

대규모 복잡 정보에서 신뢰 클러스터를 이용한 추천 정확도 향상기법 설계

노기섭¹ · 오하영² · 이재훈^{3*}

Design a Method Enhancing Recommendation Accuracy Using Trust Cluster from Large and Complex Information

Giseop Noh¹ · Hayoung Oh² · Jaehoon Lee^{3*}

¹Republic of Korea Air Academy, Cheongju, 28187, Korea

²Department of DASAN University Colledge, Ajou University, Suwon, 16499, Korea

^{3*}Department of Computer Science, Seoul National University, 08826, Korea

요 약

최근 ICT기술의 발전과 스마트 기기의 급격한 보급으로 엄청난 양의 정보가 생성되고 있다. 추천 시스템은 과도한 정보제공(information overload)으로부터 정보 수용자의 적절한 판단을 도와주고, 정보 제공자에게는 기업의 이윤과 업체홍보 효과를 증대시킬 수 있는 해결책으로 등장하였다. 추천 시스템은 다양한 접근법으로 구현이 가능하지만, 소셜 네트워크 정보로 성능을 향상시킬 수 있는 방법으로 제시되었다. 그러나 추천 시스템 내의 사용자간에 형성되는 신뢰 클러스터의 정보를 활용하는 방안은 연구되지 못하였다. 본 논문에서는 온라인 리뷰에서 생성되는 클러스터에서 클러스터 내부 객체 간 영향성과 트러스터-트러스티 간 정보를 이용하여 추천 시스템의 성능을 향상시키는 방식을 제안하였다. 제안하는 방식을 구현하고 실제 데이터를 활용하여 실험한 결과 기존의 방식들보다 예측 정확도가 향상됨을 확인하였다.

ABSTRACT

Recently, with the development of ICT technology and the rapid spread of smart devices, a huge amount of information is being generated. The recommendation system has helped the informant to judge the information from the information overload, and it has become a solution for the information provider to increase the profit of the company and the publicity effect of the company. Recommendation systems can be implemented in various approaches, but social information is presented as a way to improve performance. However, no research has been done to utilize trust cluster information among users in the recommendation system. In this paper, we propose a method to improve the performance of the recommendation system by using the influence between the intra-cluster objects and the information between the trustor-trustee in the cluster generated in the online review. Experiments using the proposed method and real data have confirmed that the prediction accuracy is improved than the existing methods.

키워드 : 추천 시스템, 온라인 소셜 관계망, 소셜 관계 클러스터, 소셜 관계 분석

Key word : Recommender system, Online social relation, Social relation cluster, Social network analysis

Received 21 August 2017, Revised 01 September 2017, Accepted 10 October 2017

* Corresponding Author Jaehoon Lee(E-mail:jhlee@popeye.snu.ac.kr, Tel:+82-2-880-1858)

Department of Computer Science, Seoul National University, Seoul 08826, Korea

Open Access <http://doi.org/10.6109/jkice.2018.22.1.17>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

최근 정보통신 기술의 발달은 다양한 분야에서 혁신적인 변화를 가져왔다. 특히 스마트 기기의 보급과 확산은 사용자들이 실시간으로 연결되어 통신하는 환경을 만들어 내고 있다. 유무선 스마트 기기들의 확산의 기반이 되는 플랫폼은 사회 관계망 서비스(social network service, SNS)이다. 스마트 기기를 이용한 카카오톡, 밴드, 카카오톡스토리, 페이스북, 링크드인 등과 같은 SNS의 활용은 일반인들의 삶의 양식과 문화를 바꾸고 있다. SNS의 발전에 따라 수많은 사용자들이 사이버 공간으로 집중하게 되었고, 각종 온라인 기업들도 SNS를 새로운 엔터프라이즈 플랫폼(Enterprise Platform)으로 적용하기 시작하였다.

온-오프라인의 유통 기업들이 SNS 기반의 새로운 사업 영역으로 확장하면서 수많은 사용자들에게 다양한 상품을 효과적으로 제안하고 인지도를 높이려는 요구가 지속적으로 증가하였다. 이러한 상황에서 사용자들에게 제공되는 정보량이 개인별 정보처리 수준과 속도를 넘어서는 정보 과부하(information overload) 문제가 발생하였다[1]. 정보 과부하는 사용자가 적절하게 판단하는 것에 장애물로 작용한다.

