 정보를 제공하고자 등장한 것이 추천 시스템(recommender system, RS)이다. RS는 사용자들의 과거 이력 및 특성 등을 자동으로 분석하여 개별 사용자에게 필요한 정보만을 우선적으로 제공하는 시스템이다.

RS는 사용자 특성에 최적화된 솔루션을 제공한다는 의미에서 사용자 성향을 고려한 경로탐색, 음반/도서 등과 같은 상품 구매, 개별 사용자에게 어울리는 친구 추천 서비스 등 다양한 분야에 활용되고 있다. 최근에는 단순한 전자 상거래(E-commerce)가 OSN을 활용한 소셜 커머스 (social commerce)로 발전하였고, SNS와 결합하여 더 큰 효과를 내고 있다. RS는 Amazon, eBay, Etsy와 같은 온라인 마켓, Netflix, Naver Movie 같은 영화제공 서비스, Facebook, LinkedIn과 같은 인맥관리 등 다양한 분야에서 활용되고 있다.

전통적인 RS의 설계 접근법은 사용자가 과거에 생성한 문자 정보의 유사성을 이용하는 콘텐츠 기반(content based: CB) 방식과 다른 사용자들의 패턴을 파악하는 collaborative filtering (CF) 방식이 대표적이다. 그러나 기존 방식들은 RS를 구성하는 매트릭스에 사용자의 정보가 충분하지 않거나 계산 정확도가 부족하다는 한계점이 있다. 결과, 최근에는 SNS 기능이 대부분의 online service에 도입되면서 social information을 추가적으로 활용하는 연구가 진행되어 왔다. Social information은 명시적 정보와 암시적 정보로 구분 가능하다. 명시적 관계는 친구(friend) 관계, 신뢰(trust) 관계 및 팔로우(follow) 관계 등이 대표적이다. 암시적 관계는 리뷰에 대한 평점부여 및 찬반 의견(thumb up/down) 등의 관계가 존재한다. Social information으로부터 발생하는 social relation을 분석하면 상호 관계를 형성하는 사용자들 사이에 영향을 미치게 된다[2]. 예를 들면 물건을 구매할 경우 전혀 다른 사용자들 보다는 친구 관계에 있는 사용자의 의견이나 구매 성향에 영향을 받는 경우를 예로 들 수 있다. 이러한 social information 정보를 적절히 활용하면 RS의 예측

정확도를 향상할 수 있다.

그러나 이전의 연구들은 실제 social relation에도 다양한 형태가 존재한다는 점을 간과하였다. Social relation 상에서 사용자들의 의견은 신뢰관계(trust relation)에 따라 그룹 (group) 또는 클러스터 (cluster)를 형성할 수 있다. 이러한 cluster 내부에서는 특정 사용자를 신뢰하는 사용자들인 truster가 존재하고, 반대로 타인으로부터 trust를 받는 trustee가 존재하게 되는데 이러한 관계는 쌍(pair)로 존재하게 되며 trustee와 truster의 관계는 일대다(one to many)의 형태로 존재한다.

본 논문에서는 RS의 정확도를 높이기 위해서 cluster 정보를 활용하는 방법론을 제시하고, cluster의 규모에 따라 RS에 미치는 영향력을 처음으로 측정한다. 주요 성과는 클러스터를 인식하는 RS를 구현하고 실제 데이터로 실험한 결과 RS 클러스터의 3가지 특징을 발견한 것이다: 1) RS에서 trust cluster를 포함하는 경우 클러스터 사용자는 Power Law를 따른다. 2) 클러스터의 사이즈가 작은 경우라도 RS 예측 정확도에 기여하는 정도는 크기 때문에 제안하는 기법은 큰 의미가 있다. 3) 사용자의 대다수를 포함하는 대형 클러스터 보다는 작은 클러스터의 영향력을 포함하는 것으로 RS 예측 정확도를 충분히 향상시킨다.

