

사용자 인정 기준 변화에 따르는 신뢰 관계 예측의 정확도 분석

김대훈*, 허유정*, 하지운**, 김상욱*
*한양대학교 컴퓨터공학부
**한양대학교 전자컴퓨터통신공학과
e-mail : ghsdh3409@hanyang.ac.kr

An Analysis on Accuracy in Trust Relationship Prediction with Changes of User Selection Threshold

Dae-Hoon Kim*, Yu-Jung Hur*, Jiwoon Ha**, Sang-Wook Kim*
*Dept. of Computer Science, Hanyang University
**Dept. of Electronics and Computer Engineering, Hanyang University

요 약

온라인 사회연결망의 신뢰 관계를 활용한 양질의 서비스를 개발하기 위해서는 신뢰 관계가 존재하지 않는 사용자들 간의 신뢰 관계를 예측할 수 있는 모델에 대한 연구가 필요하다. 신뢰 관계 예측 모델의 정확도를 향상시키기 위해서는 사용자 인정 기준을 통해 대상 사용자를 선정할 필요가 있다. 본 논문에서는 사용자 인정 기준 변화에 따른 신뢰 관계 예측의 정확도 변화를 분석하여 사용자 인정 기준을 강화하여 신뢰 예측 결과의 정확도를 향상시킬 수 있음을 보인다. 본 연구의 결과가 최적의 사용자 인정 기준을 찾기 위한 실마리를 제공할 수 있을 것으로 기대한다.

1. 서론

온라인 사회연결망에는 다른 사용자들에 대한 신뢰를 명시적으로 표현 할 수 있는 기능이 존재한다. 이에 따라 온라인 사회연결망의 신뢰 관계를 활용한 사용자 특화 서비스 또는 상품 추천 서비스 또한 활발하게 연구되고 있다.

그러나 온라인 사회연결망의 신뢰 관계의 수는 온라인 사회연결망 내의 사용자 수에 비해 그 수가 매우 적어 이를 활용한 분석 및 서비스 개발에 어려움이 있다. 따라서 이를 해결하기 위해 신뢰 관계가 존재하지 않는 사용자들 간의 신뢰 관계를 예측하기 위한 연구가 진행되어 왔다[1][2].

[1]은 사용자 간의 상호작용 정보를 이용하여 신뢰 관계를 예측하였다. 전체 사용자를 대상으로 하는 신뢰 관계 예측 모델, 클러스터링 된 사용자 집단을 대상으로 하는 신뢰 관계 예측 모델, 사용자 개개인에 특화된 신뢰 관계 예측 모델을 각각 구축하여 대상이 구체화될수록 신뢰 관계 예측 모델의 정확도가 향상됨을 보였다. 또한, 위 논문에서는 신뢰 관계 예측의 정확도를 향상시키기 위해 사용자 인정 기준을 설정하고, 해당 기준 이하의 활동량을 갖는 사용자들을 제거하는 과정을 거쳤다.

위에서 언급한 것과 같이 온라인 사회연결망의 신뢰 관계는 그 분포가 매우 희박하고 그 분포가 한 쪽으로 편중된 경향을 보이므로 사용자 인정 기준에 따라 신뢰 관계 예측의 결과에 차이가 나타날 수 있다.

기존 연구들에서는 사용자 인정 기준 변화에 따른 신뢰 관계 예측의 정확도 변화에 대한 심도 있는 연구가 이루어지지 않았다. 따라서 본 논문에서는 사용자 인정 기준 변화에 따른 신뢰 관계 예측의 정확도를 측정하고, 그 이유를 분석하고자 한다.

2. 실험 데이터 및 방법

2.1. 실험 데이터

본 논문에서는 신뢰 관계 예측을 위해 Epinions.com의 데이터를 사용하였다. Epinions.com은 제품에 대한 리뷰를 다루는 웹 사이트로, 사용자는 제품의 리뷰를 작성하거나 다른 사용자가 작성한 리뷰를 읽을 수 있으며, 다른 사용자의 리뷰에 대한 평가를 점수로 부여할 수 있다. 리뷰 평가 점수는 1 점부터 6 점 사이의 점수로 이루어져 있다.