인터넷 정보 제공자(internet service provider, ISP)에게는 과도한 정보 제공으로 고객의 이탈 및 이윤 감소로 이어질 수 있기 때문에 문제 해결을 위해 추천 시스템(recommender system, RS)이 등장하였다. RS는 개별 사용자에게는 맞춤형 정보를 제공하여 정보홍수 속에서 최적화된 정보를 제공할 수 있게 도와주며, ISP에게는 잠재 고객을 유지하고 이윤을 높일 수 있도록 도와준다. 온라인 마켓(아마존, 이베이, 오티), 영화 제공(넷플릭스, 네이버 영화), 여행정보(엘프, 트립 어드바이저), 인맥관리(카카오 스토리, 페이스북, 링크드인), 등 다양한 분야에서 RS 서비스를 이미 제공하고 있다.

RS는 전통적으로 2가지 접근법으로 설계되었다. 하나는 content based(CB) 방법이고 다른 하나는 collaborative filtering(CF) 방법이다. 전자는 유저나 아이템이 가지고 있는 고유 특성을 이용한 방식이고, 후자는 유저/아이템에 대해 유사한 점수 패턴을 나타내는 다른 유저/아이템을 추천해주는 기술이다. CB와 CF 방법의 장점을 결합하여 설계하는 것은 하이브리드 방법이다[2]. 최근 SNS의 확산에 따라 소셜 네트워크 정

보를 활용하는 기술이 추가적으로 개발되었다[3].

소셜 네트워크 정보를 활용하는 기술의 핵심은 친구, 팔로우, 추천 등으로 발생하는 명시적 또는 암시적 관계를 소셜 관계로 해석하는 것이다. 이러한 소셜 관계가 있는 사용자들 사이에서는 서로에게 영향을 주게 되는데, 이 영향력을 계산하여 RS에 반영하는 방식이다. 그러나 이전의 연구들은 단순하지만 현재까지의 연구는 RS가 작동하는 단계에서 친구관계에 따라 적절한 가중치를 부여하는 방식으로 진행되어 왔다.

실제 친구 관계에도 다양한 형태가 존재한다는 것이 간과되어 왔다. 소셜 관계가 있는 사용자들의 의견들도 여러 가지 방향으로 나타나거나 신뢰 관계(trust relation)에 따라 그룹 또는 클러스터화 될 수 있다. 파워 블로거나 유명인과 같이 영향력이 높은 친구가 존재할 경우 영향력은 친구관계의 쌍(pair) 단위로 전파되는 것이 아니라 일대다(one-to-many)의 형태로 신뢰 관계 정보에 따른 클러스터링 효과는 고려되지 못했다. 이러한 추가적인 정보를 활용할 경우 RS의 성능을 개선할 수 있다.

본 논문에서는 기존의 연구 방법에서 간과 되었던 클러스터링 효과를 이용하는 새로운 RS 알고리즘(Reinforced Cluster Recommendation, ReinClus-Rec)을 최초로 제안하고 실제 RS 데이터를 활용하여 기존 방법들과 성능을 비교 평가한다. 본 논문의 주요 성과는 다음과 같다. (1) 리뷰 클러스터를 이용하는 새로운 온라인 추천 접근법을 제안 한다; (2) ReinClus-Rec을 수학적으로 모델링하여 구현 한다; (3) 실제 RS 데이터를 활용하여 제안한 방법론을 기존의 소셜 관계 기술들과 비교하여 성능을 입증한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 기존 연구에 대하여 살펴본다. 3장에서는 ReinClus-Rec의 모델링과 시스템 디자인에 대하여 설명한 이후, 4장에서는 제안하는 방법과 비교 가능한 기술들을 실험 및 성능 측정을 통하여 성능을 입증한다. 5장의 결론으로 본 논문을 마무리 한다.

II. 본 론

RS의 초기 단계는 CF 기술을 중심으로 연구가 진행되었다. CF는 사용자 중심으로 RS의 주요 알고리

즘이 작동한다. 핵심 아이디어는 사용자들의 이전 이력 (history 또는 record)을 활용하는 것이다. 유사도 분석 등을 통해 현재 사용자와 유사한 패턴을 보이는 다른 사용자를 찾아내고, 그 유사한 사용자의 패턴을 현재 사용자에게 적용하여 행동 또는 취향을 예측한다. CF 방식은 구현 방법에 따라 메모리 기반 방식과 모델 기반 방식으로 구분할 수 있다.

메모리 기반 방식은 다른 사용자의 평점 또는 리뷰의 패턴에 초점을 두는 방법이다. 메모리 기반 방식은 다시 시스템의 설계 방식에 따라 유사한 사용자에게 중심을 두는 방법[4]과 아이템에 초점을 두는 방식[5]이 대표적이다. 주요 특징으로는 이전의 기록이 많을 경우 추천 정확도가 높다. 그러나 새로 등장한 아이템에는 기존 사용자들이 남긴 정보가 부족하여 RS를 구성하는 사용자-아이템 매트릭스가 희소한 경우가 발생하고, 이에 따라 추천 예측 정확도가 감소하는 콜드 스타트 문제(cold start problem)가 발생한다[6].