본 논문의 구성은 다음과 같다. 1장 서론에 이어 2장에서는 본 논문에서 제안하는 방법과 관련된 기존 연구에 대하여 살펴본다. 3장에서는 본 논문에서 제안하는 방법론의 시스템 설계 과정에 대하여 설명하고, 4장에서는 시스템의 구현 및 실험, 그리고 cluster 규모에 따른 영향력을 분석한다. 5장의 결론을 마지막으로 본 논문을 마무리 한다.

## II. 시스템 설계

### 2.1 RS 모델링

전자 상거래(E-commerce)에서 대부분의 RS에는 사용자들이 아이টে를 평가(evaluation) 하거나, 후기(review)를 남길 수 있도록 하고 구매 후기에 대한 추천 기능을 제공하여 해당 아이টে를 구매하려는 잠재 고객에게 구매 관련 정보를 제공하고 있다. 또 다른 기능으로 power buyer 또는 power reviewer에게 신뢰(trust)를 형성할 수 있다. 이렇

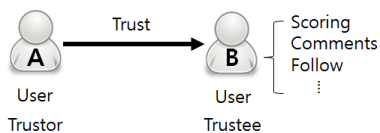


Fig. 1. Trust relation of a social network in RSs

듯, 사용자들이 후기에 대해서 추천을 하거나 사용자들끼리 trust 관계가 형성되는 경우 RS내부에서는 social relation을 기반으로 하는 가상의 네트워크가 또 다시 형성된다.

Fig. 1에서와 같이 사용자 A가 사용자 B의 평점, 리뷰, 또는 follow 정보를 신뢰하는 경우 사용자 A와 사용자 B 사이에는 trust 관계가 형성된다. 이때 "사용자 A는 사용자 B를 신뢰한다"고 표현하고 사용자 A를 truster, 사용자 B를 trustee로 정의한다. Epinions<sup>1)</sup> 또는 Ciao<sup>2)</sup>와 같이 명시적으로 follow 관계가 형성하는 경우에도 당연히 trust relation이 형성된다.

Fig. 1에서 생성되는 trust relation은 모든 RS사용자에게서 발생 가능하다. 결과, 모든 사용자의 특성을 반영하면 하나의 social network를 구성할 수 있으며, 본 논문에서는 이러한 네트워크를 trust aware network (TAN)라 명명한다. Fig. 2는 trust aware network를 개념적으로 설명한다.

TAN의 특징을 살펴보면 trustee는 다수의 truster를 가질 수도 있고 단수의 truster를 가질 수도 있다. 하나의 trustee와 관련된 단수 혹은 다수의 truster들이 형성하는 network를 trust cluster라 하면 trust cluster의 최소 사용자 수는 2명이며 최대의 경우 RS의 전체 사용자 수이다. 주의할 점은 하나의 truster는 다수의 사용자를 trust 할 수 있으므로 다수의 trust cluster의 멤버가 될 수 있다. Fig. 2 에서 동일 trust cluster 내의 truster들은 trustee의 영향을 받는다. 일반적으로 친구 관계의 사용자들은 서로 영향을 받게 되는데[3], trust cluster의 truster들의 경우에도 동일한 영향력이 존재하게 된다. 즉, truster들은 동일 trustee의 영향력을 받게 되어 truster들 사이에는 유사한 성향이 존재하게 된다.

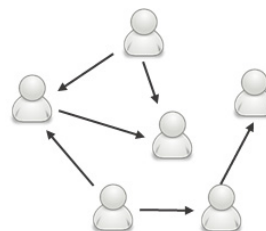


Fig. 2. Trust aware network (TAN) in RSs

### 2.2 Trust Cluster를 이용한 RS 모델링

RS를 구성하는 핵심적인 2가지 요소는 사용자와 아이템이다. RS 내부에 m명의 사용자와 n개의 아이템이 존재한다고 하면 RS는 하나의 매트릭스,  $R^{m \times n}$ 으로 표현이 가능하다.  $R^{m \times n}$ 에서 사용자 i가 아이템 j에 남긴 평점은  $R_{ij}$ 이다.  $R_{ij}$ 는 'good' 또는 'bad'와 같이 binary 값을 가질 수도 있으며, 일정 구간에서의 점수(score)로 표현될 수도 있다. 평점을 일정 구간 최소 0부터 최대 5까지 가질 수 있는  $R^{m \times n}$ 를 개념적으로 표현하면 Fig. 3과 같다.