또한 사용자들은 신뢰할만한 다른 사용자에 대한 명시적 신뢰 관계를 설정할 수 있는데, 다른 사용자를 신뢰하는 사용자를 신뢰자(trustor), 다른 사용자로부터 신뢰를 받는 사용자를 피신뢰자(trustee)라 한다.

본 논문에서 사용한 Epinions.com 데이터는 8 개월간 수집된 데이터로 131,828 명의 사용자, 717,667 개의 신뢰 관계, 755,722 개의 리뷰, 그리고 4,835,208 개의 평가로 이루어져 있다.

2.2. 실험 방법

본 논문에서는 사용자 간의 신뢰 관계를 예측하기 위해 SVM을 이용한 클래시피케이션을 수행하였다.

클래시피케이션을 위해 표 1 의 총 8 가지 속성을 사용하였다. U_i 는 사용자를 의미한다.

사용자 인정 기준으로는 자신이 평가한 사용자의 수(평가 기준), 자신이 신뢰하는 피신뢰자의 수(피신뢰자 기준)를 사용하였다. 자신이 평가한 사용자 수의 평균은 41 이고, 자신이 신뢰하는 피신뢰자의 수는 6 이었다. 본 논문에서는 각 기준의 평균 값을 간격으로 0 부터 20 까지 5 개의 평가 기준과 0 부터 160 까지 5 개의 피신뢰자 기준을 설정하였다.

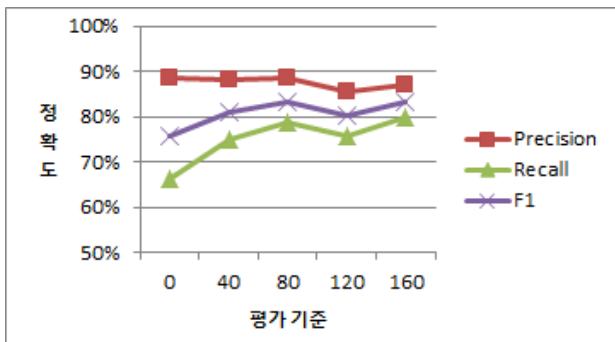
<표 1> 클래시피케이션에 사용한 속성

No	설명
1	U_i 이 U_j 에게 5 점 이상의 평가를 한 횟수
2	U_i 에게 평가를 받은 U_j 의 리뷰 중 전체 사용자에게 받은 평가의 평균 점수가 5 점 이상인 리뷰의 수
3	U_j 의 리뷰가 받은 평가 중 5 점 이 넘는 평가의 수
4	U_i 이 U_j 를 평가한 총 횟수
5	U_j 가 작성한 리뷰의 수
6	U_j 의 리뷰가 받은 평가의 표준 편차
7	U_i 에게 평가 받은 U_j 의 리뷰 중 전체 사용자에게 가장 많은 평가를 받은 리뷰의 평가 받은 횟수
8	U_j 의 리뷰 중 전체 사용자에게 가장 많은 평가를 받은 리뷰의 평가 받은 횟수

또한, 사용자 인정 기준을 만족하는 사용자 중 평가 수를 기준으로 상위 500 명과 하위 500 명의 사용자를 선정하였다. 선정된 사용자들이 갖는 신뢰 관계로 양성집합(positive set)을 생성하고, 신뢰 관계가 존재하지 않는 사용자 쌍 중에서 양성집합 내의 신뢰 관계 수만큼 임의의 사용자 쌍을 만들어서 해당 관계를 음성집합(negative set)을 생성하였다. 생성된 양성집합과 음성집합을 시간 순에 따라 5 등분하여 앞의 네 개의 집합을 모델 훈련(training set)에 사용하고 마지막 집합을 모델 정확도 측정에 사용(test set)하였다.