모델 기반 방식은 메모리 기반 방식의 콜드 스타트 문제를 보완하기 위하여 고안되었다. 모델 기반 방식에서는 사용자 간 유사성을 찾거나 아이템에 남겨진 정보들의 유사성을 발견하기 보다는 주어진 정보에 숨겨진 잠재 특성(latent feature)을 찾아 활용한다. 잠재 특성을 계산하기 위한 방식으로 매트릭스 인수분해 (matrix factorization, MF) 방법이 주로 사용된다. 매트릭스 인수분해 계산량을 줄이기 위해 다양한 저차원 근사 방법이 제안 되었다[7].

최근 SNS의 일반화에 따라, 매트릭스 인수분해 방식의 예측 정확도를 향상을 위해 도입된 기술은 소셜 네트워크 정보를 활용하는 접근법이다. 온라인 소셜 네트워크 또는 유무선 서비스 도구에서 생성되는 RS 사용자간 관계를 분석하여 매트릭스 인수분해 수행 시 부가적인 정보 분석 결과를 반영하거나 추가적인 가중치를 부여하는 방법이다. 대표적인 소셜 네트워크 기반 RS 기술로는 사용자의 성향에 친구들의 성향을 합성하는 SocialMF[8] 방식과 친구들의 신뢰 관계 구조와 사용자간의 유사성을 동시에 합성하는 Sorec[9]이 있다. 아이템 카테고리를 이용하는 방식[10]과 커뮤니티 탐지를 이용하는 방식[11]이 추가적으로 제안되었다.

그러나 트러스터-트러스티 구조에서 발생하는 소셜 네트워크 정보를 추가적으로 활용하여 리뷰 클러스터를 생성하고 이를 RS 예측 정확도를 향상시킬 수 있음

에도 불구하고 이전의 연구들은 트러스터-트러스티 구조에서 발생하는 클러스터링을 활용한 연구는 전무하다.

III. 시스템 설계

본 장에서는 ReinClus-Rec 설계에 필요한 신뢰 관계에 대하여 먼저 설명하고, 이를 바탕으로 ReinClus-Rec를 수학적 방법으로 설계한다.

3.1. 소셜 네트워크 관점의 신뢰 네트워크

SNS에는 대부분 전자상거래 기능이 탑재되어 있다. 전자상거래 시스템에서는 사용자들이 아이템을 평가 (평가점수 또는 등급부여 등)하고 구매 후기를 남길 수 있다. 또한 구매 후기에 대한 추천 기능을 제공하기도 한다. 또한 전자상거래 시스템에 사용자들 간 의견을 교환하고 서로 추천하고 팔로우하는 SNS의 기능이 제공된다. 본 논문에서는 트위터에서 팔로우-팔로워 관계가 명시적으로 존재하거나, 타인의 결과를 신뢰하거나 추천하는 경우 해당 사용자들 가운데 암시적 팔로우 관계가 형성 되는데, 이러한 소셜 관계를 신뢰 관계가 존재하는 것으로 가정한다.

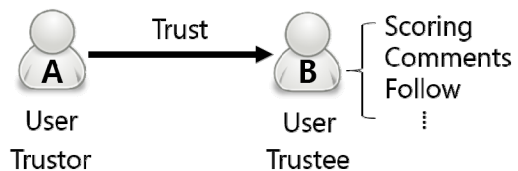


Fig. 1 Trust relation in social information.

그림 1에서 유저 B는 SNS 또는 전자상거래 시스템 구현에 따라 아이템에 대하여 점수를 부여하거나 구매 후기/사용 후기 등 과 같은 코멘트를 남기는 것처럼 유저 정보를 생성한다. 유저 A가 유저 B의 정보를 신뢰할 경우(‘Good/Bad’, 추천, ‘좋아요(like)’ 등의 시그널을 남길 경우) 유저 A는 유저 B를 신뢰(trust) 하는 관계가 생성된다. 이 경우 유저 A는 트러스터, 유저 B는 트러스티로 정의한다. 또한 트위터나 이피니언처럼 명시적으로 유저 B를 팔로우하는 경우에도 신뢰 관계가 발생한다.