Trust cluster를 이용하기 위해서는 사용자들 사이에서 형성된 trust 정보가 추가적으로 필요하다. Trust 정보의 경우 사용자와 사용자 간에 형성되므로 Trust matrix  $T$ 는  $T^{m \times m}$ 로 표현 가능하다.  $T^{m \times m}$ 의 개별 원소인  $T_{ij}$ 는 사용자 i가 사용자 j를 trust하는 경우  $T_{ij} = 1$  이며, 그렇지 않은 경우는  $T_{ij} = 0$  으로 표현된다.  $T_{ij} = T_{ji}$ 가 성립할 수도 있고, 그렇지 않은 경우도 존재한다.  $T_{ij} = T_{ji}$ 가 성립하는 경우는 사용자 i와 사용자 j 상호간에 trust를 하는 경우이다.

앞서 언급했듯이  $R^{m \times n}$  만을 활용하여 특정 사용자가 특정 아이템에 부여할 점수를 예측하는 방법

	Item <sub>1</sub>	Item <sub>2</sub>	...	Item <sub>n</sub>
User <sub>1</sub>	0	4	...	5
User <sub>2</sub>	5	0	...	3
⋮	⋮	⋮	...	⋮
User <sub>m</sub>	1	3	...	2

Fig. 3. User-Item matrix ( $R^{m \times n}$ )in RSs

1) www.epinions.com  
2) www.ciao.co.uk

으로는 MF 방식이 주로 활용한다.  $R^{m \times n}$ 에 MF를 적용하기 위해서는 2개의 잠재적 특성 매트릭스(latent feature matrix)로 분해해야 하며, 수학적으로는 Equation (1)과 같다.

$$R = UV^T \quad (1)$$

Equation (1)에서 매트릭스  $U$ 는 사용자의 잠재적 특성 매트릭스(latent feature matrix)이며  $V$ 는 아이템의 잠재적 특성 매트릭스(latent feature matrix)이다. 매트릭스의 크기는 각각  $U^{m \times k}$ ,  $V^{n \times k}$ 이다. 여기서  $k$ 는 MF의 계산 효율을 높이기 위해 low-rank MF를 수행하는 경우에 설정하는 정수(rank)이다.  $k$ 의 범위는 Equation (2)과 같다.

$$2 \leq k < \text{Min}(m, n) \quad (2)$$

$U$ 와  $V$ 에서  $u_i$ 를 사용자  $i$ 의 feature vector,  $v_j$ 를 아이템  $j$ 의 feature vector라 하자(이때,  $|u_i| = |v_j| = k$ 이다). 모든  $R$ 값은  $\|R - UV^T\|$ 의 잔여 에러(residual error)를 최소화하여 구할 수 있다.  $U$ 와  $V$ 를 구하기 위한 목적함수는 Equation (3)과 같다.

$$\min_{U, V} L(R, U, V) = \sum_{i=1}^m \sum_{j=1}^n I_{ij}^r (u_i v_j^T - R_{ij})^2 + \gamma (\|U\|_F^2 + \|V\|_F^2) \quad (3)$$

$I_{ij}^r$ 는 indicator function으로  $R_{ij}$ 값이 존재하면 1, 존재하지 않으면 0이다. Equation (3)의 learning 과정에서의 과도함(overfitting)을 방지하기 위한 loss function을 추가하면 수식 (5)와 같다. Equation (3)에서  $\|\cdot\|_F^2$ 은 Frobenius Norm이다.  $\gamma$ 는 조절 파라미터(tuning parameter)이다.