3. 실험 결과

평가 기준에 따른 정확도의 변화를 분석하기 위해 피신뢰자 기준을 고정하고 평가 기준을 변화시키며 정밀도, 회수도, 그리고 F1 척도의 변화를 측정하였다.

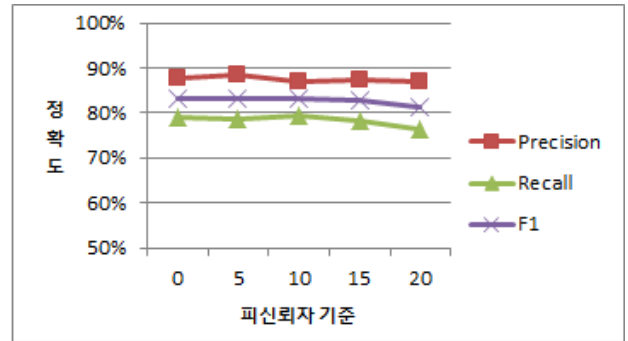


(그림 1) 평가 기준 변화에 따른 정확도 변화

그림 1 은 피신뢰자 기준을 5 로 고정하고 실험을 수행한 결과를 그래프로 나타낸 것으로, 이로부터 평가 기준을 증가시키며 따라 정밀도의 값은 다소 감소하지만 회수도와 F1 척도가 증가하는 것을 알 수 있다. 또한, 자신이 평가한 사용자 수의 평균인 41 보다 약 4 배나 큰 수를 평가 기준으로 선정하여도 정밀도

가 크게 떨어지지 않고, 회수도와 F1 척도가 유지됨을 알 수 있다.

피신뢰자 기준에 따른 정확도의 변화를 분석하기 위해 평가 기준을 고정하고 피신뢰자 기준을 변화시키며 정밀도, 회수도, 그리고 F1 척도의 변화를 측정하였다.



(그림 2) 피신뢰자 기준 변화에 따른 정확도 변화

그림 2 는 평가 기준을 80 으로 고정하고 실험을 수행한 결과를 그래프로 나타낸 것으로, 이로부터 평가 기준에 대한 실험에서와 달리 피신뢰자 기준을 변화시켜도 정확도에는 큰 차이가 나타나지 않으며, 오히려 피신뢰자 기준 10 을 초과하면 회수도가 감소하고 있음을 알 수 있다. 또한, 평가 기준과 달리 피신뢰자 기준의 경우 평균의 2 배 이상의 값을 설정하면 필요 이상으로 많은 사용자가 제거되어 신뢰 예측 결과의 정확도가 감소됨을 알 수 있다.

4. 결론

본 논문에서는 사용자 인정 기준 변화에 따른 신뢰 관계 예측의 정확도 변화를 분석하였다. 사용자 인정 기준으로는 평가 기준과 피신뢰자 기준을 사용하였으며, 사용자 인정 기준을 강화함에 따라 정확도가 향상됨을 보였다.

정밀도가 높더라도 회수도가 낮을 경우 해당 신뢰 예측 결과는 좋은 결과라고 할 수 없다. 본 논문의 결과로부터 유의미한 신뢰 관계 예측 결과를 얻기 위해서는 정밀도가 다소 감소하더라도 회수도가 높아질 수 있도록 사용자 인정 기준을 설정해야 함을 알 수 있다.

본 연구의 결과가 최적의 사용자 인정 기준을 찾기 위한 방안에 대한 연구에 도움이 될 것으로 기대한다

감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 IT 융합 고급인력과정 지원사업의 연구결과로 수행되었습니다. (NIPA-2011-C6150-1101-0001)

참고문헌

[1] N. Ma et al., "Trust Relationship Prediction Using Online Product Review Data," In *Proc. ACM Int'l Workshop on Complex Networks Meet Information and Knowledge Management*, pp. 47-54, 2009.

[2] R. Guha et al., "Propagation of Trust and Distrust," In *Proc. Int'l Conf. on World Wide Web*, pp. 403-412, 2004.