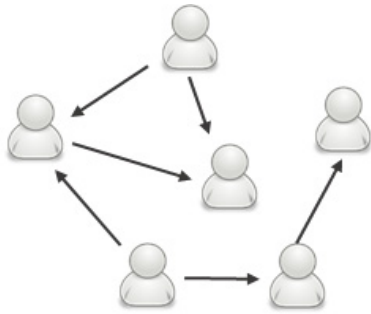


Fig. 2 Trust aware network (TAN).

그림 1에서 생성되는 신뢰 관계 모든 사용자에게서 발생이 가능하다. 특정 시스템 내의 사용자들의 특성을 모두 파악하여 반영하면 하나의 소셜 네트워크로 구성할 수 있는데, 본 논문에서는 신뢰 관계를 분석해 생성한 네트워크를 trust aware network(TAN) 이라고 정의한다. TAN의 형태는 비가중치 방향 그래프 형태이며, 개념적 표현은 그림 2와 같다.

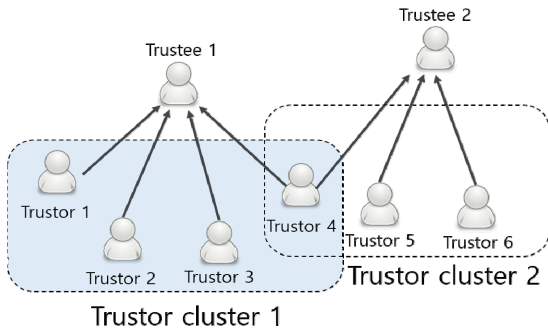


Fig. 3 An example of Trustor Clusters.

TAN의 구조를 살펴보면 파워블로거, 유명인 등과 같이 다수의 사용자로부터 신뢰를 받는 트러스티가 존재한다. 트러스티는 하나의 트러스터를 가질 수도 있고 다수의 트러스터를 가질 수도 있다. 그림 3의 경우 트러스티 1은 4명의 유저(트러스터 1~4)를 가지고 있으며, 트러스티 2의 경우 3명의 유저(트러스터 4~6)를 가지고 있다. 각각의 트러스티는 각각의 트러스터 클러스터를 가진다(그림 3의 경우 트러스터 클러스터 1과 2). 개별 트러스티는 2개 이상의 트러스터 클러스터에 동시에 속할 수 있다(예: 그림 3에서 트러스터 4의 경우). 신뢰 클러스터를 TC라 할 때, TC의 크기는 수식 (1)과 같다.

$$1 \leq TC \leq N \tag{1}$$

수식 (1)에서 N 은 RS의 전체 사용자 수이고, TC 는 양의 정수이다.

3.2. ReinClus-Rec 설계

RS의 핵심 구성요소는 사용자와 아이템이다. m 명의 사용자가 존재하고, n 개의 아이템 집합으로 구성된 RS는 그림 4와 같이 하나의 매트릭스 $R^{m \times n}$ 으로 표현하는 것이 일반적이며, $R^{m \times n}$ 의 요소 R_{ij} 는 사용자 i 가 아이템 j 에 남긴 평점이다. R_{ij} 는 'good, bad' 또는 'like, hate'와 같이 0 또는 1의 값을 가질 수도 있으며 특정 구간 내의 점수가 될 수도 있다.

	Item ₁	Item ₂	...	Item _n
User ₁	0	4	...	5
User ₂	5	0	...	3
:	:	:	...	:
User _m	1	3	...	2

Fig. 4 User-item matrix in RS.

TAN의 경우도 유저-유저 매트릭스인 신뢰관계 매트릭스 $T^{m \times m}$ 으로 표현할 수 있다. $T^{m \times m}$ 의 요소 T_{ij} 는 사용자 i 가 사용자 j 를 신뢰할 경우 1, 그렇지 않은 경우 0의 값을 갖는다. 상호간에 신뢰 관계가 형성되지 않는 한 $T_{ij} \neq T_{ji}$ 이다. 그림 5의 좌측 그림에서 TAN을 신뢰 관계 매트릭스로 표현한 예를 그림 5의 우측 그림과 같이 제시하였다.

	User ₁	User ₂	User ₃	User ₄	User ₅	User ₆
User ₁	0	1	1	0	0	0
User ₂	0	0	1	0	0	0
User ₃	0	0	0	0	1	1
User ₄	1	0	0	0	0	0
User ₅	0	1	0	0	0	1
User ₆	0	0	1	0	1	0

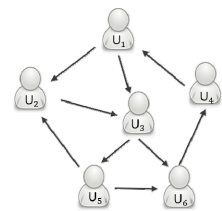


Fig. 5 A trust matrix in a TAN.