2.1절에서 기술한 바와 같이 하나의 trust cluster 내부에 존재하는 trustor 사용자들은 유사한 성향을 가지게 되는데, 이러한 성질을 RS 예측에 반영하기 위해 본 논문에서는 추가적인 matrix  $E$

를 도입한다.  $E$ 는 user feature matrix  $U^{m \times k}$ 에서 trustor cluster에 속한 사용자들의 feature만을 행(row) 단위로 추출하여 생성하며 특징은 다음과 같다.

$$E^{p \times k} \subseteq U^{m \times k}, \text{ where } p \leq m \wedge E_{i,*} \in U \quad (4)$$

Equation (4)에서  $p$ 는 특정 trustor cluster에 속한 사용자의 수이다. Equation (3)의 learning 과정에서  $E$ 의 영향성을 추가적으로 반영하기 위해서는 trustee와 trustor들 사이의 feature difference를 최소화 하도록 해야 한다. 결과, 이를 반영하면 Equation (5)와 같다. Equation (5)에서  $T^+(i)$ 는 사용자  $i$ (즉, trustee  $i$ )의 trustor cluster 멤버 집합이며,  $e_k$ 는  $E$ 의  $k$ 번째 행벡터 이고,  $\beta$ 는 trust cluster에 대한 tuning parameter 이다. 즉, Equation (5)에서 첫 번째 줄 부분은 MF에서의 실제 점수와 예측 점수의 차이를 나타내며, 두 번째 줄은  $i$  사용자 and 공통의 trustee  $k$  사용자를 trust하는 사용자들의 사용자 feature 속성을 비슷하게 하는 역할을 합니다. 세 번째 줄은 각각 사용자, 아이템, trustor cluster 매트릭스의 regularization 부분이다.

$$\min_{U, V, S} L(R, U, V, S) = \sum_{i=1}^m \sum_{j=1}^n I_{ij}^r (u_i v_j^T - R_{ij})^2 + \beta \sum_{k \in T^+(i)} \|u_i - e_k\|^2 + \gamma (\|U\|_F^2 + \|V\|_F^2 + \|E\|_F^2) \quad (5)$$

사용자  $i$ 의 경우 복수의 trustor cluster에 속할 수 있다. 단일 사용자가 다수의 trustor cluster에 속할 경우는 영향력을 감소시키고 단일 trustor cluster에 속할 경우는 영향력을 증가시키는 가중치(weight)의 추가 설계가 필요하다. 하나의 Trustor cluster에 있는 사용자를  $tc_k$ 라 하자. 단, 여기서  $k$ 는  $k \in \{1, 2, \dots, p\}$ 이다. 해당 cluster의 trustee를  $tr$ 라 하자. 만약  $tc_k$ 의 trust out-link가 2 이상이라면,  $tc_k$ 는 2개 이상의 trust

cluster에 속하게 된다. 이 경우 개별 trust에 대한 가중치를 줄여야 한다. 반면,  $tr$ 의 trust in-link가 많을 경우  $tr$ 의 영향력이 높은 경우이므로 가중치를 높여야 한다. 개별 신뢰 링크 (trust link)에 대한 가중치를  $\omega$ 라고 하면, 가중치는 Equation (6)과 같이 설정할 수 있다.

$$\omega(i, k) = \sqrt{\frac{d^-(tr_i)}{d^+(tc_k) + d^-(tr_i)}} \quad (6)$$

Equation (6)에서  $d^+(tc_k)$ 는 사용자  $k$ 의 trust out-link이며  $d^-(tr_i)$ 는 trustee  $i$ 의 in-link이다(cluster의 trustor들이  $i$ 를 trust하는 경우). 결과, Equation (5)에 Equation (6)을 반영한 수학적 표현은 Equation (7)과 같다.

$$\begin{aligned} \min_{U, V, S} L(R, U, V, E) = & \sum_{i=1}^m \sum_{j=1}^n I_{ij}(u_i v_j^T - R_{ij})^2 \\ & + \beta \sum_{i=1}^m \sum_{k \in T^+(i)} \omega(i, k) \|u_i - e_k\|^2 \\ & + \gamma (\|U\|_F^2 + \|V\|_F^2 + \|E\|_F^2) \end{aligned} \quad (7)$$

또한, 본 논문에서 제안하는 cluster size에 따른 영향성을 분석하기 위해서는 추가적인 수식의 조정이 필요하다. 추가적 조정을 위하여 truster 수에 따른 변수인  $I_{ij}^s$ 를 도입한 수학적 표현은 Equation (8)과 같다.  $I_{ij}^s$ 는 trust cluster 내의 truster 수이다.

$$\begin{aligned} \min_{U, V, S} L(R, U, V, E) = & \sum_{i=1}^m \sum_{j=1}^n I_{ij}(u_i v_j^T - R_{ij})^2 \\ & + \beta \sum_{i=1}^m \sum_{k \in T^+(i)} \omega(i, k) I_{ij}^s \|u_i - e_k\|^2 \\ & + \gamma (\|U\|_F^2 + \|V\|_F^2 + \|E\|_F^2) \end{aligned} \quad (8)$$

Equation (8)로부터  $U, V, E$ 의 값을 찾기 위해 Gradient Descent Method을 사용할 수 있으며, 각각에 대한 수학적 표현은 Equation (9)와

같다.

$$\begin{aligned} \frac{\partial L}{\partial U_i} &= 2 \left\{ \sum_{j=1}^n I_{ij}^s (u_i v_j^T - R_{ij}) v_j \right. \\ & \quad \left. + \beta \sum_{k \in T^+(i)} \omega(i, k) I_{ik}^s (u_i - e_k) \right. \\ & \quad \left. + \lambda (u_i - e_i) + \gamma u_i \right\} \\ \frac{\partial L}{\partial V_j} &= 2 \left\{ \sum_{i=1}^m I_{ij} (u_i v_j^T - R_{ij}) u_i + \gamma v_j \right\} \\ \frac{\partial L}{\partial E_k} &= 2 \left\{ \beta \sum_{k \in T^+(i)} \omega(i, k) I_{ik}^s (e_k - u_i) \right. \\ & \quad \left. + \lambda (e_k - u_i) + \gamma e_k \right\} \end{aligned} \quad (9)$$

### III. 실험 및 결과분석

#### 3.1 실험데이터 및 성능 Metric

본 논문에서 제안하는 trust cluster 기반의 RS의 성능향상을 객관적으로 평가하기 위하여 low rank MF 방식[4]과 social information을 적용한 SoRec[5]과 SoReg[6] 방식을 직접 구현 후 비교분석을 수행하였다. 본 논문에서 제안하는 trust cluster 방식과의 비교를 위하여 Ciao에서 제공하는 실제 RS 데이터를 활용하였다. Ciao 데이터는 9,385명의 사용자와 4,067개의 아이템으로 구성되어 있으며 사용자들이 남긴 평점(rating score)은 105,309개 이다. 사용자들 사이에서 발생한 Trust 관계는 121,651개 이다. 사용자들은 평균 11.2개의 평점을 남겼으며 특정 사용자가 남긴 최대 평점 수는 353개 이다. 사용자들 간 trust 관계의 평균수는 12.96개 이다.

본 논문에서 성능 측정 척도로 사용하는 metric은 mean absolute error (MAE)와 root mean square error (RMSE) 이다. 두 가지 모두 낮은 값이 높은 RS 예측 성능을 보여주며 Equation (10)과 (11)으로 각각 표현할 수 있다.

$$MAE = \frac{\sum_{i,j} |r_{ij} - \hat{r}_{ij}|}{N} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i,j} (r_{ij} - \hat{r}_{ij})^2}{N}} \quad (11)$$

Equation (10)와 (11)에서  $r_{ij}$ 는 사용자  $i$ 가 아이템  $j$ 에 남긴 실제 평점 값이고,  $\hat{r}_{ij}$ 는 사용자  $i$ 가 아이템  $j$ 에 남긴 평점을 예측한 값이다.