ReinClus-Rec에서 매트릭스 인수분해를 사용하기 위해서는 $R^{m \times n}$ 를 사용자와 아이템으로 구성된 2개의 잠재 특성 매트릭스로 수식 (2)와 같이 분해할 수

있다.

$$R = UV^T \quad (2)$$

수식 (2)에서 U 는 사용자의 잠재 특성 매트릭스이며, V 는 아이템의 잠재 특성 매트릭스이다. 매트릭스의 크기는 각각 $U^{m \times k}$, $V^{n \times k}$ 이다. 여기서 k 는 매트릭스 인수분해의 계산 효율을 높이기 위해 저차원 매트릭스 인수분해를 수행하는 경우에 설정하는 차원(rank)이다. k 의 범위는 수식 (3)과 같다.

$$2 \leq k < \text{Min}(m, n) \quad (3)$$

U 와 V 의 값을 알고 있다면, 점수 정보가 없는 어떠한 R_{ij} 값도 쉽게 구할 수 있다. u_i 를 사용자 i 의 잠재 벡터, v_j 를 아이터మ్ j 의 잠재 벡터라 하자(이때, $|u_i| = |v_j| = k$ 이다). 모든 R 값은 $\|R - UV^T\|$ 의 잉여 오차를 최소화하여 구할 수 있다. R 을 구하기 위한 목적함수는 수식 (4)와 같다.

$$\min_{U, V} \mathcal{L}(R, U, V) = \sum_{i=1}^m \sum_{j=1}^n I_{ij} (u_i v_j^T - R_{ij})^2 \quad (4)$$

수식 (4)에서 I_{ij} 는 표시 함수로써 R_{ij} 값이 존재하면 1, 존재하지 않으면 0이다. 수식 (4)의 학습 과정에서의 과적합(overfitting)을 방지하기 위한 목적 함수를 추가하면 수식 (5)와 같다. 수식 (5)에서 $\|\cdot\|_F^2$ 은 Frobenius Norm이다.

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

수식 (5)는 소셜 네트워크 정보를 사용하지 않는 순수한 매트릭스 인수분해 방식이다. TAN에서 TC 정보를 활용하기 위한 수학적 설계를 위해 2개의 추가적인 개념을 도입한다.

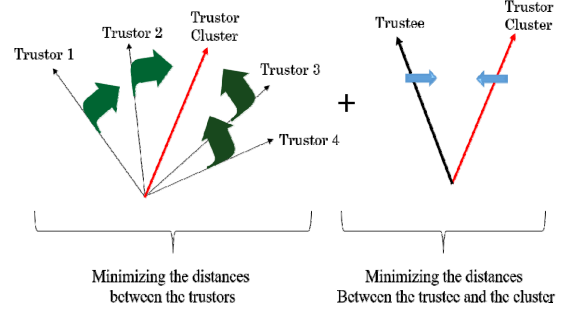


Fig. 6 A trust matrix in a TAN.

그림 6에서 보는 바와 같이 클러스터의 영향력을 반영하기 위해서는 2가지 고려사항이 반영되어야 한다. (i) 동일한 트러스터 클러스터 내부에 존재하는 트러스터들은 유사한 성향을 가진다. 즉, 유사한 잠재 벡터를 갖는다. (ii) 동일한 트러스터 클러스터 내부에서는 트러스터의 특성이 대표적인 성향을 나타낸다. 위 2가지의 특성을 모두 충족하기 위한 설계 방법은 이어지는 2개의 절에서 설명한다.

3.2.1. 동일 클러스터 내 트러스터 유사성

3.2절에서 설명한 바와 같이 동일 트러스터 클러스터에 존재하는 사용자들(즉, 트러스터들)은 유사한 성향을 가진다는 점을 반영하기 위하여 새로운 매트릭스 E 를 생성한다. E 는 유저 잠재 매트릭스 $U^{m \times k}$ 에서 클러스터에 속한 사용자들의 특성만을 행(row) 단위로 추출하여 생성하며 특징은 다음과 같다.

$$E^{p \times k} \subseteq U^{m \times k}, \text{ where } p \leq m \wedge E_{i,*} \in U \quad (6)$$

수식 (6)에서 e 는 특정 트러스터 클러스터에 속한 사용자의 수이다. 트러스터와 트러스터는 수식 (5)의 학습 과정에서 잠재 특성 벡터의 차이가 최소화 되도록 반영하면 수식 (7)과 같다.