### 3.2 Cluster Size에 따른 영향력

Trust cluster 내부의 truster들의 분포를 분석한 결과 최대 cluster size는 100 (즉, 1명의 trustee가 100명의 truster를 가짐) 이었다. 80% 이상 대부분의 trustee들은 19개 이하의 cluster size를 가지고 있었다. 특히 약 70% 정도의 cluster들은 10 미만의 cluster size를 보였다. Trust cluster 크기에 따른 분포를 Fig. 4에 그림으로 표시하였다.

Trust relation이 존재하는 RS에서는 Fig. 4에서 볼 수 있는 바와 같이 대부분의 cluster는 소규모인 반면 상위 20%의 cluster가 대부분의 trust relation을 소유하고 있음을 알 수 있다. WWW에서의 link 형성, HTTP size, Internet degree 등에서 나타나는 ‘rich-get-richer’ 현상인 power-law 분포가 존재한다는 것을 발견하였다. 본 논문에서 사용된 Ciao 데이터에서는 Fig. 5에 제시된 바와 같이 scaling parameter 1.000093를 가지는 power-law를 따른다.

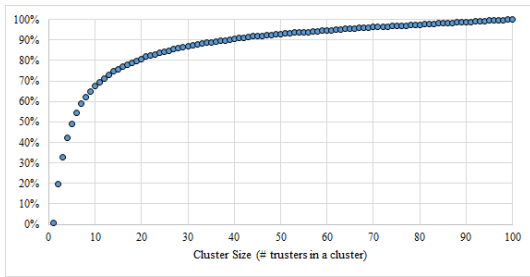


Fig. 4. Cumulative # of Cluster Users

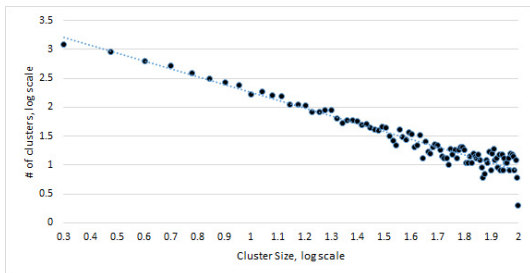


Fig. 5. Cluster size vs. # of clusters in log scale.

즉, Trust cluster의 신뢰 링크 (trust link) 분포는 power-law를 따른다는 것을 확인하였다. 본 논문에서는 많은 신뢰 링크 (trust link)를 가지는 cluster들이 RS 예측에 어떤 영향을 미치는지 파악하기 위하여 Equation (7)에서 도입한  $I_{ij}^s$  값을 변경시키며 실험하였다. 결과, Cluster size를  $C$ 라 할 때 Fig. 5와 6에서는  $I_{ij}^s$ 의 작동 방식은  $C$ 가 특정 숫자일 경우에만 Equation (7)이 작동하고, 그렇지 않은 경우에는 작동하지 않음을 발견했다. 이를 수식으로 표현하면 Equation (12)와 같다.

$$I_{ij}^s = \begin{cases} 1, & \text{if } |T^+(i)| = C \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

실험 결과에 대한 성능 측정치를 Fig. 6과 7에 제시하였다. 실험 결과 trust cluster size가 예측 정확도에 미치는 영향성은 적은 것으로 확인하였다. 예를 들어 MAE의 경우 cluster size와 무관하게 대부분의 값은 0.844와 0.848 사이의 값이었다. 다만, trust cluster size가 작은 영역에서는 MAE의 변동 폭이 약간 커지는 것을 확인하였다.