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

수식 (7)에서 $T^+(i)$ 는 사용자 i (즉, 트러스티 i)의 트러스티 클러스터 멤버 집합이다. 그러나 사용자 i 의 경우 복수의 트러스티 클러스터에 속할 수 있다. 단일 사용자가 다수의 트러스티 클러스터에 속할 경우는 영향력을 감소시키고 단일 트러스티 클러스터에 속할 경우는 영향력을 감소시키는 가중치의 추가 설계가 필요하다. 하나의 트러스티 클러스터에 있는 사용자를 tc_k 라 하자. 여기서 단, $k \in \{1, 2, \dots, p\}$ 이다. 해당 클러스터의 트러스티를 tr 라 하자.

만약 tc_k 의 신뢰 관계 아웃링크가 2 이상이라면, tc_k 는 2개 이상의 신뢰 관계 클러스터에 속하게 된다. 이 경우 개별 신뢰관계에 대한 가중치를 줄여야 한다. tr 의 신뢰 관계 인링크가 많을 경우 tr 의 영향력이 높은 경우이므로 가중치를 높여야 한다. 개별 신뢰 관계 링크에 대한 가중치를 ω 라고 하면, 가중치는 수식 (8)과 같이 설정할 수 있다.

$$\omega(i, k) = \sqrt{\frac{d^-(tr_i)}{d^+(tc_k) + d^-(tr_i)}} \quad (8)$$

수식 (8)에서 $d^+(tc_k)$ 는 사용자 k 의 신뢰 관계 아웃링크이며 $d^-(tr_i)$ 는 트러스티 i 의 인링크이다(클러스터의 트러스티들이 i 를 신뢰하는 경우). 만약 특정 클러스터의 사용자 k (tc_k)들이 모두 하나의 트러스티 i 만을 신뢰 하고 있다면, ω 값은 1에 근접한다. 그러나 여러 개의 클러스터에 속해 있는 경우 0의 값으로 근접해 간다. 모든 사용자를 고려하면서, 가중치를 추가적으로 고려한 동일 클러스터 내 트러스티 유사성은 수식 (9)와 같다.

$$\begin{aligned} \min_{U, V, E} L(R, U, V, E) = & \sum_{i=1}^m \sum_{j=1}^n I_{ij} (u_i v_j^T - R_{ij})^2 \\ & + \beta \sum_{i=1}^m \sum_{k \in T^+(i)} \omega(i, k) \|u_i - e_k\|^2 \\ & + \gamma (\|U\|_F^2 + \|V\|_F^2 + \|E\|_F^2) \end{aligned} \quad (9)$$

3.2.2. 동일 클러스터와 트러스티 유사성

그림 6의 좌측 그림에서와 같이 트러스티의 영향력을 고려하면서, 신뢰 관계 클러스터가 매우 적은 사용자들로 구성될 경우(k 값이 작은 경우)를 추가적으로 고려하여야 한다. 후자의 경우 클러스터 내부 사용자들의 특성 상관성이 떨어진다. 이 경우를 적절히 보완하기 위한 방법은 트러스티 (tc_k)의 특성을 트러스티 (tr_i)와 유사하게 조정한다. 동일 클러스터와 트러스티 유사성을 추가로 고려하면 그림 6에서 제시하는 2가지 경우를 모두 적용할 수 있다. 최종적인 목적함수는 수식 (10)으로 표현할 수 있다.

$$\begin{aligned} \min_{U, V, E} L(R, U, V, E) = & \sum_{i=1}^m \sum_{j=1}^n I_{ij} (u_i v_j^T - R_{ij})^2 \\ & + \beta \sum_{i=1}^m \sum_{k \in T^+(i)} \omega(i, k) \|u_i - e_k\|^2 \\ & + \lambda \sum_{i=1}^m \|u_i - e_i\|^2 \\ & + \gamma (\|U\|_F^2 + \|V\|_F^2 + \|E\|_F^2) \end{aligned} \quad (10)$$

수식 (10)에서 β, λ, γ 는 튜닝 파라미터이다. 수식 (10)으로부터 최소 목적 함수값은 기울기 하강법 (Gradient Descent)를 통해 최소값을 찾을 수 있으며, 수식 (11)과 같다.

$$\begin{aligned} \frac{\partial L}{\partial U_i} = & 2 \left\{ \begin{aligned} & \sum_{j=1}^n I_{ij} (u_i v_j^T - R_{ij}) v_j \\ & + \beta \sum_{k \in T^+(i)} \omega(i, k) (u_i - e_k) \\ & + \lambda (u_i - e_i) + \gamma u_i \end{aligned} \right\} \\ \frac{\partial L}{\partial V_j} = & 2 \left\{ \sum_{i=1}^m I_{ij} (u_i v_j^T - R_{ij}) u_i + \gamma v_j \right\} \\ \frac{\partial L}{\partial E_k} = & 2 \left\{ \begin{aligned} & \beta \sum_{k \in T^+(i)} \omega(i, k) (e_k - u_i) \\ & + \lambda (e_k - u_i) + \gamma e_k \end{aligned} \right\} \end{aligned} \quad (11)$$

IV. 실험 및 성능측정

4.1. 실험 데이터

본 논문에서는 실제 RS에서 사용하는 데이터는 이피니언[12]을 활용한다. 이피니언 데이터는 RS 연구에 널리 사용되고 있는 실제 데이터이며, 사용자들이 아이템(영화, 도서, 음악, 가전제품 등)에 대한 리뷰를 작성하고 1점~5점 사이의 점수를 부여할 수 있는 기능을 제공하고 있다. 또한, 모든 사용자들은 사용자들 간에 신뢰 관계를 생성하고 개별 이용자는 자신의 신뢰 관계 리스트를 유지한다. 본 논문에서 사용한 데이터의 요약은 표 1에 제시 하였다.

Table. 1 Dataset description

Dataset		Epinions
Rating Matrix	# of users	36,070
	# of items	10,590
	# of ratings	401,099
Trust Relations	# of relations	424,410
	Avg. relations	11.76

4.2. 성능측정 Metric

ReinClus-Rec 구현 이후 RS 예측 정확도를 측정하기 위해서 본 논문에서 사용하는 metric은 root mean square error (RMSE)와 normalized distance-based performance measure (NDPM)이다.

$$RMSE = \sqrt{\frac{\sum_{i,j} (r_{ij} - \hat{r}_{ij})^2}{N}} \quad (12)$$

수식 (12)에서 N 은 테스트 데이터에 있는 전체 점수의 수, r_{ij} 는 실제 점수, \hat{r}_{ij} 는 RS가 예측한 점수 값이다. RS의 예측 성능이 좋을수록 RMSE는 작은 값을 갖는다.

$$NDPM = \frac{c^- + 0.5 \times c^+}{c^u} \quad (13)$$

NDPM은 RS에서 순서에 따른 예측 값이 얼마나 정확한지를 측정한다. 수식 (13)에서 c^+ 는 아이템 쌍에서

주어진 점수의 크기 순서와 예측값에 대한 점수 크기 순서가 일치하는 모든 경우의 수이며, c^- 는 점수의 순서가 일치하지 않는 경우의 수이다. c^u RS내부에 존재하는 아이템의 모든 가능한 쌍(pair)의 수이다. 예를 들어 임의의 아이템 쌍에 대한 실제 값이 ($r_{i,j} = 4, r_{x,y} = 5$)라 하면, $r_{i,j} < r_{x,y}$ 이므로, 만약 RS의 예측값에 대한 값의 순서 또한 $\hat{r}_{i,j} < \hat{r}_{x,y}$ 와 같아야 한다. 모든 가능한 예측 순서를 측정함으로써 RS의 정확도를 평가한다. 수식 (12)의 경우 절대적 예측치를 측정하고, 수식 (13)의 경우는 상대적인 정확도를 측정한다.

4.3. 실험 결과

본 논문에서는 ReinClus-Rec의 성능을 측정하기 위하여 소셜 네트워크 정보를 사용하지 않는 기본적인 MF[13] 방식과 소셜 네트워크 정보를 이용하는 SoRec [9] 을 추가적으로 직접 구현하였다. 성능 측정 결과는 그림 7와 같다.

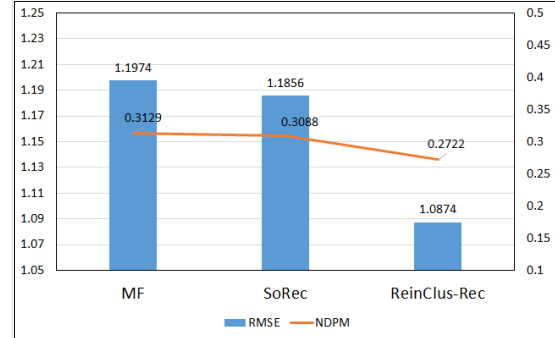


Fig. 7 Experimental Results.

그림 7에서 보여주는 바와 같이 ReinClus-Rec은 MF 방식과 비교하여 RMSE 9.19%, NDPM 13.01% 정확도 향상이 있으며, 소셜 네트워크 정보를 활용하는 SoRec 과 비교 시 RMSE 8.28%, NDPM 11.85%의 성능 향상을 달성하였다. RMSE가 조금만 향상된다 하더라도 실제 시스템의 성능에는 많은 영향을 미친다고 한다[14]. RMSE의 의미가 오차를 제공해서 평균한 값의 제공근이고 표준 편차와 같은 의미로 사용되고 있다. 따라서 해당 수치를 낮추는 것이 실제 추천 시스템에서의 정확도 향상과 관계가 있으며 정밀도가 좋다고 볼 수 있다.