따라서, RS 내부에 trust cluster 구조를 가지는 경우 상위 20%의 cluster들이 80% 이상의

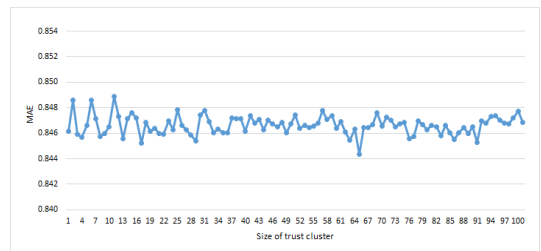


Fig. 6. MAE values with cluster sizes ( $C$ )

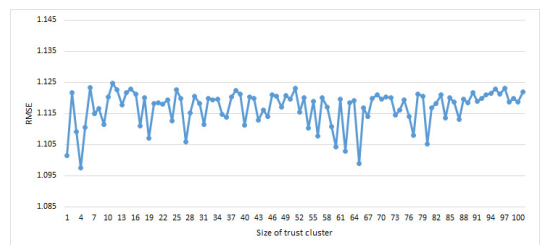


Fig. 7. RMSE values with cluster sizes ( $C$ )

trust relation을 가지고 있지만, 이렇게 큰 cluster들이 RS의 예측 정확도에 미치는 영향은 크지 않다는 것을 알 수 있다.

#### IV. 결 론

본 논문에서는 RS의 예측 정확도를 향상시키기 위하여 trust cluster를 활용하는 접근법을 제시하였다. Trust cluster를 social network로 해석하여 RS의 MF 방식을 향상시키는 방법을 수학적으로 제시하였으며, 이를 통해 trust cluster size에 따른 영향성을 분석하였다.

#### References

- [1] B. Xiao and I. Benbasat, "E-commerce product recommendation agents: use, characteristics, and impact," *MIS quarterly*, vol. 31, no. 1., pp. 137-209, Mar. 2007
- [2] D. Easley and J. Kleinberg, "Networks, crowds, and markets: Reasoning about a highly connected world:" Cambridge University Press, pp. 1-744, Sep. 2010
- [3] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Effects of user similarity in social media," in *Proceedings of the fifth ACM international conference on Web search and data mining*, pp. 703-712, Feb. 2012
- [4] J. D. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the 22nd international conference on Machine learning*, pp. 713-719, Aug. 2005
- [5] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: social recommendation using probabilistic matrix factorization," in *Proceedings of the 17th ACM.*, pp. 931-940, Oct. 2008
- [6] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 287-296, Feb. 2011

### 〈저자소개〉



노 기 섭 (Giseop Noh) 정회원  
 1998년 3월: 공군사관학교 산업공학과 학사  
 2009년 8월: University of Colorado Denver 컴퓨터과학과 석사  
 2014년 8월: 서울대학교 컴퓨터공학과 박사  
 2017년 1월~2018년 1월: 공군사관학교 전산정보과학과 조교수  
 2018년 2월~현재: 청주대학교 소프트웨어 융합학부 조교수  
 <관심분야> 소셜 정보망, 추천시스템, 네트워크 보안, 사이버전, 합동 전술 네트워크



오 하 영 (Hayoung Oh) 정회원  
 2002년 2월: 덕성여자대학교 전산학 졸업  
 2006년 2월: 이화여자대학교 컴퓨터공학과 석사  
 2013년 2월: 서울대학교 컴퓨터공학과 박사  
 2013년 9월~2016년 8월: 숭실대학교 전자정보공학부 조교수  
 2016년 9월~현재: 아주대학교 다산학부대학 조교수  
 <관심분야> 소셜 정보망, 추천시스템, 무선 네트워크 및 비디오 스트리밍



이 재 훈 (Jaehoon Lee) 정회원  
 2011년 2월: 동국대학교 정보통신공학과 졸업  
 2013년 2월: 서울대학교 컴퓨터공학과 석사 졸업  
 2013년 3월~현재: 서울대학교 컴퓨터공학과 박사 재학  
 <관심분야> 소셜 네트워크 분석, 추천시스템, 네트워크 보안