V. 결 론

본 논문에서는 기존의 RS 성능을 향상시키기 위하여 트러스티 클러스터의 소셜 네트워크 관계를 트러스티와 트러스티 관점을 활용하는 새로운 RS인 ReinClus-Rec을 제안하였다. 시스템 모델링을 통한 RS 디자인을 통해 소셜 네트워크 정보를 활용하지 않는 단순 MF 방식을 개선하는 방법을 수학적으로 설명하였으며, 이를 통해 ReinClus-Rec 을 최종적으로 제안하였다. 또한 실제 RS에서 사용하는 데이터셋을 활용하여 기존 모델과 성능을 비교하였다. 성능측정 실험 결과 본 논문에서 제안하는 방법이 기본적인 MF 대비 9.19% 정확도 향상을 확인하였으며, 소셜 네트워크 정보를 활용하는 기존의 방법보다 RS 예측 정확도(RMSE 및 NDPM) 측면에서 우수한 성능을 보임을 확인하였다

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2017R1D1A1B03035557).

REFERENCES

- [1] H. Yi and F. Zhang, "Robust recommendation method based on suspicious users measurement and multidimensional trust," *Journal of Intelligent Information Systems*, vol. 46, no. 2, pp. 349-367, Apr. 2016.
- [2] I.-Y. Jeong, X. Yang, and H.-K. Jung, "A Study on Movies Recommendation System of Hybrid Filtering-Based," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 19, no. 1, pp. 113-118, Jan. 2015.
- [3] J-H Kim, J-H Kim, S-W Jeong, S-J Kang, "Multi-level Song Recommendation System based on SNS Posts and Sentiment Analysis," *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, vol.7, no.3, pp. 283-290, Mar. 2017.
- [4] H. Ma, I. King, and M. R. Lyu, "Effective missing data prediction for collaborative filtering," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 39-46, 2007.
- [5] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, et al., "Personalized recommendation of social software items based on social relations," in *Proceedings of the third ACM conference on Recommender systems*, pp. 53-60, 2009.
- [6] S. Feil, M. Kretzer, K. Werder, and A. Maedche, "Using gamification to tackle the cold-start problem in recommender systems," in *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, pp. 253-256, 2016.
- [7] W. X. Zhao, S. Li, Y. He, E. Y. Chang, J.-R. Wen, and X. Li, "Connecting social media to e-commerce: Cold-start product recommendation using microblogging information," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 5, pp. 1147-1159, May 2016.
- [8] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the fourth ACM conference on Recommender systems*, pp. 135-142, 2010.
- [9] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: social recommendation using probabilistic matrix factorization," in *Proceedings of the 17th ACM conference on Information and knowledge management*, pp. 931-940, 2008.
- [10] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1267-1275, 2012.
- [11] H. Li, D. Wu, W. Tang, and N. Mamoulis, "Overlapping community regularization for rating prediction in social recommender systems," in *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 27-34, 2015.
- [12] Epinions. (2017, July 7). *Epinions, Unbiased Reviews by Real People*. Available: <http://www.epinions.com/>.
- [13] 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, 2005.
- [14] Y. Koren, "Factor in the neighbors: Scalable and accurate collaborative filtering," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 4, no. 1, pp. 1-24, Jan. 2010.



노기섭(Giseop Noh)

1993.3. 공군사관학교 산업공학 학사
2009.8. University of Colorado Denver, 컴퓨터과학 석사
2014.8. 서울대학교 컴퓨터공학 박사
※ 관심분야 : 추천 시스템, 소셜 네트워크 분석, 네트워크 보안, 사이버전



오하영(Hayoung Oh)

2002.2. 덕성여자대학교 컴퓨터공학과 학사
2006.2. 이화여자대학교 컴퓨터공학과 석사
2013.2. 서울대학교 컴퓨터공학과 박사
2016.8. 숭실대학교 전자정보공학부 조교수
현재: 아주대학교 다산학부대학 조교수
※ 관심분야 : 추천 시스템, 소셜 네트워크 분석, 유무선 네트워크



이재훈(Jaehoon Lee)

2011.2. 동국대학교 정보통신공학과 학사
2013.2. 서울대학교 컴퓨터공학부 석사
2013.3-현재 서울대학교 컴퓨터공학부 박사과정
※ 관심분야 : 추천 시스템, 소셜 네트워크 분석